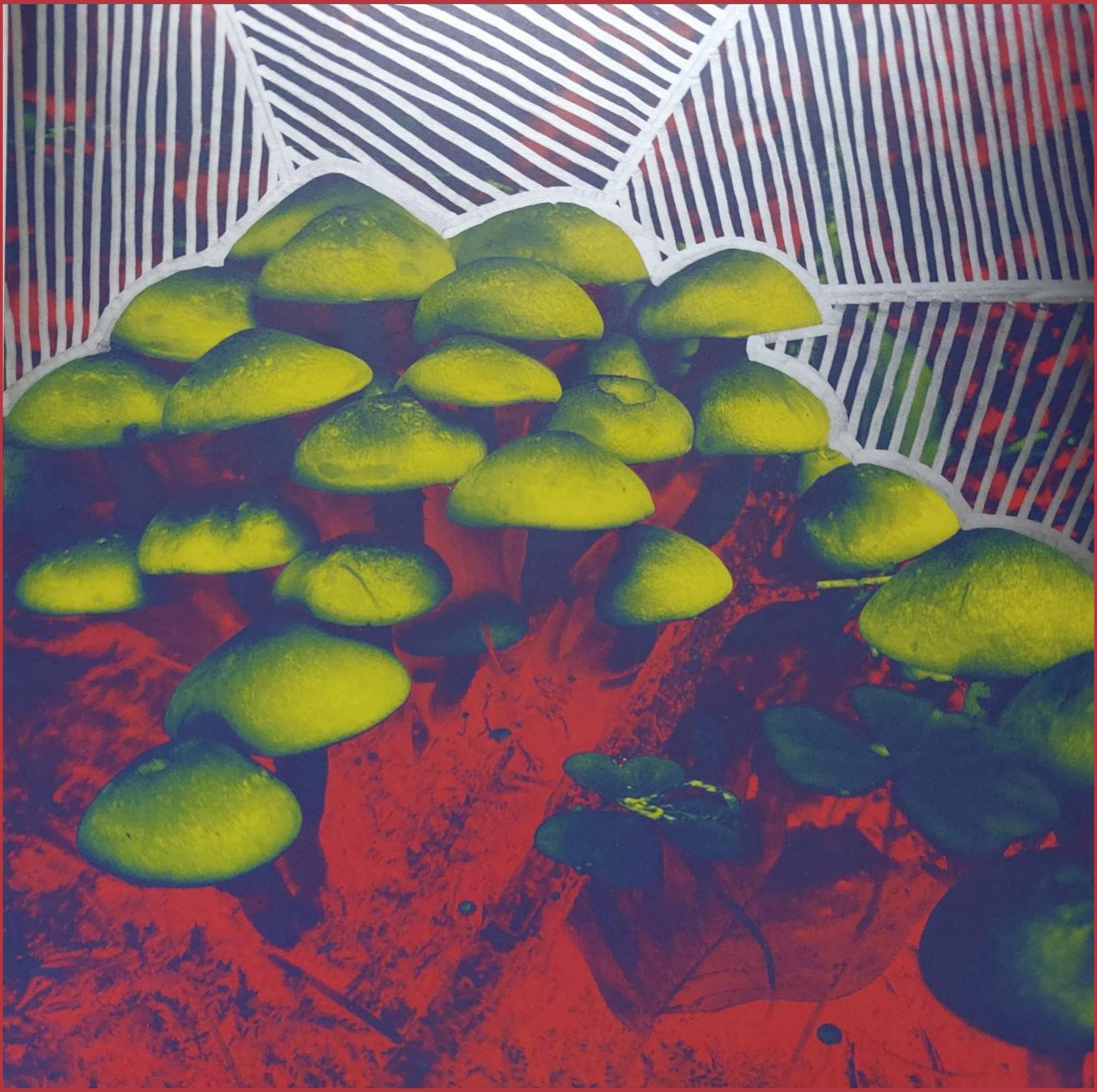
Annals of Computer Science and Information Systems
Volume 37

Communication Papers of the 18th Conference on Computer Science and Intelligence Systems

September 17–20, 2023. Warsaw, Poland



Maria Ganzha, Leszek Maciaszek, Marcin Paprzycki,
Dominik Ślęzak (eds.)

PTI

Annals of Computer Science and Information Systems, Volume 37

Communication Papers of the 18th Conference on Computer Science and Intelligence Systems

Maria Ganzha, Leszek Maciaszek, Marcin Paprzycki,
Dominik Ślęzak (eds.)

POLSKIE TOWARZYSTWO INFORMATYCZNE
POLISH INFORMATION PROCESSING SOCIETY

2023, Warszawa, Polskie
Towarzystwo Informatyczne

Annals of Computer Science and Information Systems, Volume 37

Communication Papers of the 18th Conference on Computer Science and Intelligence Systems

USB: 978-83-969601-4-6
WEB: 978-83-969601-3-9

ISSN 2300-5963
DOI 10.15439/978-83-969601-3-9

Ul. Solec 38/103
00-394 Warsaw
Poland

Contact: secretariat@fedcsis.org
http://annals-csis.org/

Cover art: Idealne nieidealnie 1 (Perfect not perfect 1)
Beata Branicka,
Elbląg, Poland

Also in this series:

Volume 36: Position Papers of the 18th Conference on Computer Science and Intelligence Systems, **ISBN WEB: 978-83-969601-1-5, ISBN USB: 978-83-969601-2-2**
Volume 35: Proceedings of the 18th Conference on Computer Science and Intelligence Systems, **ISBN WEB 978-83-967447-8-4, ISBN USB 978-83-967447-9-1, ISBN ART 978-83-969601-0-8**
Volume 34: Proceedings of the Third International Conference on Research in Management and Technovation **ISBN 978-83-965897-8-1**
Volume 33: Proceedings of the Seventh International Conference on Research in Intelligent and Computing in Engineering, **ISBN WEB: 978-83-965897-6-7, ISBN USB: 978-83-965897-7-4**
Volume 32: Communication Papers of the 17th Conference on Computer Science and Intelligence Systems, **ISBN WEB: 978-83-965897-4-3, ISBN USB: 978-83-965897-5-0**
Volume 31: Position Papers of the 17th Conference on Computer Science and Intelligence Systems, **ISBN WEB: 978-83-965897-2-9, ISBN USB: 978-83-965897-3-6**
Volume 30: Proceedings of the 17th Conference on Computer Science and Intelligence Systems, **ISBN WEB: 978-83-962423-9-6, ISBN USB: 978-83-965897-0-5**
Volume 29: Recent Advances in Business Analytics. Selected papers of the 2021 KNOWCON-NSAIS workshop on Business Analytics**ISBN WEB: 978-83-962423-7-2, ISBN USB: 978-83-962423-6-5**
Volume 28: Proceedings of the 2021 International Conference on Research in Management & Technovation, **ISBN WEB: 978-83-962423-4-1, ISBN USB: 978-83-962423-5-8**
Volume 27: Proceedings of the Sixth International Conference on Research in Intelligent and Computing in Engineering, **ISBN WEB: 978-83-962423-2-7, ISBN USB: 978-83-962423-3-4**
Volume 26: Position and Communication Papers of the 16th Conference on Computer Science and Intelligence Systems, **ISBN WEB: 978-83-959183-9-1, ISBN USB: 978-83-962423-0-3**
Volume 25: Proceedings of the 16th Conference on Computer Science and Intelligence Systems, **ISBN WEB 978-83-959183-6-0, ISBN USB 978-83-959183-7-7, ISBN ART 978-83-959183-8-4**

DEAR Reader, it is our pleasure to present to you Communication Papers of the 18th Conference on Computer Science and Intelligence Systems (FedCSIS 2023), which took place in Warsaw, Poland, on September 17-20, 2023.

In the context of the FedCSIS conference series, the *communication papers* were introduced in 2017, as a separate category of contributions. They report on research topics worthy of immediate communication. They may be used to mark a hot new research territory, or to describe work in progress, in order to quickly present it to scientific community. They may also contain additional information omitted from the earlier papers, or may present software tools and products in a research state.

FedCSIS 2023 was chaired by Jarosław Arabas and Sławomir Zadrożny, while Przemysław Biecek acted as the Chair of the Organizing Committee. This year, FedCSIS was organized by Polish Information Processing Society (Mazovia Chapter), IEEE Poland Section Computer Society Chapter, Systems Research Institute of Polish Academy of Sciences, as well as Faculty of Electronics and Information Technology and Faculty of Mathematics and Information Sciences, of Warsaw University of Technology.

FedCSIS 2023 was technically co-sponsored by IEEE Poland Section, IEEE Czechoslovakia Section Computer Society Chapter, IEEE Poland Section Systems, Man, and Cybernetics Society Chapter, IEEE Poland Section Computational Intelligence Society Chapter, Committee of Computer Science of Polish Academy of Sciences, and Mazovia Cluster ICT. Moreover, two years ago, the FedCSIS conference series formed strategic alliance with QED Software, a Polish software company developing AI-based products, and this collaboration has been continued.

In 2023, FedCSIS was sponsored by QED Software, Samsung, Hewlett Packard Enterprise, Łukasiewicz Research Network – Institute of Innovative Technologies EMAG, MDPI, Sages, Efigo, and CloudFerro.

During FedCSIS 2023, the keynote lectures were delivered by:

- Lipika Dey, Tata Consultancy Services, India, keynote title: *Deciphering Clinical Narratives – Augmented Intelligence for Decision Making in Health Care Sector*
- Marta Kwiatkowska, University of Oxford, UK, keynote title: *When to trust AI...*
- Andrea Omicini, Alma Mater Studiorum – Università di Bologna, Italy, keynote title: *Measuring Trustworthiness in Neuro-Symbolic Integration*
- Roman Słowiński, Poznań University of Technology, Poland, keynote title: *Multiple Criteria Decision Aiding by Constructive Preference Learning*

Moreover, two special guests delivered invited presentations:

- Gianpiero Cattaneo, Retired from Department of Informatics, Systems and Communications, University of Milano-Bicocca, Italy, invited presentation title: *Abstract Approach to Entropy and Co-Entropy in Measurable and Probability Spaces*
- Jerzy Nawrocki, Poznań University of Technology, Poland, invited presentation title: *Towards reliable rule mining about code smells: The McPython approach*

FedCSIS 2023 consisted of Main Track, with five Topical Areas and Thematic Tracks. Some of Thematic Tracks have been associated with the FedCSIS conference series for many years, while some of them are relatively new. The role of Thematic Tracks is to focus and enrich discussions on selected areas, pertinent to the general scope of the conference.

Each contribution, found in this volume, was refereed by at least two referees. They are presented in alphabetic order, according to the last name of the first author. The specific Topical Area or Thematic Track that given contribution was associated with is listed in the article metadata.

Making FedCSIS 2023 happen required a dedicated effort of many people. We would like to express our warmest gratitude to the members of Senior Program Committee, Topical Area Curators, Thematic Track Organizers and to members of FedCSIS 2023 Program Committee. In particular, we would like to thank those colleagues who have refereed all of the 358 submissions.

We thank the authors of papers for their great contributions to the theory and practice of computer science and intelligence systems. We are grateful to the keynote and invited speakers for sharing their knowledge and wisdom with the participants.

Last, but not least, we thank Jarosław Arabas, Sławomir Zadrożny, and Przemysław Biecek. We are very grateful for all your efforts!

We hope that you had an inspiring conference. We also hope to meet you again for the 19th Conference on Computer Science and Intelligence Systems (FedCSIS 2024), which will take place in Belgrade, Serbia, on September 8-11, 2024.

Co-Chairs of the FedCSIS Conference Series:

Maria Ganzha, *Warsaw University of Technology, Wand Systems Research Institute Polish Academy of Sciences, Poland*

Leszek Maciaszek (Honorary Chair)*, Macquarie University, Australia and Wrocław University of Economics, Poland*

Marcin Paprzycki, *Systems Research Institute Polish Academy of Sciences, and Warsaw University of Management, Poland*

Dominik Ślęzak, *University of Warsaw, Poland and QED Software, Poland and DeepSeas, USA*

Annals of Computer Science and Information Systems,
Volume 37

Communication Papers of the 18$^{\text{th}}$ Conference on Computer Science and Intelligence Systems

September 17–20, 2023. Warsaw, Poland

TABLE OF CONTENTS

COMMUNICATION PAPERS

Analysing Perceptions of South African Digital Artists towards Non-Fungible Token (NFT) Use 1
Nathier Abrahams, Pitso Tsibolane, Jean-Paul Van Belle

Exploring M-Commerce Vendors' Perspectives in Post-Saudi Vision 2030: A Thematic Analysis 9
Yahya Alqahtani, Natalia Beloff, Martin White

Population-less Genetic Algorithm? Investigation of Non-dominated Tournament Genetic Algorithm (NTGA2) for multi-objective optimization 21
Michał Antkiewicz, Pawel Myszkowski

Real-Time Detection of Small Objects in Automotive Thermal Images with Modern Deep Neural Architectures 29
Tomasz Balon, Mateusz Knapik, Bogusław Cyganek

Polyhedral Tiling Strategies for the Zuker Algorithm Optimization 37
Wlodzimierz Bielecki, Marek Palkowski, Piotr Blaszynski, Maciej Poliwoda

Can ChatGPT Replace a Template-based Code Generator? 43
Adam Bochenek

An Evaluation of a Zero-Shot Approach to Aspect-Based Sentiment Classification in Historic German Stock Market Reports 51
Janos Borst, Lino Wehrheim, Andreas Niekler, Manuel Burghardt

Proof-of-Work CAPTCHA with password cracking functionality 61
Szymon Chadam, Paweł Topa

Formal verification of BPMN diagrams in Integrated Model of Distributed Systems (IMDS) 65
Wiktor Daszczuk, Jakub Jałowiec

Star-critical Ramsey numbers for hexagon 75
Tomasz Dzido

Towards modelling and analysis of longitudinal social networks 81
Jens Dörpinghaus, Vera Weil, Martin W. Sommer

Comparative Study: Defuzzification Functions and Their Effect on the Performance of the OFNbee Optimization Algorithm 91
Dawid Ewald, Wojciech Dobrosielski, Jacek M. Czerniak, Hubert Zarzycki

Use of Dynamic Neural Networks for Modeling Nonlinear Objects with Significant Nonlinearity 97
Oleksandr Fomin, Sergii Polozhaenko, Valentyn Krykun, Andrii Orlov, Andrii Prokofiev

Compilation through interpretation 103
Adam Grabski, Ilona Bluemke

Waiting list procedure improvements for master program courses in Information and Computing Sciences 111
Tomas van Groningen, Sietse Overbeek

Conceptualizing sustainability in the context of ICT. A literature review analysis. 121
Tobias Hassmann, Markus Westner

FAS-CT: FPGA-Based Acceleration System with Continuous Training 131
Manuel Luis González Hernandez, Jorge Ruiz, Randy Lozada, Erik Sebastian Skibinsky Gitlin, Ángel M. García-Vico, Javier Sedano, José Ramón Villar

Automatic Construction of Knowledge Graph of Tea Diseases and Pests 141
Qiang Huang, Youzhi Tao, Shitao Ding, Yongbo Liu, Francesco Marinello

Harness Old Media: a cross-disciplinary approach to utilizing television data for media content analysis. 147
Piotr Jabłoński

Enhancing Image Quality through Automated Projector Stacking 153
Franciszek Jełowicki

Mushroom Picking Framework with Cache Memories for Solving Job Shop Scheduling Problem 157
Piotr Jędrzejowicz, Izabela Wierzbowska

Ausklasser - a classifier for German apprenticeship advertisements 165
Kai Krüger

Developing an interval method for training denoising autoencoders by bounding the noise 173
Bartłomiej Jacek Kubica

An Empirical Framework for Software Aging-Related Bug Prediction using Weighted Extreme Learning Machine 181
Lov Kumar, Vikram Singh, Lalita Bhanu Murthy, Sanjay Misra, Aneesh Krishna

Mathematical Modeling of Water Regimes on Drained Lands 189
Lyudmyla Kuzmych, Halyna Voropai, Stepan Kuzmych

Research Paper Blockchain (RPB) 195
Nicholas Kniveton, Navid Shaghaghi

Using Features of PLRs to Chromatic Light Pulse Irradiations of Either Eye to Detect Symptoms of Dementia 201
Minoru Nakayama, Wioletta Nowak, Anna Zarowska

A machine learning approach for automatic testing 209
Felix Petcusin, Cosmin Stoica Spahiu, Liana Stanescu

Generation of Benchmark of Software Testing Methods for Java with Realistic Introduced Errors 215
Tomas Potuzak, Richard Lipka

Efficient Feature Selection On Adversarial Botnet Detection 223
Raqeebir Rab, Farsha Bindu, Sheikh Sanjida, Nafiza Tabassoum, Tamima Binte Wahab, Anonnya Ghosh

To propose an attack detection model for enhancing the security of 5G-enabled Vehicle-to-Everything (V2X) communication for smart vehicle 229
Prince Rajak, Pavan Kumar Mishra

Parallelized Population-Based Multi-Heuristic System with Reinforcement Learning for Solving Multi-Skill Resource-Constrained Project Scheduling Problem with Hierarchical Skills 237
Ewa Ratajczak-Ropel, Piotr Jędrzejowicz

Ensemble-based versus expert-assisted approach to carbon price features selection 245
Bogdan Ruszczak, Katarzyna Rudnik

Architecture Analysis, Cloud-Based Using MQTT Protocol for Braille Literacy 251
Josimar dos Santos, Gilton José Ferreira da Silva, Admilson de Ribamar Lima

A simple algorithm for computing a multi-dimensional the Sierpiński space-filling curve generalization **259**
Ewa Skubalska-Rafajłowicz

Knowledge-Based Creation of Industrial VR Training Scenarios **265**
Paweł Sobociński, Jakub Flotyński, Michał Śliwicki, Mikołaj Maik, Krzysztof Walczak

Interactive, Personalized Decision Support in Analyzing Women's Menstrual Disorders **273**
Łukasz Sosnowski, Soma Dutta, Iwona Szymusik, Wojciech Chaber, Paulina Kasprowicz

Data science to identify crimes against public administration **281**
Luan Souza, Methanias Colaço Júnior, Raphael Fontes, Rodrigo Silva, Gabriel Lins, Caldeira Silva, Jailton Paiva, Ricardo Valetim

Evolutionary k-means Graph Clustering Method to Obtain the hub&spoke Structure for Warsaw Communication System **289**
Jarosław Stańczak, Jan Owsiński, Barbara Mażbic-Kulma, Aleksy Barski, Krzysztof Sęp

Forecasting migration of EU citizens to Germany using Google Trends **295**
Nicholas Steinbrink

The Grammar and Syntax Based Corpus Analysis Tool For The Ukrainian Language **303**
Daria Stetsenko, Inez Okulska

The Use of AI to Determine the Condition of Corn in a Field Robot that Meets the Requirements of Precision Farming **313**
Justyna Stypułkowska

Blockchain-based certification of research outputs and academic achievements: A case of scientific conference **323**
Robert Susik, Robert Nowotniak, Emanuel Kulczycki

An Elliptic Intuitionistic Fuzzy Portfolio Selection Problem based on Knapsack Problem **329**
Velichka Traneva, Petar Petrov, Stoyan Tranev

An Elliptic Intuitionistic Fuzzy Model for Franchisor Selection **337**
Velichka Traneva, Stoyan Tranev

A modular and verifiable software architecture for interconnected medical systems in intensive care **345**
Marc Wiartalla, Frederik Berg, Florian Ottersbach, Jan Kühn, Mateusz Buglowski, Stefan Kowalewski, André Stollenwerk

Experimental Assessment of MPTCP Congestion Control Algorithms for Streaming Services in Open Internet **353**
Łukasz Łuczak, Przemysław Ignaciuk, Michał Morawski

Author Index **359**

DOI: 10.15439/2023F4823
ISSN 2300-5963 ACSIS, Vol. 37

Analysing Perceptions of South African Digital Artists towards Non-Fungible Token (NFT) Use

Nathier Abrahams
0000-0002-2719-5878
Dept. of Information Systems,
University of Cape Town,
Rondebosch, 7701, South Africa
Email: abrmog027@myuct.ac.za

Pitso Tsibolane
0000-000 2-7888-1181
Dept. of Information Systems,
University of Cape Town,
Rondebosch, 7701, South Africa
Email: pitso.tsibolane@uct.ac.za

Jean-Paul Van Belle
0000-0002-9140-0143
Dept. of Information Systems,
University of Cape Town,
Rondebosch, 7701, South Africa
Email: jean-paul.vanbelle@uct.ac.za

***Abstract*—Digital art has many major pitfalls, ranging from issues around tracking ownership to piracy. Non-fungible tokens (NFTs) can solve these issues and bring new benefits, such as access to larger markets. Despite this, South Africa's digital artists have slowly adopted NFTs. This research aims to understand the values-based perceptions of South African digital artists toward NFTs. Fifteen South African digital artists were interviewed using semi-structured interviews guided by the updated Holbrook's Typology of Consumer Value framework. Ten positive perceptions, three negative perceptions, three risks and one benefit were identified, explored and analyzed using the framework. This research can assist digital artists and other stakeholders in the NFT ecosystem to understand the values-based perceptions of South African digital artists. It can be used to help assist decision-makers, artists, intermediaries and other stakeholders in South Africa and potentially elsewhere. Additionally, the validated and updated Typology of Consumer Value can benefit researchers using this framework in future research.**

***Index Terms*—Non-fungible tokens (NFTs), blockchain, consumer value, digital artists, updated Typology of Consumer Values.**

I. Introduction

DIGITAL art encounters significant challenges stemming from the intrinsic nature of its digital medium, which enables effortless replication, thereby blurring the distinction between the copy and the original artwork [1]. Another issue faced by digital art revolves around the difficulty in accurately tracking and verifying ownership of the artwork [2]. To address these challenges, non-fungible tokens (NFTs) have emerged as a potential solution, offering digital artists access to novel markets and alternative methods of selling their creations while bypassing conventional art institutions, such as galleries [3].

NFTs hold the promise of effectively tracking ownership of digital assets like digital art and providing artists with new opportunities in the art market. Despite these benefits, the adoption of NFT technology has been slow among South African artists. Hence, this study aims to explore the perceptions of South African digital artists toward NFTs from a values-based perspective.

The primary research question for this research is: *"What are the values-based perceptions of South African digital artists toward NFTs?"* The secondary questions for this research are: *(a) What are the perceived values-based benefits of NFT use by digital artists in South Africa? (b) What are the perceived values-based risks of NFT use by digital artists in South Africa?*

This research focused on the perceptions of South African digital artists toward NFTs. NFTs have a lot of applications outside of digital art, which is not part of this research. This research will not cover digital artists outside of South Africa, nor will other stakeholders within the blockchain ecosystem.

II. Literature Review

The literature review discusses four key concepts: blockchain, non-fungible tokens, digital art (including perceptions of South African digital artists), and perceptions of value.

A. Blockchain, Fungible and Non-Fungible Tokens

Blockchain serves as the underlying technology for Non-Fungible Tokens (NFTs) [22]. It operates as a cryptographically secured decentralized shared ledger, where data is transparent and visible to all network participants, and its contents are verified by all [4][21]. Utilizing blockchain eliminates the need for third-party management, as network participants collectively manage the ledger.

Tokens, generated by the blockchain system, act as digital assets representing products, services, or currencies through tokenization [5]. Tokens can be classified into three types: fungible, non-fungible, and semi-fungible. Fungible tokens are interchangeable, uniform, and divisible, while non-fungible tokens are unique, non-interchangeable, and indivisible, serving as a unique digital certificate for ownership of a digital or physical asset that cannot be replicated [5]. Non-fungible tokens adhere to the ERC721 standard, enabling the creation of tokens for both physical and digital assets, facilitating their transfer between crypto wallets, checking wallet balances, determining token ownership, and ascertaining the total supply of tokens on the blockchain [4].

The integration of smart contracts within NFTs provides digital artists with novel revenue streams, such as programming the NFT to automatically pay them a percentage of each sale as a built-in royalty in the token's metadata, offering previously unavailable revenue opportunities [5].

Recently, semi-fungible tokens have emerged, which operate more like wallets than individual tokens. Tokens of the same type are grouped together and considered fungible within their group [7]. In the context of digital art, these tokens can represent entire art collections or galleries containing multiple art pieces.

B. *Digital Art*

Digital art, especially in South Africa, is a relatively new domain, resulting in limited literature on digital artists' perceptions of new technologies. In Africa, particularly in South Africa, technology adoption has followed a unique trajectory, with artists adopting mobile technology before acquiring desktop or laptop computers [8]. South African artists have utilized platforms like Instagram to promote their artwork, conduct business, and generate commission leads [9]. The authors concur that social media platforms, designed for ease of use on smartphones, have witnessed higher adoption rates in South Africa compared to desktops and laptops, indicating that South African digital artists are open to embracing new technologies.

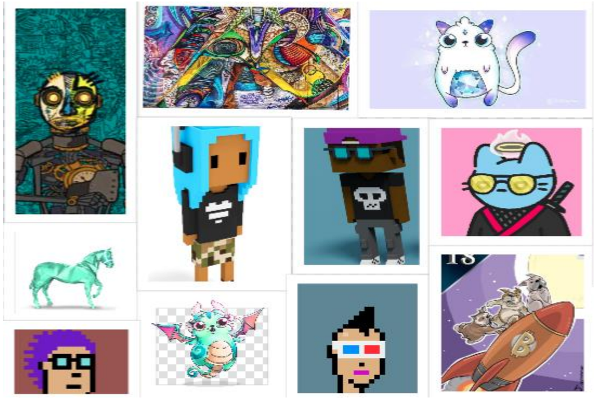


Fig 1. Sample popular NFTs (Sharma et al., 2022).

Before the advent of Non-Fungible Tokens (NFTs), selling digital art involved risks, with customers purchasing files from artists and potentially misusing them without consent. Crypto art refers to "limited-edition digital art, cryptographically registered with a token on a blockchain" [10] [11] . When an NFT is minted, it is added to the blockchain, and the art piece is linked to the token, serving as proof of ownership, origin, and a catalogue raisonné [1]- a record of ownership and the history of the art piece [1].

One of the main advantages of crypto art over traditional art and traditional digital art lies in artists' independence from galleries and art brokers, granting them self-determination. Crypto art facilitates the formation of artist communities and enables a stable income through royalties from art resales facilitated by smart contracts [12]. Marketplaces like SuperRare and OpenSea have emerged, enabling direct buying, selling, trading, and exchanging of NFTs without third-party verification [12]. These platforms offer both primary and secondary markets for NFTs.

C. *Perceptions of Value*

Holbrook's Typology of Perceived Consumer Value [13] served as the framework to gain insights into the perceptions of South African digital artists regarding NFTs. The framework examines the perceived benefits that digital artists experience through their use or non-use of NFTs. In this context, the digital artist acts as the consumer "consuming" NFT technology by utilizing it. Analyzing the perceptions of digital artists can be achieved by evaluating the perceived value they derive from employing or not employing NFTs. Holbrook's original Typology of Consumer Value comprised three dimensions, each with two options: self-versus other-oriented, active versus reactive, and extrinsic versus intrinsic (Table 1).

TABLE I. HOLBROOK'S TYPOLOGY OF CONSUMER VALUE [13].

Orientation	**Activity**	**Extrinsic**	**Intrinsic**
Self-oriented	Active	Efficiency: input/output, convenience…	Play: fun
	Reactive	Excellence: Quality	Aesthetics: beauty
Other-oriented	Active	Status: success, impression …	Ethics: virtue, justice…
	Reactive	Esteem: reputation, possession…	Spirituality: faith, ecstasy, magic…

Holbrook's framework provides a comprehensive understanding of consumer value [14]. Over the years, the framework has evolved and gained strength, incorporating new value types and adaptations as fresh insights into consumer value emerged [15]. Table II illustrates the updated typology, categorizing value types into positive and negative, comprising 14 positive value types and 10 negative value types. The updated framework serves as a flexible "menu card," allowing for the selection of relevant value types that apply to specific contexts, as not all value types may be applicable [15].

TABLE II. UPDATED TYPOLOGY OF CONSUMER VALUE [15].

Value Type	**Brief Description**	**Source**
Positive value types	***The (perceived) extent to which the object:***	
Convenience (efficiency)	Makes the life of the customer easier	O T C
Excellence	Is of high quality. Depending on the context, this can relate to the quality of the product(s), service(s), or both. Depending on the context, this can include reliability, empathy, responsiveness, interactional quality, etc.	O E T H
Status	Makes a positive impression on others and thus leads to social acceptance	O T C

Self-esteem (esteem)	Positively affects the customer's attitude toward or satisfaction with oneself	O E
Enjoyment (play)	results in fun and pleasure	O T
Aesthetics	Is appealing. This involves the attraction of the object's design and atmospheric aspects such as layout, colour, etc. This can be related to all the senses (sight, smell, touch, taste, hearing)	O
Escapism (spirituality)	allows the customer to relax and escape from reality or daily routine	O E
Personalization	Is adapted to the individual customer	T C
Control	Can be commanded or influenced by the customer. This can relate to the timing, content, and/or sequence of the service delivery process or outcome	T
Novelty	Creates curiosity and/or satisfies a desire for knowledge (i.e. wanting to know more about it). This is only applicable for new objects (such as new technologies)	T
Relational benefits	Results in a better relationship with the service provider	T H
Social benefits	Results in a better relationship with other customers	C
Ecological benefits (ethics)	Has a positive impact on environmental well-being	O R C
Societal benefits (ethics)	Has a positive impact on societal well-being. This can involve CSR initiatives such as fair trade, community support, employee fairness, etc.	O R
Negative value	***The (perceived) extent to which the object:***	
Price	Is expensive	E T C
Time	Requires time to prepare, use, understand, etc.	E
Effort	Requires effort to prepare, use, understand, etc.	E T
Privacy risk	Can result is a loss of privacy	T
Security risk	can result in security issues such as losing personal information to criminals or hacking	T
Performance risk	Can result in a loss of performance: the object does not perform as expected or intended	T C
Financial risk	can result in a loss of money	T C
Physical risk	Can result in health issues or injuries	T C
Ecological costs	Has a negative impact on environmental well-being (pollution)	C R
Societal costs	Has a negative impact on societal well-being. This can involve issues such as child labour, poor working conditions, etc.	R

Source: O= Original value type mentioned by Holbrook; E= update of original value type in empirical work using Holbrook's typology; T= value type related to technology; H= value type related to human contact; C= value type related to collaborative consumption; R= value type related to transformative service research

From Table II, the following propositions ("P") are proposed for assessment in this study.

1) Positive value types

- P1: digital artists find that it is more convenient to sell their art online as an NFT than to sell their works through a traditional art gallery.
- P2: digital artists find that they receive both service excellence by using blockchain-based services to buy and sell their art and product excellence by perceiving that their crypto art sold as NFTs is of higher quality compared to other digital art.
- P3: digital artists enjoy the fame (status) that comes with being a popular or fast-selling artist as their sales are trackable on the digital marketplaces.
- P4: digital artists derive enjoyment from the process of minting and selling their NFTs.
- P5: digital artists value being in control of the programming of their NFTs, such as being able to build in that they receive royalties from every subsequent sale of their art.
- P6: digital artists are creating NFTs because they are curious and want to learn about this new technology (novelty).
- P7: digital artists are using NFTs because they believe in the decentralization of art away from exploitative art galleries and middlemen (ethics).

2) Negative value types

- P8: The price of minting an NFT inhibits digital artists from creating NFTs.
- P9: The time and effort required to learn how to create NFTs make it difficult for digital artists to use and gain value.
- P10: Digital artists are worried that their sales and, thus, income were visible to the public due to the transparent nature of the blockchain (privacy risk).
- P11: Digital artists might be concerned about the security risk of not creating their NFT correctly or getting scammed.
- P12: Digital artists are worried that the NFT they create might not get sold or do not behave as intended (performance risk).
- P13: digital artists are concerned that they will make a loss when creating their NFTs (financial risk).
- P14: digital artists are concerned about the ecological impact of blockchain technology on the environment.

III. Research Design and Methodology

The research employed an exploratory approach to describe the perceptions of South African digital artists towards NFTs. Descriptive research, based on categorical schemes, was used to observe and understand the phenomenon [16]. An interpretive philosophy was adopted to gain insight into the subjective perspectives and behavior of the participants, specifically to comprehend the perceptions of South African digital artists regarding NFTs.

A qualitative strategy was implemented through semi-structured interviews, allowing for open exploration and adjustment of questions to suit each interviewee's understanding [17]. The study utilized the updated framework of Holbrook's Typology of consumer value, designed to understand consumer perceptions of value [14]. This framework, a synthesis of other value typologies, served as a guide for framing interview questions to explore digital artists' perceptions of NFT use [15].

The research applied purposive and snowball sampling methods. Initial participants were digital artists accessible to the researcher, who subsequently referred other artists for interviews. Additionally, purposive sampling was utilized to reach out to digital artists, resulting in interviews with fifteen South African digital artists aged above eighteen [18]. Thematic analysis, based on the phases developed by [18], was employed to analyze the data. The research protocol and study instruments were approved by the Ethics in Research Committee.

IV. Analysis and Findings

This section will present the results found by analyzing the interviews.

A. Demographic Results

Table II. Demographics of Interview Respondents

Nr	G	Age	Overall Awareness of NFTs	Used NFTs	Minted NFTs	Sold NFTs	Digital artist experience
R1	F	33	Heard about it through twitter	No	No	No	12 years
R2	F	34	Heard about it through networks	No	No	No	7 years
R3	M		Uses NFTs	Yes	Yes	Yes	1 year
R4	M	40	Uses NFTs	Yes	Yes	Yes	20 years
R5	M	33	Uses NFTs	Yes	Yes	Yes	10 years
R6	F	18	Not much, just read a bit in the news	No	No	No	1.5 yrs
R7	F	23	Has traded in NFTs	Yes	No	Yes	5 years
R8	M	27	Does not know much	No	No	No	4 years
R9	F	29	Read a little	No	No	No	2 years
R10	M	26	Has sold artwork as NFTs	Yes	Yes	Yes	1 year
R11	M	32	Knows a little, has created art that was later sold on as NFTs	Yes	No	No	10 years
R12	M	23	Does not use NFTs	No	No	No	1.5 yrs
R13	M	32	Occasionally sells his art as an NFT	Yes	Yes	Yes	13 years
R14	F	30	Just aware of the term NFT	No	No	No	8 years
R15	M	30	Does not use NFTs	No	No	No	8 years

B. Positive Value Types

1) Convenience

Proposition 1: Digital artists find that it is more convenient to sell their art online as an NFT than to sell their works through a traditional means such as an art gallery.

Respondents said that they would get better service by using the blockchain instead of a gallery for selling their digital art. R3 mentioned the faster turnaround time that they receive when selling through a gallery: “*I get better service through selling through the blockchain. In one weekend I could have a fresh project up for sale and sell it immediately for however much. With the gallery, you'd need to go through like a waiting process. They will take a cut. They're going to see if it's a good fit for the space. If they are waiting, list all that stuff on the blockchain. You can be ready in 24 hours or even less*”- R3. “*I think that I would have a lot more control of the process. Galleries have more control around the selling process*”- R15. “*Decentralization allows you to take the power into your own hands. It eliminates the middleman and replaces it with a marketplace. As a freelance artist you get screwed over a lot or people don't pay you on time*”- R13. *“Access to a broader market*”- R2. “*The blockchain itself is geared toward a target market that wants collectibles and investments. Instead of brick and mortar*”- R4. *“I'd get more value as I can sell more pieces. Whereas if you sell it through a more traditional means like Behance, you usually only sell it once off* “– R7. *“I think that digital art is more accessible than a gallery. You won't be geographically limited.*”– R6. *“The bar of entry is lower when compared with a traditional institution. You only need a phone and wifi”*- R5. “*My work has been sold much better as NFTs than a traditional means such as a gallery. Galleries are difficult to get into as they have their own methodology for who's art they display. You can attend Virtual Reality galleries in the metaverse to view digital art. During the lockdown, traditional artists struggled to sell their art due to galleries being closed.*” – R10

Less resources are perceived to be required when selling through the blockchain. “*I think its more accessible if you sell it through on the blockchain. Anyone with a phone and an internet connection could buy and sell the art. Its hard to get into a gallery”- R12. I imagine the experience would be better if selling through the blockchain, and it would give you more access to the niche demographic of NFT art purchasers. You can also get more exposure, if you don't get lost in the sea of other people*. - R9.

2) Excellence

Proposition 2: Digital artists find *that* they receive both service excellence by using the blockchain-based services to buy and sell their art, and product excellence by perceiving that their crypto art sold as NFTs are of higher quality compared to other digital art.

Excellence can relate both to products (product excellence) and service (service excellence). Non-fungible tokens have both a service component and a *product* component. In terms of service excellence, many respondents have said that they receive better service through the blockchain when compared with a gallery: “*I get better service through selling through the blockchain*”- R3.

R10 mentioned that his product sold better as an NFT, as they felt that galleries are *harder* to get into. They also mentioned that galleries were closed during the COVID-19 lockdowns in South Africa, and artists could not make an income selling their art as a result. “*My work has been sold much better as NFTs than a traditional means such as a gallery. Galleries are difficult to get into as they have their own methodology for whose art they display.” “During the lockdown, traditional artists struggled to sell their art due to galleries being closed”- R10.*

In terms of product excellence, some digital artists that used NFTs felt that their art was of the same quality regardless of whether their art was an NFT or not. *“It hasn't influenced my style in any sense. I sold art that I created before I started using NFTs in 2019 as an NFT in 2021. People are open-minded, depending on who is willing to collect your artwork. Collectors aren't stuck in a specific style. You don't need to change yourself to make it. If you really on trends, you won't* *have a style that people will appreciate. You can end up becoming a one hit wonder."- R10.*

3) Status

Proposition 3: Digital artists enjoy the fame of being a popular or fast-*selling* artists as their sales are trackable on the digital marketplaces.

Two of the respondents felt that they are enjoying the anonymity they get *from* using a pseudonym, as they do not like that their transaction history is traceable but felt that it didn't improve their status as digital artists. *"I don't think it improves my status, just puts lots of money in your bank account, you get known in the NFT circles. I use an alias, and people know my alias and not me personally, which is great"*- R4. "*So far, I can't say it improves my status as an artist. Some of the online accolades I received, doesn't translate into real life accolades. I enjoy the anonymity that I get."* – R10.

Two respondents felt that it negatively affected the social status of *an* NFT artist. The first mentioned that it can negatively affect a digital artist's status due to negative perceptions surrounding NFTs: "*I don't think it improves your status as an artist. People will categorize you as an NFT artist. I think it limits you to a certain community. A lot of people don't see NFTs as a positive thing."* - R12. The second mentioned that NFTs have a negative impact on the planet *and* being associated with that impact can have a negative impact on a digital artist's status: *"It leaves a negative impression, people think NFTs are destroying the planet"*- R3.

Some digital artists said that they would be perceived in a more positive light: "Not many people understand Blockchain and NFTs, so if you say you're an NFT artist, they see you as a tech-savvy person that's ahead *the times"*- R7. "*I don't know. Maybe they would say that you are more in tune with what's going on society today with Bitcoin and NFTs and the digital space"*- R8. *"If I look at the more professional audience, their perceptions of me might become more positive like I'm trying something new. Putting your eggs in many baskets."* - R11. Even though this might seem like it contradicted his previous statement, in this case he was speaking hypothetically: *"You may be perceived as more forward thinking or on the cusp of something...* "*"It could improve your status as an artist. I think it's quite niche, however."*- R14.

Although the data did not fully support proposition 3, it did support the construct and the literature surrounding status as a value type as respondents felt that using the technology would improve their status.

4) Enjoyment (play)

Proposition 4: Digital artists derive enjoyment from the process of minting and selling their NFTs.

Many respondents felt that minting and selling NFTs brings some level of joy. Two respondents mentioned the minting aspect of the proposition. One mentioned the joy of free minting *"Platforms that offer free minting, give me a lot of joy"- R10.* Another respondent mentioned the joy of minting in and of itself. *"There's that instant gratification, and impulse where you've worked on something from scratch. Like a farmer taking his crops to market"- R4.*

Another respondent mentioned that they enjoy having control of the *process* and being able to create art that they enjoy creating to be sold on the marketplace. *"Just having the ability to have control of the process, allowing for self-expression as opposed catering to what the client would like..."*- R15.

5) Control

Proposition 5: Digital artists value being in control of the programming of their NFTs, *such* as being able to build in that they receive royalties from every subsequent sale of their art.

Proposition P5 was supported by a respondent that mentioned that enjoy the flexibility and control that digital artists get when *programming* the smart contracts for their NFTs. *"So far, I'm extremely happy about it, you can also develop your own smart contracts. You can put your own stipulations in the contract, that gives you the power to mint from that contract and release the art on multiple blockchains at once. It also allows you to reward the people that support you through exclusive collections. You can airdrop art to your collectors, which you can view. There's also no censorship when it comes to content, as an artist you need to know which red lines not to cross."*- R10.

Other respondents mentioned the protection and control that artists can get when *writing* smart contracts. *"Designers are short changed when it comes to work as work is often stolen and if there are processes in place to control who it is shared with and if the original creator has some control, it's a good thing."- R1. "Because there's no middleman, you have a lot more control of the decision-making process, such as pricing and royalties, and galleries can be greedy"- R15.*

6) Novelty

Proposition 6: Digital artists are creating NFTs because they are curious and want to learn about this new technology. Although none of the respondents mentioned that they are creating NFTs specifically because they are interested in learning about the technology, most of the respondents were curious about the technology, for various reasons.

Some respondents were interested in the technology itself. "*...technology that's growing behind it, especially smart contracts. When you wanted to create a smart contract, you had to pay around $5000 dollars for a developer to create it. Now you get platforms that make the contract easier to create."- R10.* Another respondent was interested in applications of NFTs outside the scope of art. *"Mostly the application aspect and what we can use it for in the future, outside of art"*- R3.

Word of mouth was mentioned by respondent R12. *When I first started as a digital artist, NFTs were popular, and everyone was telling me to make NFTs. Then I went to do my research on it- R12.* Similar to word of mouth, respondent R13 *became* interested in NFTs due to online groups that he belongs to on social media. "*I'm in a couple of NFT groups and spaces and I pay attention to when people are talking* *about it on Clubhouse. Nowadays it moves so fast. It's interesting, it's fun, we need to get educated about it and everybody needs to learn."*-R13.

7) Societal benefits (ethics)

Proposition 7: Digital artists are using NFTs because they believe in the *decentralization* of art away from exploitative art galleries and middlemen.

Decentralization enables self-determination: *"Decentralization allows you to take the power into your own hands. It eliminates the middleman and replaces it with a marketplace. As a freelance artist you get screwed over a lot or people don't pay you on time"* – R13. One respondent mentioned protection against piracy and plagiarism. *"If you can protect people's intellectual property, how their work gets distributed and if it can prevent plagiarism, then I am very pro NFTs. Piracy is a big problem for us."*- R1. Another respondent mentioned the transparency that comes through selling on the blockchain instead of a gallery. *"There's 100% transparency, you know what you're getting yourself into. You know what the commissions, and there are no hidden clauses. There are no trade secrets, you can divulge information without having to worry about breaching contract."*- R10.

C. Negative Value Types

1) Price

Proposition 8: The price of minting an NFT inhibits digital artists *from* creating NFTs. The exchange rate from rands (ZAR) to dollars (USD), as well as the exchange rate from rands to the various cryptocurrencies affected the price of minting. "*That's a barrier, the minting process and the minting fees can be very expensive. The exchange rate in dollars can also affect the cost as well".* - R12. A similar sentiment was shared by *another* respondent *"...the exchange rate also needs to be taken into account."*- R4.

To work around the high cost of minting, R13 mentioned that artists would get investors fund their NFT art projects and split the risk and subsequent profit or loss. *"It's insane, it's good that the price tanked, we can all afford Ethereum, Solana and Tezos now. At the peak, minting and NFT would have costed around R4000. It made it less accessible. Artists are going to investors and splitting the risk and profits, which takes things back to the way it was before with freelance contracts"*- R13.

Another respondent mentioned that they work around high minting fees by using blockchains with very low gas fees. *"Theres some blockchains like Solana use very little gas fees. On Ethereum, the gas fees are much higher which can affect the minting cost. The cost of creating the smart contract, is like buying painting materials"* -R7.

Respondent R6, who does not use NFTs, mentioned that the costs were inhibitive for using the technology. *"I was quite shocked when I saw how much it was to upload and even just create a profile for the NFT, so I think that cost is a bit out of out of reach, especially for smaller and newer digital artists, especially since it's in dollars as well, it's quite pricey, which makes it unattainable for some people"- R6.*

2) Time and Effort

Due to *time* and *effort* being similar in the context of the study, the two dimensions were combined into one. *Proposition 9:* The time and effort required to learn how to create NFTs makes it difficult for digital artists to use and gain value from.

The language and jargon take time and effort to learn and is a barrier to using the technology. *"You would need to put in a substantial amount of time to understand it, such as terms and conditions, what is allowed, terms of trade etc", "When I tried to Google it, the language used was not comprehensible was not useful for a layman. If you can't explain it simply, I am less likely to be interested in it. A lot of time, people use convoluted language to exclude people. Like academia, it separates on a class basis."*- R1.

There were a wide variety of responses relating to time and effort it would take to learn the technology. Answers ranged from hours to weeks and to *months*. Some respondents felt that NFTs did not require a lot of effort to understand. One respondent compared understanding NFTs with understanding banks, *"I don't think NFTs are more difficult to understand than banks."* – R7. Another respondent said that social media can give digital artists the perception that it requires a lot of effort to understand the technology and compared it with cryptocurrency. "*I don't think it's hard to understand, just people and social media makes it hard to understand. Its just like crypto."*- R12.

3) Risks

Privacy risk

Proposition 10: Digital artists are worried that their sales and, thus, income are visible to the *public* due to the transparent nature of the blockchain. Proposition 10 was supported by a respondent, saying that they did not like that their hypothetical purchase history would be visible to others. *"If the purchase of the NFT is made public, the transparency is helpful when making a purchase decision, but if other people can see what you have bought, then it isn't nice."*- R1.

Another respondent also mentioned transaction history, but provided a solution to the problem, by suggesting using a secret alias." *If someone has access to your Wallet ID, they were able to access your transaction history. You can use a secret alias to cover your wallet ID, and no one would know who you are."* -R10. Respondent R10 also mentioned that they use an alias so that they can protect their privacy. *"I don't know, the whole point of the blockchain is a public ledger does not result in a loss of privacy, it depends on what you put out. I use a pseudonym"*- R5.

Other digital artists shared a similar sentiment regarding transaction history but did not mention using pseudonyms/ aliases. *"You can track the owners of the art, as everything is on the chain."*- R3. *"People have access to other people's wallet addresses and can see their transaction history. You must give away some information, but you also get the benefit of receiving ease of access."* - R13. *"People can see where my money is going, its public information on a ledger"*- R11.

Security Risk

Proposition 11: Digital artists might be concerned about the security risk of not creating their NFT *correctly* or getting scammed.

Respondent R3 mentioned that there are bugs in the security of the NFT system that can have negative implications. *"Losing information to criminals is not relevant. The hacking issue is relevant. If there's issue in the code then you could maybe steal someone's NFT, maybe steal the money, you could maybe change the code, you could maybe destroy the art- that is one of the bugs that I found where I could basically take someone else's NFT and redeem it for money..."*-R3.

Many of the respondents felt that the security risks attached to NFTs are the same as using any other online platform, service, or product. "*The possibility of getting hacked is there with anything online"*- R11. Respondent R12 also shared a similar view *"I think it's the same risk as using the internet. You can protect yourself in many ways, but a criminal will always try to get around the protection you have."*- R12.

Phishing was another issue brought up by respondents. Respondent R10 mentioned how criminals could potentially access your NFTs, *"Should someone send you a link, you should be cautious about opening that link. It can be very difficult for someone to access your account without your knowledge. There's a 12-word secret phrase to gain access to your wallet in the case you forget. If someone access those words, then they could potentially gain access."* -R10. Respondent R13 mentioned using 2-factor authentication and not using public Wi-Fi as it is unprotected, "*A lot of phishing situations and results hacks, you just need to cover yourself with 2 factor authentication. Don't use public Wi-Fi for personal banking or NFTs, it's not secure" -R7.*

Performance Risk

Proposition 12: Digital artists are worried that the NFTs they create might not get sold or does not behave as intended.

There were a wide variety of responses to the question, with most answers not relating to the proposition or value type. Respondent R8, however, made a statement that related to the question, stating, "*It depends on the NFT you purchase. What I do know is that some NFTs are linked to real world things such as events and groups. The negative thing is that if the NFT does not allow you to do those things then it would be disappointing*". Another respondent felt that the performance of NFTs is tied to people's interests in it, and people manipulating the market.

Financial Risk

Proposition 13: Digital artists are concerned that they will make a loss when creating their NFTs.

Bad *purchasing* or minting decisions were mentioned by a few respondents. Purchasing or minting at the wrong time, will cause a loss if the item is sold at a price lower than what it costed, or even worse if there is no buyer for it. *"It depends if there are costs involved, and if someone doesn't buy it and you would sit there with money you spent on something no wants to buy"*-R1.

Buying on speculation was mentioned by respondent R10, *"When you get into NFTs as a collector, you can lose money if you buy abruptly without making crucial decisions, trying to make quick money. If you're trying to buy and sell through speculation, it can cause a huge loss of money"*-R10.

Exchange rates when converting from rands (ZAR) to dollars (USD) when selling can be expensive. "*I think as South Africans, the rand to dollar conversion is quite a lot. Even if you do manage to sell your NFT's for a reasonable price, but like when you're converting that currency back to the dollars and then back to Rands."*-R6. The same *problem* was mentioned when converting from rands to Bitcoin "*...If you buy it in Bitcoin and the price of Bitcoin drops, then you can make a loss"* -R8

4) Ecological Costs

Proposition 14: Digital artists are concerned about the ecological impact of *blockchain* technology on the environment.

Many of the respondents cited a concern for the ecological impact of blockchain technology on the environment. *"The mining of NFTs. For the transactions to be processed, it gets done through electronic mining rigs, which takes a huge amount of electricity. If that electricity is generated through things like coal, it can have a negative impact on the environment."*- R10. Respondent R13 mentioned using alternate blockchains with less environmental impact: "*It's a big problem, with Bitcoin or Ethereum. Solana and Tezos are alternatives that are not as expensive, they have a lower impact on the environment, and are less of a burden on your wallet."*- R13. Respondent R15 echoed the same sentiments as respondents R10 and R13, *"The power usage it takes to mint NFTs can have a negative impact on the environment. I do think that a power-hungry platform is very damaging. The extra need to generate power will damage the planet "*-R15.

D. Discussion

South African digital artists validated the updated Typology of Consumer Value, confirming the plausibility of the identified value types in their perceptions towards NFTs. The framework encompassed positive and negative values-based perceptions and values-based perceptions related to risk and benefits. Even digital artists who had not directly interacted with NFTs offered valuable insights, considering their roles as creators who can potentially benefit from the technology.

The primary research question aimed to explore the values-based perceptions of South African digital artists towards NFTs. Data analysis revealed ten positive perceptions, including convenience, excellence, status, self-esteem, enjoyment, aesthetics, escapism, control, novelty, and societal benefits, along with three negative perceptions related to price, time and effort, and ecological costs. Time and effort were significant inhibiting factors.

Regarding the perceived values-based benefits of NFT use, the data indicated that digital artists valued societal benefits, as the technology decentralizes art away from exploitative art galleries and intermediaries.

Digital artists identified privacy, security, performance, and financial risks as perceived values-based risks of NFT use. Privacy risk was significant, leading some artists to conceal their real identities or use pseudonyms to maintain separation from their art on the blockchain. While security risk was a concern, it was not uniquely associated with NFTs, resembling risks encountered on other internet platforms.

V. Conclusion

The research explored the values-based perceptions of South African digital artists towards NFTs, analyzing both positive and negative perceptions. Additionally, the study aimed to understand the perceived benefits and risks associated with NFT use among digital artists in South Africa, using the Updated Typology of Consumer Value framework.

The findings contribute to the literature by demonstrating the robustness of the Updated Typology of Consumer Value framework for analyzing values-based perceptions in the context of digital art, shedding light on the reasons for the low adoption of NFT technology among South Africans.

This research holds significance for South African digital artists and other stakeholders in the NFT ecosystem, helping them understand the values-based perceptions of digital artists and aiding in decision-making and policy formulation. It provides a deeper understanding of the views of South African digital artists, identifying opportunities to reduce inhibiting factors and demystify NFT technology for them.

Furthermore, this study addresses a gap in the literature concerning the perceptions of South African digital artists and NFT users. It stands as one of the first papers to apply the updated Typology of Consumer Value framework in this context, validating its utility for analysis.

Despite these contributions, the research has certain limitations. Firstly, the study was confined to South African digital artists, and conducting a similar study in different locations could yield diverse results. Secondly, due to the cross-sectional approach, a longitudinal study could be conducted to observe how values-based perceptions of South African digital artists evolve over time. Thirdly, the sample size was limited to fifteen digital artists, and expanding the sample to include other stakeholders in the NFT or blockchain ecosystem could offer broader insights. Lastly, the study was confined to the updated Typology of Consumer Value framework, and exploring alternative frameworks with different respondent groups may yield different outcomes.

In conclusion, this research makes a significant contribution to understanding the values-based perceptions of South African digital artists towards NFTs. Nonetheless, the outlined limitations provide opportunities for future research to deepen our understanding of the subject matter.

References

[1] M. McConaghy, G. McMullen, G. Parry, T. McConaghy, and D. Holtz-man, "Visibility and digital art: Blockchain as an ownership layer on the Internet." *Strategic Change*, vol. 25, no. 5, pp. 461-470, 2017, doi: 10.1002/jsc.2146. [Online] Available: https://doi.org// 10.1002/jsc.2146

[2] A. Park, J. Kietzmann, L. Pitt, and A. Dabirian, "The Evolution of Non-fungible Tokens: Complexity and Novelty of NFT Use-Cases." IT Professional, vol. 24, no. 1, pp. 9-14, 2022, doi: 10.1109/ MITP.2021.3136055. [Online]. Available: https://doi.org/10.1109/ MITP.2021.3136055

[3] T. Sharma, Z. Zhou, Y. Huang, and Y. Wang, "'It's A Blessing and A Curse': Unpacking Creators' Practices with Non-Fungible Tokens (NFTs) and Their Communities." arXiv, vol. 1, no. 1, pp. 1-20, 2022, doi: 2201.13233. [Online]. Available: https://arxiv.org/pdf/ 2201.13233.pdf.

[4] R. O'Dwyer, "Producing artificial scarcity for digital art on the blockchain and its implications for the cultural industries. Convergence:, 26(4), 874-894. https://doi.org/10.1177/1354856518795097 ." Convergence: The International Journal of Research into New Media Technologies, vol. 26, no. 4, pp. 874-894, 2020, doi: 10.1177/1354856518795097. [Online]. Available: https://doi.org/ 10.1177/1354856518795097

[5] P. A, "Non-Fungible Tokens (NFT)-Innovation Beyond the Craze." 5th International Conference on Innovation in Business, Economics and Marketing Research, vol. 66, pp. 26-30. [Online]. Available: https://www.researchgate.net/profile/Andrei-Dragos-Popescu/publication/353973149_Non-Fungible_Tokens_NFT_-_Innovation_beyond_the_craze/links/611ceede0c2bfa282a514be9/ Non-Fungible-Tokens-NFT-Innovation-beyond-the-craze.pdf.

[6] N. Mofokeng and T. Matima, "Future tourism trends: Utilizing non-fungible tokens to aid wildlife conservation." African Journal of Hospitality, Tourism and Leisure, vol. 7, no. 4, pp. 2-20, 2018. [Online]. Available: http://www.ajhtl.com/uploads/7/1/6/3/7163688/article_21_vol_7_4__ 2018.pdf.

[7] Q. Wang, R. Li, Q. Wang, and S. Chen, "Non-Fungible Token (NFT): Overview, Evaluation, Opportunities and Challenges." arXiv, vol. 1, no. 1, 2021, doi: 2105.07447. [Online]. Available: https://arxiv.org/ pdf/2105.07447.pdf.

[8] L. Bisschoff, "The Future is Digital: An Introduction to African Digital Arts," Critical African Studies, vol. 9, no. 3, pp. 261-267, 2017. https://doi.org/10.1080/21681392.2017.1376506

[9] S. A. Xaba, X. Fang, and S. P. Mthembu, "The Impact of the 4IR Technologies in the Works of Emerging South African Artists," Art and Design Review, vol. 09, no. 01, pp. 58-73, 2021. https://doi.org/10.4236/adr.2021.91005

[10] M. Franceschet, G. Colavizza, T. A. Smith, B. Finucane, M. L. Ostachowski, S. Scalet, J. Perkins, J. Morgan, and S. Hernández, "Crypto Art: A Decentralized View," Leonardo, vol. 54, no. 4, pp. 402-405, 2021. https://doi.org/10.1162/leon_a_02003

[11] A. Abid, S. Cheikhrouhou, S. Kallel and M. Jmaiel, "A Blockchain-Based Self-Sovereign Identity Approach for Inter-Organizational Business Processes," 2022 17th Conference on Computer Science and Intelligence Systems (FedCSIS), Sofia, Bulgaria, 2022, pp. 685-694, doi: 10.15439/2022F194.

[12] S. Bsteh, "From Painting to Pixel: Understanding NFT Artworks."

[13] M. B. Holbrook, "Consumer Value: A Framework for Analysis and Research," Routledge, London, UK, 1999.

[14] R. Sánchez-Fernández and M. Á. Iniesta-Bonillo, "The Concept of Perceived Value: A Systematic Review of the Research," Marketing Theory, vol. 7, no. 4, pp. 427-451, 2007. https://doi.org/10.1177 /1470593107083165

[15] S. Leroi-Werelds, "An Update on Customer Value: State of the Art, Revised Typology, and Research Agenda," Journal of Service Management, ahead-of-print, 2019. https://doi.org/10.1108/JOSM-03-2019-0074

[16] P. Shields and N. Rangarajan, "A Playbook for Research Methods: Integrating Conceptual Frameworks and Project Management.".

[17] A. Bhattacherjee, "Social Science Research: Principles, Methods, and Practices," University of South Florida, 2012.

[18] V. Braun and V. Clarke, "Using Thematic Analysis in Psychology," Qualitative Research in Psychology, vol. 3, pp. 77-101, 2006. https://doi.org/10.1191/1478088706qp063oa

Communication Papers of the 18th Conference on Computer Science and Intelligence Systems pp. 9–20
DOI: 10.15439/2023F7825
ISSN 2300-5963 ACSIS, Vol. 37

Exploring M-Commerce Vendors' Perspectives in Post-Saudi Vision 2030: A Thematic Analysis

Yahya AlQahtani
The Applied College
King Khalid University
dept. of Informatics
University of Sussex
Brighton, United Kingdom
ya227@sussex.ac.uk

Natalia Beloff
dept. of Informatics
University of Sussex
Brighton, United Kingdom
N.Beloff@sussex.ac.uk

Martin White
dept. of Informatics
University of Sussex
Brighton, United Kingdom
M.white@sussex.ac.uk

***Abstract*—Despite the popularity of mobile commerce (m-commerce) services in developing countries, their adoption in Saudi Arabia has been limited. Vision 2030, launched in 2016, has triggered substantial transformations in the country, prompting the need to examine its impact on the adoption of m-commerce. This paper investigates the vendors' perspective in regard to the adoption of m-commerce in Saudi Arabia. Through a thematic analysis of semi-structured interviews with ten Saudi vendors, the study explores the vendors' views on the status of m-commerce in the country and their intentions to adopt it. The findings suggest that m-commerce services are still immature in Saudi Arabia, primarily due to government regulations and technological infrastructure.**

***Index Terms*—E-commerce, Adoption of m-commerce, Vision 2030, Saudi Arabia**

I. Introduction

THE WIDESPREAD availability of Internet services and mobile phones has enabled users to access computational services from anywhere. The mobile market has grown exponentially, with over 8 billion cellular network subscribers in 2022, and this number is expected to reach 9.1 billion by 2027. Service providers are expanding their networks and service platforms to offer not just connectivity, but also services and applications [1].

M-commerce, short for Mobile Commerce, is a technology that is reliant on the Internet and mobile devices. It uses wireless devices such as smartphones to access information and conduct transactions resulting in the exchange of goods or services. Delivery services offer the convenience of goods being delivered to customers, saving them time and effort compared to in-store shopping [2]. It also allows vendors to provide prompt services to customers regardless of their location while reducing the operating costs that come with traditional in-store commerce [3].

Saudi Arabia has been investing in the Internet infrastructure since it was first introduced in the country in 1993 [4]. This provides a fertile environment for investment in new internet-based technologies such as m-commerce. Furthermore, in 2016, the Saudi government unveiled an ambitious economic reform plan, known as 'Vision 2030', aimed at transitioning the economy from an oil-based to a knowledge-based one [5]. Information Technology (IT) has been identified as a key enabler of this transformation, given its ability to create a conducive business environment [6, 7]. The success of e-commerce and m-commerce in other countries, which involve trade processes over electronic platforms and mobile devices, has also demonstrated the potential for significant revenue growth [8, 9]. Thus, IT has rightfully received significant attention as a critical component of the Vision 2030 program.

The use of m-commerce technology in Saudi Arabia has been ongoing for approximately a decade, facilitated by advancements in the online sphere. However, despite the availability of technology, the implementation of m-commerce remains hindered by a variety of obstacles, particularly concerns regarding trust and contentment, as well as insufficient proficiency in IT [10, 11]. Although significant investments have been made to promote the use of m-commerce services, limited success has been achieved in many countries, as evidenced by recent research and industry reports [12, 13, 14].

In the literature, numerous studies have examined the factors that influence the adoption of m-commerce in Saudi Arabia. Nevertheless, little attention has been given to the viewpoints of vendors regarding m-commerce adoption e.g.[15, 16, 17]. In addition, most of those studies have been conducted before the launch of Vision 2030, which makes them fall short to capture its impact on m-commerce adoption. Consequently, this study uses thematic analysis to examine the perspectives of vendors in Saudi Arabia regarding the adoption of m-commerce. Thematic analysis is a useful tool for exploring the beliefs and experiences of vendors who currently use or are interested in using m-commerce for their businesses. The aim of this analysis is threefold: (1) to identify patterns and themes in the data gathered from interviews, revealing deeper meanings and insights that may not be immediately apparent, (2) to provide a detailed description of participants' experiences, gaining a nuanced understanding of their attitudes and practices, and (3) to inform policy and practice by providing perspectives and experiences that policymakers can use to develop policies and environments that support the adoption of m-commerce.

The remainder of this paper is organised as the following. Section 2 reviews the existing literature on the adoption of m-commerce in Saudi Arabia. Section 3 describes the methodology used in this study. Section 4 presents the results and discussion of the thematic analysis of the semi-structured interviews with vendors. Section 5 provides further discussion of the findings. Finally, Section 6 concludes the paper.

II. Related Works

Several studies have investigated the factors that affect the adoption of m-commerce in Saudi Arabia. AlSuwaidan and Mirza [15] investigates the preferences of users in terms of product information display and page navigation in m-commerce mobile applications. The study targets current users of m-commerce with a high focus on female users. The study identifies the preferred options based on the percentage for each option and outlines a prototype of a user interface of a mobile application that features the identified mostly preferred options of product information display and page navigation options. However, the study falls short in addressing the factors that influence the adoption of m-commerce and limits its focus to identifying user preferences.

Algethmi [16] focuses on the airline industry in Saudi Arabia and identifies perceived usefulness, mobility, and compatibility as the main predictors of behavioural intention to use mobile services. The study's main limitation is its focus on the airline industry, which questions the generalisability of its findings.

Turki et al. [17] focuses on the acceptance of mobile ticketing services in Saudi airports and identifies mobility, compatibility, usefulness, and social influence as the main factors affecting adoption attitudes. The study's main limitation is its focus on airline ticketing services, which makes the conclusions limited to that type of m-commerce service.

AbdulMohsin Sulaiman [18] investigates the factors that affect the use of mobile social network services (MSNS) for m-commerce and identifies personal innovation, cost, performance expectancy, and effort expectancy as the main factors affecting the intention to adopt MSNS for m-commerce. However, the study focuses solely on the customer perspective and does not investigate the vendor perspective.

Makki and Chang [19, 20] investigates the impact of mobile usage and social media penetration on the use of e-commerce in Saudi Arabia. The study finds that Saudis, especially females, spend significant time on social media and mobile phones, which provides considerable potential for companies that adopt e-commerce to widen their business by reaching potential customers through social media. The study, however, limits its focus to social media for m-commerce and provides limited information on the framework that was used for defining the factors, which limits its replicability.

Al-Hadban et al. [21] conducted a review of a number of studies that investigated the factors that affect the adoption of m-commerce. The study identifies a list of factors that can be investigated in further research including usefulness, ease of use, and trust, among others. However, the study does not collect data to statistically investigate the influence of those factors on the adoption of m-commerce.

Alkhunaizan and Love [22] investigates the factors that influence Saudis' intention to use m-commerce and identifies perceived usefulness as the prime factor affecting the intention to use m-commerce. The study's main limitation is that it does not investigate other potential factors that could influence the adoption and use of m-commerce. Additionally, the study reveals two surprising results, namely that trust has no impact on the adoption of m-commerce, and that Saudis find m-commerce services more expensive than traditional services.

Overall, these studies provide valuable insights into the factors that influence the adoption of Saudi m-commerce. However, they are not comprehensive enough due to the following limitations. Firstly, the scope of the existing research is limited as most of the works were published prior to 2016, before launching vision 2030. Therefore, many of these studies fail to account for the impact of the vision on m-commerce. This is problematic because as changes emerge due to the implementation of Vision 2030, outdated research may become unreliable and inaccurate, leading to unreliable conclusions. Secondly, most of the studies on m-commerce have focused on specific groups, such as travellers, and social media users. This makes them fall short to provide a comprehensive view of the adoption of m-commerce in general regardless of the business sector it is used in. Thirdly, the above studies mainly investigated customers' perspectives in regard to the adoption of m-commerce in Saudi Arabia. As a result, there is little attention paid to providers' perspectives on the adoption of m-commerce. While these limitations are not necessarily flaws, they do raise concerns about the adequacy of research on m-commerce in general, and the reliability of the conclusions drawn from these studies. Therefore, there is still a need for further research that addresses these limitations.

III. Methodolody

This study conducted semi-structured interviews as the primary data collection method to explore vendor perspectives on m-commerce adoption in Saudi Arabia [23]. Ten firms from different cities, including Riyadh, Jeddah, AlKharj, and Abha, participated in the interviews via Skype during June/July 2020. Given that Arabic is the native language in Saudi Arabia, the interview questions were translated into Arabic, and responses were later translated back into English to adhere to institutional standards.

Each participant received written information about the study's purpose and provided informed consent. The interviews, which lasted an average of 120 minutes, were divided into two parts. In the first part, basic information about the companies, such as expertise, size, number of branches, employees, type of business, and average number of customers, was collected to gain an understanding of their context and financial status. The second part of the interviews involved posing nineteen questions to gather insights into the companies' perspectives on various aspects related to m-commerce adoption. These aspects focused on factors that could influence both user and vendor decisions. To analyze the collected data and present the study's findings, an inductive approach was employed, specifically using the thematic analysis method. This method allowed for the identification of patterns and themes, which will be elaborated on in subsequent sections of the research paper. The study aims to shed light on the viewpoints of vendors regarding m-commerce adoption, offering valuable insights into the factors that may drive or hinder the widespread acceptance of m-commerce in Saudi Arabia.

IV. Overview of Data analysis

As shown in Table I, ten representatives of various companies from various private sectors were interviewed, including those operating in telecommunications services, the retail sector, the wholesale sector, the education sector, and the logistics sector.

To gain insights into m-commerce perspectives, it's vital to explore views from different managerial levels, including corporate headquarters, sales, advertising, marketing, and IT. These representatives handle customer issues related to mobile purchases, providing valuable knowledge. All interviewed participants had worked at their respective companies for at least one year, demonstrating a strong familiarity with customer transaction concerns. By interviewing a diverse group of representatives, a comprehensive assessment of m-commerce adoption in Saudi Arabia, from vendors' viewpoint, was achieved.

A. Type of Company

The selected companies were of different types of business. However, all have the potential for using m-commerce. This potential is guaranteed based on our observation of similar types of businesses that exist in other countries and use m-commerce [24]. Table I lists the sector types of each of the interviewed participants.

B. Company Size Categorisation

The size of the companies was categorised into three groups based on the number of employees, branches, and sales volume: Small, Medium, and Large. In addition to gathering information on the average age of clients, frequency of purchases, and having a website. Surprisingly, the findings indicated that just six of the firms have websites, and four do not, although they have a high number of customers (for instance, **AKD** has over 100,000 customers).

C. Knowledge and Qualifications

The majority of the interviewees confirmed that knowledge about m-commerce and its enabling technologies is a crucial factor that stimulates the adoption of m-commerce. This applies to both customers and providers, in their opinion. However, a minority of the interviewees hold a degree/diploma that is relevant to m-commerce or information technologies. Specifically, three out of the ten interviewees (**SOA, YAS and JFP**) mentioned that they hold a BSc or an MSc in information technologies.

It is observed here that the companies of those three interviewees leverage kind of electronic means for their business. However, that cannot be considered an m-commerce-based business. They use the internet mainly for advertising and marketing rather than for selling goods. On the other hand, it has been observed that participants who don't have sufficient knowledge about technology do not use it, even though there is room for it to improve the business. In this context, one business owner **(AKD)** commented: *"Let's be straightforward: m-commerce is widely misunderstood inside Saudi society. The majority of wholesale providers [of dates] do not understand the true meaning of the word electronic commerce and, as a result, avoid using it. As with any technology, ignorance breeds fear, which breeds reluctance to utilise the technology."*

The interviewee **(RAS)** explained, *"I believe the lack of knowledge and lack of confidence are connected. When there is a lack of understanding of technology, faith in that technology decreases. "* These findings are consistent with the findings of a customer's perspective study [25] which confirmed that higher levels of IT skills, education, and technology awareness will lead to an increase in citizens' intention to adopt m-commerce because they are more likely to be more accustomed to IT technologies in general.

D. Age

The study explored the Age factor by considering both the average age of the company employees and the average age of the company clients to investigate any correlations with the intention to adopt m-commerce. Out of the ten companies interviewed, seven had an average employee age between 25 and 35, while three had an average age above 45 years. Companies adopting electronic means for their business were predominantly found to have an average employee age between 25 and 35, whereas those not leveraging m-commerce had a majority of employees above 45. Interviewees attributed this trend to a lack of expertise and resistance to change among older employees.

(RCF) commented: *"We provide a Loyalty Card, in order to get a cup of coffee for free, the client needs to collect 10 stamps, most of those who benefit from it are those under 30."*

In this context, it can be concluded that the adoption of m-commerce from the perspective of the business providers relates to the average age of the business staff. It can be observed that the younger the age to more m-commerce-related services are provided.

E. Business Conducted on The Internet

The interviewed companies primarily used electronic means for marketing their products. Some companies, such as TCR, YAS, SOA, RCF, and AEP, utilized social media and websites to advertise and engage with customers.

(TCR) reported that *"70% of our customers are on the Internet although we don't have official sales on the site.."* **(YAS)** and **(SOA)** reported that they rely highly on electronic means to reach their customers for more than 90% of their business. A spokesman for **(YAS)** remarked, *"Our job requires us to maintain regular contact with academic institutions and universities abroad, We do not offer."* The representative of

TABLE I
INTERVIEWEE CODES FOR REPRESENTATIVES OF COMPANIES

Code of Company	Company size	Branches	Employees	Sector type	Position
(SOA)	Small	1	3	Beauty Products Retail	IT Manager
(AEP)	Small	2	8	Perfumes and Incenses Retail	Director
(JFP)	Small	1	2	New and Used Mobiles Retail	IT Manager
(MAC)	Small	1	3	Women's Accessories Retail	Marketing Manager
(RAS)	Medium	6	50	Primary and Secondary schools	Director
(YAS)	Medium	3	45	Study Abroad Consultant	Marketing Manager
(AKD)	Medium	3	30	Dates Retail	Owner
(RCF)	Large	12	550	Coffee Shops	Marketing Manager
(TCR)	Large	5	180	Cars Retail	Sales Manager
(BPR)	Large	22	330	Pizza and Pasta Restaurant	Director

(RCF) informed that *"We rely on e-commerce by 30% because we present offers and competitions on the internet"*

On the other side, we see that **(MAC)**, for example, founded his company in 2014, which is before **(YAS)** and **(SOA)**, but he has no dependency on m-commerce. A possible explanation is the lack of understanding of m-commerce technologies and the limited adoption of m-commerce (and electronic services in general) by their competitors in the surrounding area. Another possibility is the tendency of avoiding costs that may be associated with deploying m-commerce services. In this context, **MAC** representative informed that *"The profit margin in our sector is low, adding online services will increase the cost, hence increasing the final price for the consumer, and the majority of our customers are seeking the lowest price."*

F. Type of Delivery Services

Although the Saudi market offers four types of delivery services, including *Third-party*, *Proprietary* services, *Customer collection* collection, and *National postal system*, it is noteworthy that the interviewees only mentioned three of these options. Surprisingly, the national postal system was not mentioned, likely due to concerns regarding its unreliability and insufficient performance.

1) Third-party delivery services: There are companies that provide delivery services in the country. They can be either domestic companies such as MRSOOL, HungerStation, and Jahez or international services such as DHL. Domestic services are considered to be popular in Saudi Arabia because they are relatively efficient and successful. Additionally, their prices are reasonable and do not exceed 50 Riyals. In some areas, customers can expect delivery within a maximum of three hours. While these services are considered a significant advancement in the realm of e-commerce in Saudi Arabia, they are limited by the requirement that both the buyer and seller must be situated in the same city, particularly in major metropolitan areas such as Riyadh or Jeddah. This restriction presents a drawback that may limit the widespread adoption of these services in other areas of the country.

2) Proprietary Delivery Services: In Saudi Arabia, large enterprises offer a delivery service in which they handle the delivery of their products directly to customers, instead of relying on third-party companies. However, this type of service is not typically provided by medium and small businesses. This is mainly due to the feasibility issue, as the cost of setting up and maintaining a delivery service is often considered prohibitively expensive for smaller companies. For example,**AEP** conveyed to us that they *don't offer this service, but it is available in major enterprises like Arabian Oud. Having a large number of consumers using the service reduces individual costs and increases revenue. However, if the number of consumers decreases, prices may rise."*

3) Customer collection: An alternative method of product delivery is by enabling customers to collect their purchases directly from the point of sale or a designated pickup location. The majority of interviewees reported that in-person collection is the dominant way of delivering products to customers. The commonly reported reason is the unreliability of the delivery services that operated in the country, especially in rural areas.

Some participants, such as (JFP), (TCR), (AEP), and (SOA), consider dependable delivery services as a significant challenge. Past efforts to improve service quality have been met with pessimism, leading to a sense of impossibility. For instance, **AEP** mentioned that *"they provide delivery services only within Riyadh, while for areas outside Riyadh, they rely on Zajil company, which often delivers late and incurs a cost of 75 Riyals for the customer"* In their questionnaire responses, the interviewees highlighted various difficulties with delivery services, including the following reasons:

- **There is no reliable infrastructure that would enable delivery services.** For example, there is no reliable house numbering system in Saudi Arabia. The current system is inefficient as it gives very long numbers (15 digits) that are not shown on houses. The system is not applied to all houses (especially in small cities and rural areas) and it is not recognised by Internet location services such as Google Maps.
- **The efficiency, reliability, and speed of the National postal system are lacking, resulting in the possibility of parcels being lost or significantly delayed.** For example, the IT Manager of **(JFP)** said: *"As a government sector, the workers of the National Saudi Post know that their employment will not be impacted whether the service they provide is poor or excellent..."*. Also, the Director of **(BPR)** confirmed what had been mentioned before. As he informed, *"How can we deal with Saudi Post for our company while we avoid utilising it for personal*

purposes? The Saudi Post's charges are expensive, given the unfair level of service."

- **There are some private delivery companies that can be more reliable.** These include international companies such as DHL and national private companies such as Mrsool and Jahez. The international companies cover more areas but they are more expensive and have longer delivery times. The national private companies are faster but they are still expensive and do not cover small cities or rural areas. The **SOA** representative provided their thoughts on this. *"Sometimes we advertise deals, and those deals will only be valid in the city of Riyadh. . . . the difficulty lies with clients who reside in other cities."*
- **Unwanted products cannot be returned by customers,** as the return postage fees can be costlier than the product price.
- **Providing a delivery service by the company itself,** is not feasible as it increases staff costs and is subject to complicated governmental requirements.

V. Thematic data analysis

In this section, we utilised thematic analysis to examine the interview data, aiming to identify recurring themes that reflect Saudi firms' perspectives on m-commerce adoption [26]. Through this method, we identified 15 distinct themes representing patterns and significant insights within the data. These themes offer valuable insights into the attitudes and beliefs of Saudi firms towards m-commerce, shedding light on the challenges and opportunities related to its implementation. The thematic analysis method involved the following main steps [27]:

i. Familiarisation: Researchers explored the data to gain a preliminary understanding and form initial ideas about its description.

ii. Generating Initial Codes: Codes were created to describe interesting aspects of the data. This process organised the collected data into groups.

iii. Searching for Themes: The codes were transformed into themes that characterised the findings. This step required active interpretation of the data, and the process was iterative, refining the themes until a final list was identified.

A. Clients Do Not Trust M-commerce

This theme focuses on aspects related to the concerns of vendors regarding the customers' trust in m-commerce. The majority of the interviewees conveyed that their customers do not really trust m-commerce. For example, the representative of **MAC** shard with us that *"my clients' trust in m-commerce is almost non-existent, I can say some people in the age between 25 and 30 do trust m-commerce but generally customers do not."* Another example is **JFP** who responded that *"To a big extent, there is no trust in m-commerce. The percentage cannot exceed 40% as there is no party to protect them."*
The reasons for the lack of trust in m-commerce –from the perspective of vendors– can be summarised as follows:

- **Level of Education.** Some views relate trust to the level of education. Vendors think that educated people have more trust in m-commerce and similar technologies. For example, **AEP** mentioned that *"education is essential for trust and it is also prestige to understand [m-commerce]"* and **MAC** mentions that some customers trust m-commerce *"because they are educated"*.
- **Lack of Protection.** Some views relate trust to the protection provided to vendors and customers. They think that there is a lack of protection for vendors and customer rights in Saudi Arabia, e.g. if the product has been lost while in transit. This causes them to prefer in-store shopping. For example, **JFP** mentioned that for payment-on-delivery sales *"..there is no party to protect users, for example, delivery companies deliver to a named person, and if they did not deliver I lose the product, and its price"*.
- **Unreliability of Delivery Services.** As mentioned above, some interviewees mention that customers find the delivery services in the country to be unreliable. That means to them that the products they purchase may get lost during delivery, which makes them not trust the service. For example, when asked if they think customers trust m-commerce, **AKD** mentioned that *"I don't think so . . . I personally made online orders that did not arrive"*.
- **Payment Service.** The type of payment service affects customers' trust in m-commerce, as inferred from some views. Interestingly, customers would trust payment-on-delivery services rather than online or over-the-phone payment services. This made **RCF** think that the *"the best way to promote m-commerce is to provide payment-on-delivery services"*.

On the other hand, few opinions find that customers trust (at least partially) m-commerce especially if 'excellent' services are provided by vendors and when vendors have a good reputation. The latter made **BPR** think that *"advertising through famous people promotes customers trust"*.

B. M-commerce Needs Higher Expertise

This particular theme centers around the concerns expressed by vendors regarding the level of expertise required to successfully adopt m-commerce. Generally speaking, this analysis found that the questioned vendors believe that they do not have sufficient expertise for providing m-commerce services. **AEP** reported that lack of expertise *"adversely affects us as following up with customers on social media and checking bank transfers are time-consuming ..."*. Similarly, **RCF** reported that they need staff who are specialised in e-commerce to manage and promote web content and enable the vendor website to be listed first in search engine search results.

Few vendors think that they have sufficient expertise in m-commerce technologies, though they are using it only for marketing. More interestingly, one of the vendors, specifically **RAC**, emphasises the importance of knowledge and experience for m-commerce and mentioned that **RAC** promote their staff technical skills through periodic training courses.

C. M-commerce Needs Better Infrastructure

The focus of this theme is on the availability of enabling technologies of m-commerce and how they shape the vendors' perspective towards m-commerce and their intention to use it.

According to most vendors, the primary technology that facilitates m-commerce is Internet speed. Fast Internet connectivity allows customers to conveniently browse through product websites. While the majority of interviewees indicated that the speed of the Internet is satisfactory in major cities, it is considered unreliable in other parts of the country. For example, **RAC** indicated that *"The internet is so good in the Saudi cities and so bad in rural services"*.

In addition to the first point, vendors have emphasized the inefficiency of transportation services, resulting in a prolonged delivery process for their products [28]. In this regard, **SOA** commented that *"... the transport services are inefficient."* A possible explanation for this perceived inefficiency may be the lack of adequate railway services to transport goods. However, the special attention that is being given by the Saudi government to the railway sector in the country – through increasing funding to expand the railway network – promises improvements in goods transport [29], which would impact the adoption of m-commerce.

Vendors are worried about cyber security as they believe that the internet is not a safe place. They are concerned that their websites and systems could be targeted by cyber attackers, resulting in substantial financial losses. **JFP** commented that *"...there are massive and continuous cyber attacks, and no one would protect a business or reimburse our loss."*

In summary, infrastructure availability significantly influences vendor adoption of m-commerce. Challenges such as unreliable internet in rural areas, inefficient transportation, and cybersecurity concerns affect vendors' perspectives. However, the government's investment in improving transportation and cybersecurity measures may address these challenges. Vendors need to stay informed and adapt to evolving technology to remain competitive.

D. Popular E-commerce Platforms are Better

When asked about whether they prefer to have their own m-commerce application or to use third-party applications (e.g. Amazon and eBay), the majority of the interviewees indicated that they do not currently use an m-commerce application but they would prefer the latter. The vendors use either social media pages or basic websites for marketing their products, as mentioned above. However, the reasons behind their preference towards third-party applications are as follows (as reported by the interviewees):

- **Cyber Security.** Vendors have the impression that third-party applications are more secure than their proprietary applications. This might reflect a lack of trust in the developers' expertise in developing applications for the vendors' online business. For example, the representative of **SOA** reported that *"using famous third-party applications is more convenient and secure... ."*
- **Production and Maintenance Cost.** Vendors reported the cost of developing a proprietary application is high as that would include implementation, deployment, maintenance, evolution, and security costs. An example is the response of SOA who continued *"...We'll not need to hire technicians for the application development, especially in the presence of excellent third-party applications that don't need high skills to manage."*
- **Popularity.** Many third-party e-commerce platforms, such as Amazon and E-bay, enable vendors to create online shops for m-commerce in an easy way that does not need a high level of experience. More importantly, these platforms are considerably popular. Almost all m-commerce and e-commerce users are familiar with these platforms. Opening an e-shop on such a popular platform would increase the reachability of business to customers. This advantage motivates vendors to prefer third-party applications over proprietary ones as the latter need time and expenses to popularise. In this sense, the representative of **MAC** said *"I prefer third-party applications over proprietary ones as people will not know the latter."*

However, though the above reason potentially explains the vendors' preference for using third-party applications, some of the mentioned reasons apply to both third-party and proprietary applications e.g. popularity. Though m-commerce users are indeed familiar with Amazon and E-bay, they will not be familiar with the online shop of a specific vendor unless the shop becomes popular. This is associated with advertising costs to make the e-shop popular, which is similar to what they try to avoid with proprietary applications.

On another note, one interviewee **RCF** mentioned that they have their own website and they are not keen on using third-party applications. The main reason for that is the control they wanted over the style of the applications and the features it supports, which enables them to be different from other vendors' websites or applications. They mentioned that *"We've got the application that was fully designed by us as we do not want to be a copy of other companies."*

E. Native Language Support

The support of the Arabic language in m-commerce platforms is no doubt crucial for m-commerce in Saudi Arabia. The reason is obviously that Arabic is the native and official language in the country. This support is necessary for many reasons. Firstly, it makes it easier for the main targeted customers – who are native Arabic speakers – to understand the product specifications and sales policies. This would help them find m-commerce applications easy and convenient to use. It would also make them trust the m-commerce application as they understand its contents. Secondly, Arabs prefer applications that have interfaces built in the Arabic language [30]. Easiness, as mentioned above, could be behind this attitude, however, cultural reasons might also exist. Thirdly, there is a legal requirement from the Saudi Ministry of Commerce and Investment that mandates shop owners use Arabic in all of their business operations, including customer invoices and restaurant menus [31]. This imposes requirements on vendors to use platforms that support the Arabic language, not only for user interfaces but also for document generation, e.g. invoices.

In this context, all of the interviewees agreed on the importance of supporting the Arabic language. The **MAC** marketing manager conveyed that *"m-commerce is simple for all customers if Arabic language support is provided though it is crucial to have a global language such as English accessible to other [non-Arab] clients.* Also, **(YAS)** commented that *"the Arabic Language is of utmost importance to have as it is the official language. Our website's support for Arabic was a major factor in attracting new customers.* In conclusion, the interviewees are linking cultural factors and Arabic language support to the ease of use and trust of m-commerce, which contributes to the intention to use m-commerce.

F. High-level Product Specification

An interesting observation from the interviewees' responses reveals their lack of interest in providing detailed specifications about their products on their websites or social media pages. They rely on the specifications that are provided by the producers, i.e. that are written on the product itself. Therefore, usually, m-commerce users would find high-level descriptions of the products rather than detailed specifications. This high-level description could include phrases like *"approved by the government"* in order to increase customers' trust in the product. In this context, the owner of the perfume shop **(AEP)** believes that *"the website should offer broad specifications not detailed, notably the name of the product should be related to what it is manufactured from. Orchid perfume, a perfume derived from the orchid flower, is a good example of this. This is a major factor in determining whether or not to purchase.* When it comes to technological goods, **(JFP)** claims that he doesn't think it's necessary to mention their specifications since *"...it is made by well-known companies like Apple or because he can find them on several websites. Displaying the equipment and colour choices accessible to customers, as well as offering promotions and discounts, is more important."*

These findings contradict the user expectations as reported by relevant research studies [32, 33, 34]. They reported that detailed product specifications and images are critical factors that customers consider when making a purchase decision online. The specifications provide customers with detailed information about a product, including its features, functions, and technical specifications. This information helps customers make informed decisions about a product's suitability for their needs, reducing the likelihood of returns and increasing customer satisfaction. This could be a reason that explains the limited interest in m-commerce for many businesses in the country, which is an interesting finding that would draw the vendor's attention to the importance of publishing product specifications online.

G. Lack of Data Protection

Business consumers, e.g. customer using m-commerce, provide vendors with their personal data each time they purchase a product or access a service. The provided personal data include credit card details, names, contact numbers and addresses, among others. Customers provide these details for the sake of completing the purchase process and they expect vendors will preserve their privacy and keep their data secure and inaccessible by unauthorised users including other vendors. This is indeed a crucial requirement that is increasingly necessary for the development of m-commerce services [35].

According to the interviewees' feedback, vendors were found to handle consumers' data in an undesirable manner. They mentioned multiple instances where customer data was not adequately protected and was either shared with other vendors or traded as part of business deals. This indicates that the privacy of customer data is not given due consideration by these vendors. For example, in response to the data protection question the representative of **SOA** said *"we exchange customer data with other companies, though this is wrong."* The representative of **AEP** responded *"I do not know if there is a data protection policy in the country, we pass our customers' data to other businesses if they agreed to do the same."* Also, the representative of **JFP** said *"There is no privacy policy implemented. In fact, you can buy a package of numbers from telecommunication companies. You can also specify the ages and locations of the customers."*

To sum up, the results from the investigation on data protection demonstrate that vendors violate data protection regulations, and there is inadequate enforcement of those regulations in the country. These findings align with the outcomes of the customer-focused survey, indicating that Saudi m-commerce users lack trust in m-commerce vendors' ability to protect their personal information from misuse [25]. Therefore, it can be inferred that the misuse of data by vendors is the underlying reason for the relatively low level of trust in m-commerce in the country.

H. Lack of Protective Regulations

Most interviewees share the belief that the Saudi government enforces strict limitations on vendors only, while refraining from imposing any restrictions on consumers. thereby hindering the widespread adoption of m-commerce. According to their perspective, the regulatory environment overwhelmingly favors customers over vendors, offering little protection to the latter in the event of customer misconduct. To support this finding, I quote the response of the representative of **(AKD)**: *"Although the regulations are generally effective, they tend to favor customers over vendors. As a result, when customers make unfounded complaints, vendors may face reputational damage without any recourse to defend themselves."*

Government regulations have come under criticism for their perceived role in causing instability that makes it challenging for vendors to devise clear strategies, such as adopting new technologies. One significant factor contributing to this instability is Saudization, the Saudi nationalization scheme, which requires Saudi companies to employ at least 30% Saudi nationals in their workforce. However, research has shown that many Saudi nationals prefer working in governmental institutions, leading to high rates of turnover in the private sector [36]. As a result, private companies are often forced to hire employees who may leave at the first opportunity to work in the public sector, leading to a high level of labor turnover and difficulty in establishing a stable workforce. This situation has been clarified by the representative of **RAS** in his response: *"Despite the fact that we could recruit a teacher with better credentials and cheaper wages from other Arab nations, the government has forced us to employ Saudi instructors. In many cases, a teacher quits after a year of service because they have been hired by the government, . . . We've suffered great damage."*

This finding could support previous findings in the literature. For example, as reported in [37], the regulations set by the Ministry of Commerce and Industry in Saudi Arabia are not clear in the way they protect customers against vendors and local against international business.

Overall, the above suggests that government policies in Saudi Arabia may be contributing to challenges in the adoption of m-commerce services for the country's business landscape.

I. Expensive Governmental Fees

The significant drop in oil prices in the last few years has had a profound impact on the Saudi Arabian economy. As a result, the government has been forced to take measures to diversify its revenue sources and reduce its reliance on oil exports. One of these measures has been to impose new taxes and fines on products and services in the country, in an effort to boost government revenues and reduce the budget deficit [38].

However, these new taxes and fines have had a negative impact on the adoption of m-commerce in the country, as noted by the interviewees. The additional costs imposed by the government have made it more expensive for vendors to offer their products and services through m-commerce platforms. As a result, many vendors have been hesitant to invest in m-commerce, as the additional costs associated with the new taxes and fines have eroded their profit margins. This impact can be summarised into the following.

- **The adoption of m-commerce services requires vendors to register a national address for their companies,** the fees for this registration are annual and are expensive as perceived by the vendors. For example, the representative of **AEP** informed that *"We are required to pay SAR1000 annually just for a national address though they do not provide any service, it is just a national code."*.
- **The government imposes expensive custom duty on imported goods,** which in the end raises costs on customers. The custom duty applies on all goods even retail sales. This affects online trading even giant online sellers such as Amazon. The increase in product costs does not give m-commerce an advantage compared to in-store shopping, especially since Saudi Arabia is not an industrial country, meaning that most of the products are imported. It makes the expectation that online shopping is cheaper than in-store – as vendors will not need much of the costs for running physical stores– not valid. In this context, the representative of **AEP** confirmed that *"... Amazon prices are higher in Saudi Arabia than other countries due to the high customs duty..."*.
- **Vendors are required to bear the costs of undelivered items**, according to the regulations set by the Ministry of Commerce. This means that if a vendor dispatches an item to a delivery company for delivery to the customer and the item is lost, the vendor must refund the customer. This regulation has been a source of frustration for many vendors, as it places the burden of responsibility for lost items on them, even if the item is lost during the delivery process. Furthermore, when a customer raises a complaint about a product, the Ministry of Commerce tends to side with the customer. This often results in vendors having to refund the customer, even if the product was received in good condition. Vendors view this as a type of penalty that is imposed on them if the customer is dissatisfied with the product, even if the issue was not caused by the vendor. For example, the representative of **YAS** pointed out that *"the regulations are very tough and fines are high. The Ministry of Commerce always supports the customer and never cares about vendors."*

In summary, it can be concluded from the above that the governmental regulations that relate to fees and taxes impose challenges on vendors and do not encourage them to adopt m-commerce in Saudi Arabia.

J. Easier to Target Female Customers

The culture of Saudi Arabia exhibits distinct features, such as a strong emphasis on privacy, particularly when it comes to women, which sets it apart from other cultures. Though the country is gradually changing and becoming more open in the light of Vision 2030, the culture is still conservative towards women. Compared to Saudi men, Saudi women still do not conduct substantial work outside the house. Even though Saudi Arabia's government legalised women to drive in June 2018, the country's culture is still putting many lifestyle limits on Saudi women [39]. This includes, for instance, that Saudi women are not encouraged to work. Figures showed that the unemployment rate for Saudi women was about 24% in 2020 compared to only 7% for men [40].

Given these cultural norms, it can be inferred that Saudi women tend to spend a significant amount of time at home. Therefore, it is imperative for merchants to utilise e-commerce and m-commerce technologies in order to effectively reach them as potential customers. In this regard, **MAC** representative confirms that social media (e.g. Instagram) is one way to reach women customers. They said, *"As they have difficulties visiting shops compared to men, this forces women to primarily focus their purchase choices on photos displayed on social media. They always contact us through WhatsApp to make orders. That's a great deal.*

MAC representative informed also that female customers have their specific style of shopping that can be dealt with through social media. They notice that female customers find it prestigious to use technology for shopping. They care more about quality, presentation, and speed of access rather than price. These can be managed through high-quality professionally taken photos of the products.

In conclusion, the findings unequivocally establish that culture plays a significant role in the adoption of m-commerce in Saudi Arabia, as demonstrated by the perspectives of both vendors and customers

K. It Can Be Useful

The focus of this theme is on the general attitude towards and the perceived usefulness of m-commerce from the vendors' perspective. The majority of vendors pointed out that in general m-commerce can be useful despite the mentioned challenges. Many of them mentioned that they have plans to increase their adoption of m-commerce, however, in various forms.

The obvious trend currently is the use of Internet and mobile technologies for marketing, especially through social media where – as mentioned above – photos of products are displayed on social media pages. However, it can be noticed that even with this attitude towards electronic marketing (as a form of e- or m-commerce), limited utilisation of mobile marketing strategies is adopted. This is evidenced by the observation that though strategies for personalised targeting exist in the literature of mobile marketing, they are not in use, at least by the interviewed vendors. None of them mentioned that they use social media adverts or Google Ads to target customers. These adverts can be very useful as they utilise contextual information – that is provided by the customer's mobile – to deliver to the customer content and products that 'best' matches their interests. This contextual information includes location, time, surrounding environment, shopping companion, and market competition. Research shows that utilising this information for personalised targeting of customers results in a significant increase in sales [41]. Perhaps the little awareness of vendors about these advantages would provide an explanation for their limited adoption of them.

Despite that, vendors find that m-commerce through social media pages help them to popularise their stores and products. In this regard, **BPR** informed that *"... electronic marketing helped us to double our sales,"* while **RCF** revealed that they have plans to increase their adoption on m-commerce by 70% and maintain small shops as contact points for online customers.

Another reason for finding m-commerce useful is the cheap running costs as compared to in-store commerce. The representative of **TCR** informed that they have plans to create an online marketing store for their car sales as that is much cheaper than opening and maintaining car showrooms.

In summary, vendors have a positive attitude towards m-commerce despite facing challenges. Vendors find that m-commerce through social media pages helps them to popularise their stores and products. Additionally, some vendors plan to create online marketing stores for their businesses to increase their adoption of m-commerce.

L. It Makes Competition Harder

Although research suggests that m- and e-commerce can be beneficial for vendors, there is a contrary perspective that applies to small businesses. While these technologies make it easier for vendors to reach their customers and increase their sales, some interviewees have expressed concern that m-commerce will make it harder for small companies to compete, as larger companies will be able to reach wider ranges of customers, even crossing borders. This means that local vendors may struggle to reach customers in the same way that larger companies like Amazon and eBay can. As a result, the representative of **MAC** stated that *"It will be a significant loss if small businesses move to online commerce. They will not be able to compete with big companies, and they should instead rely on customers who prefer in-store shopping."*

This concern is not unfounded, as research has shown that small businesses face significant challenges in competing with larger, more established companies in the online marketplace. Limited resources, lack of brand recognition, and limited access to financing are among the most common challenges that small businesses face [42]. In addition, larger companies can use their resources to offer discounts, promotions, and free shipping to customers, which can make it harder for small businesses to compete on price.

Despite these challenges, there are still opportunities for small businesses to thrive in the online marketplace. For example, small businesses can use social media platforms to engage with customers and build their brand presence, which can help them to differentiate themselves from larger competitors. Additionally, small businesses can use niche marketing strategies to target specific customer segments and provide a more personalized shopping experience. By leveraging their unique strengths and capabilities, small businesses can carve out a profitable niche in the online marketplace.

M. M-commerce is Coming

The topic at hand pertains to the vendors' anticipation of the extent to which m-commerce will be adopted in Saudi Arabia in the future. The majority of the interviewees think that m-commerce will be widely used in the country, despite the current challenges. The majority acknowledges that the Internet and mobile technologies are playing a crucial role in people's daily life. Thus utilising the services of these technologies for trading in the country is inevitable. As people spend more time on their mobile devices, it will be feasible for vendors to utilise those technologies to attract and interact with customers. This will help vendors make anywhere and anytime sales. Location and time will no longer put constraints on people to carry out their shopping.

According to the majority of vendors interviewed, m-commerce is gradually gaining ground in Saudi Arabia, despite encountering challenges and progressing slowly. They observed an increasing number of businesses adopting some form of m-commerce service. In this context, the representative of **SOA** pointed out that the *"It [m-commerce] has been real in the country and we have to consider and deal with it to* *be successful."* Also, the representative of **YAS** informed that *"The market is becoming electronic to a big extent. Many companies closed their stores and moved to the Internet."*

In summary, it is possible to conclude that Saudi vendors think m-commerce will be widely used in Saudi Arabia. Even though they are not satisfied with m-commerce in its current state due to the mentioned challenges, they think that they have to deal with it as customers spend a considerable amount of time on their mobile devices and in the Internet world.

VI. Discussion

This section offers an in-depth discussion of the results obtained from our study, highlighting key insights and observations.

A. Delivery services

The interviewees' comments revealed two other important observations. Firstly, customers have exceptionally high expectations when it comes to delivery speed. Based on the feedback, it seems that a three-day delivery period is perceived as too long by many customers. This expectation is particularly noteworthy in light of Japan's impressive delivery services, exemplified by Yamato Transport. For instance, Yamato's innovative TA-Q-BIN delivery technology allows for ultra-fast deliveries, including within a few hours, using automated delivery lockers, and even via drone [43]. Yamato's success in this area is a testament to the importance of meeting customers' expectations and providing exceptional service.

Secondly, vendors may be considering delivery fees as an extra cost to customers that may adversely affect customers' interest in their products. However, customers may accept this extra cost in return for the delivery service. As in the US, e-commerce enterprises may profit from delivery services without charging clients. Premium subscriptions with free delivery are one way. Amazon Prime's yearly membership includes free one-day delivery. Hence, clients get free delivery while the corporation generates profit from subscription fees [44]. The aforementioned observations offer valuable insights for merchants who wish to improve their m-commerce services and provide a satisfactory customer experience.

B. Perspective on Competition

Section V-L introduces a theme that sheds light on how certain vendors perceive the role of m-commerce in the marketplace in regard to competition. While some believe that m-commerce may pose a threat to competition, another perspective suggests otherwise, as it can help reach a wider customer base. Nevertheless, some research indicates the presence of negative views on the subject. Research studies such as [45, 46] found that m-commerce has the potential to increase market share for businesses by making it easier to reach more customers. This is because the increasing popularity of smartphones and tablets will result in more people using mobile devices to browse the internet, shop online, and make purchases. This will allow businesses to reach customers on their mobile devices. In addition, m-commerce can also provide businesses with valuable data insights about their customers' shopping behaviour and preferences, allowing them to better tailor their marketing strategies and product offerings to meet their customers' needs. One example is the case of TOMS shoes, which are larger ones. For instance, TOMS Shoes is an e-commerce company that sells shoes. The firm was founded in 2006, and its business concept consists of donating one pair of shoes for every pair sold. Despite competition from larger retailers, TOMS has been successful in part due to its commitment to philanthropy. In 2014, the company was valued at over 625 million [47].

C. Study limitations

The findings of this study provide interesting insights that would help understand the vendor's perspective toward the adoption of m-commerce. This can help researchers, businesses, and policymakers develop a more holistic view of m-commerce adoption and develop more effective strategies to promote its adoption. Expanding the sample size is a potential avenue for future research in this study. However, due to practical considerations such as the time-intensive process of locating and persuading vendors to participate in interviews, a sufficiently large sample size may not be feasible. Alternatively, a survey could be employed to gather quantitative data and achieve the desired sample size. Although the current study concentrates on business-to-consumer relations, it is recommended to consider business-to-business relations in future research. Addressing both aspects could allow for a comprehensive understanding of the m-commerce addition in Saudi Arabia. Furthermore, it is important to include large companies in the study, as the current research is constrained by its emphasis on small and medium-sized enterprises (SMEs). This limitation can have notable implications, particularly regarding the concept of trust, as SMEs may demonstrate unique trust dynamics compared to larger companies with specialised divisions or roles. Overcoming this limitation in future investigations can offer a more comprehensive understanding of expertise levels within the context of m-commerce.

VII. conclusion

This paper presented a thematic analysis of the vendor perspective in regard to the adoption of m-commerce in Saudi Arabia. The data were collected through semi-structured interviews conducted with ten Saudi vendors. The findings suggest that m-commerce services are still in their early stages in Saudi Arabia, and the limited adoption can be attributed to two main factors: government regulations and technological infrastructure. Specifically, data protection regulations are not enforced adequately in the country, and vendors face expensive fines and fees. Furthermore, the country's infrastructure does not yet provide a reliable delivery service, which is a crucial requirement for m-commerce adoption. Future research should consider a more scalable approach, such as a larger sample size or a different data collection method.

References

[1] F. Jejdling, "Ericsson mobility report," Ericsson, Tech. Rep., 6 2022, accessed: 2022-08-22.

[2] M. Ahmad, "Review of the technology acceptance model (tam) in internet banking and mobile banking," *International Journal of Information Communication Technology and Digital Convergence*, vol. 3, no. 1, pp. 23–41, 2018.

[3] A. B. Nassuora, "Understanding factors affecting the adoption of m-commerce by consumers," *Journal of Applied Sciences*, vol. 13, no. 6, pp. 913–918, 2013.

[4] M. Hathaway, F. Spidalieri, and F. Alsowailm, *Kingdom of Saudi Arabia Cyber Readiness at a Glance*. Potomac Institute for Policy Studies, 2017.

[5] S. A. Govermnet, "Vision2030," 2016, accessed: 2020-06-07. [Online]. Available: https://vision2030.gov.sa/en

[6] C. Antonelli, "Localized technological change, new information technology and the knowledge-based economy: the european evidence," *Journal of evolutionary economics*, vol. 8, no. 2, pp. 177–198, 1998.

[7] "Quality of life program,," 2018, accessed: 2020-06-07. [Online]. Available: https://www.vision2030.gov.sa/media/gi3l3js2/qol-en.pdf

[8] M. Niranjanamurthy, N. Kavyashree, S. Jagannath, and D. Chahar, "Analysis of e-commerce and m-commerce: advantages, limitations and security issues," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 2, no. 6, pp. 2360–2370, 2013.

[9] R. D. Anvari and D. Norouzi, "The impact of e-commerce and r&d on economic development in some selected countries," *Procedia-Social and Behavioral Sciences*, vol. 229, pp. 354–362, 2016.

[10] S. Aljarboa, "Online shopping in saudi arabia: Opportunities and challenges," *International Journal of Managing Value and Supply Chains*, vol. 7, no. 4, pp. 1–15, 2016.

[11] A. N. H. Zaied, "Barriers to e-commerce adoption in egyptian smes," *International Journal of Information Engineering and Electronic Business*, vol. 4, no. 3, p. 9, 2012.

[12] A. R. Ashraf, N. Thongpapanl, B. Menguc, and G. Northey, "The role of m-commerce readiness in emerging and developed markets," *Journal of International Marketing*, vol. 25, no. 2, pp. 25–51, 2017.

[13] J. Love, "Mobile is changing the fortunes of ecommerce leaders," Aug 2016. [Online]. Available: https://www.digitalcommerce360.com/2016/08/23/mobile-changing-fortunes-e-commerce-leaders/

[14] F. V. Morgeson III, P. N. Sharma, and G. T. M. Hult, "Cross-national differences in consumer satisfaction: mobile services in emerging and developed markets," *Journal of International Marketing*, vol. 23, no. 2, pp. 1–24, 2015.

[15] L. AlSuwaidan and A. A. Mirza, "An investigation on user preferences of mobile commerce interface design in saudi arabia," in *Mobile Web Information Systems*, I. Awan, M. Younas, X. Franch, and C. Quer, Eds. Cham: Springer International Publishing, 2014, pp. 275–285.

[16] M. A. Algethmi, "Mobile commerce innovation in the airline sector: an investigation of mobile services acceptance in saudi arabia," Ph.D. dissertation, Brunel University School of Engineering and Design PhD Theses, 2014.

[17] A. M. J. Turki, L. Pemberton, and H. Macpherson, "Adoption and acceptance of mobile commerce in saudi arabia: the case of e-ticketing in the airline industry," Ph.D. dissertation, University of Brighton, 2017.

[18] A. AbdulMohsin Sulaiman, "Factors affecting mobile commerce acceptance in developing countries: Saudi arabia," Ph.D. dissertation, Brunel University London, 2015.

[19] E. Makki and L.-C. Chang, "E-commerce acceptance and implementation in saudi arabia: previous, current and future factors," *The International Journal of Management Research and Business Strategy*, vol. 4, no. 3, pp. 29–44, 2015.

[20] ——, "The impact of mobile usage and social media on e-commerce acceptance and implementation in saudi arabia," in *Proceedings of the International Conference on e-Learning, e-Business, Enterprise Information Systems, and e-Government (EEE)*, 2015, p. 25.

[21] N. Al-Hadban, A.-G. Hadeel, T. Al-Hassoun, and R. Hamdi, "The effectiveness of facebook as a marketing tool (saudi arabia case study)," *International Journal of Management & Information Technology*, vol. 10, no. 2, pp. 1815–1827, 2014.

[22] A. Alkhunaizan and S. Love, "Predicting consumer decisions to adopt mobile commerce in saudi arabia." in *Proceedings of the Nineteenth Americas Conference on Information Systems,*, August 2013, pp. 1–9.

[23] C. Dearnley, "A reflection on the use of semi-structured interviews," *Nurse researcher*, vol. 13, no. 1, 2005.

[24] S. Alqatan, N. M. M. Noor, M. Man, and R. Mohemad, "A theoretical discussion of factors affecting the acceptance of m-commerce among smtes by integrating ttf with tam," *International Journal of Business Information Systems*, vol. 26, no. 1, pp. 66–111, 2017.

[25] Y. Alqahtani, N. Beloff, and M. White, "A study on the factors shaping customers' adoption of m-commerce in saudi arabia following vision 2030," *IEEE Access*, 2023, manuscript submitted for publication.

[26] D. Gioia, "A systematic methodology for doing qualitative research," *The Journal of Applied Behavioral Science*, vol. 57, no. 1, pp. 20–29, 2021.

[27] D. H. Mortensen, "How to do a thematic analysis of user interviews," *Interaction design foundation*, 2020.

[28] O. Alotaibi and D. Potoglou, "Introducing public transport and relevant strategies in riyadh city, saudi arabia: A stakeholders' perspective," *Urban, Planning and Transport Research*, vol. 6, no. 1, pp. 35–53, 2018.

[29] S. Alotaibi, M. Quddus, C. Morton, and M. Imprialou, “Transport investment, railway accessibility and their dynamic impacts on regional economic growth,” *Research in Transportation Business & Management*, vol. 43, p. 100702, 2022.

[30] L. Alyahyan, H. Aldabbas, and K. Alnafjan, “Preferences of saudi users on arabic website usability,” *International Journal of Web & Semantic Technology (IJWesT)*, vol. 7, pp. 1–8, 2016.

[31] “The ministry of commerce requires all enterprises to utilise arabic in their invoices and contracts.” 2012, accessed: 2022-03-26. [Online]. Available: https://mc.gov.sa/ar/mediacenter/News/Pages/news_1835.aspx

[32] T. Zhang, J. Zhang, C. Huo, and W. Ren, “Automatic generation of pattern-controlled product description in e-commerce,” in *The World Wide Web Conference*, 2019, pp. 2355–2365.

[33] H. Hassen, N. H. Abd Rahim, and A. Shah, “Analysis of models for e-commerce adoption factors in developing countries,” *International Journal on Perceptive and Cognitive Computing*, vol. 5, no. 2, pp. 72–80, 2019.

[34] J. Nielsen, R. Molich, C. Snyder, and S. Farrell, “E-commerce user experience,” *Nielsen Norman Group*, pp. 1–51, 2000.

[35] F. Shirazi and A. Iqbal, “Community clouds within m-commerce: a privacy by design perspective,” *Journal of Cloud Computing*, vol. 6, no. 1, pp. 1–12, 2017.

[36] F. Albejaidi and K. S. Nair, “Nationalisation of health workforce in saudi arabia’s public and private sectors: A review of issues and challenges,” *Journal of Health Management*, vol. 23, no. 3, pp. 482–497, 2021.

[37] A. O. Alotaibi and C. Bach, “Consumer awareness and potential market for e-commerce in saudi arabia: Challenges and solutions,” in *ASEE 2014 Zone I Conference*. Citeseer, 2014, pp. 1–5.

[38] A. Alharbi, “Economic effects of low oil prices in saudi arabia,” *International Journal of Information Technology*, vol. 13, no. 1, pp. 195–200, 2021.

[39] “Saudi arabia driving ban on women to be lifted,” Sep 2017. [Online]. Available: https://www.bbc.co.uk/news/world-middle-east-41408195

[40] G. Saudi Statistics, “Population by age groups,and gender mid year 2020,” https://www.stats.gov.sa, 2020, accessed: 2022-03-12.

[41] S. Tong, X. Luo, and B. Xu, “Personalized mobile marketing strategies,” *Journal of the Academy of Marketing Science*, vol. 48, no. 1, pp. 64–78, 2020.

[42] A. Babar, A. Rasheed, and M. Sajjad, “Factors influencing online shopping behavior of consumers,” *Journal of Basic and Applied Scientific Research*, vol. 4, no. 4, pp. 314–320, 2014.

[43] Q. Liu, M. Goh, Q. Liu, M. Goh, and M. Mouri, “Delivering b2b with ta-q-bin,” *TA-Q-BIN: Service Excellence and Innovation in Urban Logistics*, pp. 127–148, 2015.

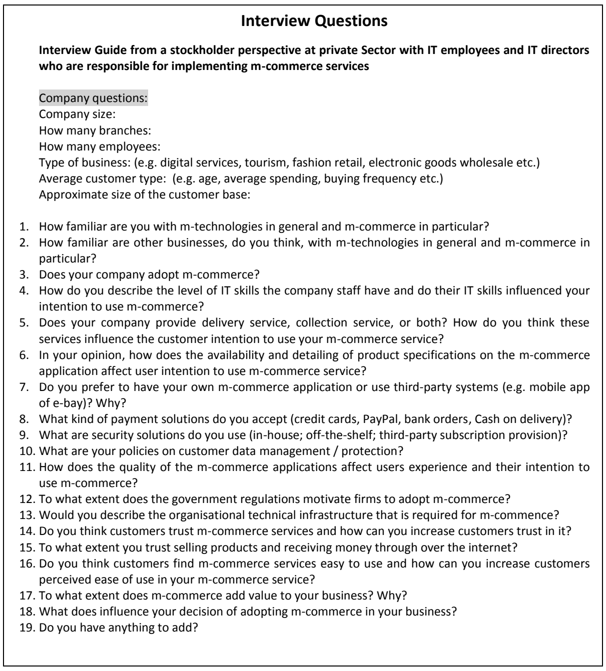
[44] A. N. DONICI, A. Maha, I. Ignat, and L.-G. MAHA, “E-commerce across united states of america: Amazon. com.” *Economy Transdisciplinarity Cognition*, vol. 15, no. 1, 2012.

[45] D. Shang and W. Wu, “Understanding mobile shopping consumers’ continuance intention,” *Industrial Management & Data Systems*, vol. 117, no. 1, pp. 213–227, 2017.

[46] J. Choi, H. Seol, S. Lee, H. Cho, and Y. Park, “Customer satisfaction factors of mobile commerce in korea,” *Internet research*, 2008.

[47] R. Abrams, “Toms’ one for one model: Is it sustainable?” *The Guardian*, Oct 2014. [Online]. Available: https://www.theguardian.com/sustainable-business/2014/oct/28/toms-one-for-one-model-is-it-sustainable

VIII. Interview Questions (English Version)



Interview Questions

Interview Guide from a stockholder perspective at private Sector with IT employees and IT directors who are responsible for implementing m-commerce services

Company questions:
Company size:
How many branches:
How many employees:
Type of business: (e.g. digital services, tourism, fashion retail, electronic goods wholesale etc.)
Average customer type: (e.g. age, average spending, buying frequency etc.)
Approximate size of the customer base:

1. How familiar are you with m-technologies in general and m-commerce in particular?
2. How familiar are other businesses, do you think, with m-technologies in general and m-commerce in particular?
3. Does your company adopt m-commerce?
4. How do you describe the level of IT skills the company staff have and do their IT skills influenced your intention to use m-commerce?
5. Does your company provide delivery service, collection service, or both? How do you think these services influence the customer intention to use your m-commerce service?
6. In your opinion, how does the availability and detailing of product specifications on the m-commerce application affect user intention to use m-commerce service?
7. Do you prefer to have your own m-commerce application or use third-party systems (e.g. mobile app of e-bay)? Why?
8. What kind of payment solutions do you accept (credit cards, PayPal, bank orders, Cash on delivery)?
9. What are security solutions do you use (in-house; off-the-shelf; third-party subscription provision)?
10. What are your policies on customer data management / protection?
11. How does the quality of the m-commerce applications affect users experience and their intention to use m-commerce?
12. To what extent does the government regulations motivate firms to adopt m-commerce?
13. Would you describe the organisational technical infrastructure that is required for m-commence?
14. Do you think customers trust m-commerce services and how can you increase customers trust in it?
15. To what extent you trust selling products and receiving money through over the internet?
16. Do you think customers find m-commerce services easy to use and how can you increase customers perceived ease of use in your m-commerce service?
17. To what extent does m-commerce add value to your business? Why?
18. What does influence your decision of adopting m-commerce in your business?
19. Do you have anything to add?

DOI: 10.15439/2023F3491
ISSN 2300-5963 ACSIS, Vol. 37

Population-less Genetic Algorithm? Investigation of Non-dominated Tournament Genetic Algorithm (NTGA2) for multi-objective optimization

Michał Antkiewicz, Paweł B. Myszkowski
Wrocław University of Science and Technology
Faculty of Information and Communication Technology
ul.Ignacego Łukasiewicza 5, 50-371 Wrocław, Poland
email: {michal.antkiewicz, pawel.myszkowski}@pwr.edu.pl

***Abstract*—This paper investigates the Non-dominated Tournament Genetic Algorithm (NTGA2) to examine how selection methods and population interact in solving multi-objective optimization problems with constraints. As NTGA2 uses tournament and GAP selections that link the current population and population' archive, the experiments' results show that the population role is significantly reduced in some cases. The study considers two benchmark problems: Multi-Skill Resource Constrained Project Scheduling Problem (MS-RCPSP) and Travelling Thief Problem. Moreover, the paper's experimental study consists of new instances for multi-objective MS-RCPSP to show some interesting results that, in some cases, the proposed Genetic Algorithm does not need population in the evolution process.**

I. Introduction

THE POPULATION in Genetic Algorithms (GA) plays a very important role. Too small a population size could significantly reduce the exploration and get stuck in local optima (a premature convergence). However, too large a population also blocks the evolution progress, where selection cannot efficiently do its work. The population size in GA also plays a crucial role in multi-objective optimization (MOO) problems, where the final results consist of a set of solutions (non-dominated, Pareto Front Approximations, PFA). GA applied to MO to be efficient can store all explored non-dominated solutions in the *archive*. Moreover, some methods use the *archive* set to select individuals under selection pressure, like Non-dominated Tournament Genetic Algorithm (NTGA2)[7]. It allows the current population to work on temporal solutions, where the *archive* stores all non-dominated solutions, which makes a ”permanent” memory. Such phenomena exist in over-constrained MOO problems, where genetic operators could make an offspring individual worse, and additional space (population) could help. In this paper, two MO NP-hard problems with constraints – Travelling Thief Problem (TTP)[3] and Multi-Skill Resource-Constrained Project Scheduling Problem (MS-RCPSP)[8] – are examined to investigate how NTGA2 explores the solution landscape and effectively uses *archive* and population in individual selection.

The rest of the paper is organized as follows. In Sec.II, a short related work is given. The investigated MS-RCPSP and TTP problems are briefly defined in Sec.III. An investigated NTGA2 is given in Sec.IV. Sec. V includes experimental results of the proposed NTGA2 MOO. Lastly, the paper is concluded in Sec.VI.

II. Related works

Effective cooperation between population, *archive*, and GAP selection works in GaMeDE2 [1] - an enhanced Multi-Modal Optimization technique (GaMeDE[6]), inspired by NTGA2, where GAP operator was introduced. While GaMeDE2 simplified the algorithm, empirical research confirmed the importance of alternating between broad exploration using the *archive* and local optimization with the population. This is achieved by triggering local optimization when the number of newly discovered optimal solutions exceeds a threshold or reverting to *archive* sampling when further optimization becomes unfeasible. Although this presents some form of adaptive operator switching, the main drawback is the requirement of fine-tuning the threshold value. It has been indicated as the area for future work to develop a fully adaptive solution that eliminates the need for manual parameter specification. For this purpose, it is necessary to answer the question of how to effectively switch between population- and *archive*-based selection, and on what basis to make this decision.

The authors of the survey [13] designed the adaptation taxonomy scheme for GA. They highlight three aspects to be considered: adaptation objects, adaptation evidence, and adaptation methods. Adaptation objects refer to the components within a GA. These objects include control parameters (crossover or mutation probability), evolutionary operators, and other elements. Adaptation evidence determines the basis on which adaptation occurs within a GA. There are four categories of adaptation evidence highlighted by the authors: deterministic factors, fitness values, population distribution, and combinations of fitness values and population distribution. There are several adaptation methods that might be implemented, including simple rules-based heuristics, or co-evolution. This paper examines how the structure and constraints of the problem affect the population- and *archive*-

based selection with a view to using it for the phase switching control.

III. Problem(s) definition

Two practical multi-objective problems have been selected to investigate the NTGA2 method. Both TTP and MS-RCPSP are constrained NP-hard problems with near-to-real-world combinatorial landscapes. Results have been compared using standard HyperVolume (HV, see Sec.V-B) measure.

A. Multi-Skill Resource-Constrained Project Scheduling Problem

The MS-RCPSP is a specific case of combinatorial NP-hard scheduling problems. It encompasses two interrelated sub-problems: task sequencing, and resource assignments. The objective of the MS-RCPSP is to determine a schedule that is both feasible and satisfies all the defined constraints. This involves assigning available resources to tasks and arranging the tasks on a timeline. In order for a schedule, denoted as PS, to be considered feasible, it must adhere to a predetermined set of constraints.

Each resource is connected to its salary r_{salary}, no salary can be negative. The set of skills S_r possessed by the resource r cannot be empty. Each task's duration d_t and finish time F_t are not negative. Tasks are constrained by a precedence relation – all task's predecessors must be finished before work on it can be started. All tasks must have assigned exactly one resource.

The skill extension of the MS-RCPSP is described in Eq. 1. The resource must have the skill at the required level or higher if it is assigned to a task.

$$\forall_{t \in T^r} \exists_{s_r \in S^r} \; h_{s_t} = h_{s_r} \wedge l_{s_t} \leq l_{s_r} \tag{1}$$

where T^r is a set of tasks assigned to a resource r, s_t is the skill required by the task t, S^r is the set of skills possessed by the resource r, h and l are the type and level of the skill respectively.

The latest definition of the MS-RCPSP is a many-objective optimization problem with five objectives (see [7][8]). The original two objectives – schedule duration (makespan) and cost – can be defined by Eq.2 and Eq.3. Further MS-RCPSP objectives tackle specific project scheduling aspects: average cash flow, skill overuse, and the average use of resources. The **Makespan** $\mathbf{f}_\tau(\mathbf{PS})$ of the project schedule PS is given as Eq.2.

$$f_\tau(PS) = \max_{t \in T} t_{finish} \tag{2}$$

where T is a set of all tasks, t_{finish} is the finish time of the task t. The **Cost** of the schedule is $\mathbf{f_C}(\mathbf{PS})$ defined as Eq. 3.

$$f_C(PS) = \sum_{i=1}^{n} R_i^{salary} * T_i^{duration} \tag{3}$$

where n is the number of all task-resource assignments, R_i^{salary} is the salary of a resource of the i'th assignment, $T_i^{duration}$ is the duration of the task of the i'th assignment.

The MS-RCPSP originally optimises 2-objectives $\mathbf{f}_\tau(\mathbf{PS})$ and $\mathbf{f_C}(\mathbf{PS})$, where all objectives must be minimized:

$$\min f(PS) = \min\left[f_\tau(PS), f_C(PS)\right] \tag{4}$$

B. Travelling Thief Problem

The TTP is a combination of two well-known optimization problems: the Traveling Salesman Problem (TSP) and the Knapsack Problem (KNP). A collection of cities is given, each characterized by its geographical coordinates, along with a set of associated items. These items are characterized by their weight and profit values. The objective is to determine an optimal route that visits all the cities while simultaneously selecting items from certain cities. The primary objective of the TTP, as expressed by Eq.5, encapsulates the main goal of this problem.

$$\min f(\pi, z) = \min f_\tau(\pi, z), \max f_P(z) \tag{5}$$

where π and z are the permutations of cities visited and the picking plan. The objective f_τ is to minimize the **traveling plan**. The f_P objective is the **profit** maximization based on the picked items. The relation between those problems is that picking items decreases travel speed.

$$f_\tau(\pi, z) = \sum_{i=1}^{n-1} \frac{d_{\pi_i, \pi_{i+1}}}{v(w(\pi_i))} + \frac{d_{\pi_n, \pi_1}}{v(w(\pi_n))} \tag{6}$$

where $d_{\pi_i, \pi_{i+1}}$ denotes the distance between two consecutive cities, n is a number of cities, $v(w(\pi_i))$ is the velocity in city π_i, which depends on weight w. As items are selected, the entire travel duration, as indicated by Eq. 6, undergoes modifications, and the velocity decreases in accordance with Eq. 7.

$$v(w) = v_{max} - \frac{W_c}{W}(v_{max} - v_{min}) \tag{7}$$

where W_c and W are the current and maximum allowed weights. The model defines the speed: maximum v_{max} and minimum v_{min} speed depending on W. The weight w is the accumulated sum of items picked up so far.

$$f_P(z) = \sum_{j=1}^{m} z_j z_j^{profit} \tag{8}$$

where m is the number of items, z_j is equal to 1 if the j'th item has been picked, 0 otherwise. z_j^{profit} is the profit of the j'th item.

The Eq.8 defines the profit as the second TTP objective. Furthermore, in order for the picking plan to be considered feasible, it is imperative that KNP constraint, as represented by Eq.9, is satisfied.

$$\sum_{j=1}^{m} z_j z_j^{weight} \leq W \tag{9}$$

where m denotes the total number of items, while z_j is defined as per Eq.8, and z_j^{weight} represents the weight of the j'th item. The aforementioned equation guarantees that the cumulative weight of the selected items remains within the W, which denotes the maximum permissible weight of the knapsack.

IV. Method

In this section, the NTGA2 method is introduced, and a greedy-based algorithm (see Sec.1) is used to get *feasible* solution in the MS-RCPSP problem.

A. Greedy–based Schedule Builder

Each investigated metaheuristic in this work to solve MS-RCPSP uses a greedy–based Schedule Builder to build the *feasible* schedule – see Algorithm 1 [7]. The method processes the tasks (in the given order). First, tasks with successors, then other tasks. The main goal is to assign each task at the earliest possible time it can be started. Namely, it is when all the predecessors of the tasks are finished, and its assigned resource finished its previous task assignment.

Algorithm 1 Greedy Schedule Builder for MS-RCPSP

```
for task t do
    $predEnd = maxFinish(t.predecessors)$
    $resEnd = t.getResource().getFinish()$
    $t.start = max(predEnd, resEnd)$
end for
```

B. Non-Dominated Tournament Genetic Algorithm 2

NTGA2[7] is an evolutionary metaheuristic promoting diversity by utilizing a *Gap* selection (GS) operator. GS works in the objective space, favoring the least explored parts of the *archive* – *Gap* in detail is given below. NTGA2 uses *archive* to store all non-dominated solutions and actively use it – see Algorithm 2. Firstly, NTGA2 initializes the population (usually a random one – see line 2). Then all individuals are evaluated (separately by each objective), and then $UpdateArchive$ takes place, where all non-dominated already found individuals are added and just dominated ones are removed.

The main loop starts (line 5) and repeats $Generations$ times. Each generation starts with a selection of individuals to the new population P_{next}. GS and the second selection (Pareto-dominance tournament selection) is used. The $gsGenerations$ parameter (line 8) switches selections a decides which one is used in the current generation. Line 15 presents the clone elimination mechanism used in NTGA2 to keep diversity in the population at a high level. Lastly, the genetic operators (e.g. mutation and crossover) should be specialized per problem. However, they can default to standard *single-point* crossover and *random bit* mutation.

The **Gap Selection** (GS) operator [7] aims to increase the *diversity* in *archive*. It operates in an objective space and considers each objective separately. The authors decided to select objectives as follows: offspring generation is divided into m parts (as the number of objectives), where each objective is selected during the corresponding part. It starts by calculating the "gap" size for each individual in the *archive*. It is calculated considering the two neighbor individuals using the minimal *Euclidean* distance. Those are the closest individuals, one with a worse objective value and one with a better value. The $GapValue$ is used as the Euclidean distance to the farther of those two neighbors. Additionally, individuals at the "edge" (i.e., the highest and lowest objectivities' values) of the *archive* have this distance set to an $infinity$ value, which is favored in selection.

Algorithm 2 Pseudocode of NTGA2 [7]

```
1:  $archive \leftarrow \emptyset$
2:  $P_{current} \leftarrow GenerateInitialPopulation()$
3:  $Evaluate(P_{current})$
4:  $UpdateArchive(P_{current})$
5:  for $i \leftarrow 0$ to $Generations$ do
6:      $P_{next} \leftarrow \emptyset$
7:      while $|P_{next}| < |P_{current}|$ do
8:          if $i$ mod (2 * $gsGen$) $< gsGen$ then
9:              $Parents \leftarrow Tour\_selection(P_{current})$
10:         else
11:             $Parents \leftarrow Gap\_selection(Archive)$
12:         end if
13:         $Children \leftarrow Crossover(Parents)$
14:         $Children \leftarrow Mutate(Children)$
15:         while $P_{next}$ contains $Children$ do
16:             $Children \leftarrow Mutate(Children)$
17:         end while
18:         $Evaluate(Children)$
19:         $P_{next} \leftarrow P_{next} \cup Children$
20:         $UpdateArchive(Children)$
21:     end while
22:     $P_{current} \leftarrow P_{next}$
23: end for
```

Thus, the GS uses a tournament selection, considering $GapValues$ instead of $fitness$ directly. In this way, GS is more likely to select those individuals that lie close to the largest "gaps" in the *archive* and also promote the spread of the result *archive*. The second parent is selected as the $random$ neighbor of the first individual. For the individuals lying on the "edge" of PF approximation, it is possible that a second parent will not be selected. It will be selected similarly to the first one.

V. Experiments

The main goal of conducted experiments is to investigate further the effectiveness of the GS operator inside the NTGA2 method, applied to different scenarios. It can be hypothesized that its effectiveness varies depending on the problem, and further - instances. To carry out the structured experiments, the following **Research Questions** have been developed:

- ***RQ0.*** How the $gsGen$ (Gap Selection %) parameter affects the effectiveness of the NTGA2 method in bi-objective MS-RCPSP?
- ***RQ1.*** How do the characteristics (size, number of constraints) of the instance affect the effectiveness of the NTGA2 method in bi-objective MS-RCPSP?
- ***RQ2.*** Are the observations made for the TTP consistent with those for the MS-RCPSP?
- ***RQ3.*** Does the size of the computational budget affect the effectiveness of the Gas Selection in the NTGA2 method?
- ***RQ4.*** What aspects (differing or connecting two problems) can be used to adapt $gsGen$ parameter control?

A. Instances

In experiments, the iMOPSE dataset [7][8] is used. The original suite contains 36+6 MS-RCPSP instances created using real-world scheduling problems. All instances have varying tasks, resources, and skills to define problems. The final suite used in this paper contains 3 small and 6 randomly selected instances from the original set. Furthermore, several new instances were prepared using the iMOPSE generator to show the influence of constraints (e.g. introducing extreme low and high values for precedence relations or no skill requirements) and a number of tasks for NTGA2 effectiveness (e.g. 500 and 1000)[1].

For TTP, the benchmark dataset [2] has been selected - 16 instances differ in varying items per city (between 51 to 100). They could be divided into three groups that show the correlation between weights and profits of items: (1) with a strong correlation, (2) completely uncorrelated, and (3) with similar weights.

B. Quality measure of multi–objective optimisation

The most popular multi-objective metric is HyperVolume (HV) [9] – measures the diversity and convergence of the Pareto Front Approximation (PFA) that includes all non-dominated solutions calculated by a given method.

Results are normalized using the $NadirPoint$ - worst possible values for all objectives. For the MS-RCPSP: makespan – total sum of all tasks' duration; cost – the cost of schedule, where the most expensive resource performs all tasks. For the TTP: time – is twice the minimum time value; profit – equal to 0. On the other side - the $Ideal\ Point$ is the point with the best possible values for all objectives. For MS-RCPSP: makespan – duration of the shortest task multiplied by the number of tasks, divided by the number of resources; cost – the cost of the schedule, where all tasks are assigned to the cheapest resource. For TTP: time – the total length of the minimum spanning tree divided by the maximum speed; profit – achieved by a brute-force algorithm starting from the items with the highest profit/weight ratio.

C. Reference methods

For a more comprehensive presentation of NTGA2 results, results of the state-of-the-art and best-known multi-objective optimization methods should also be considered.

Non-Dominated Sorting Genetic Algorithm II (**NSGA-II**) [5] is the classical method proposed in the year 2002 for MOO, utilizing the population sorting by $rank$ and $crowding\ distance$. The Strength Pareto Evolutionary Algorithm 2 (**SPEA2**) [11], a well-established method for MOO, employs environmental selection to enhance the exploration of the Pareto front. The Multi-objective Evolutionary Algorithm Based on Decomposition (**MOEA/D**) [12] is an evolutionary computation method designed for solving MOO problems by decomposing problems into a set of scalar sub-problems that are concurrently optimized.

D. Configurations

For all methods, the 5-Level Taguchi Parameter Design [10] was employed to fine-tune the parameters systematically. The best-found configurations used as the base values in the experiments are presented in Tab.I. Population Size ($PopSize$) is the constant number of individuals in a generation. For the MOEA/D, population size is derived from the number of decomposition vectors achieved using [4] algorithm for the given number of partitions ($PartNr$). The number of generations was adjusted to match the constant number of maximum births/fitness evaluations, for the MS-RCPSP computational budget was set to 50.000. For the TTP it was set to 250.000. Mutation probability (P_m) is a probability in $[0, 1]$. In the MS-RCPSP, it represents a chance of a single gene's random mutation. For the TTP, it is described using two different values: the chance of the random path segment being reversed; and the chance of a random item decision change ($bitflip$). Crossover probability (P_x) is the probability of two individuals crossover. All implemented methods use the same *Uniform crossover* operator for the MS-RCPSP and a combination of OX (route) with SX (knapsack) for the TTP. Tournament Size ($TourSize$) is the number is individuals considered whenever the tournament selection operator is used. Based on the original NTGA2 implementation, 2 values were found to be used, first for the Standard Tournament and second for the Gap tournament selection. The neighborhood Size ($NhSize$) is the number of adjacent decomposition vectors considered by MOEA/D when solutions are compared.

A number of generations ($gsGen$, see Tab.I) is the selection switch parameter used by the NTGA2. In the original NTGA2 paper, it has been interpreted as the number of generations that has to pass for the selection to switch, and it was set to 50. Although this parameter originally referred to the frequency of changes, a value of 50 can also be interpreted as 50 per 100 generations using the GS (Gap Selection %). Therefore, further in this paper, $gsGen$ is evaluated as the number of consecutive generations using the GS per 100 generations.

[1] All used MS-RCPSP instances and gained results are published in http://imopse.ii.pwr.edu.pl

TABLE I
THE BEST FOUND CONFIGURATIONS FOR INVESTIGATED METHODS

MSRCPSP	PopSize	P_m	P_x	TourSize	gsGen	NhSize	PartNr
NTGA2	50	0.01	0.6	6 / 20	50*		
MOEA/D	(50)	0.015	0.2			6	50
SPEA2	200	0.015	0.99				
NSGA-II	300	0.015	0.99	2			

TTP	PopSize	P_m	P_x	TourSize	gsGen	NhSize	PartNr
NTGA2	50	0.9 / 0.9	0.3 / 0.3	6 / 20	50*		
MOEA/D	(100)	0.4 / 0.3	0.5 / 1.0			3	100
SPEA2	100	0.4 / 0.3	0.1 / 0.8				
NSGA-II	300	0.4 / 0.7	0.9 / 0.3	2			

E. Experimental procedure

The research environment with the NTGA2 method and additional reference methods have been implemented in Java based on the literature: MOEA/D, SPEA2, and NSGA-II. Additionally, some methods (e.g. MOEA/D) use reference points in calculations (objectives normalization). Others do not require them (like NTGA2) as they treat objectives separately. Reference points ($NadirPoint$ and $IdealPoint$) calculations as presented in the Sec. V-B.

The result of each run is a set of non-dominated solutions found for a given instance. Solutions are saved using absolute coordinates in the objective space. The experimental results have been evaluated on all selected instances for MS-RCPSP and TTP. Due to the non-deterministic nature of evolutionary computation, all runs have been repeated 30 times, and results have been averaged. To verify the statistical significance of the presented results Wilcoxon signed-rank test is used with $p\ value = 0.05$. A simple average is not an appropriate solution as HV strongly differs across the instances. Therefore, a ranking system has been applied to compare configurations and methods in all conducted experiments. The procedure starts with descending sort by the average HV and assigning the best rank (1) to the first configuration. Then, each configuration is considered subsequently. If its result is not significantly lower than the best of the current setup, it gets assigned the same rank. If the rank is significantly lower - the rank is incremented and assigned to this configuration. Each experiments table contains three summary rows at the bottom: average rank, median rank, and the dominance information ($+$ sole-best / $\sim$ co-best / $-$ worst).

F. Results for MS-RCPSP

To address the $RQ0$, five variants with different values of the $gsGen$ parameter were examined. The configurations utilized Gap selection in 0%, 25%, 50%, 75%, and 100% of generations, respectively. The results are presented in Tab.II.

As observed in Tab.II, the configuration employing 100% GAP selection clearly dominates in nearly all original instances. However, in the case of two small instances (15_6_10_6, 15_9_12_9), only GS from the *archive* does not yield the best results. Similarly, configuration alternating two selection methods prove to be the most effective for new instances with a high number of constraints (e.g., 100_10_4096_15). While the differences are statistically significant, they are very small.

The newly added instances (with suffix _0_0), devoid of skill constraints and task orders, did not introduce noticeable changes. It can be assumed that they are sufficiently similar to the existing cases. Considering the number of precedence relations, where the maximum theoretical number of direct relationships is $n * (n - 1)/2$, for 100 tasks, it amounts to 4950. Therefore, the highest number of constraints in the set, which is 145, is still very small. Hence, the absence of constraints does not differ significantly. On the other hand, including instances with precedence relations at the level of several thousand introduce interesting cases that have not been observed before.

Fig. 1 indicates that, for dense PFA, there is no need to employ a population. It is easy to transition between solutions as they are close to one another. There is a large number of solutions that exploration based on the *archive* alone is sufficient, at least until a certain point. However, relying solely on the *archive* may become inadequate if the search space is highly constrained and all the 'low-hanging apples' are found. Theoretically, it is still unnecessary. In the current encoding, any feasible solution can transition to another in a single generation (as each gene can be modified independently), but it might not be very probable.

Using a population allows for delving "deeper" into certain areas of the sparse Pareto Front, as visible in Fig. 2. Classically, this brings about a solution to the problem of balance between exploration and exploitation. It is well illustrated in Fig. 3 containing results for the biggest instance in the suite. None of the configurations have sufficiently searched the space yet. The 'population-only' approach focused on a particular area, while the '*archive*-only' covered a wider range. As both approaches perform their role well, the latter achieves significantly better HV. It would likely be beneficial to activate the population search when relying on the *archive* ceases to yield progress instead of static parameters. To answer the second $RQ1$, the effectiveness does not explicitly depend on the instance size. However, rather constraints density and it changes over time as the *archive* saturates. This claim is supported by the results, where a lower budget (25.000, half of the original) results in better ranks for the '*archive*-only' approach.

Experimental results presented in this section showed that Gap affect the effectiveness of the NTGA2 applied to MS-RCPSP. How does such a mechanism work for TTP?

G. NTGA2 results for TTP

In order to verify if similar results can be observed for the TTP ($RQ2$), analogous experiments have been conducted using five configurations, which utilize GS in 0%, 25%, 50%, 75%, and 100% of generations, respectively. The results are presented in Tab.IV.

Compared to the MS-RCPSP, results achieved for the TTP (see Tab.IV) are more balanced, and there is no visible dominance of either configuration. The higher usage of GS has

TABLE II
RESULTS (HV) OF MS-RCPSP WITH 50K BUDGET

	Gap Selection Usage (%)				
instance	0	25	50	75	100
15_3_5_3	**1** (0.320398±1.67e-16)	**1** (0.320398±1.67e-16)	**1** (0.320398±1.67e-16)	**1** (0.320398±1.67e-16)	**1** (0.320398±1.67e-16)
15_6_10_6	**1** (0.545211±8.31e-06)	**1** (0.545213±2.51e-06)	**1** (0.545211±4.76e-06)	2 (0.545207±1.19e-05)	3 (0.544863±8.89e-05)
15_9_12_9	**1** (0.595148±1.30e-04)	**1** (0.595115±1.75e-04)	**1** (0.595092±2.05e-04)	2 (0.595009±2.58e-04)	3 (0.594575±2.96e-04)
100_5_20_9_D3	4 (0.429317±3.45e-03)	3 (0.435507±2.45e-03)	2 (0.436230±2.94e-03)	2 (0.436880±2.22e-03)	**1** (0.438390±1.94e-03)
100_5_48_9	3 (0.171241±2.63e-03)	2 (0.175926±8.50e-04)	**1** (0.176215±9.45e-04)	**1** (0.176347±8.66e-04)	**1** (0.176715±9.23e-04)
100_10_65_15	3 (0.458327±2.84e-03)	2 (0.469056±1.98e-03)	2 (0.469367±1.80e-03)	**1** (0.470235±1.74e-03)	**1** (0.470540±1.68e-03)
100_20_0_0	5 (0.683409±5.12e-03)	4 (0.726044±2.19e-03)	3 (0.733052±1.55e-03)	2 (0.734951±1.51e-03)	**1** (0.736693±1.19e-03)
100_40_0_0	5 (0.656597±7.34e-03)	4 (0.728598±5.49e-03)	3 (0.752332±3.74e-03)	2 (0.763022±2.99e-03)	**1** (0.769273±2.62e-03)
100_10_4096_9	2 (0.181034±2.51e-05)	**1** (0.181041±3.64e-06)	**1** (0.181042±2.93e-06)	2 (0.181039±5.45e-06)	3 (0.181026±7.46e-06)
100_10_4096_15	4 (0.250829±7.23e-05)	3 (0.250859±3.14e-05)	**1** (0.250865±0.00e+00)	**1** (0.250865±0.00e+00)	2 (0.250859±4.42e-06)
100_20_1024_9	4 (0.517370±2.50e-03)	3 (0.527683±9.73e-04)	**1** (0.528931±3.82e-04)	**1** (0.528962±3.47e-04)	2 (0.528699±2.60e-04)
100_20_2048_15	4 (0.380175±5.93e-04)	**1** (0.380732±2.20e-05)	**1** (0.380737±1.16e-05)	2 (0.380730±1.32e-05)	3 (0.380689±3.39e-05)
100_20_4096_9	3 (0.268188±9.21e-05)	**1** (0.268304±1.78e-06)	**1** (0.268299±3.12e-05)	**1** (0.268305±1.92e-06)	2 (0.268295±4.01e-06)
100_20_4096_15	3 (0.259517±1.60e-04)	2 (0.259638±9.18e-05)	**1** (0.259664±4.59e-05)	**1** (0.259678±3.28e-05)	2 (0.259650±3.17e-05)
100_40_1024_9	4 (0.570528±8.13e-04)	3 (0.574741±5.45e-04)	2 (0.575724±4.14e-04)	**1** (0.576005±3.15e-04)	**1** (0.576010±2.64e-04)
200_10_84_9	5 (0.636125±2.91e-03)	4 (0.670534±2.13e-03)	3 (0.678506±1.42e-03)	2 (0.681494±1.19e-03)	**1** (0.682893±1.26e-03)
200_20_97_9	5 (0.629058±7.46e-03)	4 (0.696804±4.35e-03)	3 (0.713753±2.85e-03)	2 (0.720315±1.71e-03)	**1** (0.724045±1.59e-03)
200_20_145_15	5 (0.571673±4.72e-03)	4 (0.617373±3.41e-03)	3 (0.627285±2.49e-03)	2 (0.630914±1.26e-03)	**1** (0.632386±1.49e-03)
200_20_0_0	5 (0.649978±5.06e-03)	4 (0.737694±5.42e-03)	3 (0.763551±4.12e-03)	2 (0.778068±2.52e-03)	**1** (0.784937±2.58e-03)
200_40_0_0	5 (0.719199±5.39e-03)	4 (0.778408±4.13e-03)	3 (0.798334±3.72e-03)	2 (0.808493±2.45e-03)	**1** (0.815440±2.66e-03)
500_10_512_5_A	5 (0.398676±1.58e-03)	4 (0.412768±1.24e-03)	3 (0.418172±1.13e-03)	2 (0.421138±7.19e-04)	**1** (0.422460±8.19e-04)
500_10_2048_5_A	5 (0.536240±1.56e-03)	4 (0.555732±1.62e-03)	3 (0.563177±1.46e-03)	2 (0.567528±1.55e-03)	**1** (0.570074±1.23e-03)
500_20_512_5_A	5 (0.553193±3.07e-03)	4 (0.587305±3.20e-03)	3 (0.601718±2.95e-03)	2 (0.609410±2.19e-03)	**1** (0.614964±1.90e-03)
500_20_0_0	5 (0.591966±4.23e-03)	4 (0.643410±3.08e-03)	3 (0.660743±3.52e-03)	2 (0.668717±2.71e-03)	**1** (0.675124±2.36e-03)
500_40_0_0	5 (0.619020±5.67e-03)	4 (0.681662±4.88e-03)	3 (0.705443±3.41e-03)	2 (0.717776±3.53e-03)	**1** (0.725970±3.46e-03)
1000_20_1024_5_A	5 (0.585542±3.56e-03)	4 (0.617928±3.22e-03)	3 (0.634121±3.98e-03)	2 (0.644301±4.14e-03)	**1** (0.650263±2.98e-03)
1000_20_4096_5_A	5 (0.564774±3.19e-03)	4 (0.583085±2.47e-03)	3 (0.594726±2.26e-03)	2 (0.599670±2.44e-03)	**1** (0.600960±2.24e-03)
1000_40_1024_10_A	5 (0.506093±4.80e-03)	4 (0.545322±4.19e-03)	3 (0.565749±4.58e-03)	2 (0.580091±4.65e-03)	**1** (0.587114±4.27e-03)
1000_20_0_0	5 (0.450345±3.67e-03)	4 (0.503184±4.13e-03)	3 (0.527720±3.98e-03)	2 (0.540095±3.83e-03)	**1** (0.549674±3.52e-03)
1000_40_0_0	5 (0.535800±5.89e-03)	4 (0.594065±4.69e-03)	3 (0.619015±4.71e-03)	2 (0.633995±4.26e-03)	**1** (0.643587±3.53e-03)
avg rank	4.067	3.067	2.233	1.733	1.4
med rank	5	4	3	2	1
dominance	(+0/∼ 3/−27)	(+0/∼ 6/−24)	(+0/∼ 10/−20)	(+0/∼ 8/−12)	(+18/∼ 4/−8)

TABLE III
METHODS COMPARISON (HV) OF MS-RCPSP WITH 50K BUDGET

method	ntga2 0%	ntga2 50% [7]	ntga2 100%	moea/d	nsgaii	spea2
avg rank	4.1	2.033	1.367	2.033	4	3.7
med rank	4	2	1	2	4	4
+	0	5	12	3	0	0
∼	3	6	8	7	1	1
−	27	19	10	20	29	29

a slightly better average ranking, but none reaches above 2.5 rank. This is most likely due to the difference in the encoding. Permutation-based (ordering) genotype encoding with inverse mutation does not allow for free transition between any solution – the transition from one solution to another might require multiple inverse operations on the genotype, which requires 'temporal' individuals - population.

Fig. 4 presents some similarities to the previous results. For the dense PFA, the best configuration uses an '*archive*-only' approach, which scans the space wider, while the 'population-only' is focused in a single direction. On the other end, Fig. 5 presents an instance with sparse PFA, where the 'population-based' approach significantly finds better results. Furthermore, to verify whether the budget has an impact on the GS effectiveness, other 'low-budget' experiments were carried out. Results achieved for lowered budget (50.000, i.e. one-fifth of the original). The effect is significant, as configuration for 75% GS improves from rank 2.5 to 2, and the 100% Gap configuration from 2.5 to 1.5. It supports the hypothesis of the increasing importance of 'population-based' selection (or decreasing importance of '*archive*-based' selection).

H. Summary

Experiments presented in previous sections showed that for MS-RCPCP and TTP, the computation budget plays an important role. For MS-RCPSP, a lower budget (25.000, half of the original) results in better ranks for the '*archive*-only' approach. Respectively, for TTP effect is also significant, as configurations that use the *archive* more 'frequently' (i.e. 75% or 100% Gap) improve their rank, which answers to $RQ3$.

Except for highly constrained instances, utilizing 100% GS yields the best results for the MS-RCPSP. This dominance of a single configuration is not that clear in the TTP. The potential reason could be the encoding difference since association encoding provides an easier transition between solutions than permutation encoding. Which is related to the constrainedness of the search space. The expected result for both problems is the better effectiveness of the 100% '*archive*-based' GS in less constrained instances - having dense PFA. In the most sparse PFA, 'population-based' selection provides a significant

TABLE IV
NTGA2 RESULTS (HV) OF TTP WITH 250K BUDGET

	Gap Selection Usage (%)				
instance	0	25	50	75	100
eil51_n50_bounded-strongly-corr_01	**1** (0.786307±2.22e-16)	5 (0.784464±0.00e+00)	4 (0.784500±2.22e-16)	2 (0.785861±1.11e-16)	3 (0.784563±2.22e-16)
eil51_n50_uncorr_01	5 (0.880994±1.11e-16)	4 (0.881789±2.22e-16)	**1** (0.884046±2.22e-16)	2 (0.883577±2.22e-16)	3 (0.881914±3.33e-16)
eil51_n50_uncorr-similar-weights_01	2 (0.732453±1.11e-16)	3 (0.731714±0.00e+00)	**1** (0.733751±3.33e-16)	4 (0.731639±2.22e-16)	5 (0.731452±2.22e-16)
eil51_n150_uncorr-similar-weights_01	4 (0.851645±2.22e-16)	**1** (0.856558±1.11e-16)	3 (0.855674±1.11e-16)	5 (0.851625±2.22e-16)	2 (0.856388±2.22e-16)
berlin52_n51_bounded-strongly-corr_01	**1** (0.884818±2.22e-16)	5 (0.880486±2.22e-16)	4 (0.881869±3.33e-16)	3 (0.883038±2.22e-16)	2 (0.883507±3.33e-16)
berlin52_n51_uncorr_01	2 (0.838511±1.11e-16)	5 (0.833224±2.22e-16)	3 (0.837703±2.22e-16)	4 (0.837138±1.11e-16)	**1** (0.838970±0.00e+00)
berlin52_n51_uncorr-similar-weights_01	**1** (0.722264±2.22e-16)	4 (0.720164±1.11e-16)	3 (0.720849±2.22e-16)	2 (0.721720±2.22e-16)	5 (0.719742±0.00e+00)
pr76_n75_bounded-strongly-corr_01	4 (0.813571±1.11e-16)	2 (0.818288±2.22e-16)	5 (0.813531±1.11e-16)	3 (0.816954±2.22e-16)	**1** (0.821745±0.00e+00)
pr76_n75_uncorr_01	5 (0.841923±2.22e-16)	3 (0.854713±1.11e-16)	**1** (0.858719±2.22e-16)	4 (0.854683±2.22e-16)	2 (0.855487±0.00e+00)
pr76_n75_uncorr-similar-weights_01	2 (0.767799±0.00e+00)	**1** (0.771500±1.11e-16)	5 (0.762926±0.00e+00)	3 (0.765856±1.11e-16)	4 (0.765030±1.11e-16)
kroA100_n99_bounded-strongly-corr_01	5 (0.866310±2.22e-16)	4 (0.874710±1.11e-16)	3 (0.881084±2.22e-16)	2 (0.884684±0.00e+00)	**1** (0.885007±3.33e-16)
kroA100_n99_uncorr_01	4 (0.844482±2.22e-16)	5 (0.838833±2.22e-16)	2 (0.852879±2.22e-16)	**1** (0.854227±3.33e-16)	3 (0.846971±0.00e+00)
kroA100_n99_uncorr-similar-weights_01	4 (0.886399±2.22e-16)	3 (0.887283±0.00e+00)	5 (0.883876±2.22e-16)	**1** (0.897105±0.00e+00)	2 (0.889732±2.22e-16)
rd100_n99_bounded-strongly-corr_01	5 (0.882900±2.22e-16)	**1** (0.892562±3.33e-16)	3 (0.889671±0.00e+00)	2 (0.892136±2.22e-16)	4 (0.889643±1.11e-16)
rd100_n99_uncorr_01	5 (0.851845±2.22e-16)	4 (0.856888±2.22e-16)	2 (0.857302±2.22e-16)	3 (0.857012±2.22e-16)	**1** (0.862809±2.22e-16)
rd100_n99_uncorr-similar-weights_01	4 (0.892621±3.33e-16)	5 (0.890683±2.22e-16)	2 (0.898613±2.22e-16)	**1** (0.898724±2.22e-16)	3 (0.893712±2.22e-16)
avg rank	3.375	3.438	2.938	2.625	2.625
med rank	4	4	3	2.5	2.5
dominance	(+3/∼ 0/−13)	(+3/∼ 0/−13)	(+3/∼ 0/−13)	(+3/∼ 0/−13)	(+4/∼ 0/−12)

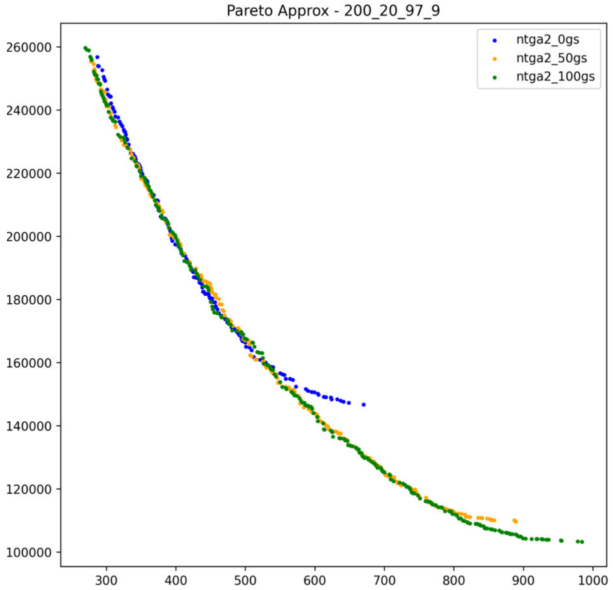


Fig. 1. Comparison of PFA for MS-RCPSP – 200_20_97_9 for GS configs.: 0%, 50% and 100%.

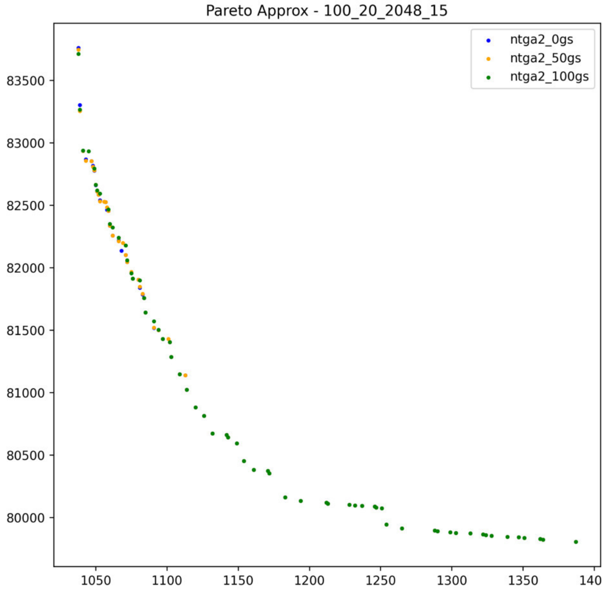


Fig. 2. Comparison of PFA for MS-RCPSP – 100_20_2048_15 for GS configs.: 0%, 50% and 100%.

TABLE V
METHODS COMPARISON (HV) OF TTP WITH 250K BUDGET

method	ntga2 0%	ntga2 50% [7]	ntga2 100%	moea/d	nsgaii	spea2
avg rank	1.938	1.312	1.312	3	3.75	2.812
med rank	2	1	1	3	4	3
+	0	2	5	0	0	0
∼	5	9	6	0	0	2
−	11	5	5	16	16	14

boost. Another potential cause is the computational budget vs the instance size. The importance of the 'population-based' approach improves over the execution time as the *archive* becomes more saturated. To answer the $RQ4$, the potential evidence for the $gsGen$ value adaptations are constraint density, encoding (transition freedom), and *archive* saturation. Where some or all of the above might be entangled.

VI. Conclusions and Future Work

This paper shows the results of an investigation of how NTGA2 with *Gap* selection effectively uses *archive*s in solving multi-objective problems with constraints (MS-RCPSP and TTP). There are five answered research questions: the size of the problem instance and number and how aspects, and problems (TTP and MS-RCPSP) differ to determine the *Gap* (%) selection parameter. The main conclusion is that in some cases (instances), the *archive* (and the Gap selection)

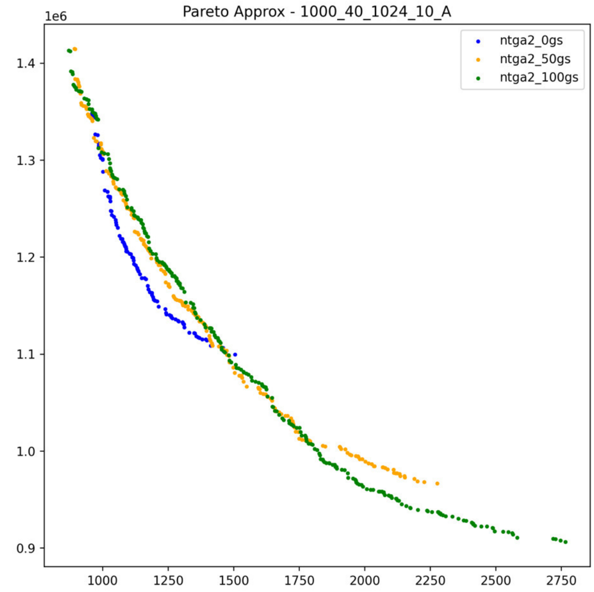


Fig. 3. Comparison of PFA for MS-RCPSP – 1000_40_1024_10_A

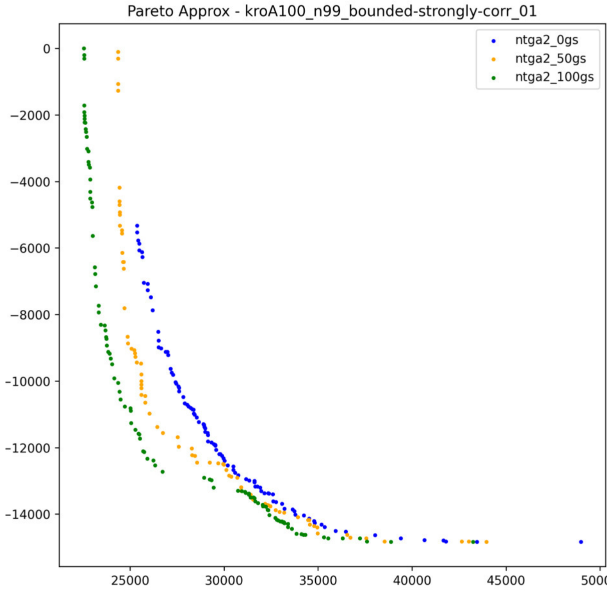


Fig. 4. Comparison of PFA for TTP – kroA100_n99_bounded-str.-corr_01

plays a crucial role, and the population could be eliminated. Experimental results presented a correlation between the *Gap* selection effectiveness and constraints density, as well as optimization progress.

There are several promising future directions of research. The *GS* in most cases (TTP and MS-RCPSP) prefers gap selection 100%, but in some cases (i.e. a large number of constraints) reduces to 50%. It encourages further work on Gap adaptation in NTGA2 to increase final NTGA2 effectiveness. Moreover, we empirically showed that such a situation occurs in two benchmark problems (TTP and MS-RCPSP), and plan to investigate other multi-objective problems (including many-objective).

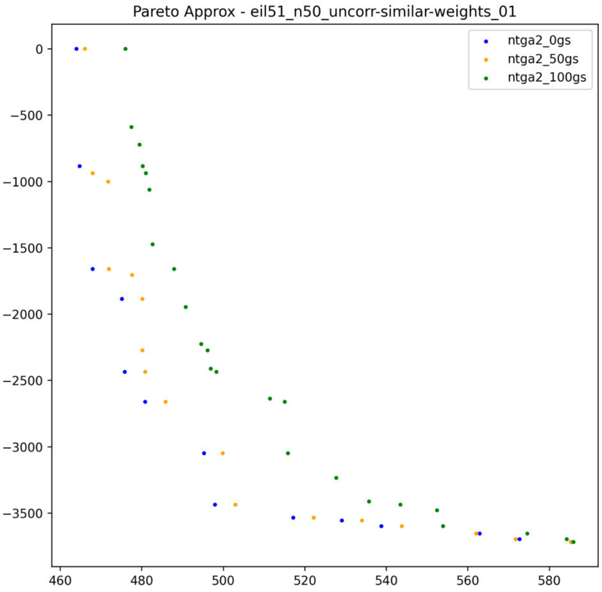


Fig. 5. Comparison of PFA for TTP – eil51_n50_uncorr-similar-weights_01

References

[1] Antkiewicz, M., and Myszkowski, P.B, and Laszczyk M. "GaMeDE2—improved Gap–based Memetic Differential Evolution applied to Multimodal Optimization." 2022 17th Conference on Computer Science and Intelligence Systems (FedCSIS). IEEE, 2022.

[2] Blank, J., Deb, K. & Mostaghim, S. Solving the Bi-objective Traveling Thief Problem with Multi-objective Evolutionary Algorithms. *Evolutionary Multi-Criterion Optimization*. pp. 46-60 (2017)

[3] Bonyadi, M., Michalewicz, Z. & Barone, L. The travelling thief problem: The first step in the transition from theoretical problems to realistic problems. *2013 IEEE Congress On Evol. Comp.*. pp. 1037-1044 (2013)

[4] Das, I. & Dennis, J. Normal-Boundary Intersection: A New Method for Generating the Pareto Surface in Nonlinear Multicriteria Optimization Problems. *SIAM Journal On Optimization*. **8**, 631-657 (1998)

[5] Deb, K., Pratap, A., Agarwal, S. & Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions On Evolutionary Computation*. **6**, 182-197 (2002)

[6] M.Laszczyk, P.B. Myszkowski, "A gap–based memetic differential evolution (gamede) applied to multi–modal optimisation–using multi–objective optimization concepts", in: Intel. Inf. and Database Systems: 13th Asian Conf., ACIIDS 2021, Phuket (Thailand) 2021, Proc. 13, pp. 211–223.

[7] Myszkowski, P. & Laszczyk, M. "Diversity based selection for many-objective evolutionary optimisation problems with constraints". *Information Sciences*. **546** pp. 665-700 (2021)

[8] Myszkowski, P.B. & Laszczyk, M. "Investigation of benchmark dataset for many-objective Multi-Skill Resource Constrained Project Scheduling Problem". *Applied Soft Computing*. **127** pp. 109253 (2022)

[9] Myszkowski, P.B. & Laszczyk, M. Survey of quality measures for multi-objective optimization: Construction of complementary set of multi-objective quality measures. *Swarm And Evolutionary Computation*. **48** pp. 109-133 (2019)

[10] Vijayan, N. & Al. Taguchi's Parameter Design: A Panel Discussion. *Technometrics*. **34**, 127-161 (1992)

[11] Zitzler, Eckart, Marco Laumanns, and Lothar Thiele. "SPEA2: Improving the strength Pareto evolutionary algorithm." TIK-report 103 (2001).

[12] Zhang, Qingfu, and Hui Li. "MOEA/D: A multiobjective evolutionary algorithm based on decomposition." IEEE Transactions on evolutionary computation 11.6 (2007): 712-731.

[13] Zhang, Jun, et al. "A survey on algorithm adaptation in evolutionary computation." Frontiers of Electrical and Electr. Eng. 7 (2012): 16-31.

Communication Papers of the 18th Conference on Computer Science and Intelligence Systems pp. 29–35
DOI: 10.15439/2023F8409
ISSN 2300-5963 ACSIS, Vol. 37

Real-Time Detection of Small Objects in Automotive Thermal Images with Modern Deep Neural Architectures

Tomasz Balon*, Mateusz Knapik†‡ and Bogusław Cyganek†
*Department of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering,
Email: *tbalon@student.agh.edu.pl*
†Department of Computer Science, Electronics and Telecommunication
Email: *mknapik@agh.edu.pl, cyganek@agh.edu.pl*
AGH University of Science and Technology,
Al. Mickiewicza 30, 30–059 Kraków, Poland
‡MyLED Inc.
Email: *m.knapik@myled.pl*
Ul. W. Łokietka 14/2, 30–016 Kraków, Poland

***Abstract*—Thermal imaging has shown great potential for improving object detection in automotive settings, particularly in low light or adverse weather conditions. To help and further develop this industry, we extend our previously shared Thermal Automotive Dataset by more than 2000 new images and 2 novel object detecting models based on YOLOv5 and YOLOv7 architecture. We point how important is the size of the dataset. Additionally, we compare the performance of both models, to see which is more reliable and superior in terms of detecting small objects in thermal spectrum. Furthermore, we analysed how preprocessing affects thermal imaging dataset and models basing on it. The new dataset is available free from the Internet.**

I. Introduction

THE introduction of deep learning has revolutionized the computer vision field, bringing about remarkable advancements in object recognition and paving the way for significant progress in various domains. One particularly crucial area that greatly benefits from accurate and efficient object detection is the development of self-driving vehicles. With the ability to analyze the surrounding environment in real-time, these vehicles rely heavily on robust object detection algorithms to make informed and safe decisions [1], [2]. Although the prevailing source of information still constitute digital cameras, operating in visual spectrum, in the recent years the far infrared, so called thermovision cameras are gaining on importance [3]. In this paper we focus on this type of signals.

In a previous article [4], a thermal automotive dataset was introduced, specifically designed for object detection using the YOLOv5 model [5]. However, the presented dataset had certain limitations, as it contained only images captured during winter conditions. Nonetheless, even with this constraint, the dataset proved to be valuable for training object detection models and laying the foundation for further advancements in the field. Additionally, the previous article introduced the model based on YOLOv5 architecture.

In order to overcome mentioned limitations and push the boundaries of object detection in thermal automotive applications, we present an expanded thermal automotive dataset. This enhanced dataset incorporates over 2,000 new images, capturing a broader range of scenarios and weather conditions. By expanding the dataset, we aim to provide a more comprehensive and diverse collection of images, better reflecting the challenges faced in real-world automotive environments.

Furthermore, we introduce a novel object detection model, the YOLOv7, which builds upon the foundation of its predecessor, the YOLOv5. The YOLOv7 model incorporates improvements in architecture and training strategies, aiming to enhance object detection accuracy and speed [6]. By comparing the performance of the new YOLOv7 model with the previous YOLOv5 model, using the expanded dataset for evaluation, we can assess which model is superior in terms oh object detection in thermal imaging.

Moreover, we delve into the impact of dataset size on model training by conducting experiments with both the YOLOv5 and YOLOv7 models. We compare the performance of the models trained on the entire expanded dataset against those trained on only half of the dataset. This analysis allows us to examine the influence of dataset size on the training outcomes, shedding light on the relationship between dataset scale and object detection performance.

By undertaking this study, our objective is to contribute to the ongoing efforts aimed at enhancing object detection accuracy and speed in the automotive industry. Through the utilization of an expanded thermal automotive dataset and the introduction of the YOLOv7 model, we aspire to facilitate the development of safer, more efficient, and more reliable self-driving cars. Ultimately, our research aims to propel the advancement of autonomous driving systems, intelligent transportation, and the broader field of computer vision in the automotive sector. Our new dataset is available free from the Internet [7].

II. Network architectures

A. You Only Look Once v5

The YOLOv5 deep convolutional neural network introduces novel advancements building upon breakthroughs in computer vision, particularly inspired by YOLOv4 [8] and other state-of-the-art approaches. Notably, YOLOv5 adopts the New CSP-Darknet53 structure as its backbone, an evolved version of the Darknet architecture used in previous iterations.

Furthermore, both YOLOv4 and YOLOv5 employ the CSP Bottleneck, originally proposed by WongKinYiu in the Cross Stage Partial Networks (CSP) paper [9], for feature formulation. The CSP architecture, built upon DenseNet [10], is designed to overcome challenges such as vanishing gradients in deep networks, facilitate feature propagation, encourage feature reuse, and reduce the number of network parameters. In CSPResNext50 and CSPDarknet53, the DenseNet structure has been tailored to separate the feature map of the base layer, thereby mitigating computational bottlenecks and enhancing learning by directly passing an unedited feature map to the subsequent stage.

YOLOv5 draws insights from YOLOv4's research inquiry to determine the optimal neck architecture. Both YOLOv4 and YOLOv5 feature the PA-NET neck for effective feature aggregation, where each "Pi" represents a feature layer in the CSP backbone. Other improvement is the auto-learning of YOLO anchor boxes when custom data is input, eliminating the need for manual anchor box tuning.

One of the main contributions of YOLOv5 repository is an introduction of a model scaling, first proposed in EfficientNet paper [11]. In contrast to conventional approach, that employ arbitrary changes in model architecture, proposed scaling method uniformly adjusts the network in depth, by changing the number of convolutional blocks repetitions, as well as in width, by changing number of filters in selected layers, using a set of fixed scaling coefficients. The rationale behind the compound scaling method is grounded in the intuitive understanding that to improve performance, the network requires additional layers to expand the receptive field and more channels to capture finer patterns in the larger image. YOLOv5 offers different pre-trained model sizes (e.g., YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x), which are variants of the same architecture, but with different scaling parameters, balancing computational costs and memory requirements.

B. You Only Look Once v7

The main focus of the introduced advancements in YOLOv7 was to achieve a superior balance between performance and efficiency in real-time object detection.

One of the key advancements in YOLOv7 is the adoption of the Efficient Layer Aggregation Networks (ELAN) architecture as its backbone. ELAN considers memory access cost and analyzes factors such as input/output channel ratio, number of branches, and element-wise operations. This careful analysis leads to reducing gradient propagation path, resulting in faster and more accurate network inference, significantly improving the overall efficiency of the model. Moreover, gradient flow propagation paths also aids the module level re-parameterization.

YOLOv7 also revisits the idea of auxiliary head proposed in the Inception paper [12], that aids the initial model training as well as reduces the vanishing gradient problem. Authors experiment with varying degree of supervision for aux head, settling on a coarse-to-fine definition where supervision is passed back from the lead head at different granularities.

The concept of model scaling is further refined by the authors, by compound scaling depth and width as well as layer concatenation. As shown by ablation studies, this technique keeps the model architecture optimal while scaling for different sizes. Based on this, YOLOv7 provides different models (e.g. YOLOv7-tiny, YOLOv7-X, YOLOv7-E6, YOLOv7-W6), that have various size and scaling parameters. Each version is tailored to different hardware configurations and requirements, allowing users to choose the one that best suits their specific needs and computing resources.

III. Data acquisition and description

A. Data acquisition

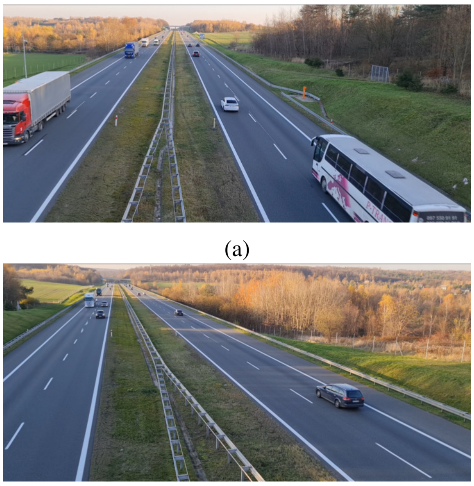
The video footage used in this study was captured using the FLIR® A35 thermal imaging camera. The data acquisition process was conducted during an autumn afternoon, specifically between 2:30 PM and 3:15 PM, when the ambient temperature ranged from 12°C to 14°C under clear weather conditions. The recording setup involved capturing real-life traffic scenes at high speeds. To achieve this, the camera was strategically positioned on an elevated bridge overlooking the road. In Figure 1, the provided images illustrate camera's field of view, showcasing the prevailing weather conditions and providing an approximate depiction of the time of day. These meticulous details ensure that the dataset encompasses realistic scenarios and accurately represents the thermal imaging perspective in a dynamic traffic environment.

B. Dataset description

Provided dataset extension consists of approximately 2000 annotated images with bounding boxes for 4 classes: car, motorcycle, bus and truck. In order to maintain consistency with previous version of dataset, all possible photos parameters were kept as they were - resolution of 320x256 pixels and 8-bit grayscale colors. For annotation we used DarkLabel [13] software and kept the same class IDs.

In total, the dataset contains over 8000 images and approximately 35000 annotations divided into 5 different classes, as shown on Figure 3.

New images introduce not only new weather conditions, but also other factors. The inclusion of highway traffic in our dataset brings forth several key factors that differentiate it from traditional city traffic datasets. Firstly, the higher speeds at which vehicles are traveling introduce motion blur, making



(a)

(b)

Figure 1: Camera's field of view

the detection task more demanding. Additionally, the bigger presence of larger vehicles, such as trucks compared to regular city traffic.

Dataset, together with object detecting models, is publicly available under the link: https://home.agh.edu.pl/~cyganek/AutomotiveThermo2_0.zip.

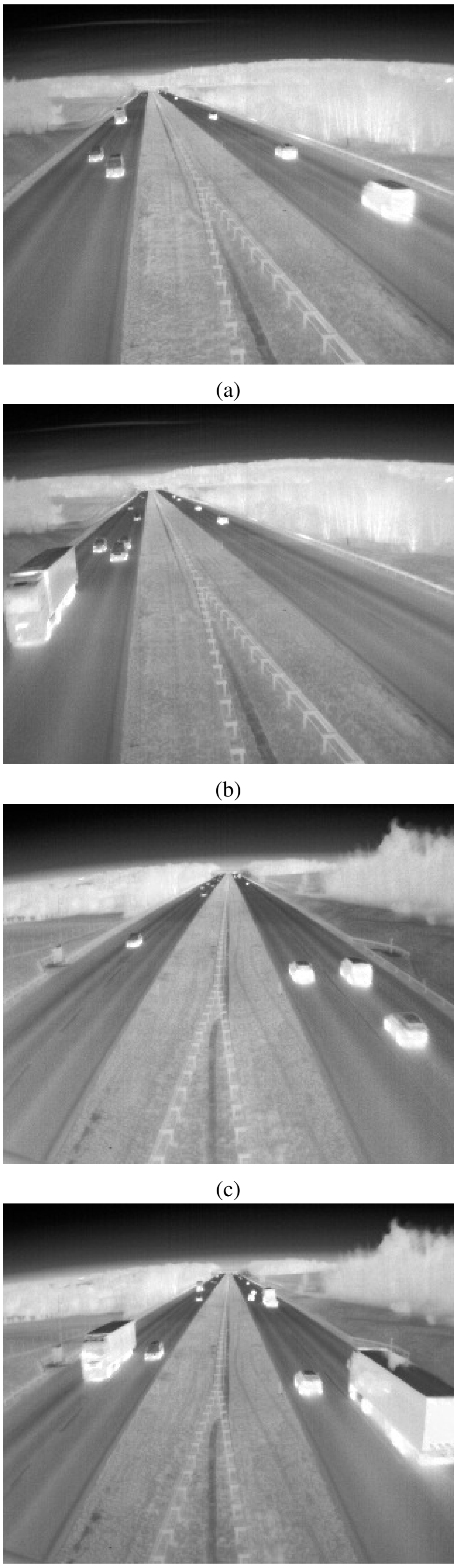
C. Data structure

Alongside with this paper, dataset and object detection models are provided. Dataset contains a total of over 8000 images divided into train, val and test subsets, each being separate folder. Additionally, trained YOLOv5 and YOLOv7 based models are published.

D. Object detection model training

To ensure a fair and comprehensive comparison between the YOLOv5 and YOLOv7 models, we adopted a systematic training approach. For the YOLOv5-based model, we retrained the previously published model, based on YOLOv5-M size architecture, on the complete thermal automotive dataset. This enabled us to evaluate the model's performance on the same dataset used for the YOLOv7 model.

Similarly, for the YOLOv7 model, we aimed to maintain consistency in the training process. As default YOLOv7 model size is comparable to YOLOv5-L, we have decided to scale it down, so that the final models have similiar number of parameters and FLOPS. Therefore, for all our tests, we set depth scaling parameter to 0.67 and width scaling parameter to 0.75. This results in a model that has computational requirements of 60.4 GFLOPS (49.2 GFLOPS for YOLOv5-M), 21.26 million parameters (21.19 million for YOLOv5-M) spread across 415 layers (291 layers in YOLOv5-M). We initially trained it using a subset of the dataset to establish a baseline performance. This step allowed us to gauge the model's initial capabilities before incorporating the expanded dataset. Subsequently, we retrained the YOLOv7 model, taking advantage of the new



(a)

(b)

(c)

(d)

Figure 2: New sample images from the dataset

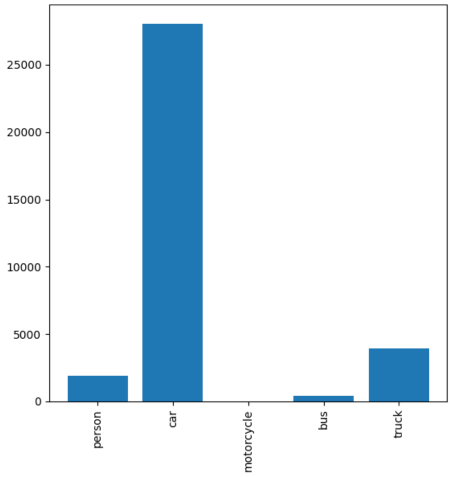


Figure 3: Number of instances in each class.

training examples in the dataset. This sequential training procedure facilitated a comprehensive analysis of the model's performance improvement with the addition of more data.

It is worth noting that both models underwent an initial pretraining phase on the COCO dataset, a widely used benchmark in computer vision. This pretraining step provided a foundation for the models to learn general object detection capabilities before being fine-tuned on the specific thermal automotive dataset. By leveraging the pretrained models, we harnessed the prior knowledge gained from the COCO dataset to enhance the object detection performance of both the YOLOv5 and YOLOv7 models on the thermal automotive dataset.

IV. Experimental part and models comparison

To evaluate the performance of the new dataset and compare the YOLOv5 and YOLOv7 models, we conducted several experiments using different training configurations.

A. Size of training dataset

Firstly, we tested how does dataset size affect training results. We took pretrained models on previous dataset and trained them using only half of the new dataset. Then we again took previously trained models and trained them on the entire new dataset. To avoid overfitting and yet to achieve best results in models training, all were trained for 50 epochs.

Results of training all four models are presented in Figure 4. These images present precision, recall, and mean average precision (mAP). As clearly visible, each model was increasing its' accuracy as with successive epochs. At first, advancements were made rather rapidly to then slow down while coming to the end of training, which was expected. Final numeric results are summarized in Tables I and II.

Although the differences between using only half or entire dataset are not very substantial, they display the overall trend – the more data available, the more accurate the model is. These numbers also show that using only part of the dataset, provides us with acceptable results which might be enough for object detection. However, we aim higher than that. The main goal is for the model to be as accurate and as robust as possible.

Model	Dataset	Precision	Recall	mAP	
				0.5	**0.5:0.95**
YOLOv5	Half	0.951	**0.971**	0.990	0.715
	Entire	**0.984**	0.965	**0.992**	**0.726**
YOLOv7	Half	0.834	0.899	0.933	0.587
	Entire	0.913	0.864	0.945	0.602

Table I: YOLOv5 and YOLOv7 models training results

Model	Dataset	Precision	Recall	mAP	
				0.5	**0.5:0.95**
YOLOv5	Half	0.976	0.985	0.994	0.722
	Entire	**0.989**	**0.995**	**0.995**	**0.749**
YOLOv7	Half	0.982	0.895	0.987	0.610
	Entire	0.903	0.970	0.984	0.626

Table II: YOLOv5 and YOLOv7 evaluation results on test subset

B. YOLOv5 vs YOLOV7

After examination of what impact does dataset size have on training results, we compared the two mentioned architectures. Head-to-head numeric results are stored in Tables I and II.

Advancements made to YOLO architecture between v5 and v7, would suggest newer version to be more accurate and have better results than it's predecessor. However it is not reflected in our results. According to outcome received after training both models, YOLOv5 outperforms YOLOv7. Particularly in mAP_0.5:0.95 – 0.726 for YOLOv5 in contrary to 0.602 for YOLOv7. The remaining results, although also in favor of YOLOv5, are not as substantial as mean average precision. These lead to a conclusion that in case of small 8-bit grey scale images, YOLOv5 would be more reasonable to use, rather than the newer YOLOv7.

During the training process of the YOLOv7 model on our thermal automotive dataset, we observed a sudden drop in precision, recall, and mAP scores. This unexpected decline in performance raised the need for investigation to identify the potential reasons behind this phenomenon. We search through known issues with the YOLOv7 implementation (and YOLOv5, as v7 codebase is heavily based on a code released by Ultralytics) code repository. We discovered that similiar problem was present in v5 code [14] and was possibly a code error triggered by very small objects present in our dataset. It was subsequently fixed in later releases, but it seems that it was transfered to the v7 repository when forked [15].

We tried to mitigate it, firstly by changing different hyperparameters such as different losses, learning rate ComputeLossOTA, as those might have lead to miscalculating loss function. Unfortunately though, changing values of these hyperparameters did not result in great improvement. It only shifted the sudden drop in epochs (e.g. drop happening in 3rd epoch, not 23rd).

C. Virtual High Dynamic Range

In our pursuit of further enhancing the quality and information content of our thermal automotive dataset, we explored the

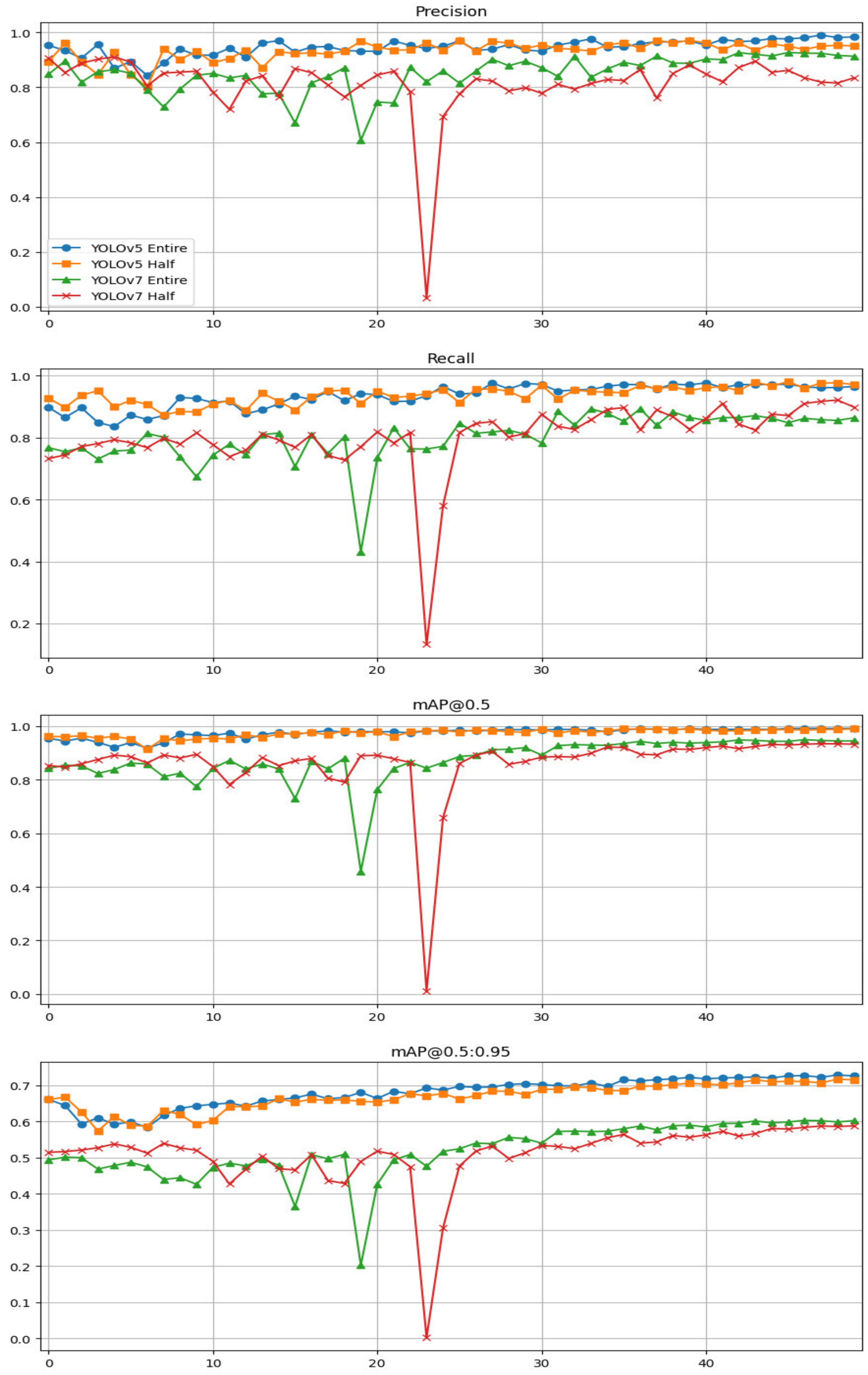


Figure 4: YOLOv5 and YOLOv7 models train results

implementation of the Virtual High Dynamic Range (VHDR) technique. Cyganek et al. [16] proposed another approach towards the VHDR method for images enhancement. To receive an VHDR image, an LDR image is taken as an input and is being processed by a set of tone adjustment curves to potentially reveal hidden details. Then it is fused to HDR image. Lastly image range conversion and contrast enhancement is done.

Based on our previous positive results with this kind of image preprocessing [1], [2], we applied the VHDR technique to our dataset and trained both the YOLOv5 and YOLOv7 models on this augmented dataset. However, in this case the results did not demonstrate a significant improvement in object detection performance. The mAP scores for both models remained relatively unchanged when compared to the models trained on the original dataset.

V. Discussion

Our results show that the YOLOv5 model outperforms YOLOv7 in terms of object detection accuracy on our thermal automotive dataset. This is an interesting observation, because v7 is both faster and achieves better results on regular datasets [6]. However, this improvements were generated by means of training procedure optimization and techniques like model re-parametrization and dynamic label assignments [6]. These can lead to increase in performance, but it also needs sufficiently big dataset to achieve that. When other modalities are used, such as long wave infrared, obtaining large scale training datasets is often unfeasible or even impossible. Older methods, such as YOLOv5, are less prone to such problems as their architecture is less data-specific. Additionally, the spatial resolution of thermal images is relatively low, posing a significant challenge for object detection, similar to the task of detecting small objects in RGB images. Furthermore, the limited input channel in thermal images further decreases the availability of extracted features during the initial stages of the network. However, the YOLOv5 algorithm addresses this issue by incorporating a unique first layer known as the Focus layer [17]. The primary purpose of this layer is to mitigate the impact of the small number of input channels compared to the significantly larger number of feature maps in deeper layers of the network. This is achieved by dividing the input layers into odd columns and rows, which are then redistributed as additional channels, enhancing the representation of features, similarly conceptually to dilated convolutions. Interestingly, we also found that YOLOv5 achieved good performance when trained on half the dataset, suggesting that it could be a more practical choice for those with limited computational resources. Furthermore, when the entire dataset is used, YOLOv5 also performs better, indicating that it is the better choice for smaller datasets and less demanding applications.

In a parallel investigation, Yang [18] conducted a comprehensive analysis comparing the performance of YOLOv5, YOLOv6, and YOLOv7 models. Interestingly, Yang's findings align closely with our own research, as he observed that the YOLOv6 model exhibited superior performance compared to its counterparts. This convergence in results reinforces the efficacy of the YOLOv6 model and underscores its potential for advancing object detection capabilities in various domains.

Olorunshola et al. [19] conducted comparable investigations in the field, focusing on the performance of the YOLOv5 and YOLOv7 models. Their study employed the Google Open Images Dataset, incorporating specific classes such as Person, Handgun, Rifle, and Knife. Although their dataset comprised slightly larger color images in contrast to our thermal dataset, their findings echoed our own observations: YOLOv5 exhibited superior performance across various metrics, with the exception of Recall. These parallel outcomes indicate a consistent trend in the comparative analysis of YOLOv5 and YOLOv7, further affirming the potential advantages of YOLOv5 in object detection tasks.

VI. Conclusion

In this article, we introduced an expanded thermal automotive dataset with approximately 2,000 new images and classes, and a new object detection model based on YOLOv7. We compared the performance of the new model with the previous YOLOv5 model using the expanded dataset, and provided insights into the acquisition process of the dataset. We made our newest dataset available free from the Internet [7].

The results showed that the YOLOv5 model outperformed the YOLOv7 model in terms of accuracy. This study contributes to the development of safer and more efficient self-driving cars by providing a better tool for object detection.

Future work will involve expanding the dataset further, including adding images taken in different weather conditions such as summer, which presents a harsher environment for thermal imaging. These additions will help to improve the robustness and versatility of our dataset and enable the development of more accurate and reliable object detection models for thermal automotive images.

In summary, this study provides valuable insights into the use of advanced object detection models and thermal imaging for object detection in the automotive industry. The expanded thermal automotive dataset and both YOLOv5 and YOLOv7 models introduced in this article can be used as a benchmarks for future research in this field.

Acknowledgments

This work has been supported by the AGH University of Science and Technology under subvention funds no. 16.16.230.434.

References

[1] M. Knapik and B. Cyganek, "Fast eyes detection in thermal images," *Multimedia Tools and Applications*, vol. 80, no. 3, pp. 3601–3621, 2021.

[2] ——, "Driver's fatigue recognition based on yawn detection in thermal images," *Neurocomputing*, vol. 338, pp. 274–292, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231219302280

[3] M. A. Farooq, W. Shariff, D. O'callaghan, A. Merla, and P. Corcoran, "On the role of thermal imaging in automotive applications: A critical review," *IEEE Access*, vol. 11, pp. 25 152–25 173, 2023.

[4] T. Balon, M. Knapik, and B. Cyganek, "New thermal automotive dataset for object detection," *Annals of Computer Science and Information Systems*, vol. 31, pp. 43–48, 2022.

[5] G. Jocher, A. Chaurasia, A. Stoken, J. Borovec, NanoCode012, Y. Kwon, TaoXie, J. Fang, imyhxy, K. Michael, Lorna, A. V, D. Montes, J. Nadar, Laughing, tkianai, yxNONG, P. Skalski, Z. Wang, A. Hogan, C. Fati, L. Mammana, AlexWang1900, D. Patel, D. Yiwei, F. You, J. Hajek, L. Diaconu, and M. T. Minh, "ultralytics/yolov5: v6.1 - TensorRT, TensorFlow Edge TPU and OpenVINO Export and Inference," Feb. 2022. [Online]. Available: https://doi.org/10.5281/zenodo.6222936

[6] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," 2022.

[7] "Automotivethermo2_0," 2023. [Online]. Available: https://home.agh.edu.pl/~cyganek/AutomotiveThermo2_0.zip

[8] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," 2020.

[9] C.-Y. Wang, H.-Y. M. Liao, I.-H. Yeh, Y.-H. Wu, P.-Y. Chen, and J.-W. Hsieh, "Cspnet: A new backbone that can enhance learning capability of cnn," 2019.

[10] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," 2018.

[11] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," 2020.

[12] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," 2014.

[13] "Darklabel annotation software," https://github.com/darkpgmr/DarkLabel, accessed: 2022-06-07.

[14] "Sudden performance decrease in training · Issue #5721 · ultralytics/yolov5 — github.com," https://github.com/ultralytics/yolov5/issues/5721, 2021, [Accessed 31-07-2023].

[15] "Unstable training · Issue #974 · WongKinYiu/yolov7 — github.com," https://github.com/WongKinYiu/yolov7/issues/974, 2022, [Accessed 31-07-2023].

[16] B. Cyganek and M. Woźniak, "Virtual high dynamic range imaging for underwater drone navigation," *Proceedings of the 6th IIAE International Conference on Industrial Application Engineering*, pp. 393–398, 2018.

[17] A. Song, Z. Zhao, Q. Xiong, and J. Guo, "Lightweight the focus module in yolov5 by dilated convolution," in *2022 3rd International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications (CVIDL & ICCEA)*, 2022, pp. 111–114.

[18] L. Yang, "Investigation of you only look once networks for vision-based small object detection," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 4, 2023. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2023.0140410

[19] O. E. Olorunshola, M. E. Irhebhude, and A. E. Evwiekpaefe, "A comparative study of yolov5 and yolov7 object detection algorithms," *Journal of Computing and Social Informatics*, vol. 2, no. 1, pp. 43–48, 2023.

Communication Papers of the 18th Conference on Computer Science and Intelligence Systems pp. 37–41
DOI: 10.15439/2023F7645
ISSN 2300-5963 ACSIS, Vol. 37

Polyhedral Tiling Strategies for the Zuker Algorithm Optimization

Wlodzimierz Bielecki, Marek Palkowski, Piotr Blaszynski, Maciej Poliwoda
West Pomeranian University of Technology in Szczecin
ul. Zolnierska 49, 71-210 Szczecin, Poland
Email: mpalkowski@zut.edu.pl

***Abstract*—In this paper, we focus on optimizing the code for computing the Zuker RNA folding algorithm. This bioinformatics task belongs to the class of non-serial polyadic dynamic programming, which involves non-uniform program loop dependencies. However, its dependence pattern can be represented using affine formulas, allowing us to automatically employ tiling strategies based on the polyhedral method. We use three source-to-source compilers Pluto, Traco, and Dapt based on affine transformations, transitive closure of dependence relation graph and space-time tiling, respectively, to automatically generate cache-efficient codes. We evaluate the speed-up and scalability of optimized codes and check their performance employing applying two multi-core machines. We also discuss related approaches and outline future work in the conclusion of the paper.**

I. Introduction

RNA secondary structure prediction is a fundamental and noticeably time-consuming problem in the biological computing. For a given RNA sequence, the secondary non-crossing RNA structure is predicted such that the total amount of free energy is minimized. Smith and Waterman [1], and Nussinov et al. [2] first defined a dynamic programming algorithm for RNA folding. Instead of using the free energy, the algorithms of [1], [2] aim to maximize the number of complementary base pairs.

Zuker et al. [3] first proposed a complex dynamic programming algorithm to predict the most stable secondary structure for a single RNA sequence by computing its minimal free energy. It uses a ”nearest neighbor” model. The algorithm estimates of thermodynamic parameters for neighboring interactions. The main idea is that the loop entropies are used to score all possible structures and the secondary structure of an RNA sequence consists of four fundamental independent substructures: stack, hairpin, internal loop, and multi-branched loop. The energy of a secondary structure is assumed to be the sum of the substructure energies.

Zuker's algorithm consists of two steps. The first step, which is the most time-consuming, involves calculating the minimal free energy of the input RNA sequence using recurrence relations as outlined in the provided formulas. The second step involves performing a trace-back to recover the secondary structure with the base pairs. While the second step is not a computationally demanding task, optimization of the energy matrix calculation in the first step is crucial for improving the overall performance of the algorithm [4].

Zuker defines two energy matrices, $W(i,j)$ and $V(i,j)$, with $\mathcal{O}(n^2)$ pairs (i,j) satisfying the constraints $1 \leq i \leq N$ and $i \leq j \leq N$, where N is the length of a sequence. $W(i,j)$ represents the total free energy of a sub-sequence defined by indices i and j, while $V(i,j)$ represents the total free energy of a sub-sequence starting at index i and ending at index j if i and j form a pair, otherwise $V(i,j) = \infty$.

The main recursion of Zuker's algorithm for all *i, j* with $1 \leq i < j \leq N$, where N is the length of a sequence, is the following.

$$W(i,j) = \begin{cases} W(i+1,j) & (1) \\ W(i,j-1) & (2) \\ V(i,j) & (3) \\ \min\limits_{i<k<j} \{W(i,k) + W(k+1,j)\} & (4) \end{cases}$$

Below, we present the computation of V.

$$V(i,j) = \begin{cases} eH(i,j) & (5) \\ V(i+1,j-1) + eS(i,j) & (6) \\ \min\limits_{\substack{i \leq i' \leq j' \leq j \\ 2 < i'-i+j-j' < d}} \{V(i',j') + eL(i,j,i',j')\} & (7) \\ \min\limits_{i<k<j-1} \{W(i+1,k) + W(k+1,j-1)\} & (8) \end{cases}$$

eH (hairpin loop), *eS* (stacking) and *eL* (internal loop) are the structure elements of energy contributions in the Zuker algorithm.

The computation of Equations 1, 2, 3, 5, 6 takes $\mathcal{O}(n^2)$ steps. Equations 4 and 8 requires $\mathcal{O}(n^3)$ steps. The time complexity of a direct implementation of this algorithm is $\mathcal{O}(n^4)$ because we need $\mathcal{O}(n^4)$ operations to compute Equation 7. This formulation as a computational kernel involves float arrays and operations.

The computation domain and dependencies for Zuker's recurrence cell (i,j) are more complex than those of Nussinov's recurrence. Equations 3, 4, and 8 generate long-range (non-local) dependencies for cell (i,j), while the other equations have short-range (local) dependencies. The computation of the element *V(i',j')* in Equation 3 spans a triangular area of several dozens to hundreds of cells.

Listing 1 shows the affine loop nest for finding the minimums of the V and W energy matrices.

Thematic track: Computer Aspects of Numerical Algorithms

Listing 1. Zuker's recurrence loop nest

```
for (i = N-1; i >= 0; i--){
   for (j = i+1; j < N; j++) {
      for (k = i+1; k < j; k++){
         for(m=k+1; m <j; m++){
            if(k-i + j - m > 2 && k-i + j - m < 30)
               V[i][j] = MIN(V[k][m] + EL(i,j,k,m), V[i][j]);                // Eq. 3
         }
         W[i][j] = MIN ( MIN(W[i][k], W[k+1][j]), W[i][j]);                  // Eq. 8
         if(k < j-1)
            V[i][j] = MIN(W[i+1][k] + W[k+1][j-1], V[i][j]);                 // Eq. 4
      }
      V[i][j] = MIN( MIN (V[i+1][j-1] + ES(i,j), EH(i,j), V[i][j]);         // Eq. 1,2
      W[i][j] = MIN( MIN ( MIN ( W[i+1][j], W[i][j-1]), V[i][j]), W[i][j]);  // Eq. 5,6,7
   }
}
```

In this paper, we focus on studying the performance of tiled Zuker loop nests codes generated by chosen automatic optimizers based on the polyhedral model.

Loop tiling, also known as loop blocking or loop partitioning, is a program transformation technique used in compiler optimization to enhance cache utilization and improve the performance of loop-based computations [5]. It involves dividing a loop into smaller, tile-sized sub-loops or blocks that fit into the cache effectively. The main idea behind loop tiling is to exploit spatial locality, which refers to accessing data elements that are close together in memory. By dividing a loop into smaller tiles, the loop iterations within each tile can reuse data elements, reducing cache misses and improving memory access patterns.

The polyhedral model represents loop nests as polyhedra with affine loop bounds and schedules. It enables advanced loop transformations and analysis of data dependences. By leveraging this model, compilers can automatically optimize loops, improve performance (especially locality employing loop tiling), and exploit parallelism [6].

II. Related work

The Zuker kernel, as well as the Nussinov RNA folding, involves mathematical operations over affine control loops whose iteration space can be represented by the polyhedral model [7]. However, the Zuker RNA folding acceleration is still a challenging task for optimizing compilers because that code is within non-serial polyadic dynamic programming (NPDP), which is a particular family of dynamic programming with non-uniform data dependences, and it, as mentioned above, is more difficult to be optimized [8]. In addition, the loop structure of Zuker's algorithm is definitely more complicated for automatic tiling strategies than that of Nussinov's algorithm, i.e. the loops are quadruple nested with more instructions which also implies a larger number of data dependencies (including non-uniform ones).

There are other RNA folding numerical approaches which can be presented within the polyhedral model. To enhance the accuracy of structure prediction for a given RNA sequence, the algorithm devised by Zhi J. Lu and colleagues in 2009 incorporates the concept of Maximum Expected Accuracy (MEA), utilizing base pair and unpaired probabilities [9]. This method employs a Nussinov-like recursion, drawing upon the probabilities obtained through John S. McCaskill's algorithm [10]. Numerical sources and aspects of the Nussinov, Zuker, and MEA algorithms can be found in the NPDP Benchmark Suite [11]. It is a collection of NPDP tasks which cannot be effectively optimized using commonly employed tiling strategies, such as diamond tiling [12], [13].

An interesting cache-efficient manual solution for Nussinov's RNA folding algorithm was proposed by Li and colleagues in [14]. Using lower and unused part of Nussinov's, they changed column reading to more efficient row reading. Diagonal scanning exposes parallelism in the output code. The method is known also as *Transpose* technique. In our previous paper [15], we adopted the *Transpose* to optimize Zuker's code. In equations 4 and 8, there are not cache-efficient column reading of the W array, $W[k+1][j]$ and $W[k+1][j-1]$, respectively. The transpose method changes these array accesses to the row reading and adds the following statement $W[col][row] = W[row][col]$ to make a transposed copy of the cells in the lower-left triangle.

Zhao et al. [16] improved the *Transpose* method and performed the experimental study of the energy-efficient codes for Zuker's algorithm. The approach based on the LRU cache model requires about half as much memory as does Li's Transpose. However, the authors did not present parallel codes for the *ByBox* strategy and any automatic optimization was not proposed.

Pluto is a widely-used, advanced tool for optimizing C/C++ programs through the use of polyhedral code generation. It transforms the source code into parallelized, coarse-grained code that is optimized for data locality, primarily using the affine transformation framework (ATF). This state-of-the-art source-to-source compiler is highly regarded in the field for its effectiveness in improving the performance of parallel software. Unfortunately, Pluto fails to achieve maximal code locality and performance for the well-known NPDP problems [8]. It is unable to tile the innermost loop of Nussinov's RNA folding, which is a key to cache locality optimization [7], [17].

It cannot produce parallel code for the McCaskill probabilistic RNA folding kernel [18]. For the Zuker code presented in Listing 1, the approach is unable to tile the 3rd loop nest.

Authors of Pluto, Bondhugula and et al. [7] presented dynamic tiling for the Zuker's optimal RNA secondary structure prediction [7]. 3-d iterative tiling for dynamic scheduling is calculated using reduction chains. Operations along each chain can be reordered to eliminate cycles in an inter-tile dependence graph. Their approach involves dynamic scheduling of tiles, rather than the generation of a static schedule.

Wonnacott et al. introduced 3-d tiling of "mostly-tileable" loop nests of RNA secondary-structure prediction codes in paper [17]. This approach extracts non-problematic statement instances in the loop nest iteration space, i.e., those that can be safely tiled by means of well-known techniques. The reminding statement instances should be run serially to preserve all the dependences available in the loop nest. Unfortunately, the approach is limited to serial codes only. The idea is presented only for simpler Nussinov's RNA folding which maximizes the number of complementary base pairs.

In past, we developed the tiling technique [8] aimed to transform (corrects) original rectangular tiles into target ones, which are valid under lexicographic order. Tile correction is performed by means of the transitive closure of loop dependence graphs. Loop skewing is used to parallelize code. We achieved a higher speed-up of generated tiled code in comparison with that produced with state-of-the-art source-to-source optimizing compilers. However, the transitive closure is a NP-difficult problem and is not always computable in general case.

Tiling correction [8] and Four-Russian RNA Folding [19] were deeply studied by Tchendji and et al. and they proposed a parallel tiled and sparsified four-Russians algorithm for Nussisov's RNA Folding [20]. They claim that this approach computation is more cache-friendly because it applies the blocks of Four-Russians mustered into parallelogram-shaped tiles. The experimental study for CPUs and massively GPUs architectures shows the out-performance in comparison to the results of [8] and [19]. Although, the authors considered manually the Nussinov loop nest only, they promised to study other NPDP problems in future.

The space-time loop tiling approach presented in paper [21] generates target tiles using the intersection operation to sets representing sub-spaces and time slices is applied. Each time partition comprises independent iterations, which can be executed in parallel while time partitions should be enumerated in lexicographical order. The presented approach is a continuation of the work on space-time tiling, which shows promising possibilities in developing new polyhedral optimizing compilers. The codes were generated with the Dapt compiler introduced in paper [22].

III. Experimental study

To carry out experiments, we used a machine with a processor AMD Epyc 7542, 2.35 GHz, 32 cores, 64 threads, 128MB Cache, and machine with a processor Intel Xeon Gold 6240, 2.6GHz (3.9GHz turbo), 18 cores, 36 threads, 25MB Cache. The optimized codes were compiled by means of the GNU C++ compiler version 9.3.0 with the -O3 flag.

TABLE I
Execution times (in seconds) for AMD Epyc 7542 and 64 threads.

Size	Classic	Transpose	Pluto	TileCorr	Space-time
1000	26.90	3.31	3.26	7.88	1.80
1500	132.87	14.08	12.33	25.97	8.22
2000	415.15	42.65	30.99	58.70	22.33
2500	1013.99	100.60	69.19	118.45	53.08
3000	2093.22	202.43	137.49	217.01	109.73
3500	3871.90	370.90	245.93	356.61	201.75
4000	6589.03	626.56	407.87	578.37	342.78
4500	10544.30	998.58	644.83	874.90	550.97
5000	15686.70	1515.55	977.20	1272.33	853.73

TABLE II
Execution times (in seconds) for Intel Xeon Gold 6240 and 36 threads.

Size	Classic	Transpose	Pluto	TileCorr	Space-time
1000	27.31	4.28	2.96	6.87	2.23
1500	135.38	17.16	19.54	29.53	14.29
2000	393.88	43.56	23.87	41.75	18.55
2500	954.87	102.06	50.39	85.09	40.95
3000	1970.48	206.51	97.42	156.89	81.35
3500	3644.10	378.81	180.23	266.01	151.19
4000	6209.09	654.14	300.47	426.64	259.77
4500	9944.50	1048.81	489.30	649.73	416.07
5000	15133.74	1589.30	819.77	968.87	634.46

Tests were conducted using ten randomly generated RNA sequences with lengths ranging from 1000 to 5000. Discussion in papers [8], [14] shows that cache-efficient code performance does not change based on strings themselves, but it depends on the size of a string.

We compared the performance of tiled codes generated with the presented approaches i) *Pluto* parallel tiled code (based on affine transformations) [23], ii) tile code based on the *Space-time* technique [21] generated with Dapt, iii) tiled code based on the correction technique *TileCorr* [8] generated with Traco, iv) Li manual cache-efficient implementation of Zuker's RNA folding *Transpose* [14]. All codes are multithreaded within the OpenMP standard [24].

The tile size 16×16×1×16 for Pluto code was chosen empirically (Pluto does not tile the third loop) as the best among many sizes examined. The tile 16×16×16×16 size for tile correction technique was chosen according to paper [15]. For the space-time tiled code, we chose the same tile sizes. Our preliminary empirical testing did not yield improved tile sizes for this algorithm.

Table 1 presents execution times in seconds for ten sizes of

Fig. 1. Speed-up for AMD Epyc 7542 and 64 threads.

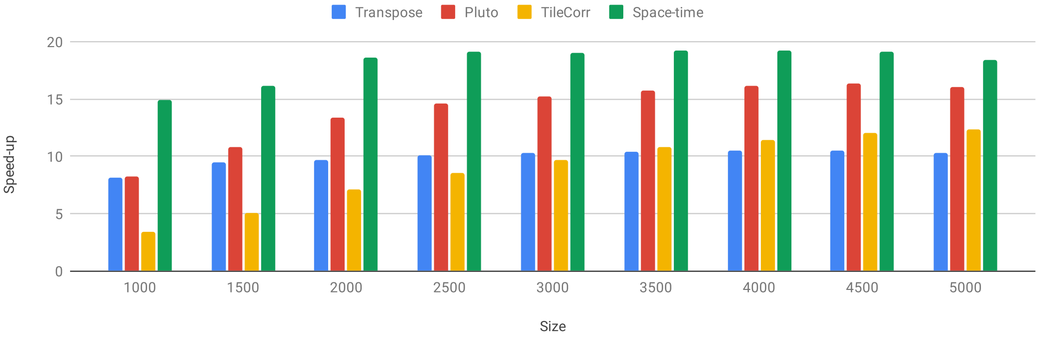
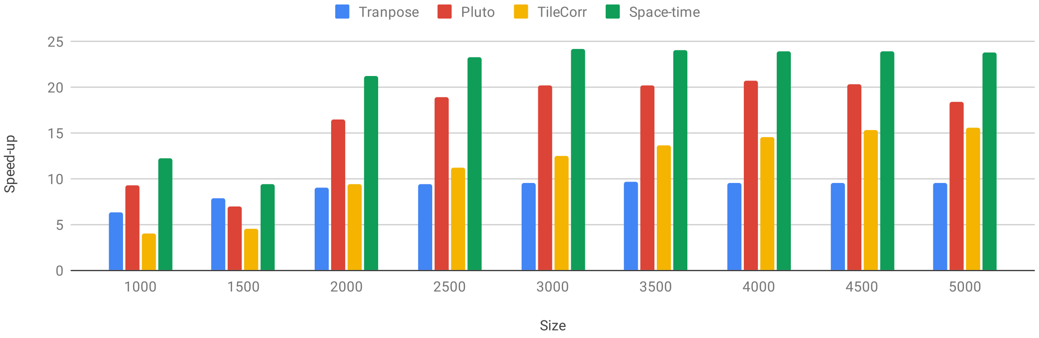


Fig. 2. Speed-up for Intel Xeon Gold 6240 and 36 threads.



RNA sequence using AMD Epyc 7542. Problem sizes from 1000 to 5000 (roughly the size of the longest human mRNA) were chosen to illustrate advantages for smaller and larger instances. Output codes are executed for 64 threads. We can observe that the presented space-time tiling approach allows for obtaining cache-efficient tiled code, which outperforms significantly the other examined implementations for each RNA strands lengths. The second most efficient code is loop tiling produced by the Pluto compiler. Figure 1 depicts the speed-up for times presented in Table I.

Table 2 presents execution times in seconds using two processors Intel Xeon E5-2695 v2 and 48 threads. The presented space-time tiling strategy outperforms strongly the other studied techniques for all RNA strands lengths. Transpose technique allows us to obtain faster code than the ATF tiled code and the tile correction code with this machine. Figure 2 depicts speed-ups for time executions in Table 2.

At the address https://github.com/markpal/zuker, all source codes used in the experimental study are available.

IV. Conclusion

Summing up, the space-time tiled code we introduced allows for improved and scalable performance on both of the multi-core processors, regardless of the number of threads or problem size. The output codes were generated automatically based on the input serial code. The space-time tiling strategy implemented within the polyhedral compiler Dapt appears to be a promising solution for optimizing NPDP tasks, and we plan to examine its use on other NPDP bioinformatics problems.

References

[1] T. Smith and M. Waterman, “Identification of common molecular subsequences,” *Journal of Molecular Biology*, vol. 147, no. 1, pp. 195 – 197, 1981.

[2] R. Nussinov *et al.*, “Algorithms for loop matchings,” *SIAM Journal on Applied mathematics*, vol. 35, no. 1, pp. 68–82, 1978.

[3] M. Zuker and P. Stiegler, “Optimal computer folding of large rna sequences using thermodynamics and auxiliary information.” *Nucleic Acids Research*, vol. 9, no. 1, pp. 133–148, 1981.

[4] G. Lei, Y. Dou, W. Wan, F. Xia, R. Li, M. Ma, and D. Zou, “CPU-GPU hybrid accelerating the zuker algorithm for RNA secondary structure prediction applications,” *BMC Genomics*, vol. 13, no. Suppl 1, p. S14, 2012. doi: 10.1186/1471-2164-13-s1-s14

[5] J. Xue, *Loop Tiling for Parallelism*. Norwell, MA, USA: Kluwer Academic Publishers, 2000. ISBN 0-7923-7933-0

[6] S. Verdoolaege, “Integer set library - manual,” Tech. Rep., 2011. [Online]. Available: www.kotnet.org/~skimo//isl/manual.pdf,

[7] R. T. Mullapudi and U. Bondhugula, “Tiling for dynamic scheduling,” in *Proceedings of the 4th International Workshop on Polyhedral Compilation Techniques*, Vienna, Austria, Jan. 2014.

[8] M. Palkowski and W. Bielecki, “Parallel tiled Nussinov RNA folding loop nest generated using both dependence graph transitive closure and loop skewing,” *BMC Bioinformatics*, vol. 18, no. 1, p. 290, 2017. doi: 10.1186/s12859-017-1707-8

[9] Z. J. Lu, J. W. Gloor, and D. H. Mathews, “Improved RNA secondary structure prediction by maximizing expected pair accuracy,” *RNA*, vol. 15, no. 10, pp. 1805–1813, Aug. 2009. doi: 10.1261/rna.1643609. [Online]. Available: https://doi.org/10.1261/rna.1643609

[10] J. S. McCaskill, "The equilibrium partition function and base pair binding probabilities for RNA secondary structure," *Biopolymers*, vol. 29, no. 6-7, pp. 1105–1119, may 1990. doi: 10.1002/bip.360290621

[11] M. Palkowski and W. Bielecki, "NPDP benchmark suite for loop tiling effectiveness evaluation," in *Parallel Processing and Applied Mathematics*. Springer International Publishing, 2023, pp. 51–62. [Online]. Available: https://doi.org/10.1007/978-3-031-30445-3_5

[12] T. Malas, G. Hager, H. Ltaief, H. Stengel, G. Wellein, and D. Keyes, "Multicore-optimized wavefront diamond blocking for optimizing stencil updates," *SIAM Journal on Scientific Computing*, vol. 37, no. 4, pp. C439–C464, Jan. 2015. doi: 10.1137/140991133. [Online]. Available: https://doi.org/10.1137/140991133

[13] U. Bondhugula, V. Bandishti, and I. Pananilath, "Diamond tiling: Tiling techniques to maximize parallelism for stencil computations," *IEEE Transactions on Parallel and Distributed Systems*, vol. 28, no. 5, pp. 1285–1298, May 2017. doi: 10.1109/tpds.2016.2615094

[14] J. Li, S. Ranka, and S. Sahni, "Multicore and GPU algorithms for Nussinov RNA folding," *BMC Bioinformatics*, vol. 15, no. 8, p. S1, 2014. doi: 10.1186/1471-2105-15-S8-S1. [Online]. Available: http://dx.doi.org/10.1186/1471-2105-15-S8-S1

[15] M. Palkowski and W. Bielecki, "Parallel tiled cache and energy efficient code for zuker's RNA folding," in *Parallel Processing and Applied Mathematics*. Springer International Publishing, 2020, pp. 25–34.

[16] C. Zhao and S. Sahni, "Efficient RNA folding using zuker's method," in *2017 IEEE 7th International Conference on Computational Advances in Bio and Medical Sciences (ICCABS)*. IEEE, oct 2017. doi: 10.1109/iccabs.2017.8114309

[17] D. Wonnacott, T. Jin, and A. Lake, "Automatic tiling of "mostly-tileable" loop nests," in *5th International Workshop on Polyhedral Compilation Techniques*, Amsterdam, 2015.

[18] M. Palkowski and W. Bielecki, "Parallel cache-efficient code for computing the McCaskill partition functions," vol. 18, pp. 207–210, 2019. doi: 10.15439/2019F8

[19] Y. Frid and D. Gusfield, "An improved Four-Russians method and sparsified Four-Russians algorithm for RNA folding," *Algorithms for Molecular Biology*, vol. 11, no. 1, aug 2016. doi: 10.1186/s13015-016-0081-9

[20] V. K. Tchendji, F. I. K. Youmbi, C. T. Djamegni, and J. L. Zeutouo, "A parallel tiled and sparsified Four-Russians algorithm for Nussinov's RNA folding," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, pp. 1–12, 2022. doi: 10.1109/tcbb.2022.3216826

[21] M. Palkowski and W. Bielecki, "Tiling Nussinov's RNA folding loop nest with a space-time approach," *BMC Bioinformatics*, vol. 20, no. 1, apr 2019. doi: 10.1186/s12859-019-2785-6

[22] W. Bielecki, M. Palkowski, and M. Poliwoda, "Automatic code optimization for computing the McCaskill partition functions," in *Annals of Computer Science and Information Systems*. IEEE, sep 2022. doi: 10.15439/2022f4

[23] U. Bondhugula *et al.*, "A practical automatic polyhedral parallelizer and locality optimizer," *SIGPLAN Not.*, vol. 43, no. 6, pp. 101–113, Jun. 2008. [Online]. Available: http://pluto-compiler.sourceforge.net

[24] OpenMP Architecture Review Board, "OpenMP application program interface version 5.2," 2022. [Online]. Available: https://www.openmp.org/specifications/

DOI: 10.15439/2023F5691
ISSN 2300-5963 ACSIS, Vol. 37

Can ChatGPT Replace a Template-based Code Generator?

Adam Bochenek
0009-0005-1259-3418
National Information Processing Institute
al. Niepodległości 188b, 00-608 Warsaw, Poland
Email: adam.bochenek@opi.org.pl

***Abstract*—This article examines whether a large language model (LLM) tool, such as ChatGPT, can replace a template-based source code generator. To this end, we conducted an experiment in which we attempted to replace an existing template-based DAO class generator (which creates entity classes and a repository for a specified database table) with a solution in which templates of target classes were presented to ChatGPT alongside the source model. We then instructed ChatGPT to generate new classes. A novelty in this work is an attempt at two-stage cooperation with ChatGPT: first we provide the pattern, then we fill it. The experiment proved that, at present, such a solution yields results that are neither predictable nor replicable, and successive attempts to execute the same commands returned wildly varying results. ChatGPT randomly recognises the rules that are present in templates, and complex instructions impact the generated results negatively. At present, classic code generation methods yield markedly superior results.**

I. Introduction

THIS article aims to determine whether, and if so, to what degree, large language model (LLM) can work as a source code generator [2]. Can the manual creation and subsequent population of a template be replaced by an LLM tool, such as ChatGPT, which has been taught to read patterns and treat them as templates to be populated with specific data?

Software engineers strive to ensure that the degree of abstraction in which they operate is as close as possible to the concepts that are present during the stages of application analysis and modelling. This approach streamlines work, increases productivity, and reduces the number of potential errors. Typically, this involves increasing the degree of abstraction and applying concepts related to a specific domain directly and as broadly as possible during the software development stage.

One method of achieving this goal involves the application of a model-driven development (MDD) methodology, such as domain-specific modelling (DSM). The essence of this methodology lies in its capability to model and specify the target application at an abstraction degree that caters to the needs of experts and analysts who are familiar with the given domain [1]. When such specification is completed, the final product (i.e. application code) should be generated automatically. The manner in which the code is generated is the focal point of our study.

The findings of this article should be considered an experiment. We examine whether the classic method of code generation based on available templates can be replaced by an LLM-based tool (for the purposes of this article, we used ChatGPT in its 2023, Mar 14 version).

In this work, we first list the areas of knowledge that will be of interest to us, i.e. Model Driven Development, code generation and the use of LLM in the field of programming. Then we move on to the description of the SDSM method, the Osfald tool and the experiment itself, which will consist in using ChatGPT as a template-based code generator. The most important point is the description of the results of the experiment and the conclusions drawn from it.

II. Related work

A. Model-driven development, domain-specific modelling, and template-based code generators

Model Driven Development (MDD) is a set of application development methodologies in which models are used as the basis of the entire software development process. It involves creating a model that describes the complete system, or a fragment thereof, which is then used as a base for generating source code (source model -> source code).

In MDD, models are typically created using formal notation, such as unified modelling language (UML), business process model and notation (BPMN), or domain-specific language (DSL). These models are then used in automated code generation. The advantages of MDD include the capability to use domain-specific concepts during the design stage, automated software production, increased effectiveness, higher quality, and streamlined change management. The weaknesses include the high cost of implementation and mandatory specialist knowledge of formal languages.

DSM is a variant of MDD. In DSM, models are created to describe specific domains. We use languages that are specific for the given domain and serve as the basis for the generation of the code and other project artefacts. DSM's chief strength lies in domain-specific languages and models typically being easier to understand than general-purpose ones [1].

The DSM approach comprises three key elements:

- the model
- the code generator
- the framework.

Template-based code generation (TBCG) is a method of automated source code generation based on templates or patterns [14], which are powered by models. In the TBCG approach, software developers define source code templates that contain particular variables or parameters; subsequently, such templates are populated with specific values to create the source code for the given project.

Despite its obvious advantages, such as exceptionally quick generation of source code and improvement of the code's quality by minimising errors caused by manual typing, TBCG also entails its own set of flaws. The most notable include:

- the complexity of code template creation and management
- the complicated handling of complex design problems that require a more algorithmically advanced approach
- necessary knowledge of template language and associated software development libraries.

With these flaws in mind, we examined whether the TBCG method could be replaced with the capabilities offered by LLMs.

B. Codex, Copilot, GPT-3

Rapid progress in the creation and use of LLMs has created new opportunities for automation of the software development process. The current approach, domain-level modelling and automated transformation into source code [5], [6], [7], [8], has been supplemented with methodologies that utilise machine learning and LLMs [9], [10], [11]. New codes are created in response to commands formed in natural language, or as attempts to supplement or complete existing code.

Both approaches have their flaws. Due to their nature, DSL languages match specific domains, and will never become general-purpose tools; generative LLMs have difficulty extracting complex coding patterns from code corpora, and often generate codes riddled with syntax or semantic errors [12], [13]. The results returned by either model are seldom predictable or replicable.

The experiment described in this article attempted to combine both approaches. We wanted the code generated to correspond to the specified pattern, and to be predictable and replicable. With consideration for the complex and time-consuming nature of template creation, we attempted to substitute it by providing an LLM with an example or a set of examples to be used as a pattern, which could then be modified after the provision of a new, different set of parameters.

Our experiment is different from the typical use of ChatGPT as a developer helper. Usually, ChatGPT is supposed to generate the source code in a given programming language based on the given natural language prompt.

C. Code generation vs. security

While analysing the methods of automated code generation, we must not omit one crucial aspect: security. Language models and tools that are based on them, such as GitHub Copilot, have been trained on tremendous quantities of open source code. This code contains errors, so the concern that the code suggested by Copilot may potentially contain errors that affect application security is a valid one. The experiment described in [4] demonstrated that in a trial that covered eighty-nine scenarios in which Copilot was used to generate 1,689 applications, as many as 40% of them contained security vulnerabilities.

We believe that a template- or pattern-based method is a significantly safer solution. When creating a template, we can verify its safety; the code generated on the basis of a safe template will also be safe, in most cases.

III. Methods

A. The simplified domain-specific modelling (SDSM) method

At OPI PIB (National Information Processing Institute - National Research Institute), we have developed our own, in-house application development method. It is based on the DSM approach, but is simplified. We call it simplified domain-specific modelling (SDSM). As with DSM, this method is also based on three key elements:

- the model
- the code generator
- the organisation- and domain-specific environment (framework).

The perception of the first component, the model, differs from that of the classic DSM. We treat the model as input data for the code generator. We neither require nor define any formal language that describes the solution at a higher degree of abstraction. We do not define any rules or syntaxes. We do not use DSL at all; instead:

- we create simple models that are understood as embedded at the level of the data structure domain
- we search for existing models. We often discover that raw or processed data, which can act as a model (input data) for the code generator, already exists.

In the SDSM method, application generation comprises four stages:

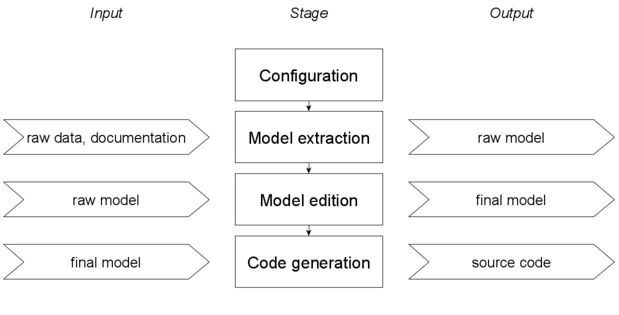


Fig. 1. SDSM stages and input and output artifacts.

Stage one, **configuration**, is an auxiliary step which enables all operations that prepare the development process to be conducted properly. This might involve specifying the place from which the data that serves a model will be read, establishing a link to a database, or specifying where documentation is stored. This step is optional. During stage two, the **model extraction** stage, we read information from an external source that has been configured in the previous step and use it to build the model. Stage three is **model edition**: the model obtained during stage two may require modifications or additions. Stage three may also be used to create a new model from scratch if there is no source from which the model can be obtained. The data structure used to generate the code is created and edited during this stage. Stage four involves **source code generation** based on the model prepared during the previous stages. Stages (and their input/output artifacts) are shown in Fig. 1.

B. The Osfald tool

OPI PIB's SDSM method is implemented by the Osfald tool, an application that acts as a framework, and offers ready-to-use, universal components and functions that are required to create code generators. Osfald is also a set of interfaces that are a recipe for an SDSM-type generator. Here, a generator is understood as an implementation (a set of classes that implement created interfaces) that enables us to progress through all stages of application development; in other words, it extracts and edits the model, and generates new code based on the model. Generators may pertain to different application elements and offer various degrees of complexity. Defining a new generator chiefly involves implementing such previously-developed interfaces.

C. TBCG-type DAO generator vs. LLM

One of the best and most complete examples that demonstrates the SDSM concept in action is the generation of the data access layer for a typical business application written in Java. Although implementation details differ depending on the libraries used, this layer typically handles two basic class types: entity classes and repository classes. On the application side, the entity class represents one row in a database table. Its primary component is a field list. The repository class is a set of methods that includes basic methods that correspond to the create, read, update, and delete (CRUD) functions, complemented with additional functions used to search for entities in accordance with specific criteria.

The generator fulfils the following tasks (by SDSM stage):

1) Configuration: in this case, establishing a connection to the database
2) Model extraction: by using standard JDBC mechanisms in Java, we read the structure details of selected database tables (field names, their types, lengths, and requirements)
3) Model edition: for each table field, a corresponding entity field is generated that bears a default (albeit editable) name and type. During stages two and three, a data model, which acts as input for the generator, is created
4) Code generation: based on previously-defined templates and the model prepared during stages two and three, entity source code and a repository (and, optionally, a test class) are generated in the specific part of the application. The existing TBCG-based generator uses the Apache Velocity engine and templates to produce code.

Example 3.1: **EnGptUser** as an example of an entity class.

```
public class EnGptUser {
  private long idAuto;
  private String idUid;
  private String firstName;
  private String surname;
  private Integer age;
}
```

Example 3.2: **RepoGptUser** as an example of a repository class.

```
public class RepoGptUser extends BaseRepo
 ↪  {
  public RepoGptUser(Trx trx)
  public void create(EnGptUser en)
  public void update(EnGptUser en)
  public List<EnGptUser> findAll()
  public EnGptUser findByKey(String key)
  private void entity2Stmt(EnGptUser en,
   ↪  PreparedStatement stmt, boolean
   ↪  update)
  protected void rs2Entity(ResultSet rs,
   ↪  EnGptUser en)
}
```

The entity class consists of fields that correspond to database table fields for which a specific entity has been created. The **EnGptUser** class is an example of an entity class (Example 3.1).

The repository class **RepoGptUser** (Example 3.2) is more complex, as it requires a base class that it expands (e.g. a repository that is based on a specific library).

D. Template table, entity, and repository

The experiment described in this article involved attempting to replace the existing template-based entity and repository class generator (the TBCG method) for a specific database table with an LLM-based solution. ChatGPT (Mar 14 version) was the LLM tool used in this test. In place of templates, we prepared a template database table with corresponding template entity and repository class. The table contained all field type variants, and the entity class contained all possible mappings of those variants to Java types. The term 'variants of mapping' refers to the principle according to which table fields that allow null values are mapped to Java object types, while table fields that do not allow null values are mapped to Java primitive types. We apply this principle in the TBCG templates, and we expected the LLM model to detect and apply the principle similarly.

Example 3.3: The **template_table_01** template table.

```
CREATE TABLE template_table_01
(
  id_auto bigint NOT NULL,
  id_uid character varying(32) NOT NULL,
  string_field_a character varying(255)
   ↪  NOT NULL,
  string_field_b character varying(10000),
  text_field_a text NOT NULL,
  text_field_b text,
  bool_field_a boolean NOT NULL,
  bool_field_b boolean,
  ...
  ...
  CONSTRAINT template_table_01_pkey
   ↪  PRIMARY KEY (id_uid),
  CONSTRAINT template_table_01_id_auto_key
   ↪  UNIQUE (id_auto)
)
```

Example 3.4: The **EnTemplateTable01** template entity class.

```
public class EnTemplateTable01 {
  private long idAuto;
  private String idUid;
  private String stringFieldA;
  private String stringFieldB;
  private String textFieldA;
  private String textFieldB;
  private boolean boolFieldA;
  private Boolean boolFieldB;
  ...
  ...
}
```

The template database table **template_table_01** (Example 3.3) contained all field type variants that we wanted our generator to handle.

Class **EnTemplateTable01** (Example 3.4) corresponds to the template table. Each field in this class correspond to a table field. Note that the NOT NULL fields in the table correspond to primitive Java types (e.g. int, long, double), while the places where the database permits NULL values correspond to object types (Integer, Long, Double). We expected ChatGPT to detect and recognise this rule.

The repository class is the most complex. Below, we present one of its key methods, which include mapping the result of SQL query to entity: **rs2Entity** (Example 3.5).

In the **rs2Entity** (and **entity2Stmt**) methods, the most important thing is to match the types correctly (for example: **getInt**, **getIntNull**, **setInt**, **setIntNull**).

The methods responsible for calling the SQL queries are used to add or modify new rows, and to run searches based on criteria entered are the basic repository methods: **create** (Example 3.6), **update**, **findAll**, and **findByKey**.

IV. Experiments

We tested whether TBCG mechanism can be replaced by ChatGPT. This task was divided into two stages: generation of entity classes and generation of repository classes.

A. Stage one: generation of entity classes

Stage one involved attempting to use ChatGPT to generate an entity (a Java class) based on the provided table structure (in SQL). The entity class generated in this way was to correspond to a specific template. First, we provided ChatGPT with the **template_table_01** template table structure and its corresponding **EnTemplateTable01** template entity class. We then asked ChatGPT to use the template to create a new entity for a different SQL structure provided.

To this end, we prepared five different database tables; for each of them, ChatGPT generated an entity. The following criteria were applied to verify the generated entity's correctness and consistency with the template:

- correct class syntax in Java, including compilation readiness
- completeness (whether all fields were included and in the correct order)
- field types (allowing for correct separation into simple and object types, which depends on whether the database allows null values)
- result replicability (whether repeated generation of the entity yields identical source code; three attempts).

When designing the experiment, we ensured that the tables tested would be sufficiently diverse. Our findings are presented below (Table I. Stage 1, entity class generation).

The **Table** column contains the names of the database tables for which an entity class was generated. The **Syntax** column contains information on whether the code generated was correct syntax-wise. For all tables, we received code that could be compiled. The values in the **Completeness** column inform us whether the generated class contained all database table fields. The values in the **Field order** column inform us whether the fields in the entity appear in the same order as in the source structure.

The next two columns pertain to entity field types. The values in the **Field types** column demonstrate whether all generated entity class fields had the expected type that resulted from the template. Inconsistent types appeared during the generation of the class for the **time_and_bool** table. The values in the **Field types (null/not null)** column tell us whether the generated code is divided correctly into simple and object Java types, as defined in the template—which depends on whether the database allows empty field values. The rightmost column shows whether repeated entity generation by ChatGPT yielded identical code. In the last two tables, we observed differences pertaining to field types.

B. Stage two: generation of repository classes

Stage one was completed with moderate success. Although most tasks were completed correctly, some errors occurred. In stage two, the bar was raised higher. Based on the database

Example 3.5: The **rs2Entity** method of the template repository.

```
void rs2Entity(ResultSet rs, EnTemplateTable01 en) {
    en.setIdAuto(StmtGet.getLong(rs, "id_auto"));
    en.setIdUid(StmtGet.getString(rs, "id_uid"));
    en.setStringFieldA(StmtGet.getString(rs, "string_field_a"));
    en.setStringFieldB(StmtGet.getString(rs, "string_field_b"));
    en.setTextFieldA(StmtGet.getString(rs, "text_field_a"));
    en.setTextFieldB(StmtGet.getString(rs, "text_field_b"));
    en.setBoolFieldA(StmtGet.getBoolean(rs, "bool_field_a"));
    en.setBoolFieldB(StmtGet.getBooleanNull(rs, "bool_field_b"));
    ...
    ...
}
```

Example 3.6: The **create** method of the template repository.

```
public void create(EnTemplateTable01 en) {
  executeWrite(
    " insert into public.template_table_01 ( " +
    " id_uid, string_field_a, string_field_b, text_field_a, text_field_b, " +
    " bool_field_a, bool_field_b, int_field_a, int_field_b, long_field_a, " +
    " long_field_b, float_field_a, float_field_b, double_field_a, double_field_b, " +
    " numeric_field_10_4_a, numeric_field_10_4_b, numeric_field_8_2_a, " +
    " numeric_field_8_2_b, date_field, time_field, timestamp_tz_field, " +
    " timestamp_field " +
    " ) values ( " +
    " ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, " +
    " ?, ? " +
    " )"
    ,
    (stmt) -> {
      entity2Stmt(en, stmt, false);
    }
  );
}
```

TABLE I
STAGE 1, ENTITY CLASS GENERATION

Table	Syntax	Completeness	Field order	Field types	Field types (null / not null)	Replicability
the simplest	ok	ok	ok	ok	ok	ok
two ints	ok	ok	ok	ok	ok	ok
lot of numbers	ok	ok	ok	ok	ok	ok
time and bool	ok	ok	ok	ERROR	ok	ERROR
complete reversed	ok	ok	ok	ok	ERROR	ERROR

table structure, we attempted to generate a repository class. This class's methods must ensure correct communication with the database table, and the class itself must also correctly convert the data stored in and read from the table to the values stored in the entity object. Unlike in stage one, three elements must be compatible: the repository class, the entity class, and the database table. Stage two was completed in two steps.

1) Stage two, step one – verification of task feasibility: The purposes of step one were to determine whether ChatGPT was capable of completing this task, and to decide what approach should be adopted. Several variants of this solution were tested (each variant was tested three times):

Variant 1. The first attempt involved applying the method used during stage one to generate entities. In the first prompt addressed to ChatGPT, we defined the structure of the template database table; in the second prompt, we defined the template repository class for the table; in the third prompt, we instructed ChatGPT to generate a repository class for another table with defined parameters. In response to the third prompt, we always received a repository class written in the Python language—despite the template being written in Java.

Variant 2. In this variant, we specified that we wanted the class to be generated in Java. Although the class returned was in Java, ChatGPT ignored the requirement to match the predefined template. The generated class corresponded to typical repository templates found in the literature. ChatGPT ignored the class template that we had provided.

Variant 3. We decided not to ask ChatGPT to generate a new class, but to modify the existing class in such a manner that it handled a table of a different structure. Once again, the class returned was in the Python language and its operation failed to meet our requirements (Example 4.1).

Variant 4. This variant comprised a single, complete prompt that contained the template table, entity, and repository class, as well as the target entity and table. We asked ChatGPT to generate the target repository. We received a Java repository class that was correct, but did not correspond to our model in any way; instead, ChatGPT used a template that utilised the popular Spring library and the JdbcTemplate class (Example 4.2).

Variant 5. In this variant, we changed our approach and stopped trying to generate a complete repository class; instead, we reduced our scope to single methods. We divided the class into its individual methods and attempted to generate the repository class, one method at a time. The order of method generation was important. This variant was the only one that returned satisfactory results that merited more thorough examination and description.

Example 4.1: Class **RepoTheSimplest** in Python.

```
import psycopg2

class RepoTheSimplest:
  def __init__(self, conn):
    self.conn = conn

  def create_table(self):
    cur = self.conn.cursor()
    cur.execute("""
      CREATE TABLE IF NOT EXISTS
       ↪ the_simplest (
        id_auto BIGINT NOT NULL GENERATED
         ↪ BY DEFAULT AS IDENTITY (
         ↪ INCREMENT 1 START 1 MINVALUE 1
         ↪ MAXVALUE 9223372036854775807
         ↪ CACHE 1 ),
        id_uid CHARACTER VARYING(32)
         ↪ COLLATE pg_catalog."default"
         ↪ NOT NULL,
        one_single_field INTEGER,
        CONSTRAINT the_simplest_pkey
         ↪ PRIMARY KEY (id_uid),
        CONSTRAINT
         ↪ the_simplest_id_auto_key
         ↪ UNIQUE (id_auto)
        );
    """)
    self.conn.commit()
    cur.close()
```

2) Stage two, step two: generation of successive class methods: Step two of the stage two was completed by applying Variant 5 in the following way: the entire task of generating the repository class was divided into sub-tasks, each of which corresponded to the generation of one specific method. For each method, we defined separate correctness criteria. All of the operations described below were performed three times (iterations 1, 2, and 3) to study the replicability of the results.

Replicability in our experiments is one of the key features. When we have a template and the data to fill it, we want to make sure that the fill operation gives exactly the same result every time. Without this certainty, the use of templates makes little sense.

Below is the list of the criteria for the correct generation of each method:

1) The **entity2Stmt** method
   a) K1.1 Correct method syntax (Java)
   b) K1.2 Correct field list (names, order)
   c) K1.3 Correct types (including null/not null).
2) The **rs2Entity** method

Example 4.2: Class **RepoTheSimplest** using Spring.

```
@Repository
public class RepoTheSimplest {

  private final JdbcTemplate jdbcTemplate;

  @Autowired
  public RepoTheSimplest(JdbcTemplate
   ↪  jdbcTemplate) {
    this.jdbcTemplate = jdbcTemplate;
  }

  public void create(EnTheSimplest entity)
   ↪  {
    String sql = "INSERT INTO the_simplest
     ↪  (id_uid, one_single_field) VALUES
     ↪  (?, ?)";
    jdbcTemplate.update(sql,
     ↪  entity.getIdUid(),
     ↪  entity.getOneSingleField());
  }
```

   a) K2.1 Correct method syntax (Java)
   b) K2.2 Correct field list (names, order)
   c) K2.3 Correct types (including null/not null).

3) The **create** method
   a) K3.1 Correct method syntax (Java)
   b) K3.2 Correct query syntax (SQL)
   c) K3.3 Correct field list (names, order)
   d) K3.4 Correct en parameter type
   e) K3.5 Correct use of entity2Stmt.
4) The **update** method
   a) K4.1 Correct method syntax (Java)
   b) K4.2 Correct query syntax (SQL)
   c) K4.3 Correct field list (names, order)
   d) K4.4 Correct en parameter type
   e) K4.5 Correct use of entity2Stmt.
5) The **findAll** method
   a) K5.1 Correct method syntax (Java)
   b) K5.2 Correct query syntax (SQL)
   c) K5.3 Correct type of the entity list returned
   d) K5.4 Correct use of rs2Entity.
6) The **findByKey** method
   a) K6.1 Correct method syntax (Java)
   b) K6.2 Correct query syntax (SQL)
   c) K6.3 Correct type of the entity list returned
   d) K6.4 Correct use of rs2Entity.

TABLE II
RESULTS FOR THE **THE_SIMPLEST** TABLE (ERRORS ONLY)

Criterion	Result (iter. 1)	Result (iter. 2)	Result (iter. 3)
K1.3	ERROR	ERROR	ERROR
K2.3	ERROR	ERROR	ERROR
K6.4	ok	ERROR	ok

TABLE III
REPLICABILITY OF THE CODE GENERATED FOR THE **REPOTHESIMPLEST** CLASS METHODS

Method	Replicability
entity2Stmt	yes
rs2Entity	yes
create	yes
update	yes
findAll	yes
findByKey	NO

C. Analysis of the results

The results are in Tables II, III, IV and V. Stage one, the creation of entity classes, resulted, for all test tables, in the generation of syntactically correct and complete Java classes that contained all fields expected in the target structure. Field order was also correct. However, for one of the classes, ChatGPT selected wrong field types; for another, it misapplied the type selection rule (primitive vs. object) relative to whether the database permits null values. The results of the experiment in stage two were less optimistic. First, multiple attempts were necessary to formulate the commands in a manner that would result in ChatGPT generating a new repository based on the template provided. Eventually, we were forced to compromise: instead of having ChatGPT generate a complete repository class, we opted to have it create the individual methods of the class. This resulted in ChatGPT having to generate these methods in the correct order, because some of them depended on methods created previously. When attempting to create a repository for a very simple structure (the **the_simplest** table), we encountered problems with type matching: in one method, ChatGPT ignored the dependency on the auxiliary method (this issue did not occur again during repeated attempts). We encountered a considerably higher number of errors in the case of the repository for the **complete_reversed** table. This table had the most complex structure—although, compared to the template table, it differed only with respect to field names and field order. A number of the generated methods had incorrect syntax due to erroneously generated field types, and could not be compiled. In the cases of these methods, the generated code's replicability was very low. Summary - ChatGPT:

- does not fully learn the patterns given to it
- does not recognize all the rules contained in the patterns
- confuses programming languages
- the size of the standard causes deterioration of the result
- the code generated on the basis of the pattern is not replicable.

V. CONCLUSION

It seems that at present, LLM-based code generation methods are unable to replace TBCG. ML- and LLM-based tools,

TABLE IV
RESULTS FOR THE **COMPLETE_REVERSED** TABLE (ERRORS ONLY)

Criterion	Result (iter. 1)	Result (iter. 2)	Result (iter. 3)
K1.1	ERROR	ERROR	ERROR
K1.3	ERROR	ERROR	ERROR
K2.1	ERROR	ERROR	ERROR
K2.3	ERROR	ERROR	ERROR

TABLE V
REPLICABILITY OF THE CODE GENERATED FOR THE **REPOCOMPLETEREVERSED** CLASS METHODS

Method	Replicability
entity2Stmt	NO
rs2Entity	NO
create	yes
update	yes
findAll	yes
findByKey	yes

such as ChatGPT, provide us with tremendous, previously-unknown capabilities, and are highly likely to change our current perspective on the software development process. At this juncture, however, we were unable to obtain results that would enable us to replace the classic template-based code generation methods. The primary stumbling blocks are ChatGPT's unpredictability and lack of replicability: successive attempts at generating code based on the same commands can yield wildly varying results. Even when the code presented is correct, successive versions differ with regard to details. These differences occur at various levels. In some cases, it is the way in which the code is formatted; in others, it is differences in how variables and methods are named or in the libraries used in the code. Another issue lies in how ChatGPT handles increased complexity, which is demonstrated by our attempts to generate repository methods. When provided with a template and a simple data structure, ChatGPT used the rules defined in the template correctly; with complex structures, however, the result was flawed. Step one of stage two also demonstrated that the formulation of effective commands demands considerable effort. In our case, it took five attempts, and we had to stop trying to generate complete repository classes and be satisfied with only the individual methods.

However, taking into account that tools such as ChatGPT are only at the beginning of their development path, we should watch them closely and hope that soon, in the next versions, they will meet our requirements and will be able to work as a full equivalent of traditional code generators.

And because at OPI PIB we deal with the topic of automatic application generation, we intend to check and test the possibilities of subsequent LLM tools on an ongoing basis.

REFERENCES

[1] Steven Kelly and Juha-Pekka Tolvanen, "Domain-Specific Modeling: Enabling Full Code Generation," John Wiley & Sons, 2008.

[2] Sven Jörges, "Construction and Evolution of Code Generators," Springer-Verlag Berlin Heidelberg, 2013.

[3] Priyan Vaithilingam, Tianyi Zhang, and Elena L. Glassman, "Expectation vs. experience: Evaluating the usability of code generation tools powered by large language models," In Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems, CHI EA '22, New York, NY, USA, 2022, Association for Computing Machinery. DOI https://doi.org/10.1145/3491101.3519665

[4] Hammond A. Pearce, Baleegh Ahmad, Benjamin Tan, Brendan Dolan-Gavitt, and Ramesh Karri, "Asleep at the keyboard? assessing the security of github copilot's code contributions," 2022 IEEE Symposium on Security and Privacy (SP), pages 754–768, 2021. DOI https://doi.org/10.48550/arXiv.2108.09293

[5] Rajeev Alur, Rastislav Bodík, Garvit Juniwal, Milo M. K. Martin, Mukund Raghothaman, Sanjit A. Seshia, Rishabh Singh, Armando Solar-Lezama, Emina Torlak, and Abhishek Udupa, "Syntax-guided synthesis," 2013 Formal Methods in Computer-Aided Design, pages 1–8, 2013. DOI 10.1109/FMCAD.2013.6679385

[6] Allen Cypher, "Eager: programming repetitive tasks by example," Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1991. DOI https://dl.acm.org/doi/10.1145/108844.108850

[7] Sumit Gulwani, "Automating string processing in spreadsheets using input-output examples," In ACM-SIGACT Symposium on Principles of Programming Languages, 2011. DOI https://doi.org/10.1145/1925844.1926423

[8] Vu Le and Sumit Gulwani, "Flashextract: a framework for data extraction by examples," Proceedings of the 35th ACM SIGPLAN Conference on Programming Language Design and Implementation, 2014. https://doi.org/10.1145/2666356.2594333

[9] Matej Balog, Alexander L. Gaunt, Marc Brockschmidt, Sebastian Nowozin, and Daniel Tarlow, "Deepcoder: Learning to write programs," ArXiv, abs/1611.01989, 2016. DOI https://doi.org/10.48550/arXiv.1611.01989

[10] Tonglei Guo and Huilin Gao, "Content enhanced bert-based text-to-sql generation," ArXiv, abs/1910.07179, 2019. DOI https://doi.org/10.48550/arXiv.1910.07179

[11] Shirley Anugrah Hayati, Raphaël Olivier, Pravalika Avvaru, Pengcheng Yin, Anthony Tomasic, and Graham Neubig, "Retrieval-based neural code generation," In Conference on Empirical Methods in Natural Language Processing, 2018. DOI https://doi.org/10.48550/arXiv.1808.10025

[12] Matteo Ciniselli, Nathan Cooper, Luca Pascarella, Denys Poshyvanyk, Massimiliano Di Penta, and Gabriele Bavota, "An empirical study on the usage of bert models for code completion," 2021 IEEE/ACM 18th International Conference on Mining Software Repositories (MSR), pages 108–119, 2021. DOI https://doi.org/10.48550/arXiv.2103.07115

[13] Antonio Mastropaolo, Simone Scalabrino, Nathan Cooper, David Nader-Palacio, Denys Poshyvanyk, Rocco Oliveto, and Gabriele Bavota, "Studying the usage of text-to-text transfer transformer to support code-related tasks," 2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE), pages 336–347, 2021. DOI https://doi.org/10.48550/arXiv.2102.02017

[14] Eugene Syriani, Lechanceux Luhunu, and Houari A. Sahraoui, "Systematic mapping study of template-based code generation," Comput. Lang. Syst. Struct., 52:43–62, 2017. DOI https://doi.org/10.48550/arXiv.1703.06353

Communication Papers of the 18th Conference on Computer Science and Intelligence Systems pp. 51–60

DOI: 10.15439/2023F3725
ISSN 2300-5963 ACSIS, Vol. 37

An Evaluation of a Zero-Shot Approach to Aspect-Based Sentiment Classification in Historic German Stock Market Reports

Janos Borst*, Lino Wehrheim†, Andreas Niekler*, Manuel Burghardt*
*Leipzig University, Email: [name].[surname]@uni-leipzig.de
†University of Regensburg, Lino.Wehrheim@ur.de

***Abstract*—One critical aspect that remains in the application of state-of-the-art neural networks to text analysis in applied research is the continued requirement for manual data annotation. In computer science research, there is a strong focus on maximizing the data efficiency of fine-tuning language models. This has led to the development of zero-shot text classification methods, which promise to work effectively without requiring fine-tuning for the specific task at hand. In this paper, we conduct an in-depth analysis of aspect-based sentiment analysis in historic German stock market reports to evaluate the reliability of this promise. We present a comparison of a zero-shot approach with a meticulously fine-tuned three-step process of training and applying text classification models. This study aims to empirically assess the reliability of zero-shot text classification and provide justification for the potential benefits it offers in terms of reducing the burden of data labeling and training for analysis purposes. The findings of our study demonstrate a strong correlation between the sentiment time series generated through aspect-based sentiment analysis using the zero-shot approach and those derived from the fine-tuned supervised pipeline, validating the viability of the zero-shot approach. While the zero-shot pipeline exhibits a tendency to underestimate negative examples, the overall trend remains discernible. Additionally, a qualitative analysis of the linguistic patterns reveals no explicit error patterns. Nevertheless, we acknowledge and discuss the practical and epistemological obstacles associated with employing zero-shot algorithms in untested domains.**

I. Introduction

SENTIMENT analysis plays a crucial role in the field of digital humanities, enabling researchers to uncover attitudes and emotions expressed in various forms of text. Existing sentiment analysis approaches can be broadly categorized into dictionary-based methods and machine learning-based methods [1].

With the advent of large language models like BERT [2] or GPT [3], machine learning approaches have gained popularity due to their ability to be fine-tuned rather than trained from scratch. However, the process of fine-tuning models remains laborious and time-consuming, demanding significant manual effort. As a result, there is a growing interest in exploring alternative approaches such as zero-shot learning [4, 5, 6] for sentiment analysis, particularly aspect-based sentiment analysis [5]. Zero-shot learning eliminates the need for manual data labeling, offering a promising avenue for automating sentiment analysis tasks. While zero-shot learning has shown promising results for general text classification tasks already [7, 6], and it also has been tested for sentiment analysis tasks specifically [8, 9, 10, 11, 5], its practical application in digital humanities (DH) projects remains relatively scarce.

To encourage more use of zero-shot approaches in DH, we present a first study that systematically evaluates the effectiveness of zero-shot text classification for aspect-based sentiment analysis. Our evaluation design is inspired by an ongoing research project called “More than a Feeling: Media Sentiment as a Mirror of Investors’ Expectations at the Berlin Stock Exchange, 1872-1930”, which is focused on detecting sentiment in historical German stock market reports. This research project serves as an exemplary case within the realm of digital humanities, highlighting the significant challenges associated with historic sources and languages. To provide a comprehensive evaluation, we compare the performance of the zero-shot approach against another machine learning approach that relies on manually annotated training data to fine-tune existing large language models. This approach was carried out in the initial phase of the above research project [12, 13] and now serves as a baseline to assess the quality of the fully automatic zero-shot approach.

By conducting this systematic evaluation, we aim to contribute to the understanding of zero-shot text classification for aspect-based sentiment analysis, thereby paving the way for its wider application in digital humanities research. Furthermore, we seek to address the unique challenges posed by historic sources and languages, enriching the discourse on sentiment analysis in the context of historical texts. The contributions of this paper are as follows:

- An in-depth evaluation of a zero-shot text classification pipeline for aspect-based sentiment analysis on historical German text data
- In-depth comparison of zero-shot text classification and trained text classification on a complex research application in a quantitative and qualitative manner.
- Insights into and discussion of the potential and limitations of this approach.

II. Related Work

Sentiment analysis is widely used in the field of text mining and social media analytics. In recent years, it has also gained increasing popularity in the Digital Humanities, particularly in the field of Computational Literary Studies [14]. Another

Thematic track: International Workshop on AI in Digital Humanities, Computational Social Sciences and Economics Research

field that is particularly fond of sentiment analysis is Finance and Financial Economics. In fact, it has long been known that economies are heavily influenced by moods, feelings and emotions [15]. Sentiment analysis in financial texts has first been approached by dictionary-based methods [16, 17], which are still used today in some cases [18]. Since machine learning approaches emerged, Transformers have been adapted and applied [19, 20]. Accordingly, resources such as FinBERT [21, 22] are also publicly available for application. One limitation here is that FinBERT only works with texts in the English language. Sentiment classification in these existing works is mainly regarded as sentence classification tasks. However, Sinha et al. [23] note that sentiment in these texts are often specifically entity-related, which can complicate analysis considerably. This challenge also applies to our paper, since the corpus often includes statements referring to entities at different granularity with contrary sentiment valuations, which we will explain later on. This is why we regard the sentiment analysis as an aspect-based sentiment text classification task [24].

Text classification and natural language processing (NLP) in general have made significant progress in recent years. In particular the accessibility of pretrained large language models (LLM) like BERT [2] through Huggingface [25] has had considerable impact on applications. While virtually every metric in NLP has jumped up by employing today's de-facto standard of finetuning LLMs [2, 26, 27], this comes with two caveats: Computational efficiency and data efficiency. While pretraining models has significantly reduced the amount of data needed to achieve competitive results, fine-tuning LLMs often comes with the computational cost of having to update billions of parameters, which can be rather difficult and even infeasible at times. In recent years, research has concentrated on methods that decrease the number of data points needed for training, leading to so-called few-shot models [3, 28, 29, 30] and even zero-shot models [4, 7, 31]. Zero-shot text classification models can be applied to text classification tasks without the need for task-specific fine-tuning or manual data labeling. This alleviates not only the the need for manual data annotation, but also the corresponding computational costs. The formulation of zero-shot text classification as an entailment of sentence pairs [7] serves as a very flexible approach that even can be adapted to aspect-based sentiment classification [5]. It has shown promising results in both sentiment and aspect-based sentiment classification [5]. Another way to apply sentiment analysis to a corpus without having to fine-tune is to make use of publicly shared trained sentiment models. There exists a broadly trained off-the-shelf solutions for German sentiment analysis text [32], which marks an inbetween of models that are trained for the task, but not specifically fine-tuned with domain data.

While there has been some work regarding zero-shot entity recognition in historic German newspaper [33], to the best of our knowledge, we are the first to apply zero-shot aspect-based sentiment classification to German texts. We present an in-depth comparison between the zero-shot approach and specifically trained models, fine-tuned on hand-coded data, for the application on historic German texts.

III. Approach

A. Introduction to the Corpus

We build upon previous work [12] where a corpus of German stock market reports between 1872 to 1930 was compiled for analyzing the sentiment over time. Sentiment analysis of the corpus aims to provide insight into the mood and opinions about the stock market during that period. While it is useful to consider the sentiment of an article or sentence in general as the aggregated sentiment of all statements, sentiment can also be expressed about specific aspects or entities. In the case of the stock market corpus we consider three levels of interest:

- **Individual Entities:** Sentiment towards specific entities of stocks that may be subject to a particular sentiment on a given day.
- **Sectors:** Statements towards sectors of the markets or groups of specific stocks, e.g. "the railway stocks were ...".
- **Overall:** The general mood at the stock market without specifying specific stocks or sectors.

The distinction between these different levels of sentiment analysis is crucial, since the historic texts tend to specifically emphasize opposing market movements, as can be seen in the following example:

> *Construction values dull throughout, only Deutsche Eisenbahnbau and Lindenbauverein again a little higher.*[1]

The example expresses a negative sentiment towards the construction value market, but highlights specific stocks (Deutsche Eisenbahnbau and Lindenbauverein) that traded higher. This type of sentiment analysis provides a more nuanced understanding of the sentiment towards the stock market during that time period. We regard these entity levels as the aspects of the aspect-based sentiment analysis.

B. Workflow and Data

To get a detailed understanding of the sentiment of the German historic stock market, we follow a three-step process: First we train a binary text classification model to identify if a sentence contains any sentiment at all, to filter out factual statements containing no sentiment. Second, we train a multi-label text classification model to detect which of the three levels are targeted by the expressed sentiment. Finally, we use the results of the entity-level classification to train an aspect-based sentiment model to extract the sentiment specifically with regard to the entity. This 3-step process is visualized in the left branch in Fig. 1. This enables us to analyze three sentiment time series with regard to the entity levels and also over all, if averaged.

To create a data set, a subsample of this corpus was annotated by an expert, as described in [13], and serves as

[1] Translated from German: *Bauwerthe durchweg matt, nur Deutsche Eisenbahnbau und Lindenbauverein wieder eine Kleinigkeit höher.*

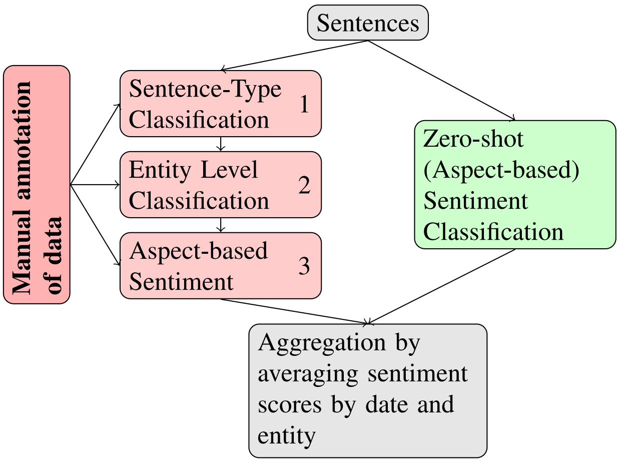

Sentences
Manual annotation of data
Sentence-Type Classification 1
Entity Level Classification 2
Aspect-based Sentiment 3
Zero-shot (Aspect-based) Sentiment Classification
Aggregation by averaging sentiment scores by date and entity

Fig. 1. Schematic drawing of the fine-tuned pipeline (left branch) and in the zero-shot pipeline (right branch). Red color indicates manual labelling or computational effort (training a modela).

training data. The data was sampled stratified over time to ensure that linguistic changes over time are represented in the data set. This results in three views on the data set:

- For sentence type classification there are 1651 examples, 609 neutral and 1042 containing a sentiment related statement.
- For sentences that contain any sentiment there are 732 sentence with at least one entity category assigned and a label density of 1.15.
- For aspect-based sentiment classification there are 1584 (sentence, aspect) pairs with an assigned sentiment of “positive", “negative" or “neutral".

Note that in our annotation scheme, “neutral" also includes calm or mixed statements, i.e. statements that have multiple contrary sentiments about an entity level or valuate it not in a positive or negative way. To simplify this into a common naming scheme we will refer to all of these as neutral, but it will be reflected in the hypothesis template of the zero-shot classification pipeline.

Using these three data sets, we build a fine-tuned pipeline of three models as shown in red in Fig. 1 that serves as a proxy of the expert solution to the task and will be the baseline to which we compare the zero-shot algorithm (shown in green).

C. Fine-tuned Pipeline

In this section we describe the fine-tuned pipeline, which consists of three separate models trained on one of the tasks corresponding to the left branch in Fig. 1. We only show a quick summary of the results that are discussed in Borst, Janos, Wehrheim, Lino, and Burghardt, Manuel [13]. As basis for every model, we use a German BERT variant pre-trained by the DBMDZ[2]. To evaluate the model, we split the annotated data into 80-20 training and validation splits, reported results are measured on the validation split. All three models use an Adam optimizer and after training the epoch with the best metric for the task is chosen.

TABLE I
Results on Validation set for the aspect classification step.

%	individual entities	sectors	market	micro avg	macro avg
precision	92.54	76.19	77.08	82.58	81.94
recall	89.86	84.21	90.24	88.02	88.11
f1-score	91.18	80.00	83.15	85.22	84.77

For sentence-type classification the transformer was fine-tuned as a binary classification model to distinguish neutral sentences from sentences containing sentiment statements, achieving 93% accuracy. The model was chosen because of the highest recall for sentences containing sentiment (96.5%). This ensures that we find most of the sentences containing sentiment statements.

Classifying which aspects the sentences contains was tackled as a multi-label classification problem with above-mentioned entity levels “individual entities", “sectors" and “market" as labels. The best model was chosen by macro average F1. The full results of this step is shown in Table Tab. I. We see quite balanced performance across all classes, with higher performance on individual entities. Individual entities have a very common linguistic pattern which makes them easy to detect.

In the third and final step we use the entity-level classification of step two as aspects and classify the combination of a (sentence, aspect)-pair into the sentiment classes “negative", “neutral" and “positive". A sentence can have multiple aspect-based sentiment annotations based on the result of the previous step. This model is trained as a single label classification task, that is, for every (sentence, aspect)-pair only one sentiment can be assigned. The best model was chosen by macro average F1 and achieves 80.7% accuracy and 80.7% macro F1.

D. Zero-Shot Pipeline

In this section we describe the pipeline to accomplish the same task without finetuning or training any model, corresponding to the right branch in Fig. 1. The aim is to perform the complex aspect-based sentiment classification process, described above, without using any of the knowledge that results from the manual coding and model training. This is especially important for aspect-based sentiment analysis, as we cannot assume knowledge about the type of entity contained in a sentence. We bypass this by classifying for the sentiments of all three entity categories and assume that, if there is no sentiment regarding any level this will result in a “neutral" label and will have no influence on the further analysis. We also classify the corpus with a zero-shot model with regard to overall sentence sentiment.

As zero-shot model, we use textual entailment classification, following the task description proposed in Yin, Hay, and Roth

[2] https://huggingface.co/dbmdz/bert-base-german-cased

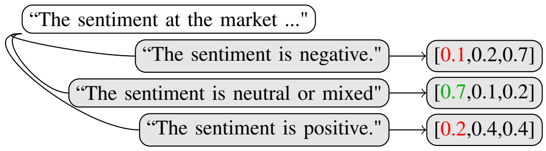

“The sentiment at the market ..."
“The sentiment is negative."
[0.1,0.2,0.7]
“The sentiment is neutral or mixed"
[0.7,0.1,0.2]
“The sentiment is positive."
[0.2,0.4,0.4]

Fig. 2. Schematic example for the formulation of the entailment task and its application to zero-shot text classification. The scores are the output for every sentence pair with regard to the categories *entailment*, *contradiction* and *neutral*. The highlighted numbers in color show the values that are compared with each other, which in this case would lead us to assign the category “neutral".

TABLE II
ZERO-SHOT EVALUATION METRICS ON THE MANUALLY LABELLED ASPECT-BASED DATA SET.

%	negativ	neutral	positiv	micro avg	macro avg
precision	75.69	60.48	81.27	67.61	72.48
recall	28.60	89.82	76.43	67.61	64.95
f1-score	41.52	72.29	78.77	67.61	64.19

[7] using a pretrained model from the huggingface hub[3]. In this approach a sentence pair, called premise and hypothesis, is classified as “entailment", “contradiction" or “neutral", based on how well the hypothesis logically entails the premise. For zero-shot classification we form hypotheses containing the label we want to classify. These hypotheses are created using a hypothesis template: *"The sentiment is [blank]"*[4]. The blank is then filled with the sentiment categories.

The model output provides a probability score for every premise and hypothesis pair and entailment class. We select the hypothesis with the highest probability of entailment as the classification result and assign the corresponding category. This leads to the formulation as show in Fig. 2. This approach is used for zero-shot sentiment classification.

For aspect-based zero-shot sentiment classification this approach can be extended by another placeholder in the hypothesis template, which is used to create the hypotheses. We use the template: *The sentiment for [aspect] is [label]*[5]. For every entity category above, we create the premise and hypothesis pairs by combining the entity category with each of the sentiment labels. Within each entity-level the procedure is the same as above. Fig. 3 shows a schematic drawing for this. The result of this step is an assignment of one sentiment label for every sentence and entity-level pair.

IV. EXPERIMENTS

Code and Data to replicate these findings can be found at https://git.informatik.uni-leipzig.de/computational-humanities/research/fedcsis-zero-shot-sentiment/

A. Quantitative Comparison

TABLE III
EVALUATION OF THE FINE-TUNED PIPELINE ON THE VALIDATION SET OF THE MANUALLY LABELLED ASPECT-BASED DATA SET.

	negativ	neutral	positiv	micro avg	macro avg
precision	81.6	90.5	85.7	86.1	85.9
recall	88.6	85.7	83.5	86.1	85.9
f1-score	84.9	88.0	84.6	86.1	85.9

TABLE IV
TABLE OF AGREEMENT BETWEEN THE ZERO-SHOT AND TRAINED PIPELINE ON THE ENTIRE CORPUS.

%	truth	negative	neutral	positive
zero-shot	negative	75.69	16.57	7.73
	neutral	30.54	60.48	8.97
	positive	9.49	9.25	81.27
trained	negative	84.76	11.43	3.81
	neutral	7.52	89.47	3.01
	positive	7.59	12.66	79.75

TABLE V
CONFUSION MATRICES OF THE TWO PIPELINES ON THE MANUALLY CODED VALIDATION SET.

	fine-tuned	negative	neutral	positive
aspect	zero-shot			
market	negative	85.79	6.79	07.42
	neutral	50.14	30.74	19.12
	positive	10.57	10.19	79.24
sectors	negative	83.86	8.76	7.37
	neutral	31.02	49.95	19.03
	positive	3.32	09.86	86.81
individual entities	negative	79.68	12.85	07.47
	neutral	33.73	42.23	24.05
	positive	2.91	09.54	87.55

1) Data set metrics: We evaluate the zero-shot algorithm on the same data used to train the fine-tuned pipeline on. Tab. II and Tab. III show the evaluation metrics for training and zero-shot respectively. Although there is a significant improvement in the F1-score of the trained model over the zero-shot model, it is noteworthy that this gap largely stems from the fact that the recall of negative sentiments is rather low. The precision for “negative" sentiments and all metrics for “positive" values are higher but a bit short of competing with the fine-tuned pipeline.

With further analysis, we find that the confusion matrices confirm the problem: Around 30% of predicted neutral labels are actually negative labels. This error is systematic, thus it may lead to an over-estimation of absolute values in the aggregated time series, but should not affect the overall trends.

2) Agreement: Tab. V shows the confusion matrix of the zero-shot pipeline and the fine-tuned pipeline. With regard to the manually coded data set, both algorithms seem to have comparable performance with strengths in classifying positive and negative examples. The confusion between neutral and

[3] https://huggingface.co/svalabs/gbert-large-zeroshot-nli
[4] Translated from German: *“Die Stimmung ist [label]."*
[5] Translated from German: *“Die Stimmung für [aspect] ist [blank]"*

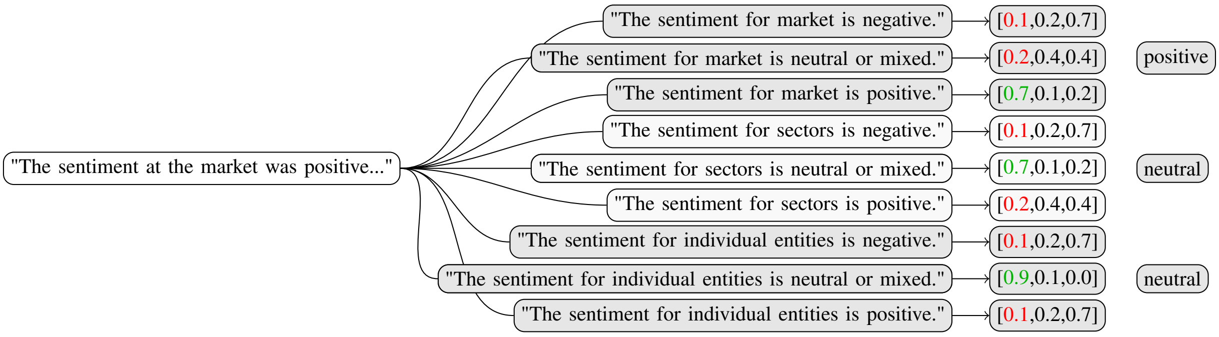

"The sentiment at the market was positive..."
"The sentiment for market is negative."
[0.1,0.2,0.7]
"The sentiment for market is neutral or mixed."
[0.2,0.4,0.4]
positive
"The sentiment for market is positive."
[0.7,0.1,0.2]
"The sentiment for sectors is negative."
[0.1,0.2,0.7]
"The sentiment for sectors is neutral or mixed."
[0.7,0.1,0.2]
neutral
"The sentiment for sectors is positive."
[0.2,0.4,0.4]
"The sentiment for individual entities is negative."
[0.1,0.2,0.7]
"The sentiment for individual entities is neutral or mixed."
[0.9,0.1,0.0]
neutral
"The sentiment for individual entities is positive."
[0.1,0.2,0.7]

Fig. 3. Schematic example for the formulation of the entailment task and its application to zero-shot text classification. The scores are the output for every sentence pair with regard to the categories *entailment*, *contradiction* and *neutral*. The highlighted numbers in color show the values that are compared with each other, which in this case would lead us to assign the category “neutral".

negative labels is the only position in the confusion matrix that seems to have substantially benefited from fine-tuning at all. All other label pairs show similar performance.

In Tab. IV we look at the confusion of agreement on entity level. In contrast to the previous table, high values now indicate that both pipelines agree to the same assignment of labels for any given (sentence, aspect) example. The table shows these statistics for the entire corpus (not only the validation data set). A very similar picture emerges: Both pipeline have a very high agreement about positive and negative examples for all entity levels. But there is considerable confusion if the zero-shot pipeline predicts neutral. For all entity levels there is a significant tendency that the fine-tuned pipeline would hand out a “negative" label where the zero-shot pipeline assigns “neutral". For the “market" level the agreement is even lower than 50%.

For the entity levels “sectors" and “individual entities" and on a global level, these errors have a systematic character that will not influence the overall trends considering that “most" prediction will still be correct.

3) Time Series Metrics: Besides an assessment of the quality of classification models, we want to compare resulting insights and possible analyses of both pipelines. To be able to use zero-shot instead of standard fine-tuning in a real-world application scenario, it should produce similar if not the same analysis result as the fine-tuned pipeline. In our case the basis of analysis are the sentiment time series that emerge from both of the pipelines. Comparing the time series expands the quantitative assessment of the classifier with the aspect of time. So the question we want to investigate here is: Would these pipelines create the same insights into the data?

The time series are generated by grouping the data over seven days and by summing up the sentiment labels. We evaluate negative, neutral and positive as “-1",“0", and “1" respectively. After that, time series are created by computing the rolling average over half a year (26 weeks). Time series are created for every entity level separately and for the overall sentiment. For overall sentiment we averaged the trained pipeline’s output per sentence to create one score and then did the same as above. Since we are not interested in absolute values, and we are dealing with a systematic error of the negative values, we also normalize each time series by mean-normalization. This normalization has no influence on trends or on the correlation factor, which we calculate below.

For comparison, we also applied an off-the-shelf neural network sentiment classification network for German language model for overall sentiment. Guhr et al. [32] train a “general-purpose German sentiment classification model". Since this model is off-the-shelf, it is not adapted to *historic* German language. However, we regard the comparison still as useful, since the availability of specific models is still one of the most common problems when researching textual data.

In the following, we want to evaluate the time series qualitatively and quantitatively. Fig. 4 shows a visual comparison of all these time series and will be discussed in the next section. For quantitative evaluation, we calculate the Pearson correlation of the zero-shot time series to the time series of the fine-tuned pipeline to verify if they have similar tendencies and trends. Tab. VI shows the Pearson correlation factors.

Fig. 4 reveals that the zero-shot and respective trained time series are often very close. In all cases except “market” the trends and tendencies are highly correlated. For “market” there is a period from 1900-1920 with a higher deviation. This leads to a Pearson correlation factor of only around 0.3. One explanation for this lies in the fact that while the concepts of “sector” and “individual entities” manifest in explicit syntactic figures and tokens, the concept of an “market” sentiment is not as easy to grasp sometimes. Another explanation lies in the fact that almost 50% of all entity mentions regard specific stocks or sectors. This might have higher influence on overall sentiment when aggregating over time. This argument is backed by the fact that the time series for individual entities shows the highest correlation of all time series. We also argue that it is not due to linguistic changes because these should affect all entity levels similarly in the same time periods, which we do not observe.

Additionally, we see that an off-the-shelve neural network sentiment classifier does not agree with the results of either zero-shot or fine-tuned pipeline. There is virtually no correlation of Guhr et al. [32] with the fine-tuned pipeline.

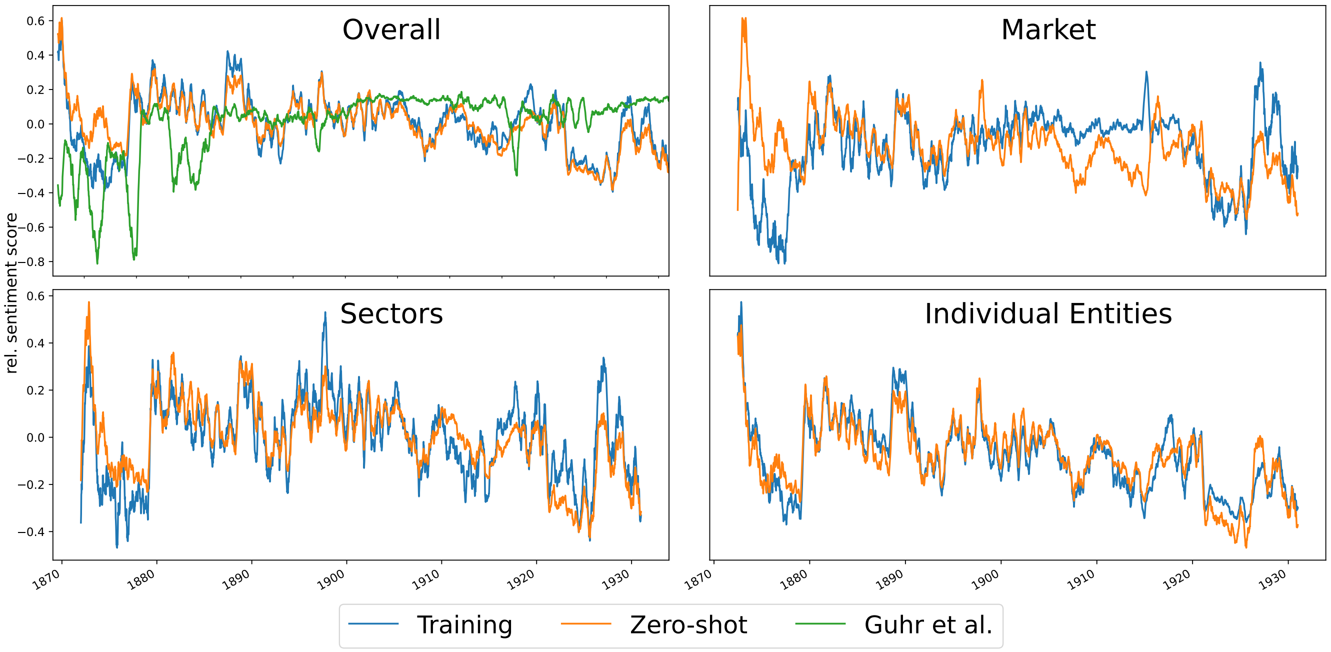


Fig. 4. Plot of the time sentiment time series for overall sentiments (top left) and for every entity level.

TABLE VI
PEARSON CORRELATION COEFFICIENT BETWEEN ZERO-SHOT AND TRAINED SENTIMENT TIME SERIES.

market	0.358472
sectors	0.822404
individual entities	0.911497
overall	0.870379
Guhr et al. [32]	0.077549

B. *Qualitative Assessment*

There has been some criticism about the application of entailment-based zero-shot recently [34]. The paper mentioned spurious correlation as a main driver of zero-shot classification performance in entailment-based solutions.

In our case this would be quite critical, because of formalisation of syntax and language in the historic realm could lead to considerable bias. We address this with a qualitative check of examples, on which the zero-shot and the trained pipeline disagree to identify patterns of errors.

As shown in Tab. IV, most of the disagreement occurs between neighbouring classes: positive and neutral , or negative and neutral, which in itself is often quite ambiguous. The only pattern we could find is that if there are opposing sentiments in one sentence such that the example should be labeled "neutral", both algorithms seem to randomly pick one of the sentiments as the predicted polarity. This is not a systematic error but rather a random choice made by both pipelines. For instance, in the following example, two opposing polarities refer to the same entity level ("Individual Entities"). The trained pipeline assigns a positive polarity, whereas the zero-shot pipeline predicts it as neutral.

> *'Among foreign currency, Dutch lay firm, ruble notes continued to decline'.* [6]

In order to identify examples of linguistic patterns that are more difficult to classify, we examined instances where the zero-shot pipeline and trained pipeline assigned opposite polarities. One pattern that emerged with slightly higher frequency was related to the interpretation of the terms "supply" and "demand"[7]. While in general language, "supply" might have a positive connotation, in stock market reports, a predominance of supply is often associated with a high amount of selling and thus dropping prices, which is negatively connoted in this domain.

> *For Hansa shares, supply predominated.*[8]

> *Strongly in demand without supply were the 4% Reich and government bonds.*[9]

In the first example, the report mentions more supply than demand, which refers to falling prices. In this case, the trained pipeline correctly predicts a negative polarity while the zero-shot pipeline assigns a positive polarity. In the second example, the same situation occurs with "demand", where this time the zero-shot pipeline correctly predicts a positive polarity while the trained pipeline assigns negative. Overall, the zero-shot pipeline tends to classify more sentences containing "supply"

[6] Translated from: *'Unter den fremden Devisen lagen holländische fest, Rubelnoten wurden weiter rückgängig.'*

[7] Translated from: *Angebot und Nachfrage*

[8] Translated from: *'Für Hansa-Aktien überwog Angebot.'*

[9] Translated from: *'Stark gefragt ohne Angebot wurden die 4% Reichs- und Staatsanleihen.'*

or “demand” as “neutral” (73%) than the trained pipeline (49%), which is in line with the above mentioned evaluation metrics.

While these examples highlight linguistic difficulties regarding the textual domain of historic stock market reports, there are no clear-cut patterns where one of the algorithms significantly fails to conform to our defined label scheme.

V. Practical and Epistemological Considerations

An alternative way to frame this paper's research question could be whether the zero-shot pipeline's results align with those based on human annotation, or more specific, how well the presumed annotation scheme conform with the zero-shot's definition of the annotation scheme. In our case, the results are generally encouraging thus far, but they also raise practical and epistemological follow-up questions, which we will outline in this section.

The comparison between the trained and zero-shot pipelines reveals that the assessment of the latter depends on the research question at hand. For researchers interested in sector- or entity-level sentiment, such as business historians, the zero-shot approach appears to be feasible, as evidenced by the evaluation metrics presented above. However, if one is interested in the market level, the differences between the two approaches appear to be too substantial, especially for the 1870s and 1910s. There are even differences in the degree of agreement with regards to different research objects within the same task, domain and time period.

Although we can only speculate about the reasons why the zero-shot approach does not agree with a model trained on human annotations for these periods, the lower performance at the market level, in general, is not surprising. Sentences referring to the “market” sentiment are more difficult to detect and interpret, even for human annotators, because the entity of “market” is often only mentioned implicitly. Ambiguity is a general problem in sentiment analysis. Furthermore, the fact that the zero-shot approach suffers from systematic biases may not be a problem if one is interested in time trends, but it may pose a problem in other cases. This is also true for other approaches, that can be used without training evaluation, e. g. dictionary-based methods.

Apart from these more technical aspects of our specific set-up, there are some epistemological reflections to be made. Researchers who consider using zero-shot methods because they lack the resources to create training data and fine-tuning a model, face a fundamental dilemma - as anyone using unsupervised tools does: How can we rely on the results if we do not have data for a formal evaluation and if we have data to evaluate why not also fine-tune? When looking at the time series in Fig. 4, even the most well-versed domain expert may struggle to discern whether these results are mere artifacts created by an algorithm or substantial results. Furthermore, even if the expert can discern the results, what would we learn from these results that we did not know before? In any case there is the possibility of confirmation bias, when not properly supported by close reading. In other words, we face the paradox that the very advantage of zero-shot models is also a considerable drawback for practical application.

Of course, there is general evidence for the quality and performance of zero-shot models, where especially polarity classification has been shown to work more consistently. But there is no general notion of why this should be transferable to another domain or another language with a “similar” task. The problem of distributional shift or domain adaptation often contributes to loss of performance [35, 36, 37]. Then again: How similar do these domains have to be, to safely assume generalization? These questions are particularly challenging to answer for historiographic research, which often covers very specific domains and languages or longer time periods for which there are hardly any pre-checked settings to be found.

Finally, this study raises a lot of questions regarding implicit assumptions when applying zero-shot sentiment classification, like: Is there a “correct" sentiment in these texts? If so, are expert-level humans able to identify it, correctly reflecting the historic reality? Does the technical ability of these models to generalize suffice in this scenario?

Recent research suggest that every step along the way to a trained pipeline based on human annotations involves the risk of bias [38]. Also, in the special case of sentiment, there are many studies that evidence that scholar's and crowd-sourced annotations alike have particularly low agreement between annotators in historic texts [39, 36, 37]. Even the agreement of various machine-learning algorithms seems surprisingly low even when trained and evaluated on the same sentiment data sets [40]. There is much work done in the field of domain adaptation in sentiment classification [35, 41, 42, 43], but all these implicitly rely on the human annotations as performance metric as well. All these factors contribute to the uncertainty of “correct" results in any case, but might also lead to the conclusion that a zero-shot approach may suffice to discover underlying trends.

There is no space here to provide answers to these questions, nor do we claim to have any. In the end, the usefulness of zero-shot learning will probably depend on the research question, the domain, the possibility to conduct some sort of evaluation (e.g. at least some human annotations), and maybe the general willingness to trust unsupervised approaches, which is distributed unevenly across different research communities. These issues will, however, get more pressing with the wider availability of powerful zero-shot tools like GPT-4.

VI. Conclusion

In this study, a zero-shot text classification pipeline was applied to an aspect-based sentiment analysis of German historic texts of stock market reports in the digital humanities. The goal was to get insights in how useful these methods are in a real application scenario. We provided in-depth comparisons, qualitatively and quantitatively, between a fine-tuned pipeline and a zero-shot pipeline. The results show that both can deliver usable results in our aspect-based sentiment analysis. The trends and insights produced by the zero-shot models were highly correlated with those produced by the trained models. They were found to be particularly useful for classical sentence polarity classification, but also performed well for aspect-based sentiment analysis. Even if the results may differ in some details and there are systematic errors, we can confidently say that zero-shot models provide a good exploration tool and an easy start for sentiment analysis, especially in cases where no hand-coded data exists.

The zero-shot models were also found to work better than an off-the-shelf BERT German sentiment model. Since the general availability of domain-specific models is still not fully achieved, especially for languages other than English, zero-shot approaches to sentiment analysis may help to close the gap. Also, zero-shot models for aspect-based sentiment were found to work better with concrete entities (target text) rather than general aspects, which we were not able to consider in this study. We defer this task to future work on the subject.

However, it should still be noted that correlation and data set metrics are just a hint at performance and that the factual correctness of the sentiment analysis is difficult to prove, as this would require manual examination of the textual content and sentiment analysis still contains a subjective nature. We only compared the zero-shot results to the result of the trained pipeline to answer the question, if both models would provide similar insights, while the interpretation of these results is not part of this paper.

Still, we believe that zero-shot models in sentiment analysis, and for text classification in general, seem a promising approach, especially when considering the progress and the potential of models similar to GPT-4.

Acknowledgements

This research is funded By German Research Foundation (DFG) under project number BU 3502/1-1.

References

[1] Bing Liu. *Sentiment Analysis and Opinion Mining*. Vol. 5. 2012.

[2] Jacob Devlin et al. “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1*. Minneapolis, Minnesota: Association for Computational Linguistics, June 2019, pp. 4171–4186. DOI: 10.18653/v1/N19-1423.

[3] Tom Brown et al. “Language Models are Few-Shot Learners”. en. In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 1877–1901.

[4] Y. Xian, B. Schiele, and Z. Akata. “Zero-Shot Learning — The Good, the Bad and the Ugly”. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, 2017, pp. 3077–3086.

[5] Lei Shu et al. *Zero-Shot Aspect-Based Sentiment Analysis*. 2022. arXiv: 2202.01924 `[cs.CL]`.

[6] Kishaloy Halder et al. “Task-Aware Representation of Sentences for Generic Text Classification”. en. In: *Proceedings of the 28th International Conference on Computational Linguistics*. Barcelona, Spain (Online): International Committee on Computational Linguistics, 2020, pp. 3202–3213. DOI: 10.18653/v1/2020.coling-main.285.

[7] Wenpeng Yin, Jamaal Hay, and Dan Roth. “Benchmarking Zero-shot Text Classification: Datasets, Evaluation and Entailment Approach”. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*. Ed. by Kentaro Inui et al. Association for Computational Linguistics, 2019, pp. 3912–3921. DOI: 10.18653/v1/D19-1404.

[8] Senait Gebremichael Tesfagergish, Jurgita Kapočiūtė-Dzikienė, and Robertas Damaševičius. “Zero-Shot Emotion Detection for Semi-Supervised Sentiment Analysis Using Sentence Transformers and Ensemble Learning”. In: *Applied Sciences* 12.17 (2022), p. 8662. DOI: 10.3390/app12178662.

[9] Mengting Hu et al. “Multi-Label Few-Shot Learning for Aspect Category Detection”. In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. ACL-IJCNLP 2021. Online: Association for Computational Linguistics, 2021, pp. 6330–6340. DOI: 10.18653/v1/2021.acl-long.495.

[10] Ronald Seoh et al. “Open Aspect Target Sentiment Classification with Natural Language Prompts”. In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. EMNLP 2021. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, 2021, pp. 6311–6322. DOI: 10.18653/v1/2021.emnlp-main.509.

[11] Anindya Sarkar, Sujeeth Reddy, and Raghu Sesha Iyengar. “Zero-Shot Multilingual Sentiment Analysis Using Hierarchical Attentive Network and BERT”. In: *Proceedings of the 2019 3rd International Conference on Natural Language Processing and Information Retrieval*. NLPIR 2019. Tokushima, Japan: Association for Computing Machinery, 2019, pp. 49–56. DOI: 10.1145/3342827.3342850.

[12] Wehrheim, Lino et al. “„Auch heute war die Stimmung im Allgemeinen fest.“ Zero-Shot Klassifikation zur Bestimmung des Media Sentiment an der Berliner Börse zwischen 1872 und 1930”. In: *Konferenzabstracts*. Dhd23 Open Humanities Open Culture. Trier, 2023. DOI: 10.5281/zenodo.7688632.

[13] Borst, Janos, Wehrheim, Lino, and Burghardt, Manuel. ““Money Can’t Buy Love?” Creating a Historical Sentiment Index for the Berlin Stock Exchange, 1872–1930”. In: *Book of Abstracts*. Digital Humanities. Graz, 2023.

[14] Evgeny Kim and Roman Klinger. “A survey on sentiment and emotion analysis for computational literary studies”. In: *Zeitschrift für digitale Geisteswissenschaften* (Aug. 2019). DOI: 10.17175/2019_008_v2.

[15] George Akerlof and Robert Shiller. *Animal Spirits: How Human Psychology Drives the Economy and Why It Matters for Global Capitalism*. Vol. 21. Jan. 1, 2009. ISBN: 978-0-691-14592-1. DOI: 10.2307/j.ctv36mk90z.

[16] Paul C. Tetlock. “Giving Content to Investor Sentiment: The Role of Media in the Stock Market”. en. In: *The Journal of Finance* 62.3 (2007), pp. 1139–1168. DOI: 10.2139/ssrn.685145.

[17] Diego García. “Sentiment during Recessions”. en. In: *The Journal of Finance* 68.3 (2013), pp. 1267–1300. DOI: 10.1111/jofi.12027.

[18] Alan J. Hanna, John D. Turner, and Clive B. Walker. “News media and investor sentiment during bull and bear markets”. en. In: *The European Journal of Finance* 26.14 (Sept. 2020), pp. 1377–1395.

[19] Kostadin Mishev et al. “Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers”. In: *IEEE Access* 8 (2020). ISSN: 2169-3536.

[20] Wouter van Atteveldt, Mariken A. C. G. van der Velden, and Mark Boukes. “The Validity of Sentiment Analysis: Comparing Manual Annotation, Crowd-Coding, Dictionary Approaches, and Machine Learning Algorithms”. In: *Communication Methods and Measures* 15.2 (Apr. 2021), pp. 121–140. DOI: 10.1080/19312458.2020.1869198.

[21] Zhuang Liu et al. “FinBERT: A Pre-Trained Financial Language Representation Model for Financial Text Mining”. In: *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*. IJCAI’20. Yokohama, Yokohama, Japan, 2021. DOI: 10.24963/ijcai.2020/615.

[22] Pekka Malo et al. “Good debt or bad debt: Detecting semantic orientations in economic texts”. In: *Journal of the Association for Information Science and Technology* 65.4 (Apr. 2014), pp. 782–796.

[23] Ankur Sinha et al. “SEntFiN 1.0: Entity-aware sentiment analysis for financial news”. In: *Journal of the Association for Information Science & Technology* 73.9 (2022), pp. 1314–1335.

[24] Wenxuan Zhang et al. “A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges”. In: *IEEE Transactions on Knowledge and Data Engineering* (2022). Conference Name: IEEE Transactions on Knowledge and Data Engineering, pp. 1–20. ISSN: 1558-2191. DOI: 10.1109/TKDE.2022.3230975.

[25] Thomas Wolf et al. “HuggingFace’s Transformers: State-of-the-art Natural Language Processing”. In: *Computing Resource Repository* abs/1910.03771 (2019). URL: http://arxiv.org/abs/1910.03771.

[26] Jeremy Howard and Sebastian Ruder. “Universal Language Model Fine-tuning for Text Classification”. In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*. Association for Computational Linguistics, 2018, pp. 328–339. DOI: 10.18653/v1/P18-1031.

[27] Zhilin Yang et al. “XLNet: Generalized Autoregressive Pretraining for Language Understanding”. In: *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, 8-14 December 2019, Vancouver, BC, Canada*. 2019, pp. 5754–5764.

[28] Jonathan Bragg et al. “FLEX: Unifying Evaluation for Few-Shot NLP”. In: *Neural Information Processing Systems*. 2021. DOI: 10.1162/tacl_a_00485.

[29] Yujia Bao et al. “Few-shot Text Classification with Distributional Signatures”. In: *International Conference on Learning Representations*. 2020. DOI: 10.1145/3531536.3532949.

[30] Yaqing Wang et al. “Generalizing from a Few Examples: A Survey on Few-Shot Learning”. In: *ACM Comput. Surv.* 53.3 (June 2020). DOI: 10.1145/3386252.

[31] Edgar Schonfeld et al. “Generalized Zero- and Few-Shot Learning via Aligned Variational Autoencoders”. en. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA: IEEE, June 2019, pp. 8239–8247. DOI: 10.1007/s00521-022-07413-z.

[32] Oliver Guhr et al. “Training a Broad-Coverage German Sentiment Classification Model for Dialog Systems”. English. In: *Proceedings of the Twelfth Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, May 2020, pp. 1627–1632. ISBN: 979-10-95546-34-4.

[33] Francesco De Toni et al. “Entities, Dates, and Languages: Zero-Shot on Historical Texts with T0”. In: *ArXiv* abs/2204.05211 (2022). DOI: 10.18653/v1/2022.bigscience-1.7.

[34] Tingting Ma et al. “Issues with Entailment-based Zero-shot Text Classification”. In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. ACL-IJCNLP 2021. Online: Association for Computational Linguistics, Aug. 2021, pp. 786–796.

[35] Chenggong Gong, Jianfei Yu, and Rui Xia. “Unified Feature and Instance Based Domain Adaptation for Aspect-Based Sentiment Analysis”. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. EMNLP 2020. Association for Computational Linguistics, Nov. 2020, pp. 7035–7045. DOI: 10.18653/v1/2020.emnlp-main.572.

[36] Cecilia Ovesdotter Alm, Dan Roth, and Richard Sproat. “Emotions from Text: Machine Learning for Text-based Emotion Prediction”. In: *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*. HLT-

EMNLP 2005. Vancouver, British Columbia, Canada: Association for Computational Linguistics, Oct. 2005, pp. 579–586. DOI: 10.3115/1220575.1220648.

[37] Thomas Schmidt, Manuel Burghardt, and Katrin Dennerlein. *Sentiment Annotation of Historic German Plays: An Empirical Study on Annotation Behavior*. Aug. 1, 2018.

[38] Mihir Parmar et al. "Don't Blame the Annotator: Bias Already Starts in the Annotation Instructions". In: *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*. EACL 2023. Dubrovnik, Croatia: Association for Computational Linguistics, May 2023, pp. 1779–1789.

[39] Rachele Sprugnoli et al. "Towards sentiment analysis for historical texts". In: *Digital Scholarship in the Humanities* 31.4 (July 2015), pp. 762–772. DOI: 10.1093/llc/fqv027.

[40] Frank Xing et al. "Financial Sentiment Analysis: An Investigation into Common Mistakes and Silver Bullets". In: *Proceedings of the 28th International Conference on Computational Linguistics*. COLING 2020. Barcelona, Spain (Online): International Committee on Computational Linguistics, Dec. 2020, pp. 978–987. DOI: 10.18653/v1/2020.coling-main.85.

[41] Mohammad Rostami and Aram Galstyan. *Domain Adaptation for Sentiment Analysis Using Increased Intraclass Separation*. July 4, 2021. arXiv: 2107.01598[cs]. URL: http://arxiv.org/abs/2107.01598 (visited on 05/22/2023).

[42] Guoliang Kang et al. "Contrastive Adaptation Network for Unsupervised Domain Adaptation". In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2019, pp. 4888–4897.

[43] Jeremy Barnes, Roman Klinger, and Sabine Schulte im Walde. "Projecting Embeddings for Domain Adaption: Joint Modeling of Sentiment Analysis in Diverse Domains". In: *Proceedings of the 27th International Conference on Computational Linguistics*. COLING 2018. Santa Fe, New Mexico, USA: Association for Computational Linguistics, Aug. 2018, pp. 818–830.

Communication Papers of the 18th Conference on Computer Science and Intelligence Systems pp. 61–64 DOI: 10.15439/2023F5820 ISSN 2300-5963 ACSIS, Vol. 37

Proof-of-Work CAPTCHA with password cracking functionality

Szymon Chadam*, Paweł Topa†
Faculty of Computer Science, Electronics and Telecommunications,
AGH University of Kraków
Kraków, Poland
Email: *szymon@chadam.pl,
†topa@agh.edu.pl

***Abstract*—This document proposes an alternative CAPTCHA system that implements a proof-of-work mechanism to protect resources (usually web services) from being accessed by automatic entities called bots. Normally, CAPTCHA forces the user to do some work in order to prove that he is not a machine. The proposed system utilizes a novel alternative to Proof-of-Work algorithm that utilizes the user's computing power to crack password hashes.**

I. Introduction

THE PROBLEM of unwanted traffic and automated applications and scripts (commonly referred to as bots) has plagued the Internet almost since its inception. Netacea estimates that unwanted bot traffic costs companies up to 250 million annually [1]. Existing solutions such as reCAPTCHA capture almost 97% of the market [2]. However, Google's solution raises privacy concerns [3] as it collects and stores information about the user's browser.

This paper presents an alternative CAPTCHA-like system. This system uses a Proof-of-Work mechanism to impose a computational cost on the user. The user's computational power is used to attempt to crack password hashes provided by the system. Based on other metrics, the computing power required to access an online service can be arbitrarily increased to combat unwanted traffic.

A common method of storing passwords securely is to convert the password phrase into a hash, which is generated by a hash function or key derivation function, and store it in a database. The properties of the hash function or KDF function ensure that the input phrase cannot be reconstructed from the hash. The only effective method of attacking such protected passwords is dictionary cryptanalysis, i.e. calculating hashes for a huge dictionary of potential passwords and comparing the calculated hashes with the attacked password.

Normally, it is undesirable for users' passwords to be cracked. However, in the case of law enforcement, we often need to obtain suspects' passwords in order to access encrypted evidence. The obvious solution is to build powerful (and expensive) dictionary cryptanalysis computers. A less obvious approach is to use the distributed power of web users' computers, as has been done in the Seti@Home (https://setiathome.berkeley.edu/ — suspended project) or Folding@Home projects (https://foldingathome.org/). The proposed approach can therefore support law enforcement activities while providing the desired functionality to the web community.

The paper is organized as follows. The next section shortly reminds the history of CAPTCHA mechanisms. Section III presents the general idea of the Proof-of-Work and its applications. Chapter IV cites examples of several solutions to stop unwanted web traffic. The next section presents details of the proposed solution. At the end, we add some concluding remarks.

II. Overview of the current CAPTCHA solutions

CAPTCHA ("Completely Automated Public Turing test to tell Computers and Humans Apart") is a type of challenge–response test used in computing to determine whether the user is human [4]. Since its inception, there have been many implementations with varying characteristics and effectiveness.

Released in 2007, the first iteration of reCAPTCHA (reCaptcha v1) asked the user to re-type a blurry and distorted set of words to gain access to the web resource. With the increasing rise and accuracy of optical character recognition software, a second version of reCAPTCHA has been introduced, taking a different approach. This time, the user was asked to select pictures containing a given object out of a set of images. Finally, in 2017 Google released reCAPTCHA v3 which allows verifying whether the user is a bot without any additional user interaction. The mechanism tracks user's interaction with the website and returns a score determining how likely it is that this user is a bot.

While all the solutions mentioned above work in stopping unwanted traffic, they also significantly cause reduced User Experience. Selecting images of traffic lights or fire hydrants has become an inseparable and frustrating part of numerous online forms. Additionally, even the most recent reCAPTCHA solutions implementing artificial intelligence can be bypassed with more than 90% success rate [5]. Moreover, there are numerous Captcha Solving Services [6] that provide correct solutions for given CAPTCHAs for a small fee making it trivial for sufficiently motivated opponents to bypass all of the most popular CAPTCHA implementations.

III. Proof-of-Work mechanism

Proof-of-Work is a mechanism designed to provide the verifying party a cryptographic proof that the user utilized a specified amount of computing power to perform a task. While there are many approaches to implementing a Proof–of–work algorithm, the most popular approach is to perform a digest (or hash) function of a given data alongside a nonce value until certain criteria have been met. This approach is based on the fact that a good digest function is preimage resistant which means that it is computationally infeasible to find any input m that has a given digest $h = H(m)$. The only way of finding such a message is brute-force. An example of such criteria can be the number of zeroes in the binary representation of the resulting hash. If the output of the hash function does not satisfy it, the nonce value is incremented and the whole process starts from the beginning. Upon finding the correct nonce value that produces a hash output described by the verifying party — the user sends the nonce value to be verified. The party in charge of verifying the work performed by the user needs to perform only one hash function with the nonce value sent by him to determine whether the proof is correct or not. This property makes it easy for the system to scale for a large number of users while keeping only one party responsible for the verification of the Proof–of–Work. Worth noting here is the fact that current Proof-of-Work algorithms require the solver to find an exact nonce value rather than a range of acceptable solutions. Additionally, Proof–of–Work has been criticized for the high energy consumption required to perform the task.

IV. Related work

There have been numerous attempts to stop unwanted web traffic by utilizing users' computing power. Some of them are described below.

A. Hashcash

Proposed in 1997 by Adam Back, Hashcash[7] is a mechanism originally proposed to combat email spam abuse. When sending an email, the user performs a Proof-of-Work algorithm on the whole body of the email message. Additional data is appended such as timestamp and string of random characters. Computation is performed until the sender reaches a desired number of zero bits in the hash output and a resulting counter value is obtained. Finally, a new header is appended to the email containing information used to prove the work. Although the proposed system did not achieve significant adoption, a version similar to it has been implemented into Bitcoin's [8] mining mechanism.

B. CoinHive CAPTCHA

Around the year of 2018, a company called CoinHive launched its Proof-of-Work captcha widget, a reCAPTCHA alternative where a user visiting the website can commit a portion of his device's computing power to mine cryptocurrency for the website owner instead of selecting a set of images containing a specified item. Unfortunately, CoinHive's solution became widely used for attackers to perform cryptojacking [9] — an attack when a user's computing power is used to mine cryptocurrency for the attacker.

C. Cloudflare Turnstile

Cloudflare, a company specializing in DDoS attack mitigation has introduced its own CAPTCHA alternative with Proof-of-Work mechanism. Turnstile [10] aims to replace frustrating CAPTCHAs by utilizing a set of non-interactive JavaScript challenges. Some of that challenges require the user's device to perform computations similar to how the Proof-of-work mechanism works. Although Cloudflare's solution is relatively similar to the one proposed in this paper, the computing power used for its challenges is not used in a way that can be used for other purposes.

V. Proposed solution

As shown above, a sufficiently motivated opponent should not have any difficulties bypassing the most popular CAPTCHA implementations. Taking that into consideration, an alternative approach has been taken. The proposed solution borrows some of its features from Hashcash [7]–style Proof–of–Work with additional mechanisms. Instead of requiring from user to select a set of images, the proposed solution utilizes a small portion of user's computing power to solve a cryptographic puzzle. Moreover, rather than utilizing CPU power to compute meaningless hash functions that can be seen as a waste of electricity, the system uses it to perform brute-force attacks on password hashes stored within the system's database.

Instead of specifying a strict value for the target as is usual in Proof-of-Work implementations, the system provides an *upper* and *lower* bounds of target values of the resulting hash. These bounds can be arbitrarily changed by the system to reduce the puzzles complexity or make it more difficult for suspicious traffic.

This approach can be visualized by plotting all possible target values with the length of *n* bytes in a circle as shown in Figure 1. The resulting image resembles a clock, where the upper and lower bounds can be thought of as the hands of that clock. The radius of the bounds is the range for which the hash value must be found in order to complete the puzzle.

Additionally, to correctly test the entire set of characters, the system provides the user with the *starting_point* value which binary representation should be considered as starting point for the puzzle and any value below it will be rejected by the system as an incorrect solution.

A sample CAPTCHA puzzle request has been presented below. The *token* value can be treated as a random ID used to identify the puzzle within the system.

```
{
    "hash_type": "SHA256",
    "starting_point": "20AA",
    "lower_target": "11FF41",
    "upper_target": "3228AF",
    "token": "c41...4e9"
}
```

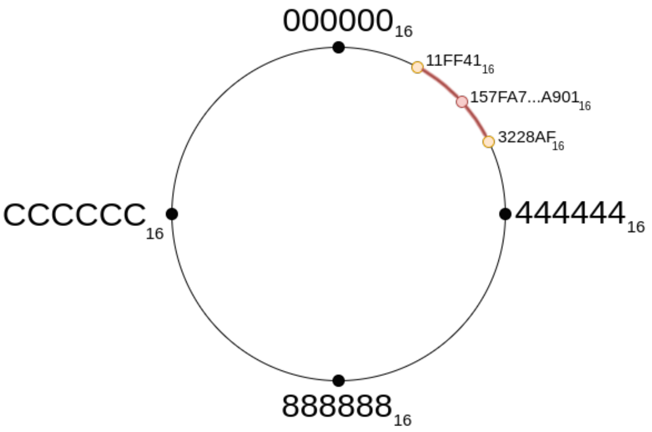


Figure 1. Visualistion of 3-byte space of values encoded in hex. Lower and upper target bound values have been highlighted in yellow while the target hash value has been represented by the red dot.

Next, the user attempts to solve the puzzle by incrementing the value of *starting_point* decoded to its binary format. The resulting binary data is then hashed using the digest algorithm negotiated between server and client and a hash value is obtained. Finally, the users compare the first *n* bits of resulting hash where *n* is the length of the *upper_target* in its binary form with the upper and lower target values. If the resulting value fits within the boundary selected by the system, the puzzle has been completed successfully.

Once the correct solution has been found, a preimage of the resulting hash is sent to the server for verification. If the verification process is completed, access is granted. A sample response to the system has been presented below.

```
{
    "token": "c41...4e9",
    "value": "20AA221F"
}
```

In short, the system guides the solver to the hash it is looking for, while implementing certain character requirements that must be met to mark the CAPTCHA puzzle as complete. The lower and upper bounds are used to adjust the difficulty of the puzzle and guide the solver toward the system's desired hash, while the starting point value is used to ensure that all possible values have been checked.

By carefully tweaking the bounds and starting point values for every user or for a batch of users, the system can test the entire range of values by utilizing users' computing power.

A. Puzzle verification

By referencing the *token* value within the system's internal database, information about the digest function, start point, and target bounds is obtained, along with a timestamp indicating when the puzzle request was made. The system then hashes the value provided by the user with the appropriate hash function and checks that it fits within the target bounds. Next, the system checks whether the value provided by the user is greater than the given starting point. Finally, the server checks the time difference between the generation of the puzzle and the submission of the solution to enforce a time-based expiration of the CAPTCHA. Only when all checks have been successfully completed can the user be granted access.

Additionally, to detect potentially malicious behavior, the system calculates an expected average amount of hash calculations that need to be performed in order to satisfy the puzzle where n is the length of the hash function in bits.

$$x = \frac{2^{(n-1)}}{upper\_target - lower\_target} \tag{1}$$

Similarly, the amount of hashes calculated if the user was honest is derived by subtracting the user-supplied solution by the puzzle *starting_point*.

$$Hashes\_Calculated = Solution - Starting\_Point \tag{2}$$

In a case where a user-supplied solution requires much more hash functions to be calculated than what was expected by the system, access still should be granted to the user to avoid false positives, but his solution should not be taken into account when generating new starting point value for the next batch of users.

Worth noting here is the fact, that when the server is verifying the puzzle, it also compares the resulting hash with the hash the system is trying to crack. Additionally, even when the user correctly finds the hash preimage, he can not distinguish that his solution in fact cracks the hash — server responds to the correct solution value in the same manner.

B. Adjusting starting point value

In a perfect world, where all users are honest, the starting point of one user should be equal to the solution value of the previous user incremented by 1. Unfortunately, that assumption can not be made. Despite this, that approach is still viable when a larger batch of users is taken into account. The system should serve the same puzzle (although with a different *token* value) to a small group of users. Only after a sufficient number of solutions have been proposed, the server can treat the solution as correct and correctly incremented if at least 51% consensus in regards to submitted value has been achieved amongst these users. After that, *starting_point* value of the next batch of users should be the same as the solution of the previous batch incremented by 1. Ideally, the system should contain many hashes ready to be converted to CAPTCHA puzzles to avoid the situation where one user lands in the same batch and receives the same puzzle more than once.

C. Puzzle difficulty adjustment

Just like all the other CAPTCHA solutions, there is a need for a mechanism to make the puzzle easier or harder for the user to solve. In the case of the proposed solution, decreasing the radius between the lower and upper bounds acts as a tool for difficulty adjustment. By making the bounds closer together, the system decreases the number of potential values that can be used as a solution.

VI. Benefits of the proposed solution

The proposed solution provides an interesting alternative to the popular image-based CAPTCHA system. It requires no user interaction making the user experience seamless and better suited for users with accessibility issues. Additionally, the system successfully prevents attack utilizing machine learning as the best strategy for solving the cryptographic puzzle is by a brute–force attack.

Compared to solutions that use essentially meaningless computation tasks, proposed solutions manage to utilize users' computing power to achieve an ultimately beneficial goal. The ability to quickly and cheaply crack password hashes of seized, but encrypted devices, would be of great help for law enforcement bodies. By giving a purpose to the computation task that would have to be performed nonetheless, the overall energy consumption is reduced. Moreover, there is a clear need for a Proof–of–Work style verification as shown by the Cloudflare Turnstile solution (see Section IV-C). The proposed solution can be thought of as an improvement to aforementioned mechanism, adding a meaningful purpose to the computational task..

VII. Threat modeling

In this section, we analyze how the proposed solution could be attacked to gain an unfair advantage or bypass it completely.

A. Low electricity cost

The cost of electricity can be widely different across different regions of the world making it difficult to estimate the amount of computing power that would be considered sufficient to scare off the attacker. As a result, the server responsible for generating and scheduling the puzzles should dynamically adjust the difficulty based on a range of gathered metrics. Aside from the IP address metrics could include data gathered by the JavaScript code run on the solver's machine such as time zone, default language, or hardware information. The cost of electricity used to solve a standard task has not been measured and are a subject of further research.

B. Use of FPGA, ASIC or botnet network to solve cryptographic puzzles

A dedicated attacker could use a botnet network to harvest cheap and accessible computing power for solving cryptographic puzzles and abusing the system. This case can be compared to existing and commonly used services of independent CAPTCHA solvers. Worth noting is the fact that increased CPU usage and power consumption make the botnet more susceptible to detection as the victim's computers would be negatively affected by the computation. Additionally, the time needed to solve a cryptographic puzzle can be vastly reduced by specialized hardware such as FPGA or ASIC. While this attack would in fact be successful, it would require a large upfront investment by the attacker on top of the already increased electricity bill. We are not able to share performance benchmarks of the proposed solutions as proposed solution is still at the work–in–progress stage.

In general, we conclude with the estimate that the security implications of the proposed solution are similar to that of the modern CAPTCHA solutions used across the web today.

VIII. Conclusion

The system introduces a real-world cost on the attacker requiring a greater electricity use to successfully pass the puzzle. Combined with unobtrusive network traffic analysis techniques such as originating IP address it can reduce the complexity of the puzzle for users categorized as low-risk. Computing power used to solve cryptographic puzzles is used to crack password hashes stored within the system's database and can help law enforcement gain access to seized devices. The system could potentially be adapted to utilize computing power already used for Proof–of–Work style verification, thus reducing overall electricity usage and introducing an additional benefit to the computation.

Acknowledgment

The research presented in this paper was realized with funds from the Polish Ministry of Science and Higher Education assigned to AGH University of Krakow and it was supported in part by Cybercrypt@gov project (no. DOB-BIO9/32/03/2018).

References

[1] Netacea, "Businesses lose up to $250m every year to unwanted bot attacks," https://netacea.com/blog/businesses-lose-up-to-250m-every-year-bots/, [Accessed 13-March-2023].

[2] Wappalyzer, "reCAPTCHA market share compared to an alternative hCAPTCHA," https://www.wappalyzer.com/compare/recaptcha-vs-hcaptcha/, [Accessed 13-March-2023].

[3] Fastcompany, "Google's new recaptcha has a dark side," https://www.fastcompany.com/90369697/googles-new-recaptcha-has-a-dark-side, [Accessed 13-March-2023].

[4] Wikipedia, "CAPTCHA — Wikipedia, the free encyclopedia," https://en.wikipedia.org/wiki/CAPTCHA, 2023, [Accessed 04-February-2023].

[5] I. Akrout, A. Feriani, and M. Akrout, "Hacking google recaptcha v3 using reinforcement learning," 2019. doi: 10.48550/ARXIV.1903.01003. [Online]. Available: https://arxiv.org/abs/1903.01003

[6] 2Captcha, "a captcha solving solution," https://2captcha.com/, [Accessed 13-March-2023].

[7] A. Back, "Hashcash – a denial of service counter-measure," 2002, [Accessed 06-February-2023]. [Online]. Available: http://www.hashcash.org/papers/hashcash.pdf

[8] S. Nakamoto, "Bitcoin: A peer–to–peer electronic cash system," 2008, [Accessed 02-May-2023]. [Online]. Available: https://bitcoin.org/bitcoin.pdf

[9] Interpol, "Cryptojacking," https://www.interpol.int/en/Crimes/Cybercrime/Cryptojacking, [Accessed 02-March-2023].

[10] Cloudflare, "Turnstile," https://developers.cloudflare.com/turnstile/, 2023, [Accessed 25-February-2023].

Communication Papers of the 18th Conference on Computer Science and Intelligence Systems pp. 65–74
DOI: 10.15439/2023F8794
ISSN 2300-5963 ACSIS, Vol. 37

Formal verification of BPMN diagrams in Integrated Model of Distributed Systems (IMDS)

Jakub Jałowiec
Institute of Computer Science,
Warsaw University of Technology
Nowowiejska Str. 15/19, 00-665 Warsaw, Poland
Email: kuba.jal@gmail.com

Wiktor B. Daszczuk
Institute of Computer Science,
Warsaw University of Technology
Nowowiejska Str. 15/19, 00-665 Warsaw, Poland
Email: wiktor.daszczuk@pw.edu.pl

Abstract— **Business process model and notation (BPMN) is a way of describing business processes using convenient diagrams. In the last decade, it became a de-facto industry standard, widely used by software architects and business analysts to describe business requirements and the overall structure of a designed information system. Ensuring that diagrams model their intended behavior is of utmost importance for notation users. This article deals with the definition of BPMN through the conversion to the Integrated Model of Distributed Systems (IMDS) and automated verification of BPMN diagrams. The translation of a subset of BPMN preserves information about the processes in the formal model. This allows finding partial deadlocks and checking partial termination (concerning a subset of processes), verification in terms of BPMN processes, and mapping found errors onto source BPMN definition. Moreover, IMDS is tailored to model distributed systems, which is the very nature of business processes. A tool for automated translation of BPMN diagrams to IMDS, automated verification, and visualization of results is developed.**

I. Introduction

An EXAMPLE of business processes in everyday life is ordering a product in an online shop. Appropriate steps of such a business process include several steps, like filling up the online form (ordering the product), preparation of money transfer form, payment, shipping the order, delivery and final confirmation.

Business processes involve several steps and should preserve logical relationships, be efficient, reliable, and coordinate work between stakeholders. Business processes have theoretical foundations in workflow nets [1], a class of Petri nets [2] used to model business behavior and formally analyzed mathematically.

Business Process Model and Notation (BPMN [3]), is a graphical framework used to model business processes with around 100 symbols representing various aspects of process execution, communication, and dataflow. It has found commercial use with numerous supporting tools. Proper modeling is important to save time and money during implementation and maintenance. The verification using model checking and temporal logic [4] ensures that a BPMN diagram follows its intended behavior.

BPMN's rich syntax presents challenges in formalizing its semantics [5], making it difficult to verify properties of a given model. This article presents a method for verifying BPMN diagrams using the Integrated Model of Distributed Systems (IMDS [6]), which is a formalism for describing and verifying distributed computational systems. IMDS is capable of detecting partial deadlocks, which involve only a subset of processes. The proposed method maps BPMN into IMDS, giving the diagrams formal semantics, and allowing for partial (or total) deadlocks detection. The Dedan tool [7] is used to automatically detect deadlocks and check the termination of the entire model system or set of processes.

The contributions of this article are:

1. Normalization of BPMN Process and Collaboration diagrams to an intermediate representation. It provides explicit semantics for all elements and helps to avoid ambiguity in interpretation.
2. Translation rules of BPMN to IMDS. All implemented BPMN elements have been reduced to 9 constructions for which implementation in IMDS was given. These constructs can be thought of as an "intermediate language" defining the semantics of BPMN elements through their corresponding IMDS structures.
3. Verification in IMDS for the location of deadlocks (total and partial) or checking the termination (total and partial). Checking of partial properties, not present in other tools, allows the designer to find errors in diagrams with their local cooperation, and even individual diagrams. Moreover, our verification tool Dedan uses a fair verification algorithm that prevents discovering false deadlocks [8].
4. Mapping verification traces back to BPMN specification to observe errors in the original specification rather than in the verification model. The verification methods described in the literature give the result of the check in a form specific to them, leaving the mapping of the resulting trace on the source BPMN diagram for a human. It is possible because the semantics of BPMN and its IMDS translation is the same.

5. Animation of counterexamples/witnesses traces on BPMN diagrams to observe the behavior of model components leading to a deadlock or successful termination/lack of termination.

There are many publications about verifying BPMN using different kinds of formalisms. However, few present a working solution that can be used to verify real-life use cases of BPMN diagrams in an automated way and view the verification results on the source BPMN diagram. However, the most important is automatic checking of partial deadlocks, which is rare in BPMN verification. A partial deadlock can occur even in a single process or in two communicating processes, while other processes perform their work.

II. Related work

Multiple techniques of BPMN formalization have been proposed. A common way of doing it is to map BPMN into Petri nets. Dijkman et al. [9] propose a mapping into classic Petri nets. It deals with a concise subset of BPMN, containing BPMN elements with local semantics, including Pools, Sub-processes, Exclusive Gateways, Parallel Gateways, Event-Based Gateways, Tasks, Events, and exception handling, whereas Tasks with only a single outgoing or incoming Message Flow are considered. There is also presented a tool to transform BPMN diagrams into a Petri Net Markup Language specification [10] that Petri net verifier can further process. Although the article deals with translation only of BPMN 1.x diagrams, it gives a general idea of BPMN 2.0.x translation.

Rachdi [11] proposes mapping into Time Petri Nets (TPN). The considered subset is the same as in [9], but the mapping introduces time constraints on executing BPMN elements. The authors propose an algorithm for the reachability analysis of the resulting TPN but do not provide an automatic tool. Other approaches to model time in BPMN are described in [12].

Authors of [13] propose mapping BPMN into Colored Petri Nets (CPN). Additionally, they introduce a method to divide a given BPMN model into partitions and verify them hierarchically, which reduces the complexity of the resulting CPN model. Unfortunately, the method works only for BPMN model that is well-defined, that is, only for a relatively small subset of BPMN, which limits its expressiveness. Additionally, the authors have implemented a tool that automatically translates BPMN diagrams into the corresponding CPN.

Li and Die [14] propose another method of mapping BPMN into classic Petri Nets. The method is rather poorly described and can be applied to only a small subset of BPMN, but it introduces the concept of preprocessing of BPMN diagrams. A similar concept, called normalization, is proposed in this article. Other attempts to formalize BPMN include for example transformation of BPMN into Pi-Calculus [15], PROMELA [16], COWS [17], Alvis [18] and YAWL [19] with the formalization through Graph Transformation Systems [20].

Automated BPMN diagrams verification is conducted in the BProVe program [21][22]. It verifies the diagrams in terms of safeness, proper termination, dead activities, and other properties. Its underlying logic is based on translating BPMN into MAUDE – a re-writing logic implementation. Another automatic tool VBPMN [23], uses a translation of BPMN into LOTOS NT process algebra formal notation and verification using the CDAP verifier. Its authors also provide a set of benchmarks [24] that are used during tests of the proposed automatic tool. Work [25] proposes verification using process automata that can be compared to our IMDS graphical view [26]; however, this technique concerns checking a single BPMN pool. In [20], the Bogor LTL checker is used to verify the workflows. As in other approaches, only total deadlocks are caught automatically; partial deadlocks require the specification of temporal formulas.

Some verification techniques concern only a limited set of BPMN elements, for example, in [25], communication between pools and boundary links are not considered.

Modeling in rules and processes is proposed in [27], in a spreadsheet in [28], Free-Choice Nets [29], Function-Behaviour-Structure Diagram [30], and Linked Data [31], but without a verification.

The detailed comparison of verification techniques and subsets of BPMN elements served in individual methods cannot be presented due to size limitation of the article. Some overviews of the verification tools and approaches can be found in [32], [33].

III. IMDS and BPMN

A. Overview of IMDS

IMDS formalism is addressed to distributed systems modeling. Its main idea is to show interactions between the two basic concepts: servers and agents. Servers $S=\{s_1,...,s_n\}$ are distributed computing nodes offering some services. Agents $A=\{a_1,...,a_k\}$ are distributed computations modeled as sequences of messages invoking servers' services. A system configuration T is a set of current servers' *states*=(*server, value*) – one state per server – and *messages*=(*agent, server, service*) of all agents – one message per agent. An interaction between servers and agents takes the form of *actions* in set *F*. Action is the execution of a service on a server by an agent message. The action transforms one configuration into another one, in which the server state and the agent message are replaced by new ones: ((*agent message, server state*), (*next agent message, next server state*)). There are also agent-terminating actions that do not produce a new. The system starts with an initial configuration T_0 containing initial states of all servers M_{init} and starting messages of all agents R_{init}. The formal definition of IMDS can be found in [6]. For an IMDS system we will use the notation $(S, A, M_{init}, R_{init}, F)$. IMDS semantics is defined by a Labeled Transition System

(LTS) where the nodes represent system configurations, and the arcs represent IMDS actions. The LTS root is the initial configuration.

B. Business Process Model and Notation (BPMN)

The BPMN standard includes four diagram types: Process, Collaboration, Choreography, and Conversation [3]. This article focuses on verifying Process and Collaboration diagrams, which describe the control. The elements used are Pools, Swimlines, Flow Objects, Data Objects, Connecting Objects, and Artifacts. Data Objects are not considered, because they do not have any semantic meaning when verifying the behavior [5]. BPMN lacks a formal definition framework, necessitating the need for a formal verification method.

C. BPMN Process Diagram syntax

BPMN Process Diagram is a graph $PD=(N,F)$, where N are nodes: *Activities A*, *Gateways G* (*Exclusive* G_X and *Parallel* G_P), *Events E* (*Start* E_s, *End* E_e and *Interrupting Boundary Intermediate* E_i). F are *Flows*: *Sequence* F_s, *Message* F_m and *Boundary Links* F_b. They are $F⊆N×N$, $F_s⊆N×N$, $F_m ⊆A×A$, $F_b⊆A×E_i$.

In the proposed syntax, *nodes* correspond to *Flow Objects*. They play the role of building blocks of BPMN diagrams. Fig. 1 shows all major subtypes of *Flow Object* that are in the scope of this article and their graphical notation.

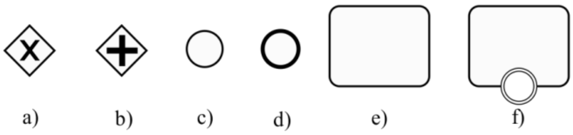


Fig. 1. The subset of BPMN elements formalized in this article: a) *Exclusive Gateway*, b) *Parallel Gateway*, c) *Start Event*, d) *End Event*, e) Activity f) Activity with attached Interrupting Boundary Int*ermediate Event*

F is a non-symmetrical relation which means that for a given element $f=(n_1,n_2)∈F$, there is a directed connection from n_1 to n_2, $source(f)=n_1$ and $target(f)=n_2$. In addition to *Sequence Flows* and *Message Flows,* which are explicitly defined in the BPMN specification, we allow *Boundary Links*. Additionally, *Sequence Flows*, *Message Flows* and *Boundary Links* are subject to the following syntactic rules: *Message Flows* are between different *Pools*, *Sequence Flows* are inside a *Pool*, *Boundary Links* connect an *Activity* to an *Interrupting Boundary Intermediate Event*. We will also use the following notation: *in* are incoming flows, *out* – outgoing, *seq* concerns *Sequence Flows* and *mes* concerns *Message Flows*. For a given element E we have $in(E)$, $out(E)$, $in_{seq}(E)$, $out_{seq}(E)$, $is_{mes}(E)$, $out_{mes}(E)$.

Each maximal connected component of the underlying undirected graph of $P=(N,F_s∪F_b)$ is called a BPMN *Pool* (an element for grouping nodes). Every node in N must be contained in some *Pool*. Additionally, we intentionally omit BPMN *Swimlanes* as the authors of the notation left its meaning to the modeler, which is ambiguous [5].

There are additional syntactical constraints that the graph (N,F) must fulfil: *Sequence Flows* cannot have the same node as source and target, *Start Events* must not have incoming *Sequence Flows*, *End Events* must not have outgoing *Sequence Flows*, and *Interrupting Boundary Intermediate Events* must have exactly one incoming flow: a *Boundary Link*.

D. BPMN Process Diagram semantics

The proposed BPMN semantics is based on the labeled transition system, called *state space*. For BPMN nodes we use the term *internal state* to distinguish it from IMDS *server state*.

While the general structure of a diagram – its nodes and flows – constitutes the static characteristics of a diagram, the concept of *tokens* and their interactions with nodes is introduced to describe the semantics of BPMN [3]. In contrast to [3], we model BPMN messages using tokens, which simplifies the semantics of *Message Flows*.

A *marking* is a pair $S = (D_{flows}, D_{nodes})$, where: D_{flows} is a distribution of tokens on flows of the given diagram, and D_{nodes} is a distribution of symbols {*neutral, activated, 1, 2, ...*} on the nodes of the given diagram. The symbol of the given node N in D_{nodes} is the *internal state of N*. In the Initial Marking, the *Start Events* are in the *activated* state, all other nodes are in *neutral*, and all D_{flows} have 0 tokens. The *End Events* play the role of sinks for tokens.

The lifecycle of a BPMN element, is just an automaton: the loop of transitions between the states *neutral, activated, 1, 2,...,* and back to *neutral*. If there exists a *Boundary Link*, then a transition from every state to *neutral* is triggered by an *Interrupting Boundary Intermediate Event*.

In order to formalize BPMN execution semantics in a concise way, the authors propose to split the dynamics of a BPMN element into 3 phases: a*ctivation*, *execution pattern* and *completion. Activation* and *completion* are atomic transitions fired at the *beginning* and at the *ending* of BPMN element execution, respectively. *Execution pattern* in turn is a sub-automaton which simulates the stateful execution semantics of a given BPMN element.

Activation can be fired only if the required number of tokens were put on the incoming *Sequence Flows* or *Boundary Links* of a given BPMN element, and the node has been the *neutral* state. During *activation* of the BPMN element, two things happen atomically: the number of tokens from its incoming flows that triggered the activation is consumed (i.e. the tokens disappear) and the element's changes its state to *activated*.

Activation it is followed by the *execution pattern* which simulates arbitrary stateful interactions that the BPMN element may be subject to. *Execution pattern* is required to mark its ending with a *completion* transition to the *neutral* state of the BPMN element. This article focuses primary on BPMN elements that follow an *execution pattern* involving: empty automata, handling of *Message Flows* in *Activities* (i.e. inter-process communication) and handling of

Boundary Events in *Activities* (i.e. exception handling) which will be discussed later in this subsection.

Completion finalizes the execution of a BPMN element. Analogously to *activation*, the following two things happen atomically at the same time during it: a specified number of tokens is emitted along a subset of its outgoing Sequence Flows and the element changes its state back to *neutral*.

1. Activation patterns

We interpret the behavior of BPMN diagram as a token game: the tokens on arrows incoming to an element represent preconditions, and the tokens on arrows outgoing from an element represent postconditions. Each BPMN element follows either of the two activation patterns: XOR or AND. If a BPMN element follows the *XOR activation pattern*, a token on any of its incoming *Sequence Flows* or *Boundary Links* will enable activation of the element. In case of the *AND activation pattern*, a token will be put on each of its incoming *Sequence Flows* to enable its activation. If the number of tokens on any incoming Sequence Flow exceeds the number of tokens required for its activation, only the required number of tokens is consumed.

2. Execution patterns

Empty execution pattern. This pattern represents the trivial case when the element does not involve a complex stateful synchronization logic. That includes elements such as *Parallel* and *Exclusive Gateways* (control flow elements) as well as normalized *Activities* that do not have incident Message Flows or incident *Interrupting Boundary Intermediate Events*.

Handling of Message Flows. We define *handling of Message Flows of a given BPMN element E* as sequential generation of tokens on the outgoing *Message Flows*, and consumption of tokens from the incoming *Message Flows* incident to *E*, provided that the element was *activated*. The *Message Flows* are processed sequentially, in a left-to-right and then upper-to-lower order implied by graphical representation of the BPMN diagram. If the given incoming *Message Flow* does not contain a token, the message exchange will be blocked until such a token appears on it. Each numeric internal state (*1, 2, ...*) denotes how many *Message Flows* have been processed by the *Activity*.

Handling of Boundary Events (Exception handling). By *exception handling,* we mean the ability of *BPMN Activities* with *Boundary Events* to change their state to *neutral* at any point of their execution and generate a token on their incident *Boundary Link*. This definition effectively treats *Boundary Events* as *interrupting exceptions*.

3. Completion patterns

Analogously to *nodes activation*, each BPMN element follows either of the two *completion patterns*: XOR and AND. If a BPMN element follows the *XOR completion pattern*, a token will be generated along exactly one arbitrary outgoing *Sequence Flows*. If the BPMN element follows the *AND activation pattern*, a token will be generated on each of its outgoing *Sequence Flows*.

4. Final remarks on execution semantics

The activation and completion patterns for individual BPMN elements are: *Parallel Gateway* (AND, AND), *Exclusive Gateway* (XOR, XOR), *Activity* (XOR, AND), *Event* (XOR, AND).

Additionally, we assume throughout this article that BPMN uses interleaving semantics. It means that if many transformations are enabled in the diagram, one of them is executed, chosen in a non-deterministic manner. Choosing other semantics is also possible, but interleaving matches the semantics of IMDS used for verification. As shown in [34], every coincidence-based system can be transformed into an interleaved system.

The concepts of reachability, initial marking, reachable markings, and marking space are introduced to complement the semantics of BPMN.

Consider marking *A*. *Reachable markings* are all markings that can be reached from *A* by executing a sequence of transformations (nodes activation, message exchange, nodes completion, *Boundary Event* handling). The initial marking sets all Start Events to the *activated* state and all others to the *neutral*.

The transformation of a BPMN diagram include:

1. Node completion changes the state of the node to *neutral* and inserts tokens to the output *Sequence Flows* following the appropriate *Completion Pattern*.
2. Node activation removes the tokens from input *Sequence Flows* following the appropriate *Activation Pattern* and changes the state of the node from activated to the next state according to its *Execution Pattern*.
3. Message sending changes the state of the node to the next state and inserts a token to the *Message Flow*.
4. Message receiving removes the token from the *Message Flow* and changes the node state to the next state.
5. Executing a *Boundary Link* resets the state of the node to *neutral* and inserts a token into the *Sequence Flow* incident to the link.

The initial marking of the diagram is implied by its structure. Namely, all nodes with no incoming *Sequence Flows* that are not *Boundary Events* are initially in the activated state, i.e., they contain tokens.

The *marking space* of the BPMN diagram is the graph $G = (S_0, S, R)$, where S_0 is initial marking (position of tokens and internal states of nodes), S is a set of all reachable markings and R is a transformation relation moving the tokens and changing states of the nodes.

To sum up, vertices of *marking space* are markings of the given BPMN diagram, that are reachable from its initial marking. Initially, the tokens are present in all *Sequence Flows* outgoing from *Start Events*, as they are *activated*. Every transformation moves activation to the output of the *Sequence Flows*. Some transformations are executed with non-deterministic choices, like *Exclusive Gateways* and *Activities* with *Boundary Events*. If $|in_{mes}(n)|+|out_{mes}(n)|>0$, the enabled transformations contain sending and receiving messages. Also, *Parallel Gateways* move the tokens to all their outgoing *Sequence Flows* atomically. All those rules are adopted in target IMDS models in a slightly different, but equivalent way. The difference lies in breaking atomicity with interleaving; see next section. The formal semantics of BPMN *marking space* is given by translation rules to IMDS system.

Apart from the braking atomicity of BPMN, the marking space of the BPMN diagram and the LTS of its translation to IMDS are the same. So both descriptions share the same semantics.

IV. Translation of BPMN into IMDS

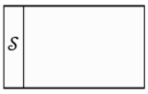
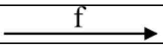
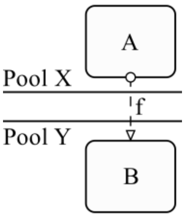
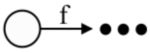
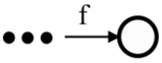
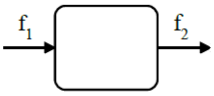
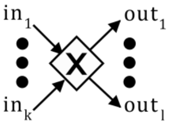
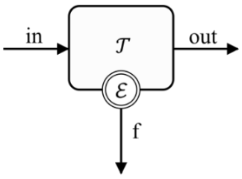
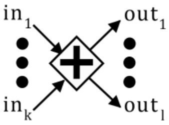
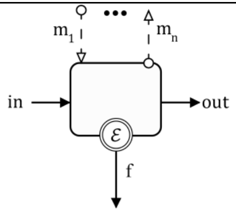
A. *Normalization*

This is a preliminary step for translation to IMDS. Normalization strips the diagram of ambiguity:

- appending *Start Events* to all elements e which are not *Events* and for which $|in_{seq}(e)|$=0,
- appending *End Events* to all elements e for which are not *Events* and for which $|out_{seq}(e)|$=0,
- refactoring all BPMN elements that are not *Exclusive Gateways* and *Parallel Gateways* into semantically

Table I.

BPMN to IMDS translation rules

Group	Element type	Graphical symbol	Translation rules
1	*Pool* $\mathcal{S}$		set of agents $A_{\mathcal{S}} \subseteq A$, server *pool*($\mathcal{S}$) = ($\mathcal{S}$, $V_{\mathcal{S}}$, $Q_{\mathcal{S}}$) , $V_{\mathcal{S}}$ = {*ready*} initial state (*pool*($\mathcal{S}$), *ready*)
2	*Sequence Flow f*		service $f \in Q_{\mathcal{S}}$
3	*Message Flow f* between *Activity A* in *Pool X* and *Activity B* in *Pool Y*,		service $f \in Q_{and(A)}$, service $f \in Q_{and(B)}$, set *agents(f)* = $\left\{agent_f^j \middle| j \in \{1,\dots,K\}\right\} \subseteq A$ initial messages for agents in *agents(f)*: $M_{\mathrm{f}} = \left\{m_f^j \middle| \left(agent_f^j, f\right) \wedge j \in \{1,\dots,K\}\right\} \subseteq M_{init}$
4	*Start Event* $\mathcal{E}$,		agent $a_{\mathcal{E}} \in agents(\mathcal{S})$, initial message $(a_{\mathcal{E}}, pool(\mathcal{S}), f) \in M_{init}$
5	*End Event* $\mathcal{E}$,		action $\{(agents(\mathcal{S}), pool(\mathcal{S}), f), (pool(\mathcal{S}), ready)\}$ $\rightarrow \{-, (pool(\mathcal{S}), ready)\}$ (agent-terminating action)
6	Nodes with 1 incoming & 1 outgoing *Sequence Flows*, no incident *Message Flows*,		action $\{(agents(\mathcal{S}), pool(\mathcal{S}), f_1), (pool(\mathcal{S}), ready)\}$ $\rightarrow \{(agents(\mathcal{S}), pool(\mathcal{S}), f_2), (pool(\mathcal{S}), ready)\}$
7	*XOR activation* / *XOR execution*, (nondeterministic choice)		set of actions: $\{(agents(\mathcal{S}), pool(\mathcal{S}), in_i), (pool(\mathcal{S}), ready)\}$ $\rightarrow \left\{\left(agents(\mathcal{S}), pool(\mathcal{S}), out_j\right), pool(\mathcal{S}), ready)\right\}$ $i \in \{1,\dots,n\}, j \in \{1,\dots,m\}$
8	*Interrupting Boundary Intermediate Event* $\mathcal{E}$ bound to *Activity* $\mathcal{T}$,		2 actions in $Pool$ ($\mathcal{S}$): $\{(agents(\mathcal{S}), pool(\mathcal{S}), in), (pool(\mathcal{S}), ready)\}$ $\rightarrow \{(agents(\mathcal{S}), pool(\mathcal{S}), \varepsilon), (pool(\mathcal{S}), ready)\}$, $\{(agents(\mathcal{S}), pool(\mathcal{S}), in), (pool(\mathcal{S}), ready)\}$ $\rightarrow \{(agents(\mathcal{S}), pool(\mathcal{S}), out), (pool(\mathcal{S}), ready)\}$
9a	*AND activation* / *AND execution*,		Tuple $\begin{pmatrix} ANDserver, andActions, \\ poolActions, agents, initialMessages \end{pmatrix}$ = *createAND*(*n*, $\mathcal{S}$, *K*)
9b	*Activity* with incident *Message Flows* and optional *Interrupting Boundary Intermediate Event*,		

equivalent constructs in which $|out_{seq}(e)|=1$ *and* $|in_{seq}(e)|=1$.

The first two rules follow directly from the BPMN specification of nodes that do not have incoming or outgoing *Sequence Flows* [3]. The third rule guarantees that elements that follow mixed XOR activation and AND completion semantics are transformed into equivalent groups of elements, following either XOR activation and XOR completion semantics or AND activation and AND completion semantics.

B. Translation of normalized graph into IMDS

The model is formally the tuple (S, A, M_{init}, R_{init}, F), The actions have the form ((*agent, server, service*), (*server, state*))/((*agent, other server, other service*), (*server, next state*)). The proposed BPMN-to-IMDS translation reflects the behavior of BPMN elements. *Server states* mimic the BPMN *internal states* while *agent messages* are tokens. Node activation, execution and completion are all mimicked by IMDS *actions*.

Elements of a BPMN diagram can be divided into ten groups with respect to the way that they are mapped into IMDS, as described in Table 1.

For Group (9) the authors propose a special procedure of translation called *createAND*(n, $\mathcal{P}$, K), transforming the given BPMN element e that is contained within a *Pool* $\mathcal{S}$ into a tuple (*ANDserver, andActions, poolActions, agents, initialMessages*) consisting of: a new *ANDserver*, its set of *andActions*, a set of *poolActions* in the corresponding *Pool* $\mathcal{S}$, new agents in *agents*($\mathcal{S}$), that become agents of the *Pool* $\mathcal{S}$, and *initialMessages* of the new agents. The new server mimics the atomic consumption and generation of multiple tokens, and to mimic consumption and generation of tokens. As proved in [34], coincident actions are equivalent to interleaved actions at the cost of adding new states between them.

Because IMDS agents cannot be created dynamically, they must be preallocated during the translation of BPMN into IMDS. Let us introduce a constant K to define how many agents are preallocated on each *Message Flow* or group (9) and used in the function, described in detail in [35].

C. Limitations of the proposed translation

The proposed translation method has some major issues which should be taken into account. The first is that IMDS agents cannot be created dynamically. In order to simulate the dynamic creation of tokens, the concept of *preallocation of agents* is introduced. The translation method cannot work for diagrams whose execution may generate infinitely many tokens (for example in a loop containing a *Parallel Gateway*). We refer to such diagrams as *unbounded*. The translation method proposed in this article can preallocate more IMDS agents than are needed. This is because the translation parameter, K, can be arbitrarily large. If K is allowed to be infinite, then the resulting IMDS system would not be static.

If the translation method is parametrized using $K = 1$, then only a single agent will be created for the *Sequence Flow f*. First execution of the *Parallel Gateway G* is correctly simulated by the translated IMDS model. The problem arises when the gateway is executed for the second time. In this case, there is no agent whose message can mimic the behavior of the token generated for the second time on the *Sequence Flow f*. Using parametrization *K=2* solves the problem, because there is a second preallocated agent, whose messages simulate the second execution. Thus, K has to be chosen carefully. It stems from the nature of IMDS. The choice of K value belong to the designer, it depends on how many token can come to a node "splitting" the behavior.

In the future, we plan to extend the plan to enrich IMDS to cover dynamic process creation, which will substitute agent preallocation.

The second major problem with the proposed translation method concerns BPMN elements with non-local semantics, like Inclusive Gateways. They were excluded from the proposed syntax, because the authors could not propose correct execution semantics for such elements. Thus, they are also not considered in the translation method. We don't plan to support this feature as non-local behavior is incompatible with distributed system.

V. Examples

A. AND activation and AND execution pattern

Here we use the graphical view of IMDS [26]. The example in Fig. 2 contains a *Pool P*, *Start Event E1*, two *End Events E2* and *E3*, *Parallel Gateway G*, and three *Sequence Flows*: *s1, s2, s3*. Those flows names are included in both *Pool server* and *AND server* definitions. We add the suffix *P* to *Pool server* services and *G* to *AND server* services, to differentiate between the two servers' services. Red dashed arrows show the transfers of agents between the servers, the arrows point to the states expected by the agents to perform their next actions.

Let $K = 2$ be the parametrization used in the translation. It can be the result of receiving more than one token acquired from the subdiagram represented by *E1* (for example tokens produced by a *Parallel Gateway*). The translation results in the following IMDS system:

```
(S, A, Minit, Rinit, F) = (
S = {(pool(P), {ready}, {s1P, s2P, s3P}),
(and(G), {0,1,2}, {s1G, s2G, s3G})},
A = {a_{s1}, a^{1}_{s3}, a^{2}_{s3}},
Minit = {(pool(P), ready), (and(G), 0)},
Rinit = {(pool(P), a_{s1}, s1P), (and(G), a^{1}_{s3},
s3G), (and(G), a^{2}_{s3}, s3G)},
F = {
((agents(P),pool(P),s1P),(pool(P),ready))→((age
nts(P),and(G),s1G),(pool(P),ready)),
((agents(P),and(G),s1G), (and(G),0)) →
((agents(P),and(G),s2G),(and(G),1)),
((agents(P),and(G),s3G), (and(G),1)) →
((agents(P),pool(P),s3P), (and(G),2)),
10. ((agents(P),and(G),s2G), (and(G),2)) →
    ((agents(P),pool(P),s2P), (and(G),0)),
11. ((agents(P),pool(P),s2P), (pool(P),ready)) →
    ((pool(P),ready)),
12. ((agents(P),pool(P),s3P), (pool(P),ready)) →
    ((pool(P),ready))})
```

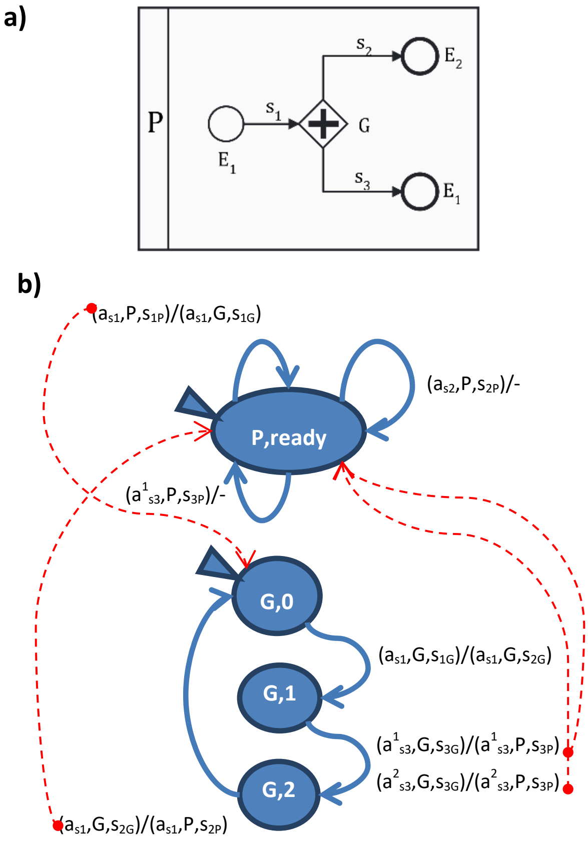

a)
P
E_1
s_1
G
s_2
E_2
s_3
E_1
b)
$(a_{s1},P,s_{1P})/(a_{s1},G,s_{1G})$
$(a_{s2},P,s_{2P})/$-
P,ready
$(a^1_{s3},P,s_{3P})/$-
G,0
$(a_{s1},G,s_{1G})/(a_{s1},G,s_{2G})$
G,1
$(a^1_{s3},G,s_{3G})/(a^1_{s3},P,s_{3P})$
$(a^2_{s3},G,s_{3G})/(a^2_{s3},P,s_{3P})$
G,2
$(a_{s1},G,s_{2G})/(a_{s1},P,s_{2P})$

Fig. 2. a) *Exclusive Gateway* b) its translation to the IMDS automaton

Individual actions are responsible for: Action 7. starts the operation of the *AND server* by transferring the agent a_{s1} from the *Pool server* to *AND server*. Action 8. Transfers the incoming agent to the outgoing Sequence Flow s_2. Action 9. activates the preallocated agent a^1_{s3}, (or a^2_{s3}), transfers it to *the Pool server*, and changes the *AND server* state to *2*. Action 10. finishes counting the flows, transfers agent a_{s1} to the *Pool server*, and changes the state of *AND server* to *0*. Actions 11. and 12. terminate both agents as they are simulating behavior of the two *End Events*.

Note that action 9. can be executed by both agents a^1_{s3} and a^2_{s3}. However, any of those actions changes the state of *AND server* to *2*, prohibiting the other agent to execute its action. As a result, only one on the agents executes its action when the gate *G* is activated. Actions 7.-10. are created by the *createAND* procedure.

B. partial deadlock

Fig. 3 presents an example of a BPMN diagram that contains a partial deadlock. It consists of two *Pools*: *X* and *Y*. *Pool X* consists of one *Start Event*, one *End Event,* two *Exclusive Gateways*, and two *Activities*, one of which is connected through a *Message Flow* to the other *Pool*. *Pool Y* consists of one *Start Event*, one *End Event*, two *Parallel Gateways* connected in a fork manner, and two *Activities*, one of which is connected to an *Exclusive Gateway, Message Flow*, and a *Interrupting Boundary Intermediate Event* of type *Timer*.

This BPMN diagram falls into partial deadlock if *Pool X* follows this execution path: f_1, f_2, f_4, f_6. In this case, *Activity R* in *Pool Y* will keep throwing timeout exceptions because the *Message Flow M* will not be fired. The proposed translation of BPMN diagrams into IMDS lets the designer keep track of such partial deadlocks using the Dedan tool, and backward way to *bpmn2imds* tool for observing the deadlock is source diagram. The agent in *Pool X* is not deadlocked since it terminates. The agent of *M* is technically deadlocked (its starting message will never cause an action), but we do not treat such situation as a real deadlock (however, the designer can draw some conclusions from this fact). One of the agents in *Pool Y* – a_{g1} is looping (g_4, g_5, g_4, g_5, ...); it does not deadlock but the question of its termination gives *false*. The question of possible termination gives *true* with the witness of the agent *X* following f_1, f_3, f_5, f_6, the "upper" agent *Y* – a_{g1} following g_1, g_2, g_4, g_6, g_8, the "lower" agent *Y* - a_{P1g3} following g_3, g_7, and the agent a_M following *Message Flow M*.

Let $K = 1$ be the parametrization used in the translation. The translation results in the following IMDS system; we omit the upper index in the services of *P* and *Q Activities*; we add the *X*, *Y*, *Q* and *R*, P_1 and P_2 suffixes to IMDS services for readability.

Let $K = 1$ be the parametrization used in the translation. The translation results in the following IMDS system; we omit the upper index in the services of *P* and *Q Activities*; we add the *X*, *Y*, *Q* and *R*, P_1 and P_2 suffixes to IMDS services for readability:

```
1.  (S, A, Minit, Rinit, F) = (
2.  S = {
3.  (pool(X), {ready}, {f1, f2, f3, f4, f5, f6}),
4.  (pool(Y), {ready}, {g1, g2, g3, g4, g5, g6, g7,
    g8}),
5.  (and(P1), {0, 1, 2}, {g1, g2, g3}),
6.  (and(P2), {0, 1, 2}, {g6, g7, g8}),
7.  (and(R), {0, 1, 2}, {g4, M, g6}),
8.  (and(Q), {0, 1, 2}, {f3, M, f5})},
9.  A = {af1, ag1, aM, aP1g3},
10. Minit = {(pool(X), ready), (pool(Y), ready),
    (and(Q), 0), (and(R), 0), (and(P1), 0),
    (and(P2), 0)},
11. Rinit = {(pool(X), af1, f1), (pool(Y), ag1, g1),
    (and(Q), aM, M),(and(P1), aP1g3, g3)},
12. F = {
13. ((agents(X),pool(X),f1),
    (pool(X),ready))→((agents(X),pool(X),f2),
    (pool(X),ready)),
14. ((agents(X),pool(X),f1),
    (pool(X),ready))→((agents(X),pool(X),f3),
    (pool(X),ready)),
15. ((agents(X),pool(X),f2),
    (pool(X),ready))→((agents(X),pool(X),f4),
    (pool(X),ready)),
16. ((agents(X),pool(X),f₃),
    (pool(X),ready))→((agents(X),and(Q),f₃),
    (pool(X),ready)),
17. ((agents(X),pool(X),f₄),
    (pool(X),ready))→((agents(X),pool(X),f₆),
    (pool(X),ready)),
```

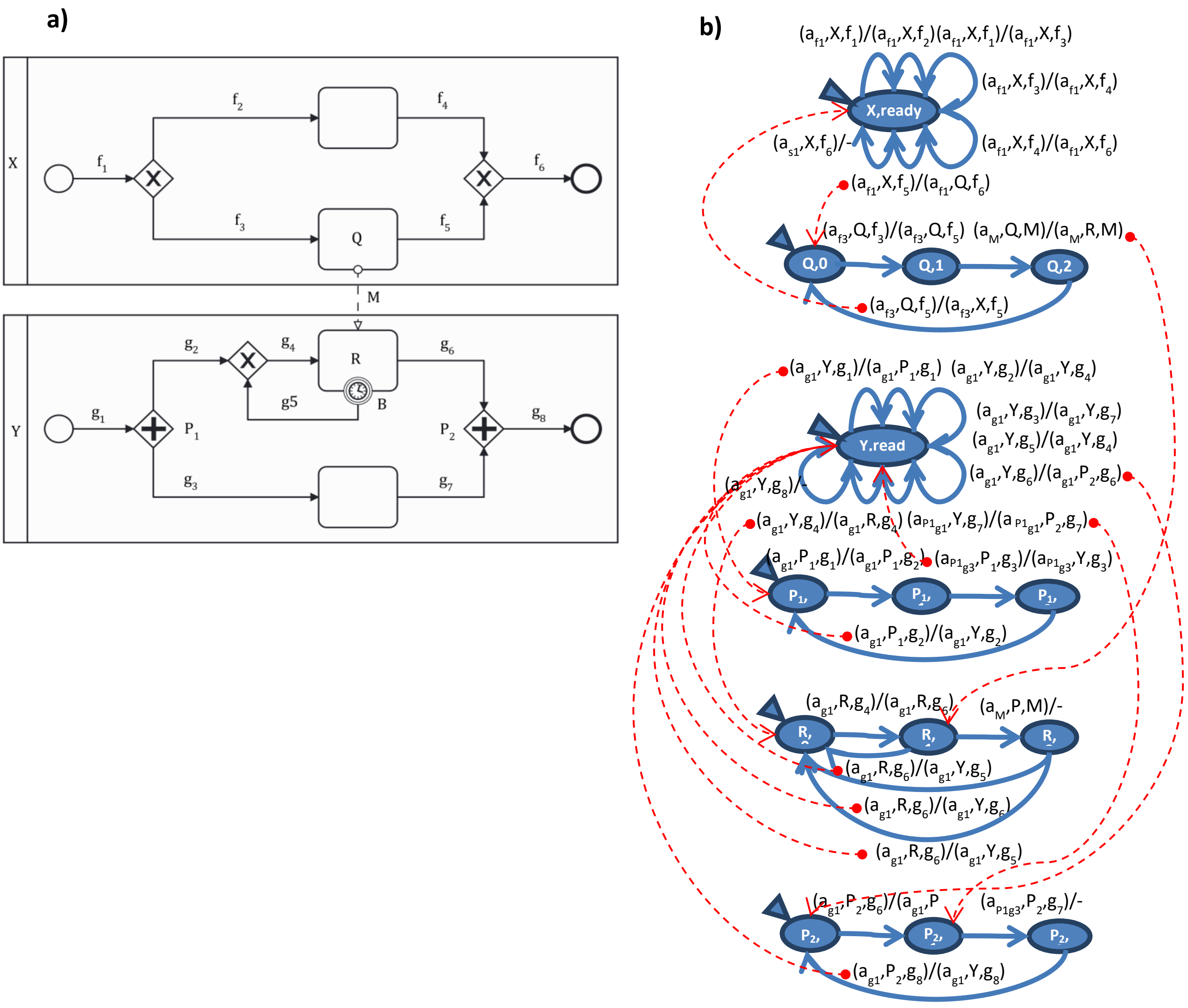
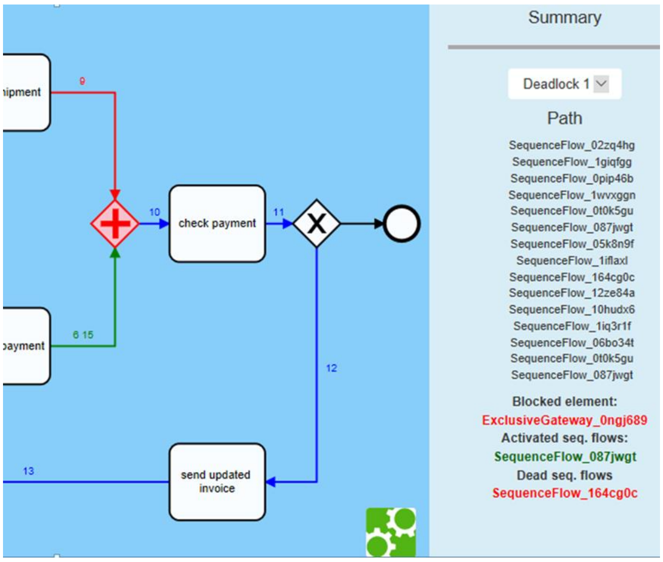

a)
X
Y
f_1
f_2
f_3
f_4
f_5
f_6
Q
M
g_1
g_2
g_3
g_4
g5
g_6
g_7
g_8
R
B
P_1
P_2
b)
$(a_{f1},X,f_1)/(a_{f1},X,f_2)(a_{f1},X,f_1)/(a_{f1},X,f_3)$
$(a_{f1},X,f_3)/(a_{f1},X,f_4)$
X,ready
$(a_{s1},X,f_6)/-$
$(a_{f1},X,f_4)/(a_{f1},X,f_6)$
$(a_{f1},X,f_5)/(a_{f1},Q,f_6)$
$(a_{f3},Q,f_3)/(a_{f3},Q,f_5)$
$(a_M,Q,M)/(a_M,R,M)$
Q,0
Q,1
Q,2
$(a_{f3},Q,f_5)/(a_{f3},X,f_5)$
$(a_{g1},Y,g_1)/(a_{g1},P_1,g_1)$
$(a_{g1},Y,g_2)/(a_{g1},Y,g_4)$
$(a_{g1},Y,g_3)/(a_{g1},Y,g_7)$
$(a_{g1},Y,g_5)/(a_{g1},Y,g_4)$
Y.read
$(a_{g1},Y,g_8)/-$
$(a_{g1},Y,g_6)/(a_{g1},P_2,g_6)$
$(a_{g1},Y,g_4)/(a_{g1},R,g_4)$
$(a_{P1_{g1}},Y,g_7)/(a_{P1_{g1}},P_2,g_7)$
$(a_{g1},P_1,g_1)/(a_{g1},P_1,g_2)$
$(a_{P1_{g3}},P_1,g_3)/(a_{P1_{g3}},Y,g_3)$
$P_1,$
$P_1,$
$P_1,$
$(a_{g1},P_1,g_2)/(a_{g1},Y,g_2)$
$(a_{g1},R,g_4)/(a_{g1},R,g_6)$
$(a_M,P,M)/-$
R,
R,
R,
$(a_{g1},R,g_6)/(a_{g1},Y,g_5)$
$(a_{g1},R,g_6)/(a_{g1},Y,g_6)$
$(a_{g1},R,g_6)/(a_{g1},Y,g_5)$
$(a_{g1},P_2,g_6)/(a_{g1},P$
$(a_{P1g3},P_2,g_7)/-$
$P_2,$
$P_2,$
$P_2,$
$(a_{g1},P_2,g_8)/(a_{g1},Y,g_8)$

Fig. 3. Example BPMN (a) to IMDS (b) translation translation

```
18. ((agents(X),pool(X),f₆),
    (pool(X),ready))→((pool(X),ready)),
19. ((agents(X),and(Q),f₃), (and(Q),0)) →
    ((agents(X),and(Q),f₅), (and(Q),1)),
20. ((agents(M),and(Q),M), (and(Q),1)) →
    ((agents(M),and(R),M), (and(Q),2)),
21. ((agents(X),and(Q),f₅), (and(Q),2)) →
    ((agents(X),pool(X),f₅), (and(Q),0)),
22. ((agents(Y),pool(Y),g₁),
    (pool(Y),ready))→((agents(Y),and(P₁),g₁),
    (pool(Y),ready)),
23. ((agents(Y),and(P₁),g₁), (and(P₁),0)) →
    ((agents(Y),and(P₁),g₂), (and(P₁),1)),
24. ((agents(Y),and(P₁),g₂), (and(P₁),1)) →
    ((agents(Y),pool(Y),g₂), (and(P₁),2)),
25. ((agents(Y),and(P₁),g₃), (and(P₁),2)) →
    ((agents(Y),pool(Y),g₃), (and(P₁),0)),
```

```
26. ((agents(Y),pool(Y),g₂),
    (pool(Y),ready))→((agents(Y),pool(Y),g₄),
    (pool(Y),ready)),
27. ((agents(Y),pool(Y),g₅),
    (pool(Y),ready))→((agents(Y),pool(Y),g₄),
    (pool(Y),ready)),
```

```
28. ((agents(Y),pool(Y),g₄),
    (pool(Y),ready))→((agents(Y),and(R),g₄),
    (pool(Y),ready)),
29. ((agents(Y),pool(Y),g₃),
    (pool(Y),ready))→((agents(Y),pool(Y),g₇),
    (pool(Y),ready)),
30. ((agents(Y),and(R),g₄), (and(Q),0)) →
    ((agents(Y),and(R),g₆), (and (Q),1)),
31. ((agents(M),and(R),M), (and(Q),1)) → ((and
    (Q),2)),
32. ((agents(Y),and(R),g₆), (and(Q),2)) →
    ((agents(Y),pool(Y),g₆), (and (Q),0)),
33. ((agents(Y),pool(Y),g₆),
    (pool(Y),ready))→((agents(Y),and(P₂),g₆),
    (pool(Y),ready)),
34. ((agents(Y),pool(Y),g_7),
    (pool(Y),ready))→((agents(Y),and(P_2),g_7),
    (pool(Y),ready)),
35. ((agents(Y),and(P_2),g_6), (and(P_1),0)) →
    ((agents(Y),and(P_2),g_8), (and(P_1),1)),
36. ((agents(Y),and(P_2),g_7), (and(P_1),1)) →
    ((and(P_1),2)),
37. ((agents(Y),and(P_2),g_8), (and(P_1),2)) →
    ((agents(Y),pool(Y),g_8), (and(P_1),0)),
38. ((agents(Y),pool(Y),g_8), (pool(Y),ready)) →
    ((pool(Y),ready))})
```



Summary

Path

SequenceFlow_02zq4hg
SequenceFlow_1giqfgg
SequenceFlow_0pip46b
SequenceFlow_1wvxggn
SequenceFlow_0t0k5gu
SequenceFlow_087jwgt
SequenceFlow_05k8n9f
SequenceFlow_1iflaxl
SequenceFlow_164cg0c
SequenceFlow_12ze84a
SequenceFlow_10hudx6
SequenceFlow_1iq3r1f
SequenceFlow_06bo34t
SequenceFlow_0t0k5gu
SequenceFlow_087jwgt

Blocked element:
ExclusiveGateway_0ngj689
Activated seq. flows:
SequenceFlow_087jwgt
Dead seq. flows
SequenceFlow_164cg0c

Fig. 4. Example operation of the *bpmn2imds* program.

Another example of a partial deadlock can concern the communication itself, when two Activities try to accept messages from each other. We would like to supplement the examples with more real-life ones, but text size constraints do not allow it.

VI. Conclusion and Future Work

The main goal of this article is to propose a translation of Business Process Collaboration and Process Diagrams into the IMDS, and verification of BPMN diagrams for deadlocks and termination. In order to achieve this, normalization of BPMN and translation of Process and Collaboration Diagrams into IMDS specifications was introduced. We identify partial (and total) deadlocks and check distributed termination. In IMDS, such checking is possible thanks to the preservation of information about component processes in the configuration space, and the development of temporal formulas independent of the structure of the analyzed system, and thus not requiring the designer to know temporal logic [6][7]. Compared to other tools, such as [9] and [13], our tool enables the visualization of diagrams and animation of their dynamics (not possible in [36]). Fig. 4 shows the example of deadlock animation. The designer is not limited to automatic verification of deadlocks/termination. The model can be automatically converted to the Uppaal tool [37], where arbitrary temporal questions can be asked, even with real-time constraints. However, reverse engineering of highlighting erroneous situations is impossible in such cases. Nevertheless, the simulation of a counterexample over the source BPMN diagram is preserved.

It should be noted that the complexity of every stage of work: conversion BPMN→IMDS, partial/total deadlock checking [38] and conversion IMDS→Uppaal are performed in linear time to size of a system (number of nodes and transitions). The example system of nearly 1 million configurations was checked in about an hour.F

What is rare in BPMN diagrams verification, we introduce *Boundary Links*, in order to formalize *Interrupting Boundary Intermediate Event* handling. The proposed semantics is characterized by *locality* – that is, the semantics of each of those elements depends only on other elements to which they are directly connected. During translation, the given BPMN diagrams are refactored into another semantically equivalent diagram to achieve consistency of activation and execution semantics.

One of the limitations of the proposed method is that it cannot handle diagrams that are unbounded, because it stems from the nature of finite state model checking. Another limitation is not considering BPMN elements with non-local semantics. Additionally, the need to use preallocated agents slows down the process of its analysis. Some elements are not covered by our translation, particularly *Event-based Gateways* and *Inclusive Gateways*. *Event-based Gateways* as they cannot be as easily translated into a model checking formalism. They need the creation of a set of agents of a purely technical nature to reset the gateway if an *Interrupting Boundary Intermediate Event* is bound to it. *Inclusive Gateways* have non-local semantics

We support the verification process with automatizing the verification process and giving run visualization and counterexample simulation properties (screenshots would take too much space, they can be found in [35]).

A possible improvement to the proposed translation method is to use preallocated agents for the entire diagram rather than for individual elements. This problem could be solved generally by introducing dynamic agent creation or agent reusability.

It may seem a controversial way of ordering messages sent and received by an *Activity* in the syntactic order of their appearance on the edge of the symbol. Other communication semantics can be envisioned. For example, first sending all outgoing messages and then waiting for all incoming. Alternatively, any order of sending and waiting for incoming messages. with the cost of exponential number of states in implementing server.

References

[1] W. M. P. van der Aalst, “The Application of Petri Nets to Workflow Management,” J. Circuits, Syst. Comput., vol. 08, no. 01, pp. 21–66, Feb. 1998. doi: 10.1142/S0218126698000043

[2] W. Reisig, Understanding Petri Nets. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. doi: 10.1007/978-3-642-33278-4

[3] Object Management Group, “Business Process Model and Notation

(BPMN) Version 2.0.2," 2013. http://www.omg.org/spec/BPMN/2.0.2

[4] C. Baier and J.-P. Katoen, Principles of Model Checking. Cambridge, MA: MIT Press, 2008. doi: 10.1007/978-3-030-12835-7_6

[5] F. Kossak et al., A Rigorous Semantics for BPMN 2.0 Process Diagrams. Cham: Springer International Publishing, 2014. doi: 10.1007/978-3-319-09931-6

[6] W. B. Daszczuk, "Specification and Verification in Integrated Model of Distributed Systems (IMDS)," MDPI Comput., vol. 7, no. 4, pp. 1–26, Dec. 2018. doi: 10.3390/computers7040065

[7] W. B. Daszczuk, "Using the Dedan Program," in Integrated Model of Distributed Systems, Cham, Switzerland: Springer Nature, 2020, pp. 87–97. doi: 10.1007/978-3-030-12835-7_6

[8] W. B. Daszczuk, "Fairness in Temporal Verification of Distributed Systems," in 13th International Conference on Dependability and Complex Systems DepCoS-RELCOMEX, 2-6 July 2018, Brunów, Poland, AISC vol.761, 2019, pp. 135–150. doi: 10.1007/978-3-319-91446-6_14

[9] R. M. Dijkman, M. Dumas, and C. Ouyang, "Semantics and analysis of business process models in BPMN," Inf. Softw. Technol., vol. 50, no. 12, pp. 1281–1294, Nov. 2008. doi: 10.1016/j.infsof.2008.02.006

[10] J. Billington et al., "The Petri Net Markup Language: Concepts, Technology, and Tools," in ICATPN 2003: Applications and Theory of Petri Nets, Eindhoven, The Netherlands, 23–27 June 2003, LNCS vol. 2679, 2003, pp. 483–505. doi: 10.1007/3-540-44919-1_31

[11] A. Rachdi, "Liveness and Reachability Analysis of BPMN Process Models," J. Comput. Inf. Technol., vol. 24, no. 2, pp. 195–207, Jun. 2016. doi: 10.20532/cit.2016.1002774

[12] K. Kluza, K. Jobczyk, P. Wiśniewski, and A. Ligęza, "Overview of Time Issues with Temporal Logics for Business Process Models," in 11 Federated Conference on Computer Science and Information Systems (FedCSIS), 11-14 Sept 2016, Gdansk, Poland, 2016, pp. 1115–1123. doi: 10.15439/2016F328

[13] C. Dechsupa, W. Vatanawood, and A. Thongtak, "Hierarchical Verification for the BPMN Design Model Using State Space Analysis," IEEE Access, vol. 7, pp. 16795–16815, 2019. doi: 10.1109/ACCESS.2019.2892958

[14] L. Li and F. Dai, "Transformation and Visualization of BPMN Models to Petri Nets," in International Conference of Green Buildings and Environmental Management (GBEM 2018), Qingdao, China, 23–25 Aug. 2018, IOP Conference Series: Earth and Environmental Science vol. 186, 2018, vol. 186, p. 012047. doi: 10.1088/1755-1315/186/5/012047

[15] R. Boussetoua, H. Bennoui, A. Chaoui, K. Khalfaoui, and E. Kerkouche, "An automatic approach to transform BPMN models to Pi-Calculus," in 2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA), Marrakech, Morocco, 17-20 Nov. 2015, 2015, pp. 1–8. doi: 10.1109/AICCSA.2015.7507176

[16] S. Yamasathien and W. Vatanawood, "An approach to construct formal model of business process model from BPMN workflow patterns," in Fourth International Conference on Digital Information and Communication Technology and its Applications (DICTAP), Bangkok, Thailand, 6-8 May 2014, 2014, pp. 211–215. doi: 10.1109/DICTAP.2014.6821684

[17] D. Prandi, P. Quaglia, and N. Zannone, "Formal Analysis of BPMN Via a Translation into COWS," in International Conference on Coordination Models and Languages COORDINATION 2008: Oslo, Norway, 4-6 June 2008, LNCS, vol. 5052, 2008, pp. 249–263. doi: 10.1007/978-3-540-68265-3_16

[18] M. Szpyrka, G. J. Nalepa, and K. Kluza, "From Process Models to Concurrent Systems in Alvis Language," Informatica, vol. 28, no. 3, pp. 525–545, Jan. 2017. doi: 10.15388/Informatica.2017.143

[19] J. Ye and W. Song, "Transformation of BPMN Diagrams to YAWL Nets," J. Softw., vol. 5, no. 4, pp. 396–404, Apr. 2010. doi: 10.4304/jsw.5.4.396-404

[20] V. Rafe and A. T. Rahmani, "A Graph Transformation-Based Approach to Formal Modeling and Verification of Workflows," in CSICC 2008: Advances in Computer Science and Engineering, Kish Island, Iran, 9-11 March 2008, 2008, pp. 291–298. doi: 10.1007/978-3-540-89985-3_36

[21] F. Corradini, F. Fornari, A. Polini, B. Re, F. Tiezzi, and A. Vandin, "BProVe: Tool support for business process verification," in 32nd IEEE/ACM International Conference on Automated Software Engineering (ASE), Urbana, IL, 30 Oct.-3 Nov. 2017, 2017, pp. 937–942. doi: 10.1109/ASE.2017.8115708

[22] F. Corradini, F. Fornari, A. Polini, B. Re, F. Tiezzi, and A. Vandin, "BProVe: A formal verification framework for business process models," in 32nd IEEE/ACM International Conference on Automated Software Engineering (ASE), Urbana, IL, 30 Oct.-3 Nov. 2017, 2017, pp. 217–228. doi: 10.1109/ASE.2017.8115635

[23] A. Krishna, P. Poizat, and G. Salaün, "VBPMN: Automated Verification of BPMN Processes," in 13th International Conference on integrated Formal Methods (iFM 2017), Turin, Italy, Sep 2017, 2017, pp. 1–8. http://convecs.inria.fr/doc/publications/Krishna-Poizat-Salaun-17.pdf. Accessed on 23.07.2023

[24] G. Salaün and P. Poizat, "VBPMN Samples." 2017. https://pascalpoizat.github.io/vbpmn-web/ Accessed on 23.07.2023

[25] N. Tantitharanukul, P. Sugunnasil, and W. Jumpamule, "Detecting deadlock and multiple termination in BPMN model using process automata," in 2010 ECTI International Confernce on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, Chiang Mai, Thailand, 19-21 May 2010, 2010, pp. 478–482. https://ieeexplore.ieee.org/document/5491443 Accessed on 23.07.2023

[26] W. B. Daszczuk, "Graphic modeling in Distributed Autonomous and Asynchronous Automata (DA3)," Softw. Syst. Model., vol. 20, no. 5, pp. 363–398, 2021. doi: 10.1007/s10270-021-00917-7

[27] K. Kluza and G. J. Nalepa, "Towards Rule-based Pattern Perspective for BPMN 2.0 Business Process Models," in 11 Federated Conference on Computer Science and Information Systems (FedCSIS), 11-14 Sept 2016, Gdansk, Poland, 2016, pp. 1359–1364. doi: 10.15439/2016F324

[28] K. Kluza and P. Wiśniewski, "Spreadsheet-Based Business Process Modeling," in 11 Federated Conference on Computer Science and Information Systems (FedCSIS), 11-14 Sept 2016, Gdansk, Poland, 2016, pp. 1355–1358. doi: 10.15439/2016F376

[29] W. M. P. van der Aalst, "Using Free-Choice Nets for Process Mining and Business Process Management," in 16 Federated Conference on Computer Science and Information Systems (FedCSIS), 2-5 Sept 2021, Sofia, Bulgaria, 2021, pp. 9–15. doi: 10.15439/2021F002

[30] S. J. Niepostyn and I. Bluemke, "The Function-Behaviour-Structure Diagram for Modelling Workflow of Information Systems," in CAiSE 2012: International Conference on Advanced Information Systems Engineering, Gdańsk, Poland, 25-26 June 2012, 2012, pp. 425–439. doi: 10.1007/978-3-642-31069-0_34

[31] S. Robak, B. Franczyk, and M. Robak, "Applying Linked Data concepts in BPM," in 6 Federated Conference on Computer Science and Information Systems (FedCSIS), 9-12 Sept 2012, Wrocław, Poland, 2012, pp. 1105–1110. url: https://ieeexplore.ieee.org/abstract/document/6354386 Accessed on 23.07.2023

[32] A. Suchenia, P. Wiśniewski, and A. Ligęza, "Overview of Verification Tools for Business Process Models," in 12 Federated Conference on Computer Science and Information Systems (FedCSIS), 3-6 Sept 2017, Prague, Czech Republic, 2017, pp. 295–302. doi: 10.15439/2017F308 doi: 10.15439/2017F308

[33] T. Lopes and S. Guerreiro, "Assessing business process models: a literature review on techniques for BPMN testing and formal verification," Bus. Process Manag. J., vol. 29, no. 8, pp. 133–162, Apr. 2023. doi: 10.1108/BPMJ-11-2022-0557

[34] Z. Manna and A. Pnueli, The Temporal Logic of Reactive and Concurrent Systems. New York, NY: Springer, 1992. doi: 10.1007/978-1-4612-0931-7

[35] J. Jałowiec, "Translation of Business Process Model and Notation into Integrated Model of Distributed Systems," B.Sc. thesis, Warsaw University of Technology, Institute of Computer Science, 2019. https://repo.pw.edu.pl/info/bachelor/WUT31de757656da422c87be61e7ede00630/?r=diploma&tab=&lang=pl Accessed on 23.07.2023

[36] F. Corradini, C. Muzi, B. Re, L. Rossi, and F. Tiezzi, "Global vs. Local Semantics of BPMN 2.0 OR-Join," in 44th International Conference on Current Trends in Theory and Practice of Computer Science, Krems, Austria, 29 Jan. - 2 Feb., 2018, LNCS, vol. 10706, 2018, pp. 321–336. doi: 10.1007/978-3-319-73117-9_23

[37] G. Behrmann, A. David, K. G. Larsen, P. Pettersson, and W. Yi, "Developing UPPAAL over 15 years," Softw. Pract. Exp., vol. 41, no. 2, pp. 133–142, Feb. 2011. doi: 10.1002/spe.1006

[38] W. B. Daszczuk, "Evaluation of temporal formulas based on 'Checking By Spheres,'" in Euromicro Symposium on Digital Systems Design, Warsaw, Poland, 4-6 Sept. 2001, 2001, pp. 158–164. doi: 10.1109/DSD.2001.952267

Communication Papers of the 18th Conference on Computer Science and Intelligence Systems pp. 75–80
DOI: 10.15439/2023F8192
ISSN 2300-5963 ACSIS, Vol. 37

Star-critical Ramsey numbers for hexagon

Tomasz Dzido
0000-0001-6712-8541
Department of Information Systems
Gdynia Maritime University
ul. Morska 81/87, 81-225 Gdynia, Poland
Email: t.dzido@wznj.umg.edu.pl

***Abstract*—Erdös and Faudree stated that it is an interesting problem to determine all the graph pairs which are Ramsey-full. For even cycles, they only showed that the pair (C_4, C_4) is Ramsey-full. It turns out that this statement cannot be applied to longer even cycles. Wu, Sun and Radziszowski obtained that the pair (C_n, C_4) for $n > 4$ is not Ramsey-full. In this article we will show that the pairs (C_n, C_6) for different values of n are also not Ramsey-full.**

We will also determine the values of some star-critical Ramsey numbers, in particular $r_*(C_6, C_6) = 6$ and $r_*(C_7, C_6) = 7$. In addition, we also show other values and bounds for star-critical Ramsey numbers for two cycles, one of which is an even cycle. These results are the beginning of the star-critical Ramsey number problem for even cycles of length 6 or more, and may help in obtaining further properties of this type.

I. Introduction

THE theorem, later called Ramsey's theorem, was proved by Ramsey and published shortly after his death in 1930. Informally written, this theorem proves that "complete disorder is impossible". In other words, any sufficiently large structure contains a substructure with the desired property. One of the popular ways looking at Ramsey's theory is in the context of graph theory, and more specifically edge coloring of graphs. To put it quite simply, we want to answer the following question: If we have a complete graph K_n on n vertices where every edge is arbitrarily colored either blue or red, what is the smallest value of n that guarantees the existence of either a subgraph G_1 which is red, or a subgraph G_2 which is blue? This smallest search n is called a 2-color Ramsey number $R(G_1, G_2)$. Initially, only the case when subgraphs G_1 and G_2 are complete subgraphs was considered. Therefore, Ramsey numbers for subgraphs other than complete and those defined analogously for more subgraphs and colors became popular very quickly. Currently, many classes of graphs are considered, such as paths, stars or cycles considered in this article.

From the informal definition of Ramsey numbers presented above, it follows that there is a critical graph, i.e. an edge coloring of a complete graph of order $n-1$, which does not contain a red copy of G_1 or a blue copy of G_2. Therefore, each 2-edge coloring of K_n contains either red G_1 or blue G_2, and there is a coloring of K_{n-1} without red G_1 or blue G_2. These facts lead us to an interesting question. For known Ramsey numbers, $R(G_1, G_2)$ = n, and a 2-coloring of the graph $K_{n-1} + v$, if we add colored edges individually from a new vertex v to vertices of K_{n-1}, then at what point must the graph have a red G_1 or a blue G_2? Alternatively, what is the largest star that can be removed from K_n so that the underlying graph is still forced to have either a red G_1 or a blue G_2? To study this, Hook and Isaak [6] introduced the definition of the star-critical Ramsey number $r_*(G_1, G_2)$.

Numerous other varieties of non-classical Ramsey numbers have been defined. For example: bipartite, planar, on-line, induced, local, diagonal, geometric, rainbow, linear and star-critical that are considered in this work. Many interesting applications of Ramsey theory arose in the field of mathematics and computer science, these include results in number theory, algebra, geometry, topology, set theory, logic, information theory and theoretical computer science. The theory is especially useful in building and analyzing communication nets of various types. Ramsey theory has been applied by Frederickson and Lynch to a problem in distributed computations [5], and by Snir [12] to search sorted tables in different parallel computation models. The reader will find more applications in Rosta's summary titled "Ramsey Theory Applications" [11].

II. Definitions and known results

In this paper we consider only finite and simple graphs. Let $G = (V(G), E(G))$. The deletion of edges of a copy of a subgraph H from G will be denoted as $G - H$ and the deletion of an edge e from G will be denoted as $G - e$. Let K_n denote a complete graph on n vertices and $K_{m,n}$ a complete bipartite graph on $m + n$ vertices. Denote by C_n a cycle of order n.

Definition 1. *The circumference $c(G)$ of a graph G is the length of its longest cycle.*

Definition 2. *The girth $g(G)$ of a graph G is the length of its shortest cycle.*

Definition 3. *A graph is called weakly pancyclic if it contains cycles of every length between the girth and the circumference.*

The following terminology, definitions and some descriptions are taken from [16].

Definition 4. *Given two graphs G_1 and G_2, we say that a graph G arrows the pair (G_1, G_2), denoted by $G \to (G_1, G_2)$, if in any red/blue coloring of the edges of G, there is a red copy of G_1 or a blue copy of G_2.*

For two given graphs G_1 and G_2, the most extensively investigated concept within Ramsey theory is *the graph Ramsey* *number* $R(G_1, G_2)$, which is the smallest integer r such that, for any graph G of order r, either G contains G_1 as a subgraph or G contains G_2 as a subgraph, where $\overline{G}$ is the complement of G. For simplicity, we now restate this definition of $R(G_1, G_2)$ in the language of arrowing.

Definition 5. $r = R(G_1, G_2) = min\{n|K_n \rightarrow (G_1, G_2)\}$.

Let r denote the Ramsey number $R(G_1, G_2)$ throughout the paper. A dynamic survey on Ramsey numbers can be found in [10].

Since $K_r \rightarrow (G_1, G_2)$, but $K_{r-1} \nrightarrow (G_1, G_2)$, a natural problem is to consider G such that $K_{r-1} \subseteq G \subseteq K_r$ and $G \rightarrow (G_1, G_2)$. To study this, Hook and Isaak [6] introduced the definition of *the star-critical Ramsey number* $r_*(G_1, G_2)$.

Definition 6 ([6]). $r_*(G_1, G_2) = min\{k|K_{r-1} \sqcup K_{1,k} \rightarrow (G_1, G_2)\}$.

The values of many star-critical Ramsey numbers have been determined. We will only recall the results for two cycles. In [16], Zhang, Broersma and Chen showed the following results.

Theorem 7 ([16]). $r_*(C_n, C_m) \geq \frac{m}{2} + 3$ *for even* $m \geq 4$, *odd* $n \geq \frac{3m}{2}$, *and for even* $m \geq 4$, *even* $n \geq m$, $n \geq 6$.

Theorem 8 ([16]). *For* m *odd,* $n \geq m \geq 3$ *and* $(m, n) \neq (3, 3)$, $r_*(C_n, C_m) = n + 1$.

Wu, Sun and Radziszowski [14] obtained that $r_*(C_n, C_4) = 5$ for $n \geq 4$. This result indicates that star-critical Ramsey number can be constant and much smaller than the corresponding classical Ramsey number. A fairly extensive and interesting summary of the all known results for star-critical Ramsey numbers can be found in the article [9]. One of the open problems appearing in various articles is the determination of the values of the numbers $r_*(C_n, C_m)$ for even m and $n \geq m \geq 6$. In this article, we focus on cycle C_6 and present a number of new values and bounds. In particular, we determine the following results: $r_*(C_6, C_6) = 6$ and $r_*(C_7, C_6) = 7$.

In the context of $G \rightarrow (G_1, G_2)$ and star-critical Ramsey numbers, some other definition was introduced.

Definition 9 ([16]). *A pair of graphs* (G_1, G_2) *is called Ramsey-full if* $K_r \rightarrow (G_1, G_2)$, *but* $K_r - e \nrightarrow (G_1, G_2)$.

Erdös and Faudree [3] stated that it is an interesting problem to determine all the graph pairs which are Ramsey-full. All the known graph pairs which are Ramsey-full are summarized in [16]. In the case of two cycles, we know that the pair (C_4, C_4) is Ramsey-full [3]. Wu, Sun and Radziszowski [14] obtained that the pair (C_n, C_4) for $n > 4$ is not Ramsey-full. The same is true for larger even cycles, as evidenced by the results obtained in this article for star-critical Ramsey numbers. In this article we will show that the pairs (C_n, C_6) for different values of n are also not Ramsey-full.

III. Preliminary results

The following notation and terminology comes from [2].

For positive integers a and b we define $r(a, b)$ as

$$r(a, b) = a - b\lfloor \frac{a}{b} \rfloor = a \bmod b.$$

For integers $n \geq k \geq 3$, we define $w(n, k)$ as

$$w(n, k) = \frac{1}{2}(n-1)k - \frac{1}{2}r(k - r - 1),$$

where $r = r(n - 1, k - 1)$.

Woodall's theorem [13] can then be written as follows.

Theorem 10 ([2]). *Let* G *be a graph on* n *vertices and* m *edges with* $m \geq n$ *and* $c(G) = k$. *Then*

$$m \leq w(n, k)$$

and this result is the best possible.

Lemma 11 ([1]). *Every nonbipartite graph* G *of order* n *with* $|E(G)| > (n-1)^2/4 + 1$ *is weakly pancyclic with* $g(G) = 3$.

For a graph G, define the Turán number $ex(n, G)$ to be the largest integer m such that there exists a graph on n vertices with m edges that does not contain G as a subgraph. In other words, if H has n vertices and more than $ex(n, G)$ edges, then H must contain G as a subgraph. A graph on n vertices is said to be *extremal with respect to* G if it does not contain a subgraph isomorphic to G and has exactly $ex(n, G)$ edges.

It is easy to see that for odd cycles, the Turán number $ex(n, C_{2t+1}) = \lfloor \frac{n^2}{4} \rfloor$ for $n > 4t - 1$, since no bipartite graph contains an odd cycle. For smaller values of n, we also know the value of $ex(n, C_{2t+1})$. Write n in the form $n = (s-1)(2t-1) + r$ where $s \geq 1$, $2 \leq r \leq 2t$ are integers. Then we have the following property.

Theorem 12 ([4]). *For any* $n \geq 1$ *and* $2t + 1 \geq 5$,

$$ex(n, C_{2t+1}) = (s-1)\binom{2t}{2} + \binom{r}{2}, \textit{ for } 2t+1 \leq n \leq 4t-1.$$

However, the problem of determining the Turán numbers for even cycles is still open. In the case of cycle C_6, we know all values of $ex(n, C_6)$ for $n < 22$ and all exstremal graphs with respect to C_6 for these numbers. These results are included in the paper [15].

We define the bipartite Turán number $ex(m, n, H)$ of a graph H to be the maximum number of edges in an H-free bipartite graph with parts of sizes m and n.

Theorem 13 ([8]). *Let* t *be an integer and* $G = (X, Y; E)$ *be a bipartite graph. Suppose* $|X| = n$, $|Y| = m$, *where* $n \geq m \geq t \geq \frac{m}{2} + 1$. *Then*

$$ex(m, n, C_{2t}) = (t-1)n + m - t + 1.$$

IV. Results

When $n \neq 4$ is even, $r = R(C_n, C_n) = \frac{3n}{2} - 1$. A new proof of this classic result was given by Károlyi and Rosta [7]. Erdös and Faudree [3] showed that (C_4, C_4) is Ramsey-full. It turns out that this is not the case for longer cycles of even length.

Theorem 14. *$K_r - e \rightarrow (C_6, C_6)$, where $r = R(C_6, C_6) = 8$.*

Proof. Let G be a graph $K_8 - e$ with $|V(G)| = 8$ and $|E(G)| = \binom{8}{2} - 1$. Let us consider an arbitrary coloring all the edges of the graph G. For any red/blue edge coloring of G, let G^R (G^B) be the graph whose vertex set is $V(G)$ and edge set consists of all red (blue) edges of G, respectively. Suppose to the contrary that neither G^R nor G^B contains a C_6. Since $|E(G)| = \binom{8}{2} - 1 = 27$, then without loss of generality we can assume that $|E(G^R)| \geq \lceil \frac{\binom{8}{2}-1}{2} \rceil = 14$. By Lemma 11, G^R is weakly pancyclic with girth 3. On the other hand, $w(8,4) = 13$ and by Theorem 10, G^R contains a C_5 as a subgraph.

Claim 14.1. *If G^B contains a monochromatic C_7, then G^R contains a monochromatic C_6.*

Proof. Let $C = x_1x_2...x_7x_1$ be a blue C_7 in G. Were some 2-chord of C blue, G would contain a blue C_6, a contradiction. Whence all 2-chords of C are red. In particular, the 2-chords of C form a red C_7 with at most one edge deleted (we consider $K_8 - e$). First, if we have a red C_7, then by pancyclicity of G^r we immediately have a red C_6. Assume we have a red $C_7 - e = x_1x_3x_5x_7x_2x_4x_6$. In order to avoid a red C_6, x_1x_4 and x_3x_6 are blue. Thus $x_1x_4x_5x_6x_3x_2x_1$ is a blue C_6, a contradiction. □

Let $C = v_1v_2v_3v_4v_5v_1$ be a cycle of length 5 in G^R. Let $C^* = \{w_1, w_2, w_3\}$ denote the set of vertices of G not in C. To avoid a red C_6, every vertex C^* is red incident to at most two vertices in C.

Claim 14.2. *If there are two red edges connecting $v \in C^*$ and C, say wv_i and wv_j, then we have $|i - j| = 2$ or 3.*

Proof. If the above condition does not hold, then it is easy to see that G contains a red C_6, a contradiction. □

Observe that the possible pairs of vertices in C that can be joined to vertex $w_i \in C^*$ are $P = \{v_1v_3, v_1v_4, v_2v_4, v_2v_5, v_3v_5\}$. Let $A_i = \{v \in C | vw_i \in G^R\}$ where $i \in \{1, 2, 3\}$.

Claim 14.3. *Given two integers $i, j \in \{1, 2, 3\}$, if $|A_i| = |A_j| = 2$ and $A_i \cap A_j = \emptyset$, then G^R contains a red C_6.*

Proof. Without los of generality, let us assume that the set $D = \{w_1v_1,\ w_1v_3,\ w_2v_2,\ w_2v_4\}$ is the set of red edges connecting the vertices w_1, w_2 with the cycle C. Then we immediately have a red $C_7 = w_1v_3v_2w_2v_4v_5v_1w_1$ and by pancyclicity of G^R, we have a red C_6. □

The rest of the proof contains all possible cases of setting the maximum number of red edges between C and C^*. Therefore, we want to consider all possible maximal structures of A_i and show that we always get a monochromatic cycle C_6. We start from the case where all vertices w_i are connected by red edges to the same vertices from cycle C (Claim 14.4). Later we consider the case where two of A_i have the same structure and the third one has a different structure (Claim 14.5). Finally, we show what other cases remain (Claim 14.6) and consider them (Claims 14.7 and 14.8). Keep in mind that we are dealing with $K_8 - e$. This means that it may happen that one of the red edges connecting C and C^* may not be there.

Claim 14.4. *For each $i \in \{1, 2, 3\}$, let $A_i \subseteq \{v_m, v_n\}$ with $v_mv_n \in P$. Then G^B contains a monochromatic C_6.*

Proof. Without los of generality, let us assume that $A_i \subseteq \{v_1, v_3\}$ for each $i \in \{1, 2, 3\}$. Consider now the blue bipartite subgraph F with parts $\{v_2, v_4, v_5\}$ and $\{w_1, w_2, w_3\}$. Then $|E(F)| \geq 8$ (we consider $K_8 - e$) $> ex(3, 3, C_6) = 7$, according to Theorem 13. □

Without loss of generality, consider the case where $A_i \subseteq \{v_1, v_3\}$ for $i \in \{1, 2\}$. Note that in this situation $A_3 \subseteq \{v_2, v_4\}$ or $A_3 \subseteq \{v_1, v_4\}$. The case $A_3 \subseteq \{v_2, v_5\}$ is the same as the first variant, and the case $A_3 \subseteq \{v_3, v_5\}$ is the same as the second.

Claim 14.5. *Let $A_i \subseteq \{v_1, v_3\}$ for $i \in \{1, 2\}$ and $A_3 \subseteq \{v_2, v_4\}$ or $A_3 \subseteq \{v_1, v_4\}$. Then G contains a monochromatic C_6.*

Proof. 1) $A_i \subseteq \{v_1, v_3\}$ for $i \in \{1, 2\}$ and $A_3 \subseteq \{v_2, v_4\}$. Let us consider all blue edges connecting C and C^* and all possible edges from A_i that form the set $R = \{w_1v_1, w_1v_3, w_2v_1, w_2v_3, w_3v_2, w_3v_4\}$. Consider the case where one of the edges in R does not exist in $G = K_8 - e$. Note that every edge in R if it is blue, then it is part of some blue C_6 in the bipartite graph $[C, C^*]$. Taking into account this fact and the thesis of Claim 3, without loss of generality, we can consider a situation where there is no edge w_3v_2. This means that the edges w_1v_1, w_1v_3, w_3v_4 are colored red, then in order to avoid red C_6, the edges v_2v_4, v_2v_5, w_1w_3 edges are colored blue. We then get the blue cycle $w_2v_2v_4w_1w_3v_5w_2$.

It remains to consider a situation in which there is no edge belonging to the rest (without edges from the set R) of the bipartite graph $[C, C^*]$. As a result of the analysis of the structure of the sets A_i, without loss of generality, we obtain the following 3 cases.

a) There is no edge w_1v_2 in graph G.
In order to avoid the following blue 6-cycles: $w_1v_3w_3v_5w_2v_4w_1$, $w_1v_1w_3v_5w_2v_4w_1$, the edges w_1v_3, w_1v_1 must be colored red. Then the edges v_2v_4 and v_2v_5 must be colored blue. Note that then the edge w_1w_2 must be blue. Suppose, on the contrary, that w_1w_2 is red. In this case, edges w_2v_1 and w_2v_3 must be blue and we get a blue 7-cycle: $w_3v_1(v_3)w_2v_2v_4w_1v_5w_3$, and by Claim 1 we get a red cycle C_6. To avoid the next two blue

cycles $w_3v_2v_4w_1w_2v_5w_3$ and $w_3v_4v_2w_2w_1v_5w_3$, edges w_3v_2 and w_3v_4 must be colored red. But then by Claim 3 we have a red cycle C_6.

b) There is no edge w_1v_5 in graph G.
In order to avoid the following blue 6-cycles: $w_2v_2w_1v_4w_3v_5w_2$, $w_3v_2w_1v_4w_2v_5w_3$, $w_1v_1w_3v_5w_2v_2w_1$ and $w_1v_3w_3v_5w_2v_2v_1$, the edges w_3v_4, w_3v_2, w_1v_1 and w_1v_3 are colored red. By Claim 3 we immediately obtain a red cycle of length 6.

c) There is no edge w_3v_1 in graph G.
In order to avoid the following blue 6-cycles: $w_2v_2w_1v_4w_3v_5w_2$, $w_3v_2w_1v_4w_2v_5w_3$ and $v_2w_2v_5w_3v_3w_1v_2$, the edges w_1v_3, w_3v_2 and w_3v_4 are red. Then the edges v_3v_5 and w_1w_3 must be blue and we have the blue 6-cycle: $v_3v_5w_2v_4w_1w_3v_3$.

2) $A_i \subseteq \{v_1, v_3\}$ for $i \in \{1, 2\}$ and $A_3 \subseteq \{v_1, v_4\}$.
Consider the blue bipartite subgraph F with parts $\{v_2, v_4, v_5\}$ and $\{w_1, w_2, w_3\}$. Then $|E(F)| \geq 8 > ex(3, 3, C_6) = 7$, according to Theorem 13. This means that only some edge of subgraph F may be missing in G. As a result of the analysis of the structure of the blue 6-cycles in F, without loss of generality, we obtain the following 2 cases.

a) There is no edge w_1v_4 in graph G
First, let's note that to avoid the blue cycle C_6, the edges w_1v_3, w_2v_3 and w_3v_4 must be colored red. Then consider the edges connecting vertex v_1 with vertices from C^*. At least two of them must be red. Suppose the edge w_1v_1 is colored red. Then the edges v_2v_4, v_2v_5, w_1w_2, w_1w_3 and w_2w_3 must be blue. We obtain the following blue 6-cycle: $w_1v_5v_2v_4w_2w_3w_1$. Finally, let us consider the case when the edge w_1v_1 is colored blue. This leads to the fact that both edges w_2v_1 and w_3v_1 are red, while the edges v_2v_4, v_2v_5, v_3v_5 and w_2w_3 are blue. Hence we have a blue cycle of length 6: $w_2v_4v_2v_5v_3w_3w_2$.

b) There is no edge w_3v_2 in graph G
The proof is identical to that of subcase (a). □

Let us now summarize the above cases and indicate which ones still remain to be considered.

Claim 14.6. *Without loss of generality, the maximum possible structures of A_i can be:*

1) $A_1 \subseteq \{v_1, v_3\}$, $A_2 \subseteq \{v_1, v_3\}$ *and* $A_3 \subseteq \{v_1, v_3\}$
2) $A_1 \subseteq \{v_1, v_3\}$, $A_2 \subseteq \{v_1, v_3\}$ *and* $A_3 \subseteq \{v_2, v_4\}$
3) $A_1 \subseteq \{v_1, v_3\}$, $A_2 \subseteq \{v_1, v_3\}$ *and* $A_3 \subseteq \{v_1, v_4\}$
4) $A_1 \subseteq \{v_1, v_3\}$, $A_2 \subseteq \{v_2, v_4\}$ *and* $A_3 \subseteq \{v_1, v_4\}$
5) $A_1 \subseteq \{v_1, v_3\}$, $A_2 \subseteq \{v_2, v_4\}$ *and* $A_3 \subseteq \{v_2, v_5\}$

Proof. Cases 1-3 have already been considered above. It remains to prove that cases 4-5 exhaust the situation when all sets A_i can be different. For this problem, let us consider situations where $A_1 \subseteq \{v_1, v_3\}$ and $A_2 \subseteq \{v_2, v_4\}$ or $A_2 \subseteq \{v_1, v_4\}$. Note that the case $A_2 \subseteq \{v_2, v_5\}$ is the same as the first variant, and the case $A_2 \subseteq \{v_3, v_5\}$ is the same as the second. For both variants let us analyze all possible maximal structures of A_3 and notice that all possible structures fall into cases 2-5.

1) $A_1 \subseteq \{v_1, v_3\}$ and $A_2 \subseteq \{v_2, v_4\}$
   a) $A_3 \subseteq \{v_1, v_3\}$ - Case 2
   b) $A_3 \subseteq \{v_1, v_4\}$ - Case 4
   c) $A_3 \subseteq \{v_2, v_4\}$ - Case 2
   d) $A_3 \subseteq \{v_2, v_5\}$ - Case 5
   e) $A_3 \subseteq \{v_3, v_5\}$ - Case 5
2) $A_1 \subseteq \{v_1, v_3\}$ and $A_2 \subseteq \{v_1, v_4\}$
   a) $A_3 \subseteq \{v_1, v_3\}$ - Case 3
   b) $A_3 \subseteq \{v_1, v_4\}$ - Case 3
   c) $A_3 \subseteq \{v_2, v_4\}$ - Case 4
   d) $A_3 \subseteq \{v_2, v_5\}$ - Case 5
   e) $A_3 \subseteq \{v_3, v_5\}$ - Case 4 □

Claim 14.7. *Let $A_1 \subseteq \{v_1, v_3\}$, $A_2 \subseteq \{v_2, v_4\}$ and $A_3 \subseteq \{v_1, v_4\}$. Then G contains a monochromatic C_6.*

Proof. Consider the blue bipartite subgraph with parts C and C^*. Note that this subgraph contains the following cycle of length 6: $v_2w_1v_5w_2v_3w_3v_2$. This means that only some edge of this cycle may be missing in graph G. We obtain the following 6 cases.

1) There is no edge w_3v_2 in graph G
In order to avoid the following blue 6-cycles: $w_1v_2w_2v_3w_3v_5w_1$, $w_1v_4w_2v_3w_3v_5w_1$, $w_1v_1w_2v_3 - w_3v_5w_1$, the edges w_2v_2, w_2v_4 and w_1v_1 must be colored red. Then the edges v_1v_3, v_3v_5 and w_1w_2 must be colored blue. If the edge w_1v_3 is blue, we obtain the following blue cycle: $w_3v_3v_1w_2w_1v_5w_3$. This means that the edge w_1v_3 is red. But then, based on Claim 3, we have a red cycle C_6.
2) There is no edge w_1v_5 in graph G
In a similar way as in the previous case, in order to avoid the following blue 6-cycles: $w_1v_2w_3v_3w_2v_1w_1$, $w_1v_2w_3v_5w_2v_3w_1$, $w_1v_4w_2v_3w_3v_2w_1$, the edges w_1v_1, w_1v_3 and w_2v_4 must be colored red. Then the edges v_2v_4, v_2v_5 and w_1w_2 must be colored blue. We have the blue 7-cycle: $w_1v_4v_2v_5w_3v_3w_2w_1$, and by Claim 1 we obtain a red cycle C_6.
3) There is no edge w_1v_2 in graph G
As before, to avoid the following blue 6-cycles: $w_1v_1w_2v_3w_3v_5w_1$, $w_2v_4w_1v_5w_3v_3w_2$, the edges w_1v_1 and w_2v_4 must be colored red. If edge w_1v_3 is colored red then similarly to case 2 we have the blue 7-cycle: $w_1v_4v_2v_5w_3v_3w_2w_1$, and by Claim 1 we obtain a red cycle C_6. If the edge w_2v_2 is red, then, as in case 1, we obtain the following blue cycle: $w_3v_3v_1w_2w_1v_5w_3$. If both edges w_1v_3 and w_2v_2 are blue, we have the blue 6-cycle: $w_2v_2w_3v_3w_1v_5w_2$. This means that these two edges are red. But then, based on Claim 3, we have a red 6-cycle.
4) There is no edge w_3v_3 in graph G

The proof of this case is analogous to the proof of previous cases. We start by noting that because of the cycles $w_1v_4w_2v_5w_3v_2w_1$, $w_1v_1w_2v_5w_3v_2w_1$, $w_1v_3w_2v_5w_3v_2w_1$, the edges w_2v_4, w_1v_1 and w_1v_3 are red. The rest of the reasoning is as in the above cases.

5) There is no edge w_2v_5 in graph G
The proof of this case is almost identical to the proof of case 3, so we leave it to the reader.
6) There is no edge w_2v_3 in graph G
Avoiding the corresponding blue cycles of length 6, we get that the edges w_1v_1 and w_2v_4 must be colored red. Then consider the possible colors of edges w_1v_3 and w_2v_2. If both of these edges are red, then by Lemma 3, we immediately have a red 6-cycle. If both are blue, we have the following blue cycle: $w_1v_3w_3v_5w_2v_2w_1$. The situation remains when both of these edges have different colors. It is similar to the situations considered in cases 1-3, so we omit it. □

Claim 14.8. *Let $A_1 \subseteq \{v_1, v_3\}$, $A_2 \subseteq \{v_2, v_4\}$ and $A_3 \subseteq \{v_2, v_5\}$. Then G contains a monochromatic C_6.*

Proof. First, note that the blue bipartite graph with partitions C and C^* contains the following two 6-cycles: $v_4w_1v_5w_2v_1w_3v_4$ and $v_3w_2v_5w_1v_4w_3v_3$. This means that only an edge occurring in both of these cycles can be missing in G. Due to this fact, we are given the following 4 cases to consider.

1) There is no edge w_1v_4 in graph G
To avoid the following blue cycles $w_1v_5w_2v_3w_3v_1w_1$ and $w_1v_3w_3v_1w_2v_5w_1$, the edges w_1v_1 and w_1v_3 must be colored red. Then the edges v_2v_5 and v_2v_4 are blue. We have the following blue 7-cycle: $w_1v_2v_4w_3v_1w_2v_5w_1$. Taking into account the thesis of Claim 1, we obtain a red cycle of length 6.
2) There is no edge w_3v_4 in graph G
Again considering the same cycles as at the beginning of the proof of case 1, we have that edges w_1v_1 and w_1v_3 are red. In order to avoid the blue 6-cycle $v_2w_3v_3w_2v_5w_1v_2$, the edge w_3v_2 will also be red. This forces edges v_2v_4, v_2v_5 and w_1w_3 to be colored blue. This all leads us to the blue cycle of length 6: $w_1w_3v_1w_2v_5v_2w_1$.
3) There is no edge w_1v_5 in graph G
As in the previous two cases, we get that the edges w_1v_1, w_1v_3 and w_2v_2 are red, while the edges v_2v_4, v_2v_5 and w_1w_2 are blue. Taking this into account, we immediately obtain the blue cycle of length 6: $w_2v_5v_2v_4w_3v_3w_2$.
4) There is no edge w_2v_5 in graph G
In this case we obtain the same red and blue edges as at the beginning of the proof of case 3. This time we have the following blue cycle of length 7: $w_1v_5v_2v_4w_3v_3w_2w_1$. From Claim 1 we also have a red cycle of length 6. □

We have already considered all possible cases and in each of them we have obtained a monochromatic C_6, so the proof of the theorem is complete. □

Corollary 15. *The pair of graphs (C_6, C_6) is not Ramsey-full.*

Corollary 16.
$$r_*(C_6, C_6) = 6.$$

Proof. We know that $R(C_6, C_6) = 8$ [7]. The lower bound follows easily from Theorem 7 in the special case $n = m = 6$. The upper bound follows directly from the conclusion of above Theorem 14. □

Theorem 17. *For even $m \geq 6$, odd $k \geq 1$ and $k \leq \frac{m}{2}$, $r_*(C_{m+k}, C_m) \geq m + 1$.*

Proof. Since $m+k$ is odd, then $r = R(C_{m+k}, C_m) = 2m-1$ [7]. Let P_1, P_2 be a partition of $V(K_{r-1})$ with $|P_1| = |P_2| = m - 1$. Assign colors to the edges of the K_{r-1} as follows: color the edges of P_1 and P_2 blue and all the other edges red. Let p_0 be an additional vertex, which is adjacent to P_1 with $m-1$ red edges and adjacent to P_2 with one blue edge. It is easy to check that there is neither a red C_{m+k} nor a blue C_m. □

By $K_{p1} * K_{p2} * ... * K_{pi}$ we denote a blockgraph, which consists of i complete blocks $K_{p1}, ..., K_{pi}$ such that exactly one vertex is contained in any of these complete subgraphs. Using this notation, for three graphs G, H and I, the graph $G * (H * I)$ consists of two graphs G and $H * I$, which have exactly one common vertex which is contained in G and H.

Theorem 18.
$$r_*(C_7, C_6) = 7.$$

Proof. From [7] we know that $R(C_7, C_6) = 11$. In oder to determine the value of $r_*(C_7, C_6)$, it is enough to prove that $F = K_{11} - K_{1,3} \rightarrow (C_7, C_6)$ because the lower bound follows from Theorem 17. Let us consider an arbitrary red/blue coloring all the edges of the graph F. For this coloring of F, let F^R (F^B) be the graph whose vertex set is $V(F)$ and edge set consists of all red (blue) edges of F, respectively. Suppose to the contrary that F^R does not contain a C_7 and F^B does not contain a C_6. The following results are taken from papers [15] and [4], respectively.

Claim 18.1 ([15]). *$ex(11, C_6) = 23$ and there are exactly three extremal graphs with respect to C_6 for this number, namely $K_5 * K_3 * K_5$, $K_5 * (K_3 * K_5)$ and $K_5 * (K_5 * K_3)$.*

Claim 18.2 ([4]). *$ex(11, C_7) = 30$ and there are exactly two extremal graphs with respect to C_7 for this number, namely $K_6 * K_6$ and $K_{5,6}$.*

Since $|E(F)| = 52$, then $|E(F^B)| \geq 23$ or $|E(F^R)| \geq 30$. Note that in the complement of each of the extremal graphs with respect to C_6 or C_7 for these numbers, we obtain a red C_7 or a blue C_6, respectively. We have a contradiction, which completes the proof of the theorem. □

Again, using the results of [15] and [4], we can easily obtain the following theorem.

Theorem 19. *$K_r - e \to (C_k, C_6)$, where $r = R(C_k, C_6)$ and $k \in \{17, 19\}$.*

Proof. Based on the works [7], [15] and [4] we have $R(C_{17}, C_6) = 19$, $R(C_{19}, C_6) = 21$, $ex(19, C_{17}) = 126$, $ex(19, C_6) = 44$, $ex(21, C_{19}) = 159$ and $ex(21, C_6) = 50$. Note that for $k \in \{17, 19\}$ the property $ex(r, C_k) + ex(r, C_6) = |E(K_r)| - 1$ holds. With a simple analysis the complements of critical graphs with respect to C_k described in [15], we have the proof. □

For graphs G_1, G_2 a coloring f is a $(G_1, G_2; n) - coloring$ if f is a red/blue edge coloring all the edges of K_n and f contains neither a red G_1 nor a blue G_2. A coloring $(G_1, G_2; n)$ is said to be *critical* if $n = R(G_1, G_2) - 1$.

Two more results can be obtained by simple computer methods.

Theorem 20.

$$r_*(C_8, C_6) = 6,$$

$$r_*(C_9, C_6) = 7.$$

Proof. On the website *https://users.cecs.anu.edu.au/∼bdm/data /graphs.html* we can find a database of all non-isomorphic graphs of order up to 11. They can be easily filtered out, yielding 24 critical colorings $(C_8, C_6; 9)$ for $r_1 = R(C_8, C_6) = 10$ and 26 critical colorings $(C_9, C_6; 10)$ for $r_2 = R(C_9, C_6) = 11$. Then we take all these critical colorings and consider all possible colorings of type $K_{r_1 - 1} \sqcup K_{1,k}$ and $K_{r_2 - 1} \sqcup K_{1,k}$ for increasing values of k, starting from $k = 1$. We are looking for the largest value of k that there is a coloring without forbidden subgraphs. □

Let's end the article with two interesting questions.

Question 1. *Let us note that $r_*(C_6, C_6) = r_*(C_8, C_6) = 6$ and $r_*(C_7, C_6) = r_*(C_9, C_6) = 7$. Will it turn out that $r_*(C_n, C_6) = 6$ or 7 depending on the parity of n?*

Question 2. *Do similar relationships hold for even cycles longer than 6?*

References

[1] S. Brandt, "A sufficient condition for all short cycles," *Discrete Appl. Math.,* vol. 79, 1997, pp. 63–66.

[2] L. Caccetta, K. Vijayan, "Maximal cycles in graphs," *Disc. Math.,* vol. 98, 1991, pp. 1–7.

[3] P. Erdös, R. J. Faudree, "Size Ramsey functions," *Sets, Graphs and Numbers,* 1991, pp. 219–238.

[4] Z. Füredi, D. S. Gunderson, "Extremal Numbers for Odd Cycles," *Combinatorics, Probability and Computing,* vol. 24(4), 2014, pp. 641–645.

[5] G. Frederickson, N. Lynch, "Electing a leader in a synchronous ring," *Journal of Association for Computing Machinery,* vol. 34, 1987, pp. 98–115.

[6] J. Hook, G. Isaak, "Star-critical Ramsey numbers," *Discrete Appl. Math.,* vol. 159, 2011, pp. 328–334.

[7] G. Károlyi and V. Rosta, "Generalized and geometric Ramsey numbers for cycles," *Theoret. Comput. Sci.,* vol. 263, 2001, pp. 87–98.

[8] B. Li, B. Ning, "Exact bipartite Turán numbers of large even cycles," *Journal of Graph Theory,* vol. 97(4), 2021, pp. 642–656.

[9] Y. Liu, Y. Chen, "Star-critical Ramsey Numbers of Wheels Versus Odd Cycles," *Acta Mathematicae Applicatae Sinica, English Series,* vol. 38(4), 2022, pp. 916–924.

[10] S. P. Radziszowski, "Small Ramsey numbers," *Electron. J. Combin.,* 2021, DS1.16.

[11] V. Rosta, "Ramsey Theory Applications," *Electronic Journal of Combinatorics*, 2004, Dynamic Survey 13.

[12] M. Snir, "On parallel searching," *SIAM Journal Computing,* vol. 14, 1985, pp. 688–708.

[13] D. R. Woodall, "Maximal circuits of graphs I," *Acta Math. Acad. Sci. Hungar.,* vol. 28, 1976, pp. 77–80.

[14] Y.L. Wu, Y.Q. Sun, S. Radziszowski, "Wheel and star-critical Ramsey numbers for quadrilateral," *Discrete Appl. Math.,* vol. 186, 2015, pp. 260–271.

[15] Y. Yuansheng, P. Rowlison, "On graphs without 6-cycles and related Ramsey numbers," *Utilitas Mathematica,* vol. 44, 1993, pp. 192–196.

[16] Y. Zhang, H. Broersma, Y. Chen, "On star-critical and upper size Ramsey numbers," *Discrete Appl. Math.,* vol. 202, 2016, pp. 174–180.

Communication Papers of the 18th Conference on Computer Science and Intelligence Systems pp. 81–89
DOI: 10.15439/2023F4965
ISSN 2300-5963 ACSIS, Vol. 37

Towards modelling and analysis of longitudinal social networks

Jens Dörpinghaus*†, Vera Weil*‡, Martin W. Sommer§
* Federal Institute for Vocational Education and Training (BIBB), Bonn, Germany
† University of Koblenz, Germany,
Email: jens.doerpinghaus@bibb.de, https://orcid.org/0000-0003-0245-7752
‡ Department of Mathematics and Computer Science, University of Cologne, Germany,
Email: weil@cs.uni-koeln.de
§ Argelander-Institut für Astronomie, Bonn, Germany

***Abstract*—There are currently several approaches to managing longitudinal data in graphs and social networks. All of them influence the output of algorithms that analyse the data. We present an overview of limitations, possible solutions and open questions for different data schemas for temporal data in social networks, based on a generic RDF-inspired approach that is equivalent to existing approaches. While restricting the algorithms to a specific time point or layer does not affect the results, applying these approaches to a network with multiple time points requires either adapted algorithms or reinterpretation. Thus, with a generic definition of temporal networks as one graph, we will answer the question of how we can analyse longitudinal social networks with centrality measures. In addition, we present two approaches to approximate the change in degree and betweenness centrality measures over time.**

I. Introduction

Social network analysis (SNA) is an important part of the social sciences and has been used in both theory and practice for several decades. It is important to understand social interactions and networks and how they affect society. In the last few years, there has been a growing interest in the use of social networks in the historical sciences. In religious studies, especially narrative studies and theology, social networks have recently received considerable attention.

Scholars have always seen SNA as part of the humanities, and in recent years there has been a rapid increase in the use of methods from the digital humanities, which includes the humanities and computer science.

Most works indicate that the described data and source problems are one of the greatest hurdles [1]. Although some preliminary work on how missing data influences a network has been carried out [2], there are still several open questions regarding the stability of social networks with respect to missing and additional data. The main question is: Can we still use the same algorithms, if we know that the data are incomplete? The need to work with temporal data makes an answer to this question even more urgent.

The three main research questions of this paper are thus:

- How can we model longitudinal social networks in one graph in the most generic way possible? (RQ1)
- How can we analyse longitudinal social networks with centrality measures? (RQ2)
- Can we approximate the change of centrality measures over time? (RQ3)

These questions cannot be answered without discussing the data schema for temporal data. Therefore, RQ1 is dedicated to the efficient storage of temporal data in a social network. While most entities such as actors and locations have a given lifetime, organisations or functions may have predecessors and successors. In other words: When an entity is detached, what relationships exist, and how can we manage their lifetimes? How can algorithms track and use these temporal data? RQ2 also contains several sub-questions: If a network G contains data for different time points $t_1, ..., t_n$, can we still apply analysis methods, e.g. centrality measures or community detection, that were originally developed for a particular time point? Or do we need to reinterpret the results or adapt the algorithms? Answering these questions is key to understanding the algorithmic challenges of temporal data in social network analysis.

This paper is divided into five sections. After this introduction, we give an overview of related work and the background of this research. We focus on historical network analysis (HNA) because it helps to highlight the challenges and is the natural habitat for longitudinal networks. Our methodological approach is described in the third section, where we discuss the modelling of longitudinal social networks, and their analysis. The fourth section is dedicated to the experimental results. Our conclusions are presented in the final section.

II. Related Work

Modeling temporal or longitudinal data in SNA is a well-known problem [3]. Temporal data lead to complex network structures and Lemercier stated in 2015: “There is no one best way for the analysis or even description of such multi-dimensional data” [4]. There are several modeling challenges, for example with synchronous and asynchronous events or relations, see [5]. Several methods have been proposed, for example, modeling with stream graphs [6], [7], Markov chains [8], [9], with network snapshots [10], or with a discrete set of time points that may contain snapshots. Most of these approaches are equivalent [11]. However, no single graphtheoretic definition currently covers all these approaches. This can be identified as the first gap in research.

Scientists are not only careful about how to model temporal networks, but also how to analyze them: “Traditional analyses of temporal networks have addressed mostly pairwise interactions, where links describe dyadic connections among individuals” [12]- Concetti et al. thus introduced “temporal hypergraphs” to address this challenge. Other researchers proposed visual analysis [13], pattern search [14], or probabilistic discrete temporal models [15]. Centrality measures, widely used in SNA, are also challenging in temporal networks. Some researchers have proposed definitions of temporal closeness, betweenness, and eigenvector centrality, see [16], [17], [18]. However, these definitions remain limited to the underlying graph topology, e.g. Sizemore et al. [18] work with a contact sequence where nodes remain static. In addition, the natural extension of centrality to groups and classes [19], [20] is usually omitted. Other authors propose MLI based on network embedding and machine learning (ML) [21]. In general, ML approaches are widely used in dynamic networks, not only in temporal networks, see [22]. However, these approaches – although providing significant insights on the networks – are not comparable to the results of centrality measures, which makes them difficult to reproduce. Thus, directly related to the first research gap – the lack of a generic definition of temporal networks – is the second gap: How can algorithms track and use this temporal data, and how does this affect the analysis of networks, e.g., with centrality measures?

These issues may be due to the fact that several aspects of knowledge graphs and the semantic web are not widely perceived in the SNA community. They have only recently been brought together [23]. Barats et al. conclude in 2020: FAIR data, a topic directly related to knowledge graphs, “remains a theoretical discussion rather than a shared practice in the field of humanities and social sciences.” [24] Thus, our work will try to address the research questions using knowledge graphs.

III. Method

We will use a definition of a knowledge graph that combines the approaches of [14], [23]:

Definition 1 (Temporal Social Network)**.** *A* Social Network *is a graph $G = (V, E, \mathcal{T})$ with vertices (nodes) $v \in V$, edges (relations) $e \in E$ and a time domain $\mathcal{T} = \{t_0, ..., t_k\}$ where $t_i \in \mathbb{R}$ and $t_i < i_j\ \forall i < j$. Every node and edge may exist at one or multiple intervals of timepoints*

$$[t_s, t_e] = \{x \in \mathcal{T} : t_s \leq x \leq t_e; t_s, t_e \in \mathcal{T}\}$$

denoted by $t(v)$ and $t(e)$. Thus, $t : V \cup E \to I \subseteq \mathbb{R}$. We denote the graph G at time t by

$$G^t = (V^t, E^t),\ where$$

$$V^t = \{v \in G\ |t \in t(v)\},\ E^t = \{v \in E\ |t \in t(e)\},$$

so that

$$\bigcup_{t\in\mathcal{T}} G^t = G.$$

Both edges and vertices are part of previously well-defined categories, $V \subseteq C_1 \cup C_2 \cup ... \cup C_n$ and $E \subseteq R_1 \cup R_2 \cup ... \cup R_m$.

Is is important to notice, that – in contrast to other definitions, e.g. [25] – both edges and nodes are temporal. Unless otherwise noted, we assume that G is an undirected graph. We will now present examples of the notation introduced above.

Each vertex $v \in V$ has a lifetime $t(v)$. In general, any edge connected to v may only exist for times $t \in t(v)$. But this rule is not strict. For example, we can define categories for successors T_s and predecessors T_p, so that these edges can indicate a predecessor of a certain position at any time. For these edges we set $t(e) = \emptyset$, they are ‘timeless’. In addition, v can be part of several categories, e.g. it can be an actor $v \in C_a$ and a politician $v \in C_p$. Thus, our approach can combine static and temporal information.

We will now prove that this definition is equivalent to stream graphs:

Theorem 1. *The temporal social network defined in 1 is equivalent to the concept of a stream graph introduced by Latapy, Magnien and Viard in [6] for discrete time instants T.*

Proof. “⇒” Let $G = (V', E, \mathcal{T})$ be a temporal social network as defined in Definition 1. We create a stream graph as follows: First, we can set the discrete time instants T to the time domain $\mathcal{T}$, thus $T = \mathcal{T}$. In addition, both node set are equal, thus $V = V'$.

The set of temporal nodes, $W \subseteq T \times V$, can be constructed as

$$W = \{(t(v), v) \forall v \in V\}.$$

The set of links $E \subseteq T \times V \otimes V$ can be constructed by

$$E = \{(t(e), e_1, e_2) \forall e = (e_1, e_2) \in E\}.$$

However, if $t(e) = \emptyset$, we define $t(e) = [\min_{t\in T}, \max_{t\in T}]$.

“⇐” Let S be a stream graph as defined by [6] with discrete time instants T, the node set V, a temporal node set $W \subseteq T \times V$ and a temporal edge set $E \subseteq T \times V \otimes V$.

We create a temporal social network $G = (V', E, \mathcal{T})$ as follows: Again, we the discrete time instants and nodes are equal and we set $\mathcal{T} = T$, $V' = V$. For each set of presence time $w = (t, (t, v)) \in W$ we define $t(v) = [\min t, \max t]$ and the same for edges $e = (t, (t, e_1, e_2)) \in E$. □

As we can see, the only difficulties are those edges and vertices that are ‘timeless’. However, extending their interval to $\mathcal{T}$ models their behaviour in the intended way. It is quite easy to see that both approaches are also equivalent to models using snapshots of time points [21]. For a detailed overview we refer to [11].

Thus, Definition 1 is well aligned with other approaches. However, it is also compatible with semantic web approaches and makes it easier to integrate analysis approaches. We will now move on to modelling longitudinal social networks with semantic web technologies.

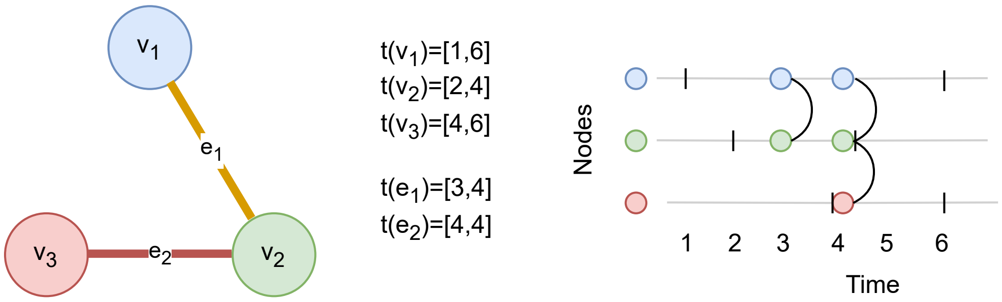


Fig. 1. Illustration of the graph in example 2 with a definition of lifetimes in the middle and a visualisation of the lifetime of edges and the sequence of edges over time (right).

A. Modelling longitudinal social networks

The initial definition of a social network in [23] corresponds to the definition of a knowledge graph. In particular, the categories for nodes $C_1, ..., C_n$ and edges $R_1, ..., R_m$ can be modelled using RDF classes. So we need to add time intervals to nodes and edges. To do this, Hobbs and Pan introduced the time ontology, see [26], [27]. Here they use a function *duration*: Intervals × TemporalUnits to express intervals. We can set $\mathit{duration}(v) = t(v)$ and $\mathit{duration}(e) = t(e)$ for any node $v \in V$ and edge $e \in E$.

Thus, any social network according to the knowledge graph definition in [23] can be easily transformed into a temporal social network, where time is modelled as a property of nodes and edges.

Example 2. *Consider the graph $G = (V, E, \mathcal{T})$ in figure 1 with $V = \{v_1, v_2, v_3\}$ and $E = \{e_1, e_2\}$ and a set of time intervals $t(v_1) = [1, 6]$, $t(v_2) = [2, 4]$, $t(v_3) = [4, 6]$, $t(e_1) = [3, 4]$ and $t(e_2) = [4, 4]$. They also provide a visualisation according to [18]: We visualise time by plotting a sequence of edges on a time scale. However, we extend the latter approach by adding information about the lifetime of nodes.*

In this case, each lifetime can be mapped according to the temporal duration.

It is worth noting that the general knowledge graph definition of a social network is open to adding a variety of additional data while maintaining the general graph structure. Thus, it is useful for modelling not only temporal social networks, but also any other temporal data, e.g. disease models.

B. Temporal graph structures

Similar to the approaches of [28], [18] we can study time-respecting structures in a graph. However, definition 1 of temporal social networks makes it easier to generalise graph structures as it keeps the generic definition of a graph.

A *path* p in a graph $H = (V, E)$ is a set of vertices $v_1, ..., v_t$, $t \in \mathbb{N}$, for example written as

$$p = [v_1, ..., v_t],$$

where $(v_i, v_{i+1}) \in E$ for $i \in \{1, \ldots, t-1\}$. However, to track the meaning of time in a temporal social network $G = (V, E, \mathcal{T})$, we define p^t, which is a path p that exists at time t. In turn, we define $t(p)$ as the interval of time in which the path p exists in G.

Unless otherwise noted, we use G for a temporal social network $G = (V, E, \mathcal{T})$ and H for any undirected graph.

We can add this generic notation for other structures as well. For example, we denote the time-respecting degree of a node v by $d^t(v)$. In this way, we get a series of *temporal degree centrality measures* (TDC) for a node $v \in V$ denoted by

$$dc^t(v) = \frac{d^t(v)}{n-1}.$$

In addition, we can analyse the *temporal degree distribution* which tells us about the network structure since we can distinguish between sparsely and densely connected networks.

Betweenness centrality (BC) was first introduced by [29] and considers other indirect links, see [30]. Given a node v, $bc(v)$ is defined as

$$bc(v) = \sum_{k \neq j, v \neq k, v \neq j} \frac{P_v(k,j)}{P(k,j)} \cdot \frac{2}{(n-1)(n-2)},$$

that is, we compute the number of all shortest paths $P_v(k, j)$ in a network for all starting and ending nodes $k, j \in V$ that pass through v. Let $P(k, j)$ denote the total number of shortest paths between k and j. Then the importance of v is given by the ratio of the two values of P_v and P. Again, for any time $t \in \mathcal{T}$ we may set $P_v^t(k, j)$ and $P^t(k, j)$ accordingly, such that

$$bc^t(v) = \sum_{k \neq j, v \neq k, v \neq j} \frac{P_v^t(k,j)}{P^t(k,j)} \cdot \frac{2}{(n-1)(n-2)}$$

defines the series of *temporal betweenness centrality* (TBC). This definition is similar to that of [18], who, however, used the concept of fastest paths.

We will proceed similarly with *closeness centrality* (CC). Given a node $i \in V$ we can compute the average distance between the first and other nodes $j \in V$ with $\sum_{j \neq i} d(i, j)$, where $d(i, j)$ denotes the length of a shortest path between i and j. Then, according to [31], we can compute closeness-centrality as follows:

$$cl(v) = \frac{n-1}{\sum_{u \in V} d(u, v)}.$$

Again, with a definition of $d^t(i,j)$ for the length of a shortest path at time $t \in \mathcal{T}$ at hand, we can define *temporal closeness centrality* (TCC) as

$$cl^t(v) = \frac{n-1}{\sum_{u \in V} d^t(u,v)}.$$

However, these definitions are currently not more than a containment of well-known centrality measures on time snapshots of the temporal social network. They allow an interpretation of these snapshots, comparable to static social networks, and they provide a series of centrality measures that can be interpreted as the progression of these measures over time.

For social networks, perceiving the world with as few snapshots as possible is most feasible. Other approaches, e.g. defining paths closely so that they could split up from one time to another, if the interval is so small that an event lasts less, is often necessary to model traffic [16]. Social interaction, on the other hand, does usually change on the basis of longer lasting events. This is a crucial observation, because computing temporal paths with increasing timestamps from one node to the next is computationally hard, see [25].

While interdisciplinary approaches are available, applications from humanities and in particular historical networks research lead to a different perspective on data. For example, a closed organization may still have an influence on parts of the network or may be referred to later. However, with our novel approach, we will evaluate the behavior of analysis methods like centrality measures and community detection and discuss limitations and challenges for further research.

C. Random graphs

For further analysis, we rely on random graphs. The *degree distribution* provides us with information about the network structure since we can distinguish between sparsely and densely connected networks. In social network analysis (SNA), the following two graphs are widely considered:

Definition 2 (Scale-Free Network)**.** *A network is scale-free if the fraction of nodes with degree s follows a power law $s^{-\alpha}$, where $\alpha > 1$.*

Definition 3 (Small World Network [32])**.** *Let $G = (V, E)$ be a connected graph with n nodes and average node degree k. Then G is a small-world network if $k \ll n$ and $k \gg 1$.*

[33] introduced a widely used graph model with three random parameters $\alpha + \beta + \gamma = 1$. These values define probabilities and thus define attachment rules to add new vertices between either existing or new nodes. This model allows loops and multiple edges, where a loop denotes one edge where the endvertices are identical, and multiple edges denote a finite number of edges that share the same endvertices. Thus, we convert the random graphs to undirected graphs. For testing putposes, we scale the number of nodes n and use $\alpha = 0.41$, $\beta = 0.54$, and $\gamma = 0.05$. This random graph model is generic and feasible for computer simulations for measuring and evaluation purposes, see [34], [35].

One of the core concepts important in social network research is the graph diameter $D(G)$. From the 1960s on, it was widely discusses whether the average path length of social networks is near six, see [36]. However, there is an ongoing discussion on this issue, see for example [37], [38]. However, it was shown that in a scale-free network the diameter is always lower than $\log(n)$, and if the fixed number m of earlier vertices is larger than 1, in general the diameter is lower than $\frac{\log(n)}{\log\log(n)}$, see [39]. Here, n describes not only the number of steps to create the random graph, but also the number of nodes in the graph. While the connection between a particular graph and a particular diameter is quite complex, see [40], we can rely on these bounds. For small-world random graphs we find [41] the almost surely upper bound $D(G) \leq \frac{72}{p} \log^2 n$ while [42] proved the diameter is usually bound by $\log(n)$.

The diameter of a scale-free graph is in general quite low, while in small-world graphs it is bound by $\log(n)$. However, we may expect random graphs to have a different behavior from real-world social networks. Thus, for some of the following proofs we will assume that $D(G) \leq 5$.

D. Analysing networks

For a detailed overview of centrality measures, we can consider the series of a particular measure, e.g. a generic c (centrality, e.g. which could refer to closeness or betweenness centrality), which is basically a vector in $\mathbb{R}^{|\mathcal{T}|}$:

$$\widetilde{c}(v) = \left(c^{t_1}(v), ..., c^{t_{|\mathcal{T}|}}(v)\right).$$

Note that $c^{t_i}(v) = \emptyset$ if $t_i \notin t(v)$. We define

$$|\widetilde{c}(v)| = \sum_{i \in \{1,...,|\mathcal{T}|,\, c^{t_i}(v) \neq \emptyset\}} l(t_{i-1}, t_i)|,$$

where $l(t_{i-1}, t_i)$ defines the length of time elapsed between two times t_{i-1} and t_i. For $x \in V$ or $x \in E$ we set

$$l(x) = \sum_{i \in \{1,...,|\mathcal{T}|,\, c^{t_i}(x) \neq \emptyset\}} l(t_{i-1}, t_i)$$

as the lifespan of x. However, if all times are equally distributed, this simplifies to

$$|\widetilde{c}(v)| = |\mathcal{T}| - |\{x \in \widetilde{c}(v) \mid x = \emptyset\}|.$$

This allows us to calculate the *average temporal centrality* of a node v over its lifetime as

$$\overline{c}(v) = \frac{1}{|\widetilde{c}(v)|} \sum_{t \in \mathcal{T}, c^t(v) \neq \emptyset} c^t(v).$$

However, for a proper analysis of centrality measures over time, we should also consider the *temporal centrality* of a node v:

$$A\left(c\left(v\right)\right) = \sum_{i \in \{1,...,|\mathcal{T}|,\, c^{t_i}(v) \neq \emptyset\}} c^{t_i}(v) l(t_{i-1}, t_i).$$

Again, for evenly distributed time, $l(t_{i-1}, t_i) = 1$ and

$$A\left(c\left(v\right)\right) = \sum_{t \in \mathcal{T}, c^t(v) \neq \emptyset} c^t(v).$$

We can also normalise this measure by life span as *normalised temporal centrality* to compare the centrality measure over time within a life span:

$$A'(c(v)) = \frac{1}{l(v)} \sum_{i \in \{1,\ldots,|\mathcal{T}|,\, c^{t_i}(v) \neq \emptyset\}} c^{t_i}(v) l(t_{i-1}, t_i).$$

In section IV we will discuss several working examples and offer an interpretation of these values in light of the current state of research on degree and betweenness centrality.

First, we consider how a centrality measure evolves over time. Since we need to plot this for n nodes, we consider a heatmap visualisation that bins the number of nodes in a given interval. Next, we can plot the average centrality measure at a particular time and the average centrality over all time points, as we show in Figure 2.

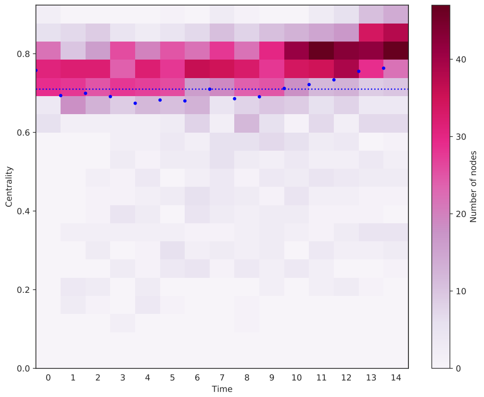


Fig. 2. Illustration of the distribution of a centrality measure over time, grouped into 20 bins between 0 and 1, as a heatmap. The blue horizontal line refers to the overall average centrality, while the blue dots refer to the average degree at a given time. This illustrates the degree centrality for $\mathcal{G}^s(100, 15, 0.1)$.

This figure gives us a good overview of how many nodes are below and above the average centrality at a given time, and whether the network at a given time is special for the scenario. To analyse and compare a particular node with this overall picture, we can plot $\widetilde{c}(v)$ and $\overline{c}(v)$, as we show in Figure 3

Some [17] considered calculating and plotting $\widetilde{c}(v)$, [16] added probabilities. Thus, in addition to the classical approach (e.g. [18]), $\widetilde{c}(v)$ and $\overline{c}(v)$ allow the study of static centrality measures at a time $t \in \mathcal{T}$, comparing the individual centrality value of a particular node with the average node degree and the distribution of node degrees. In addition, by plotting the series of centrality over time, we can compare the temporal centrality measures within a given interval or across the entire timeline. While some general measures, such as average temporal centrality, have been studied previously [3], their interpretation remains vague. If networks change significantly over time, this value is not comparable.

E. Approximating the changes over time

Let $\mathfrak{G}^p = \{G_1, \ldots G_\iota\}$ be a series of graphs and $p \in \mathbb{R}$ with $0 \leq p \leq 1$ and

$$|\,(V(G_i) \cup V(G_i + 1)) \setminus (V(G_i) \cap V(G_i + 1))\,| \leq p|V(G_i)|,$$

$$|\,(E(G_i) \cup E(G_i + 1)) \setminus (E(G_i) \cap E(G_i + 1))\,| \leq p|E(G_i)|,$$

for $i \in \{1, \ldots, \iota - 1\}$. Thus, $\mathfrak{G}^p$ is a series of graphs with a fixed set of differences and changes from one to the other.

Now we can approximate the changes over time, or the error in the centrality measures that can occur due to these changes. Unless otherwise noted, we will consider $\mathfrak{G}^p = \{G_1, \ldots G_\iota\}$.

Theorem 3. *Let $i \in \{1, \ldots \iota - 1\}$ so that $v \in V(G_i)$ and $v \in V(G_{i+1})$. Then it holds that*

$$dc^{i+1}(v) \geq \frac{d^i(v) - p|V(G_i)|}{|V(G_i)| - 1 + p|V(G_i)|},$$

$$dc^{i+1}(v) \leq \frac{d^i(v) + p|V(G_i)|}{|V(G_i)| - 1 - p|V(G_i)|}.$$

Proof. We know that

$$dc^i(v) = \frac{d^i(v)}{|V(G_i)| - 1}.$$

However, due to the definition of $\mathfrak{G}^p$, we know that at most $p|V(G_i)|$ new connections from v to other nodes can exist in G_{i+1} or may be lost. Thus, in G_{i+1} it holds that

$$d^i(v) - p|V(G_i)| \leq d^{i+1}(v) \leq d^i(v) + p|V(G_i)|.$$

In addition, we know that for G_{i+1}

$$|V(G_i)| - p|V(G_i)| \leq |V(G_{i+1})| \leq |V(G_i)| + p|V(G_i)|$$

holds. Hence the claim follows. □

For betwenness centrality, we define

$$\sigma = |N(G_i)|p,$$

$$\epsilon = \min\{D(G_i)^2, 2|V(G_i)|p\},$$

where $D(G)$ is the diameter of G. We will prove two lemmata to obtain a bound for $bc^{i+1}(v)$ for $v \in V$.

Lemma 4. *Let $i \in \{1, \ldots \iota - 1\}$ so that $v \in V(G_i)$ and $v \in V(G_{i+1})$. Then,*

$$P_v(k, j)\frac{1}{\sigma} \leq P_v^{i+1}(k, j)$$

holds.

Proof. All shortest paths between $k, j \in V(G_i)$ have the same length $\mathfrak{l} \leq D(G_i)$. For $D(G_i) \leq 5$, $\mathfrak{l} = \delta(v)$ holds: If $D(G_i) = 3$, k, j must both be adjacent to v. If $D(G_i) = 4$, we say k must be adjacent to v and $\nu \in \mathbb{N}^+$ nodes exist which are adjacent to j and v, which implies $\delta(v)$ paths. If $D(G_i) = 5$, $\nu \in \mathbb{N}^+$ nodes exist which are adjacent to j and v, and $\mu \in \mathbb{N}^+$ nodes exist which are adjacent to k and v, which implies $\delta(v)$ paths.

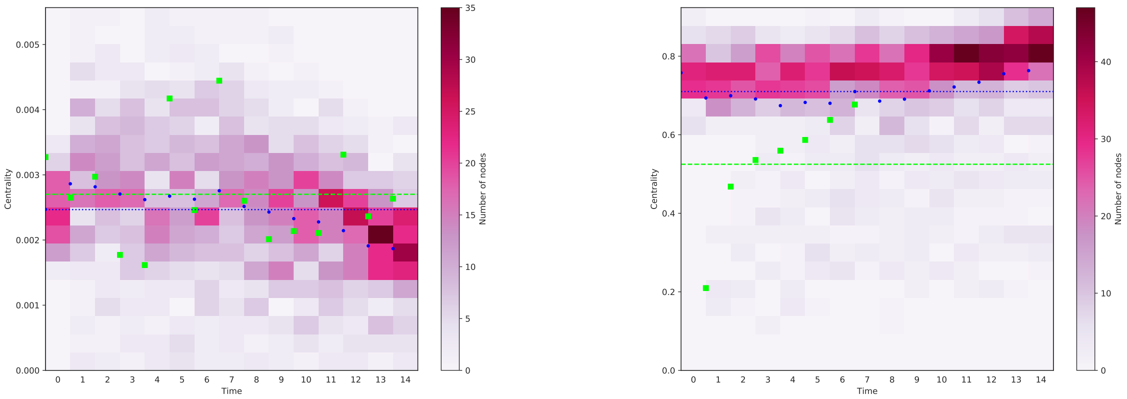


Fig. 3. Illustration of the distribution of a centrality measure over time, grouped into 20 bins between 0 and 1, as a heatmap. The blue horizontal line refers to the overall average centrality, while the blue dots refer to the average degree at one point in time. Both figures show $\widetilde{c}(v)$ and $\overline{c}(v)$ (green dots and horizontal line, respectively) for two different nodes. Left: This node exists over all 15 time points and usually shows that the betweenness centrality varies a lot. Right: This node exists from time 1 to 7 and has an increasing degree centrality value. The network is based on $\mathcal{G}^s(100, 15, 0.1)$.

Let us assume that a maximum of edges and nodes will be removed from G_i towards G_{i+1} and a maximum number of them is adjacent to v. Then, at most $|N(G_i)|p$ edges and neighbours of v can be removed in G_{i+1} which, in turn, removes one possible shortest path between k, j over v. Thus, $P_v^{i+1}(k,j)$ cannot have more than $P_v(k,j)\frac{1}{|N(G_i)|p} = P_v(k,j)\frac{1}{\sigma}$ of the initial paths through v. □

Lemma 5. *Let $i \in \{1, ...\iota - 1\}$ so that $v \in V(G_i)$ and $v \in V(G_{i+1})$. Then,*

$$P_v^{i+1}(k,j) \leq \begin{cases} P_v(k,j)\epsilon & P_v(k,j) > 0 \\ D(G_i)^2\epsilon & P_v(k,j) = 0 \end{cases}$$

holds.

Proof. As shown in the proof of Lemma 4, all shortest paths between $k, j \in V(G_i)$ have the same length $\mathfrak{l} \leq D(G_i)$ and for $D(G_i) \leq 5$, $\mathfrak{l} = \delta(v)$ holds.

Let us assume that a maximum number of edges and nodes will be added to G_{i+1}. This is at maximum $2|V(G_i)|p$. However, no more than $D(G_i) \cdot D(G_i) = D(G_i)^2$ paths between k and j may exist if $P_v(k,j) > 0$. Thus,

$$P_v^{i+1}(k,j) \leq P_v(k,j)\min\{D(G_i), |V(G_i)|p\} = P_v(k,j)\epsilon$$

holds.

If $P_v(k,j) = 0$, we know that no more than $D(G_i)^2$ paths may exist at all. Thus,

$$P_v^{i+1}(k,j) \leq P_v(k,j)\min\{D(G_i), |V(G_i)|p\} = P_v(k,j)\epsilon$$

holds. □

Theorem 6. *Let $i \in \{1, ...\iota - 1\}$ so that $v \in V(G_i)$ and $v \in V(G_{i+1})$. Then,*

$$bc^i(v)\epsilon \leq bc^{i+1}(v) \leq bc^i(v)\frac{1}{\sigma}$$

holds.

Proof. Recall that

$$bc(v) = \sum_{k \neq j, v \neq k, v \neq j} \frac{P_v(k,j)}{P(k,j)} \cdot \frac{2}{(n-1)(n-2)}.$$

We have already shown the following two inequalities with lemmata 4 and 5:

$$P_v(k,j)\frac{1}{\sigma} \leq P_v^{i+1}(k,j) \leq P_v(k,j)\epsilon$$

Thus, with Lemma 4 we can show:

$$\begin{aligned} bc^{i+1}(v) &= \sum_{k \neq j, v \neq k, v \neq j} \frac{P_v^{i+1}(k,j)}{P^{i+1}(k,j)} \cdot \frac{2}{(n-1)(n-2)} \\ &\leq \sum_{k \neq j, v \neq k, v \neq j} \frac{P_v(k,j)\frac{1}{\sigma}}{P^{i+1}(k,j)} \cdot \frac{2}{(n-1)(n-2)} \\ &= \frac{1}{\sigma} \sum_{k \neq j, v \neq k, v \neq j} \frac{P_v^{i+1}(k,j)}{P^{i+1}(k,j)} \cdot \frac{2}{(n-1)(n-2)} \\ &= bc^i(v)\frac{1}{\sigma} \end{aligned}$$

Similarly, with Lemma 5 we can show:

$$\begin{aligned} bc^{i+1}(v) &= \sum_{k \neq j, v \neq k, v \neq j} \frac{P_v^{i+1}(k,j)}{P^{i+1}(k,j)} \cdot \frac{2}{(n-1)(n-2)} \\ &\geq \sum_{k \neq j, v \neq k, v \neq j} \frac{P_v(k,j)\epsilon}{P^{i+1}(k,j)} \cdot \frac{2}{(n-1)(n-2)} \\ &= \epsilon \sum_{k \neq j, v \neq k, v \neq j} \frac{P_v^{i+1}(k,j)}{P^{i+1}(k,j)} \cdot \frac{2}{(n-1)(n-2)} \\ &= bc^i(v)\epsilon \end{aligned}$$

□

We will now continue with an experimental setting showing the results of these bounds.

IV. Experimental results

We evaluate the degree centrality and betweenness centrality on random graphs, see Section III-C. First, we consider scale-free networks with n nodes, see [31]. With this, we create a series of random Graphs $\mathcal{G}^s(n, i, p)$ which creates one initial scale-free network with n nodes and $i-1$ more random graphs with a probability of $p/2$ for each node and edge to be deleted and $p/2$ for each node and edge to be deleted and a new one created. In addition, we consider scale-free networks and create a series of random Graphs $\mathcal{G}^w(n, i, p)$ which starts with one initial small world network with n nodes and $i-1$ more random graphs with a probability of $p/2$ for each node and edge to be deleted and $p/2$ for each node and edge to be deleted and a new one created.

We will evaluate both degree centrality and betweenness centrality on the following four random graph series:

- $\mathcal{G}^s(50, 15, p)$, $p \in \{0.15, 0.05\}$
- $\mathcal{G}^w(50, 15, p)$, $p \in \{0.15, 0.05\}$
- $\mathcal{G}^s(150, 15, p)$, $p \in \{0.15, 0.05\}$
- $\mathcal{G}^w(150, 15, p)$, $p \in \{0.15, 0.05\}$

For evaluation purposes, we select several nodes and display the distribution of the centrality measure over time and the approximation of the changes over time.

A. Degree centrality

We present an evaluation of sample nodes in Figures 4-7. We show the upper and lower bounds for degree centrality introduced in Theorem 3.

First, small world random graphs are shown in Figures 4 and 5. Here the bounds on degree centrality are quite tight, but get worse for larger p. We can make a similar observations for scale-free networks in Figures 6 and 7.

Thus, the bounds introduced in Theorem 3 work well for small p and provide overall good results for estimating the evolution of degree centrality for the next time step when p is known.

B. Betwenness centrality

We will now consider the upper and lower bounds for betwenness centrality introduced in Theorem 6. We present a selected evaluation of small-world graphs in Figures 8 and 9. For the small-world graph in Figure 8 (left), the node has a lifetime between 0 and 5, but a centrality measure of zero. This figure shows how the upper bound approximates $D(G_i)^2$. For larger p in Figure 9, the node for $n = 50$ has a lifetime between 3 and 10. Compared to Figure 8, a higher value of p results in even less sharp bounds. For the larger small-world network, neither the upper nor lower bounds are sharp, although the upper bound tends to be even worse.

For the scale-free random networks in Figures 10 and 11 the situation is similar. However, the heatmap shows that most nodes have small betweenness centrality values, while there are many outliers. In Figure 10 we see that again the lower bound is sharper than the upper bound. However, for $n = 50$ we see an example that shows that in some cases the upper bound is suitable to estimate the change over time. Comparing these results to the results shown in Figure 11 again highlights that these bounds get less precise for larger p.

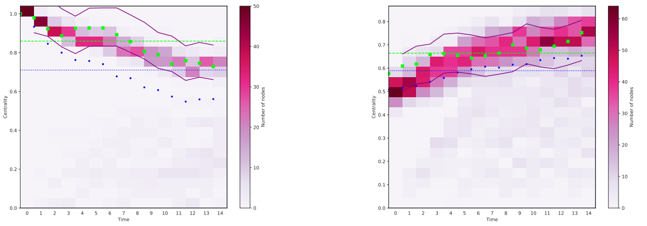


Fig. 4. $\mathcal{G}^w(n, 15, p)$, $p = 0.05$ with $n = 50$ (left) and $n = 150$ (right).

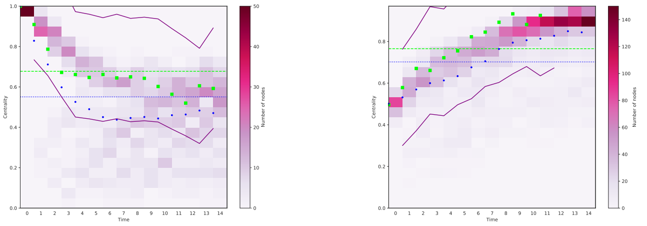


Fig. 5. $\mathcal{G}^w(n, 15, p)$, $p = 0.15$ with $n = 50$ (left) and $n = 150$ (right).

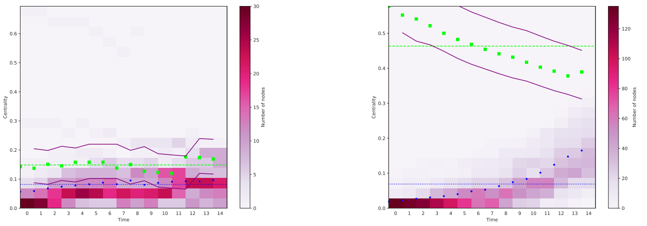


Fig. 6. $\mathcal{G}^s(n, 15, p)$, $p = 0.05$ with $n = 50$ (left) and $n = 150$ (right).

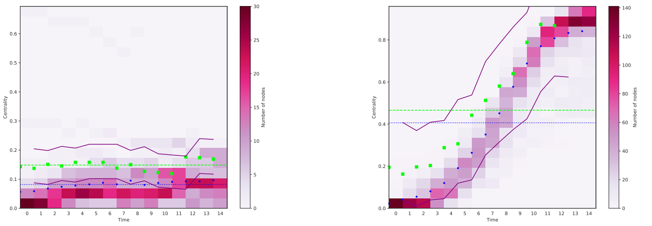


Fig. 7. $\mathcal{G}^s(n, 15, p)$, $p = 0.15$ with $n = 50$ (left) and $n = 150$ (right).

Thus, the upper and lower bounds for betwenness centrality introduced in theorem 6 are not suitable for estimating change over time in any situation. However, the lower bound tends to be sharper than the upper bound, where the behaviour is sometimes unpredictable.

V. Discussion and Outlook

Several approaches exist to manage longitudinal data in networks. All of them bias the output of algorithms analyzing the data. We presented an overview on limitations, possible solutions and open questions to different data schemas for temporal data in social networks based on a generic RDF-inspired approach. In this way, we answered out first research question: How can we model longitudinal social networks in one graph as generic as possible? While not the primary focus of our work, this approach allows the integration of further data from the semantic web making results and approaches directly available for social networks.

We also discussed a second research question. How can we analyse longitudinal social networks with centrality measures? While limiting algorithms to one particular time point or layer does not influence the output, applying them to a network comprising multiple time points does either need adjusted algorithms or reinterpretation. We presented a solution for adjusted approaches and could show that if a network G contains data for different time points $t_1, ..., t_n$, we can still apply centrality measures that were originally developed for a particular time point. We proposed the concepts of average temporal centrality and temporal centrality as core concepts to analyse the temporal development of centrality over the given time, together with a novel representation to compare an individual node against the whole graph. Indeed, we need to reinterpret these results and adapt algorithms. However, while our approach works for all centrality measures, we only considered betwenness centrality and degree centrality and more research needs to consider other centrality measures and methods like community detection. Answering these questions is key to understanding the algorithmic challenges of temporal data in social network analysis.

Our third question was concerned whether we can approximate the change of centrality measures over time. We presented upper and lower bounds for betwenness and degree centrality. However, these bounds need a prior knowledge of the change ratio p between different time points. With an increasing value of p, these bounds become less sharp. More research needs to focus on different types of bounds, in particular for other centrality measures. In addition, a detailed analysis of graph substructures having an influence on the temporal behavior of centrality measures might be fruitful, in particular if p is unknown.

However, rewriting algorithms to analyse longitudinal social networks and the re-interpretation of existing measures and algorithms demands discussion between different scientific domains. Therefore, our paper is also a plea for more interdisciplinary exchange, in particular between mathematics, computer science, social sciences and the humanities.

References

[1] J. Leidwanger, C. Knappett, P. Arnaud, P. Arthur, E. Blake, C. Broodbank, T. Brughmans, T. Evans, S. Graham, E. S. Greene *et al.*, "A manifesto for the study of ancient mediterranean maritime networks," *Antiquity*, vol. 88, no. 342, 2014.

[2] S. de Valeriola, "Can historians trust centrality?" *Journal of Historical Network Research*, vol. 6, no. 1, 2021.

[3] P. Holme and J. Saramäki, "A map of approaches to temporal networks," *Temporal network theory*, pp. 1–24, 2019.

[4] C. Lemercier, "Taking time seriously. how do we deal with change in historical networks?" in *Knoten und Kanten III. Soziale Netzwerkanalyse in Geschichts- und Politikforschung*. Transcript, 2015, pp. 183–211.

[5] S. Lehmann, "Fundamental structures in temporal communication networks," *Temporal Network Theory*, pp. 25–48, 2019.

[6] M. Latapy, T. Viard, and C. Magnien, "Stream graphs and link streams for the modeling of interactions over time," *Social Network Analysis and Mining*, vol. 8, pp. 1–29, 2018.

[7] M. Latapy, C. Magnien, and T. Viard, "Weighted, bipartite, or directed stream graphs for the modeling of temporal networks," *Temporal Network Theory*, pp. 49–64, 2019.

[8] T. P. Peixoto and M. Rosvall, "Modelling temporal networks with markov chains, community structures and change points," *Temporal network theory*, pp. 65–81, 2019.

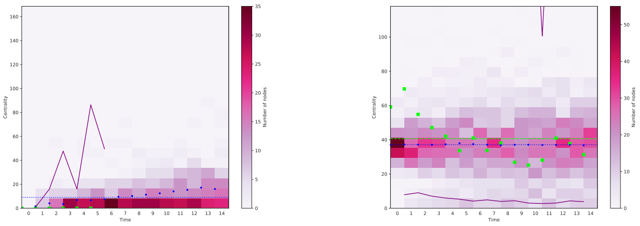


Fig. 8. $\mathcal{G}^w(n, 15, p)$, $p = 0.05$ with $n = 50$ (left) and $n = 150$ (right).

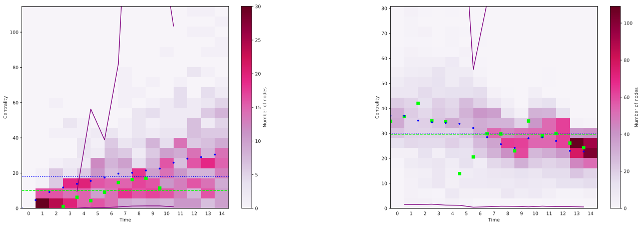


Fig. 9. $\mathcal{G}^w(n, 15, p)$, $p = 0.15$ with $n = 50$ (left) and $n = 150$ (right).

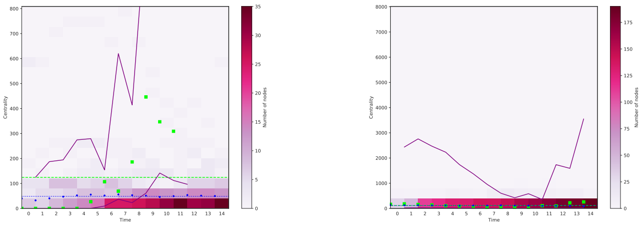


Fig. 10. $\mathcal{G}^s(n, 15, p)$, $p = 0.05$ with $n = 50$ (left) and $n = 150$ (right).

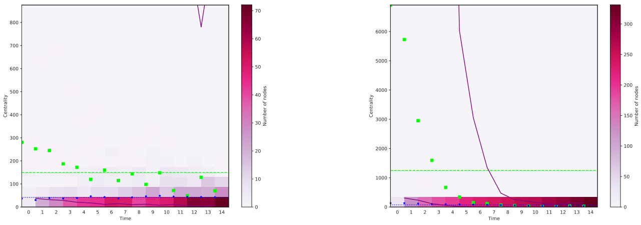


Fig. 11. $\mathcal{G}^s(n, 15, p)$, $p = 0.15$ with $n = 50$ (left) and $n = 150$ (right).

[9] I. Scholtes, N. Wider, R. Pfitzner, A. Garas, C. J. Tessone, and F. Schweitzer, "Causality-driven slow-down and speed-up of diffusion in non-markovian temporal networks," *Nature communications*, vol. 5, no. 1, p. 5024, 2014.

[10] K. S. Xu and A. O. Hero, "Dynamic stochastic blockmodels: Statistical models for time-evolving networks," in *Social Computing, Behavioral-Cultural Modeling and Prediction: 6th International Conference, SBP 2013, Washington, DC, USA, April 2-5, 2013. Proceedings 6*. Springer, 2013, pp. 201–210.

[11] P. Holme and J. Saramäki, "Temporal networks," *Physics reports*, vol. 519, no. 3, pp. 97–125, 2012.

[12] G. Cencetti, F. Battiston, B. Lepri, and M. Karsai, "Temporal properties of higher-order interactions in social networks," *Scientific reports*, vol. 11, no. 1, p. 7028, 2021.

[13] J. S. Yi, N. Elmqvist, and S. Lee, "Timematrix: Analyzing temporal social networks using interactive matrix-based visualizations," *Intl. Journal of Human–Computer Interaction*, vol. 26, no. 11-12, pp. 1031–1051, 2010.

[14] M. Franzke, T. Emrich, A. Züfle, and M. Renz, "Pattern search in temporal social networks," in *Proceedings of the 21st International Conference on Extending Database Technology*, 2018.

[15] S. Hanneke, W. Fu, and E. P. Xing, "Discrete temporal models of social networks," *Electronic Journal of Statistics*, vol. 4, pp. 585–605, 2010.

[16] R. K. Pan and J. Saramäki, "Path lengths, correlations, and centrality in temporal networks," *Physical Review E*, vol. 84, no. 1, p. 016105, 2011.

[17] D. Taylor, S. A. Myers, A. Clauset, M. A. Porter, and P. J. Mucha, "Eigenvector-based centrality measures for temporal networks," *Multiscale Modeling & Simulation*, vol. 15, no. 1, pp. 537–574, 2017.

[18] A. E. Sizemore and D. S. Bassett, "Dynamic graph metrics: Tutorial, toolbox, and tale," *NeuroImage*, vol. 180, pp. 417–427, 2018.

[19] M. G. Everett and S. P. Borgatti, "The centrality of groups and classes," *The Journal of mathematical sociology*, vol. 23, no. 3, pp. 181–201, 1999.

[20] S. Rasti and C. Vogiatzis, "Novel centrality metrics for studying essentiality in protein-protein interaction networks based on group structures," *Networks*, vol. 80, no. 1, pp. 3–50, 2022.

[21] E.-Y. Yu, Y. Fu, X. Chen, M. Xie, and D.-B. Chen, "Identifying critical nodes in temporal networks by network embedding," *Scientific reports*, vol. 10, no. 1, p. 12494, 2020.

[22] P. Cinaglia and M. Cannataro, "Network alignment and motif discovery in dynamic networks," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 11, no. 1, p. 38, 2022.

[23] J. Dörpinghaus, S. Klante, M. Christian, C. Meigen, and C. Düing, "From social networks to knowledge graphs: A plea for interdisciplinary approaches," *Social Sciences & Humanities Open*, vol. 6, no. 1, p. 100337, 2022.

[24] C. Barats, V. Schafer, and A. Fickers, "Fading away... the challenge of sustainability in digital studies." *DHQ: Digital Humanities Quarterly*, vol. 14, no. 3, 2020.

[25] D. Santoro and I. Sarpe, "Onbra: Rigorous estimation of the temporal betweenness centrality in temporal networks," in *Proceedings of the ACM Web Conference 2022*, 2022, pp. 1579–1588.

[26] J. R. Hobbs and F. Pan, "An ontology of time for the semantic web," *ACM Transactions on Asian Language Information Processing (TALIP)*, vol. 3, no. 1, pp. 66–85, 2004.

[27] M. Grüninger, "Verification of the owl-time ontology," in *The Semantic Web–ISWC 2011: 10th International Semantic Web Conference, Bonn, Germany, October 23-27, 2011, Proceedings, Part I 10*. Springer, 2011, pp. 225–240.

[28] V. Nicosia, J. Tang, C. Mascolo, M. Musolesi, G. Russo, and V. Latora, "Graph metrics for temporal networks," *Temporal networks*, pp. 15–40, 2013.

[29] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, pp. 35–41, 1977.

[30] T. Schweizer, *Muster sozialer Ordnung: Netzwerkanalyse als Fundament der Sozialethnologie*. Berlin: D. Reimer, 1996.

[31] M. O. Jackson, *Social and Economic Networks*. Princeton: University Press, 2010.

[32] D. J. Watts, "Networks, dynamics, and the small-world phenomenon," *American Journal of sociology*, vol. 105, no. 2, pp. 493–527, 1999.

[33] B. Bollobás, C. Borgs, J. T. Chayes, and O. Riordan, "Directed scale-free graphs." in *SODA*, vol. 3, 2003, pp. 132–139.

[34] B. Bollobás and O. M. Riordan, "Mathematical results on scale-free random graphs," *Handbook of graphs and networks: from the genome to the internet*, pp. 1–34, 2003.

[35] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, "Multilayer networks," *Journal of complex networks*, vol. 2, no. 3, pp. 203–271, 2014.

[36] S. Milgram, "The small world problem," *Psychology today*, vol. 2, no. 1, pp. 60–67, 1967.

[37] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world'networks," *nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[38] J. S. Kleinfeld, "The small world problem," *Society*, vol. 39, no. 2, pp. 61–66, 2002.

[39] O. Riordan *et al.*, "The diameter of a scale-free random graph," *Combinatorica*, vol. 24, no. 1, pp. 5–34, 2004.

[40] F. Ma, X. Wang, and P. Wang, "Scale-free networks with invariable diameter and density feature: Counterexamples," *Physical Review E*, vol. 101, no. 2, p. 022315, 2020.

[41] L. Gu, H. L. Huang, and X. D. Zhang, "The clustering coefficient and the diameter of small-world networks," *Acta Mathematica Sinica, English Series*, vol. 29, no. 1, pp. 199–208, 2013.

[42] C. Martel and V. Nguyen, "Analyzing kleinberg's (and other) small-world models," in *Proceedings of the twenty-third annual ACM symposium on Principles of distributed computing*, 2004, pp. 179–188.

DOI: 10.15439/2023F5828
ISSN 2300-5963 ACSIS, Vol. 37

Comparative Study: Defuzzification Functions and Their Effect on the Performance of the OFNbee Optimization Algorithm

Dawid Ewald
0000-0002-0608-0801
Faculty of Computer Science
Kazimierz Wielki Univesrity in Bydgoszcz,
ul. J.K. Chodkiewicza 30,
85-064 Bydgoszcz, Poland
Email: dawidewald@ukw.edu.pl

Wojciech Dobrosielski,
Jacek Czerniak, Hubert Zarzycki
00000-0001-7756-4259
0000-0001-9848-7192
0000-0003-2314-6152
Faculty of Computer Science
Kazimierz Wielki Univesrity in Bydgoszcz,
ul. J.K. Chodkiewicza 30,
85-064 Bydgoszcz, Poland
Email: {wdobrosielski, jczerniak}@ukw.edu.pl
Tadeusz Kosciuszko military Academy of land Forces
in Wroclaw Poland Email: hzarzycki@yahoo.com

***Abstract*—This article explores the pivotal role of defuzzification functions in the operation of the OFNBee algorithm, which employs ordered fuzzy number arithmetic to harness the inherent dynamics within a hive. Defuzzification functions serve the purpose of representing the OFN (Ordered Fuzzy Number) as a real number, while fuzzification functions convert real numbers into OFN representations. By focusing on the defuzzification function, this study investigates its impact on the performance of the OFNBee algorithm. The research demonstrates that tailoring dedicated fuzzification functions for specific optimization problems can yield substantial improvements in algorithmic performance. It is important to note that the overall performance of the algorithm relies on both the fuzzification and defuzzification functions. Consequently, this article provides valuable insights into the effects of the defuzzification function on algorithmic outcomes.**

I. Introduction

THIS article is part of a series of research focused on the issue of fuzzy logic, and more specifically, fuzzy numbers. There are several leading models of fuzzy numbers, which will be discussed later in the text. This article focuses on the OFN model, and more specifically on the specific arithmetic resulting from the use of referral in OFN. In earlier papers [1], issues related to ordered fuzzy numbers and their application were thoroughly discussed. The aim of the current research is to create optimization algorithms based on ordered fuzzy numbers. The algorithm has been presented in many publications [2], [3], [4]. However, this article discusses the impact of the defuzzification function on the operation of this algorithm. An OFNbee algorithm has already been developed that uses the flocking behavior of bees and ordered fuzzy numbers. Due to the fact that specific defuzzyfication and fuzzyfication methods sensitive to direction are needed to move from the set of real numbers to the set of ordered fuzzy numbers and vice versa, their impact on the performance of the algorithm should be investigated. It is intuitively known that the appropriate selection of methods can improve the performance of optimization algorithms based on OFN. Until then [1], [5], the impact of the fuzzyfication function on the results of OFNBee has been examined, while the following text will present the most popular defuzzyfication methods and their impact on the results of the algorithm. Algorithms based on the behavior of insects or animals are an attempt to use natural mechanisms for optimization [6]. However, the difficulty of accurately describing the behavior of living organisms leads to oversimplification. The simplification process, although effective, limits the possibilities of such methods. Therefore, it seems reasonable to use the mechanisms of fuzzy logic, which more naturally describes phenomena occurring in the real world. Given how bee optimization algorithms are evolving and the source of their inspiration, a combination of more natural fuzzy arithmetic with bee algorithms should be considered.

II. Selected elements of fuzzy set theory

To discuss the subject of fuzzy numbers, it is necessary to start with the concept of a fuzzy set. The fuzzy set was introduced in 1965 by Lotfi Zadeh [7], [8], [9], who defined that the fuzzy set A in space X is the set of pairs described:

$$A = \{(x, \mu_A(x) : x \in X)\} \tag{1}$$

gdzie: μ_A is a membership function, assigning to each element $x \in X$ (the assumed space of considerations X) its degree of membership in the set A, where: $\mu_A : X \to [0,1]$, therefore $\mu_A(x) \in [0,1]$.

As in the case of classical sets, which are described by a characteristic function, in the case of fuzzy sets, the membership function is used to describe them [10] [11]. Such a function assigns to each element of the set a real number in the interval [0,1], thus determining the degree of membership of a given element to the set. A fuzzy set must be uniquely described by its membership function, and this is the most important feature of fuzzy sets. In the theoretical model, there are no contraindications for this function to assume any shape. In the literature, however, you can find several basic functions with specific shapes [12]: h

- triangular – described by formula 2, where $a \leq b \leq c$ i and shown in 1:

$$\mu_A(x,a,b,c) = \begin{cases} 0 & , where\ x \leq a \\ \frac{x-a}{b-a} & , where\ a < x \leq b \\ \frac{c-x}{c-b} & , where\ b < x \leq c \\ 0 & , where\ x > c \end{cases} \tag{2}$$

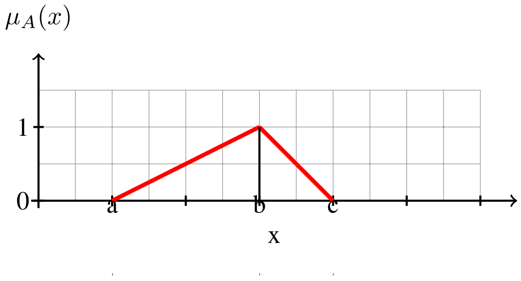


Fig. 1. Triangular

- trapezoidal – described by formula 3, where $a \leq b \leq c \leq d$ i and shown in Figure 2:

$$\mu_A(x,a,b,c,d) = \begin{cases} 0 & , where\ x \leq a \\ \frac{x-a}{b-a} & , gdzie\ a < x \leq b \\ 1 & , where\ b < x \leq c \\ \frac{d-x}{d-c} & , gdzie\ c < x \leq d \\ 0 & , where\ x > d \end{cases} \tag{3}$$

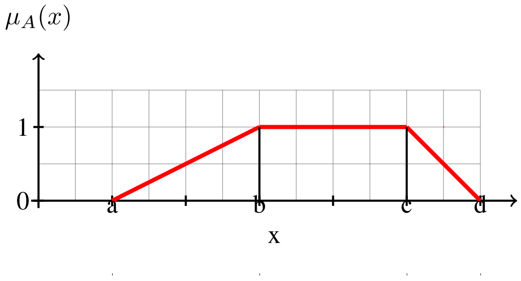


Fig. 2. Trapezoidal

- singleton – described by the formula 4, where x_0 is a parameter defining the location of the singleton, shown in 3:

$$\mu_A(x,x_0) = \begin{cases} 1 & , where\ x = x_0 \\ 0 & , where\ x \neq x \leq x_0 \end{cases} \tag{4}$$

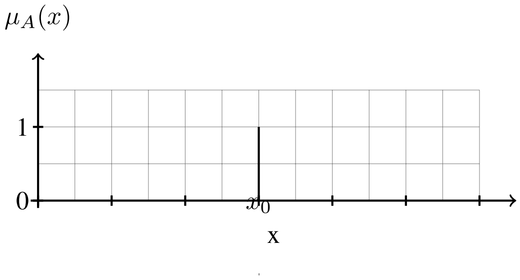


Fig. 3. Singleton

III. Fuzzy number - LR model

In 1978, Dubois and Prade proposed the LR (Left-Right) model, which was supposed to simplify quite complicated arithmetic operations performed on classical fuzzy numbers. We define a fuzzy number A of the LR type as follows: A of the LR type as follows:

$$\mu_A(x) = \begin{cases} L(\frac{m-x}{\alpha}) & , Where\ x \leq m \\ R(\frac{x-m}{\beta}) & , Where\ x \geq m \end{cases} \tag{5}$$

Where:

- m is a real number defined as an average number – $\mu_A(m) = 1$,
- $\alpha > 0$ – ufixed left-hand real number,
- $\beta > 0$ – fixed right-hand real number.

L and R are basis functions that satisfy the following conditions:

- $L(-x) = L(x), R(-x) = R(x)$,
- $L(0) = 1, R(0) = 1$,
- L and R are non-increasing functions on the interval $[0, +\infty)$.

IV. Ordered Fuzzy Numbers

The ordered fuzzy numbers OFN were proposed in 2002 by Witold Kosiński, Piotr Prokopowicz and Dominik Ślęzak. They focused on eliminating the shortcomings of classical fuzzy number algebra. The disadvantages in question are primarily the fact that by performing several operations on given L-R numbers, you can get numbers that are too fuzzy, which may make them less useful. This entails a large computational complexity and the inability to backward chaining. The creators of OFN also set themselves the goal of developing arithmetic, thanks to which it would be possible to perform operations on both triangular and trapezoidal numbers. They proposed a model defined as follows:

Definition 1 [13], [14], [15]
An ordered fuzzy number A is an ordered pair of functions

$$A = (f_A, g_A) \tag{6}$$

where:
$f_A, g_A : [0,1] \longrightarrow R$ are continuous functions. Accordingly, we call the functions $f_A$the increasing part (up), and the function g_A the decreasing part (down) of the ordered fuzzy number. The continuity of both parts shows that their images

are limited by intervals. They are given the names UP and DOWN respectively. To mark the limits (being real numbers) of these intervals, the following notations have been adopted: $UP = (l_A, 1_A^-)$ and $DOWN = (1_A^+, p_A)$.

V. Optimization method using ordered fuzzy numbers

A bee as a single individual shows almost no features that could be considered worth using in optimization. However, the collective work of these insects is very interesting and shows how nature deals with optimization. The communication mechanisms present in the hive allow bees to optimally manage resources and survive. Algorithms based on bee herd behavior use a space-searching mechanism to find nectar. Such algorithms treat the bee as a single solution or treat the found source as a solution. There are also more complicated behavioral adaptations. However, in all cases, the question of how the bees communicate information to each other is overlooked, or this step is reduced to some simple selection condition. The OFNBee method, using the arithmetic of directed fuzzy numbers, allowed to reflect the mechanisms of information transfer that are actually present among bees.

A new OFNBee optimization method was created by combining ordered fuzzy numbers with bee optimization. The use of OFN notation in bee optimization seems to be a natural way to describe the behavioral mechanisms observed in the hive and quoted above. These mechanisms are presented in the new method using dedicated fuzzification operators. The input data is information carried by a single bee (Figure 4), i.e.:

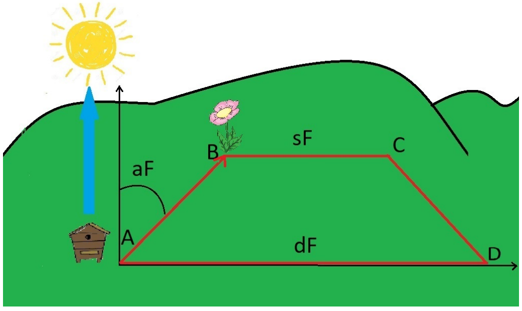


Fig. 4. Graphical interpretation of OFN in OFNBee

- the direction in which the food is located,
- navigation angle acc. sun,
- flight length,
- abundance of food source.

The determination of the directed fuzzy number A is done as follows. First, support(A) is established, which is the base of the trapezoid. Then the rising edge of f(x) is plotted at $90^o - aF$. The other base of the trapezoid is set aside from the point where the intersections of f(x) with y = 1. Finally, it remains to connect the slope g(x) to the two ends of the base of the trapezoid, as shown in Figure 4.

VI. Research methodology

In the case of checking the influence of defuzzyfication functions on the operation of the algorithm, their influence should be presented on the set of testing functions. These functions were selected due to their frequent occurrence in the literature during the verification of optimization algorithms. The OFNbee algorithm includes several configuration options. For correct operation, when running the algorithm, the fuzzyfication and defuzzyfication operator and the size of the population must be specified. In another paper [1] fuzzyfication operators were examined. Each of the two defuzzyfication functions was run with the same fuzzyfication operator and population size. Each run of the algorithm was repeated 30 times for all combinations of fuzzyfication and defuzzyfication operators, and the results are shown below.

VII. Selected OFN defuzzification operators

The new optimization method requires defuzzification operators to work. These functionals allow to represent OFN as a real number. A very important feature of OFN is number referral (direction) – this added value distinguishes ordered numbers from other solutions. That's why it's so important to use direction-sensitive defuzzification operators. The first work on defuzzification operators was undertaken by W. Kosiński [16] [17] [18] [19]. The work was continued by W. Dobrosielski in numerous articles. The functional proposed by W. Dobrosielski is described later in the document. In the operation of the new method, an important feature of the defuzzification functionals is the sensitivity to the direction of OFN, the focus was only on those methods that meet the above condition. These methods are well described in the literature and often used in control models.

A. Golden Ratio defuzzification operator

This section presents the golden ratio defuzzification functional developed by W. Dobrosielski [20]. The method from the fuzzy number A allows to determine the real value from it in accordance with the formula 7. A graphical interpretation of GR is shown in Figure 5[20]:

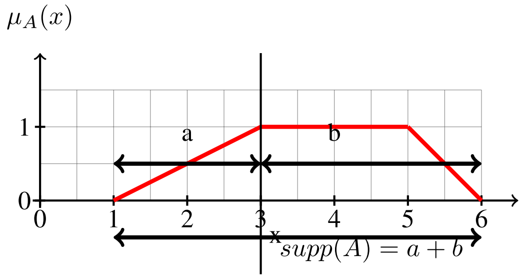


Fig. 5. Graphic interpretation of GR

$$GR = \frac{min(supp(A)) + |supp(A)|}{\Phi} \quad (7)$$
$$where\ \Phi = 1,618033998875\ldots$$

where:
GR is the defuzzification operator,
$supp(A)$ means the support of the fuzzy number A in the X universe.

Equation 8 allows to find a crisp (defuzzification) value for an ordered OFN using the GR method[20]:

$$GR(A) = \begin{cases} min(supp(A)) + \frac{|supp(A)|}{\Phi}, \\ \ for\ (A)\ positively\_directed \\ \\ max(supp(A)) - \frac{|supp(A)|}{\Phi}, \\ \ for\ (A)\ negatively\_directed \end{cases} \tag{8}$$

B. Mandala Factor Defuzzification Operator

Another operator used in the new method is the Mandala Factor operator [21] [22] proposed by J. Czerniak. Mandala is a painting composed of colorful grains of sand, arranged by Buddhist monks. The inspiration of grains of sand forming a mandala is at the heart of the Mandala Factor. Given the trapezoidal OFN A shown in Figure 6, fill the contour marked by the sides of the OFN number and the OX axis with virtual grains of sand as in Figure 7. Then a rectangle is built, which is filled with virtual grains of sand. Backfilling consists in pouring sand vertically in columns until it is exhausted. The crisp value of the number is obtained in the place where the last of the columns poured ended (Figure 8). Mathematically, the notation is as follows 9[21]:

$$MF(A) = \begin{cases} c + r & , for\ (A)\ positively \\ c - r & , for\ (A)\ negatively \end{cases} \tag{9}$$

w:

$$r = \frac{1}{d-c}\int_c^d x\,dx - \frac{c}{d-c}\int_c^d dx + \frac{f}{f-e}\int_f^e dx \\ - \frac{1}{f-e}\int_e^f x\,dx + \int_d^e dx$$

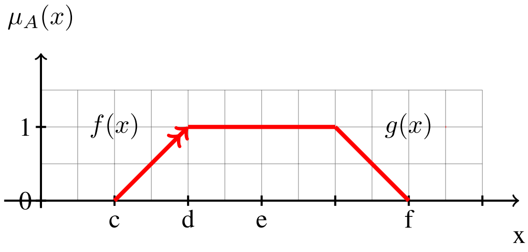


Fig. 6. OFN number A

VIII. Selected mathematical testing functions

For the purpose of the experiment, at this stage, the functions that are most often used as testing were selected. Literature results for various optimization algorithms are also available [23]. Selected mathematical functions:

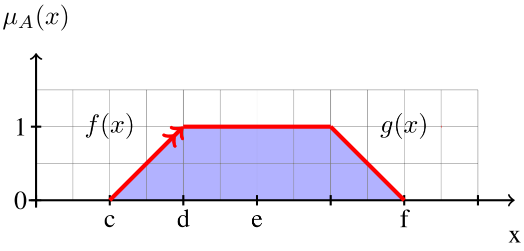


Fig. 7. Visualization of the Mandela Factor operation - step one

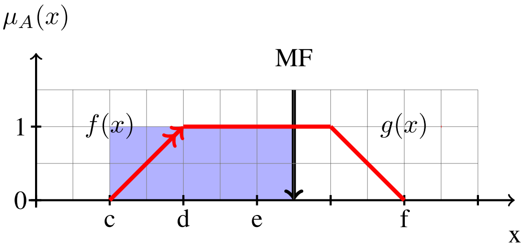


Fig. 8. Visualization of the Mandela Factor operation - step two

- The Sphere function is described by Equation 10.

$$f(x) = \sum_{i=1}^{n} x_i^2 \tag{10}$$

  - Recommended variable values: $-5.12 \leq x_i \leq 5.12$ $i = 1, 2, ..., n$
  - Global minimum: $x = (0, ..., 0), f(x) = 0$
- The Rosenbrock function is described by Equation 11.

$$f(X) = \sum_{i=1}^{d-1}[100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2] \tag{11}$$

  - Recommended variable values: $-2,048 \leq x_i \leq 2,048$ $i = 1, 2, ..., n$
  - Global minimum: $x = (1, ..., 1), f(x) = 0$
- The Rastrigin function is described by Equation 12.

$$f(X) = An + \sum_{i=1}^{n}[x_i^2 - Acos(2\pi x_i)] \tag{12}$$

  - Recommended variable values: $-5,12 \leq x_i \leq 5,12$ $i = 1, 2, ..., n$
  - Global minimum: $x = (0, ..., 0), f(x) = 0$
- The Griewank function is described by Equation 13.

$$f_n(x_1, ..., x_n) = 1 + \frac{1}{4000}\sum_{i=1}^{n} x_i^2 - \prod_{i=1}^{n} cos(\frac{x_i}{\sqrt{i}}) \tag{13}$$

  - Recommended variable values: $-600 \leq x_i \leq 600$ $i = 1, 2, ..., n$
  - Global minimum: $x = (0, ..., 0), f(x) = 0$

- The Schwefel function is described by Equation 14.

$$f(x) = \sum_{i=1}^{n} \Bigg[- x_i \sin(\sqrt{|x_i|}) \Bigg] \tag{14}$$

  The test area is usually limited to a hypercube $-500 \leq x_i \leq 500,\ i = 1, ..., n$.
  - Recommended variable values: $-500 \leq x_i \leq 500$ $i = 1, 2, ..., n$
  - Global minimum $f(x) = -n \cdot 418.9829;\ \ x_i = 420.9687,\ i = 1, ..., n.$
- The Ackley function is described by Equation 15.

$$f(x) = -aexp(-b\sqrt{\frac{1}{d}\sum_{i=1}^{d} x_i^2)} \tag{15}$$

  - Recommended variable values: $a = 20,\ b = 0.2,\ c = 2\pi$.
  - Global minimum: $x = (0, ..., 0), f(x) = 0$

TABLE I
AVERAGE RESULTS FOR THE DEFUZZIFICATION FUNCTION

Function name	Statistic data	GR	MF	GR	MF
		Trapezoid		Rectangular triangle	
Sphere	SD	0	0	0	0
	AV	0	0	0	0
	SD time	0,564357941	0,026869202	0,043654864	0,035467
	AV time	0,443666667	0,344333333	0,336666667	0,322
Rosenbrock	SD	0,004398444	0,004238917	0,004300068	0,00308
	AV	0,003596884	0,002850368	0,00267976	0,003214
	SD time	1,119912815	0,081095942	0,063162807	0,020525
	AV time	0,592333333	0,454	0,399666667	0,361667
Rastrigin	SD	0	0	0	0
	AV	0	0	0	0
	SD time	0,020899321	0,068043259	0,086659593	0,022816
	AV time	0,356666667	0,406666667	0,420666667	0,336333
Griewank	SD	0	0	0	0
	AV	0	0	0	0
	SD time	0,021890532	0,033314938	0,05339185	0,020457
	AV time	0,369666667	0,392666667	0,361	0,342333
Schwefel	SD	0,015795763	0,041012122	0,015873975	0,059517
	AV	0,003475821	0,011620559	0,004544122	0,01207
	SD time	0,041811014	0,038639209	0,055403432	0,034709
	AV time	0,469666667	0,483666667	0,441666667	0,435667
Ackley	SD	0	0	0	0
	AV	4,44E-16	4,44E-16	4,44E-16	4,44E-16
	SD time	0,020117471	0,030026808	0,088617089	0,020126
	AV time	0,394333333	0,395333333	0,504333333	0,361333

IX. RESULTS

The defuzzification operator is necessary for the new method to work, because it is used to defuzzify the OFN number - to get the crisp value in the form of a real number. Tables 1 compares the results generated by the algorithm for the GR and MF defuzzification functionals. In order to select the appropriate and optimal combination of defuzzification and fuzzification operators, the results obtained by the algorithm for the selected functions should be compared. Each of the test functions was run 30 times and the results are presented in Table I.

As can be seen, the results presented in Table I show that the defuzzification method has an impact on the results of the OFNBee algorithm. Although the method works better and its optimization result is good for both defuzzification functions presented in the article, it can be observed that we get a more accurate result for the MF. The results presented in the table are the average results for each of the mathematical functions and the fuzzification function. The algorithm was run 30 times for each math function and for each defuzzification function. Therefore, it can be considered that the algorithm has a high repeatability. Thanks to the stable operation of the method, it is possible to assess the effect of the defuzzification method on the result.

The table shows that the algorithm achieves the expected value of 0 for the functions Sphere, Rastrigin, Griewank and Ackley. For the functions Schwefel and Rosenbrock, the results are close to the expected 0, but do not reach it. However, you can see that the results for the MF function are closer to the expected value of 0. It follows that this defuzzification function will be suitable for this set of functions. Results presented

X. SUMMARY

The new hybrid OFNBee method is characterized by the use of three groups of bees, but thanks to the use of OFN notation and the use of the basic feature of OFN, i.e. referral, the algorithm finds a solution much faster when using a smaller population size - fewer bees in groups. Directed fuzzy numbers allow very well to reflect the sense of decision-making occurring in the hive during the dance of real bees. A very important part of OFN are the fuzzyfication and defuzzyfication operators, which become essential elements when using directed fuzzy numbers to solve real-world problems. Defuzzyfication methods are available in the literature, but the specificity of OFN and bee optimization algorithms does not always allow the use of existing functionals. In the experimental part, two existing fuzzyfication methods were used, i.e. Golden Ratio and Mandala Factor. They were created in the AIRLAB Artificial Intelligence and Robotics Research Laboratory at Kazimierz Wielki Univesrity.

A new optimization method using ordered fuzzy numbers is exceptionally good at optimizing mathematical functions. And thanks to the use of ordered fuzzy numbers, it is possible to accurately reproduce the natural mechanisms occurring in the hive. As can be seen in Table I, the accuracy and speed of the method are greatly influenced by the defuzzification functions. It should be noted that these functions are available in the literature, so they are not adapted to the method. However, it can be seen that the MF method returns better results than the GR method. It can therefore be concluded that the use of a dedicated defuzzification method could further increase its efficiency. In-depth research in this area may result in the creation of dedicated defuzzification functions for the OFNBee optimization method.

There are few defuzzification methods dedicated to OFN in the literature. The main reason for this is that such methods must be direction sensitive. Directing is the main characteristic of OFN, therefore the use of defuzzification methods other than those dedicated to OFN could disturb the specificity of OFNBee operation that uses the arithmetic of these numbers. The next stage of research will be the creation of further defuzzification methods, but dedicated to selected optimization problems and comparing them with the methods described in this article.

References

[1] D. Ewald, H. Zarzycki, and J. M. Czerniak, "Certain aspects of the ofnbee algorithm operation for different fuzzifiers," in *Uncertainty and Imprecision in Decision Making and Decision Support: New Advances, Challenges, and Perspectives*, K. T. Atanassov, V. Atanassova, J. Kacprzyk, A. Kałuszko, M. Krawczak, J. W. Owsiński, S. S. Sotirov, E. Sotirova, E. Szmidt, and S. Zadrożny, Eds. Cham: Springer International Publishing, 2022. ISBN 978-3-030-95929-6 pp. 241–256.

[2] D. Ewald, J. M. Czerniak, and M. Paprzycki, "Ofnbee method applied for solution of problems with multiple extremes," in *Advances and New Developments in Fuzzy Logic and Technology*, K. T. Atanassov, V. Atanassova, J. Kacprzyk, A. Kałuszko, M. Krawczak, J. W. Owsiński, S. S. Sotirov, E. Sotirova, E. Szmidt, and S. Zadrożny, Eds. Cham: Springer International Publishing, 2021. ISBN 978-3-030-77716-6 pp. 93–111.

[3] ——, "A New OFNBee Method as an Example of Fuzzy Observance Applied for ABC Optimization," in *Theory and Applications of Ordered Fuzzy Numbers. A Tribute to Professor Witold Kosinski*, ser. Studies in Fuzziness and Soft Computing, P. Prokopowicz, J. M. Czerniak, D. Mikolajewski, L. Apiecionek, and D. Ślęzak, Eds. Springer International Publishing, 2017, ch. 12, pp. 207–222.

[4] D. Ewald, J. M. Czerniak, and H. Zarzycki, "OFNBee Method Used for Solving a Set of Benchmarks," in *Advances in Fuzzy Logic and Technology 2017. IWIFSGN 2017, EUSFLAT 2017*, ser. Advances in Intelligent Systems and Computing, J. e. a. Kacprzyk, Ed. Springer, 2018, vol. 642, pp. 24–35.

[5] B. Kadda Beghdad, Bey nad Sofiane, B. Farid, and N. Hassina, "Improved virus optimization algorithm for two-objective tasks scheduling in cloud environment," *Communication Papers of the 2019 Federated Conference on Computer Science and Information Systems,ACSIS*, vol. 20, pp. 109–117, 2019.

[6] R. Pellegrini, A. Serani, G. Liuzzi, F. Rinaldi, S. Lucidi, and M. Diez, "Hybridization of multi-objective deterministic particle swarm with derivative-free local searches," *Mathematics*, vol. 8, no. 4, 2020. doi: 10.3390/math8040546. [Online]. Available: https://www.mdpi.com/2227-7390/8/4/546

[7] L. Zadeh, "Fuzzy sets," *Information and Control*, 1965.

[8] ——, "Outline of new approach to the analysis of complex systems and decision process," *IEEE on Systems, „Man and Cybernetics"*, vol. SMC-3, pp. 28–44, 1973.

[9] D. Kacprzak, "Przychód i koszt całkowity przedsiębiorstwa wyrażony przy użyciu skierowanych liczb rozmytych," *Zarządzanie i Finanse*, no. 2, pp. 139–149, 2012.

[10] M. Wagenknecht, R. Hampel, and V. Schneider, "Computational aspects of fuzzy arithmetics based on Archimedean t-norms," *Fuzzy Sets and Systems*, vol. 123, no. 1, pp. 49–62, 2001. doi: https://doi.org/10.1016/S0165-0114(00)00096-8. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165011400000968

[11] M. B. Khan, H. A. Othman, M. G. Voskoglou, L. Abdullah, and A. M. Alzubaidi, "Some certain fuzzy aumann integral inequalities for generalized convexity via fuzzy number valued mappings," *Mathematics*, vol. 11, no. 3, 2023. doi: 10.3390/math11030550. [Online]. Available: https://www.mdpi.com/2227-7390/11/3/550

[12] J. Łeski, *Systemy neuronowo-rozmyte*, WNT, Warszawa, 2008.

[13] W. Kosiński, P. Prokopowicz, and D. Ślęzak, "On algebraic oprerations on fuzzy numbers," *Inteligent Information Processing and Web Mining: proceedings of the International IIS:IIPWM'03 Conference held In Zakopane*, pp. 353–362, 2003.

[14] W. Kosiński and U. Markowska-Kaczmar, "On evolutionary approach for determining defuzzyfication operator," *Proceedings of the International Multiconferece on Computer Science and Information Technology*, pp. 93–101, 2006.

[15] W. Kosiński and P. Prokopowicz, "Algebra liczb rozmytych," *Matematyka Stosowana*, no. 5, pp. 37–63, 2004.

[16] W. Kosinski, "On Defuzzyfication of Ordered Fuzzy Numbers," in *Artificial Intelligence and Soft Computing - ICAISC 2004*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2004, vol. 3070, pp. 326–331. ISBN 978-3-540-22123-4

[17] W. Kosinski and D. Wilczynska-Sztyma, "Defuzzification and Implication Within Ordered Fuzzy Numbers," in *Fuzzy Systems (FUZZ), 2010 IEEE International Conference on Computational Intelligence*. IEEE, 2010, pp. 1–7.

[18] W. Kosinski, W. Piasecki, and D. Wilczynska-Sztyma, "On fuzzy rules and defuzzification functionals for Ordered Fuzzy Numbers," in *Proc. of AI-Meth'2009 Conference, November 2009*. AI-METH Series, Gliwice, 2009, pp. 161–178.

[19] D. Wilczyńska-Sztyma and K. Wielki, "Direction of Research into Methods of Defuzzification for Ordered Fuzzy Numbers," *XII International PhD Workshop OWD 2010, 23–26 October 2010*, 07 2019.

[20] W. T. Dobrosielski, J. Szczepański, and H. Zarzycki, "A proposal for a method of defuzzification based on the golden ratio—gr," in *Novel Developments in Uncertainty Representation and Processing*, K. T. Atanassov, O. Castillo, J. Kacprzyk, M. Krawczak, P. Melin, S. Sotirov, E. Sotirova, E. Szmidt, G. De Tré, and S. Zadrożny, Eds. Cham: Springer International Publishing, 2016, pp. 75–84.

[21] J. M. Czerniak, W. T. Dobrosielski, and I. Filipowicz, "Comparing fuzzy numbers using defuzzificators on ofn shapes," in *Theory and Applications of Ordered Fuzzy Numbers. A Tribute to Professor Witold Kosinski*, ser. Studies in Fuzziness and Soft Computing, P. Prokopowicz, J. M. Czerniak, D. Mikolajewski, L. Apiecionek, and D. Ślęzak, Eds. Springer International Publishing, 2017, pp. 99–132.

[22] J. M. Czerniak, *Zastosowania skierowanych liczb rozmytych w wybranych algorytmach optymalizacji rojowej*. Wydawnictwo Uniwersytetu Kazimierza Wielkiego w Bydgoszczy, 2019.

[23] Z. Abdel-Rahman, "Studies on metaheuristics for continous global optimization problems," *Ph.D. thesis*, Kyoto University, Japan, 2004.

DOI: 10.15439/2023F3874
ISSN 2300-5963 ACSIS, Vol. 37

Use of Dynamic Neural Networks for Modeling Nonlinear Objects with Significant Nonlinearity

Oleksandr Fomin
0000-0002-8816-0652
Odessa Polytechnic National University,
1, Shevchenko Avenue,
Odessa, Ukraine
Email: fomin@op.edu.ua

Sergii Polozhaenko
0000-0002-4082-8270
Odessa Polytechnic National University,
1, Shevchenko Avenue,
Odessa, Ukraine
Email: polozhaenko@op.edu.ua

Andrii Orlov
0000-0002-3256-5044
Odessa Polytechnic National University,
1, Shevchenko Avenue,
Odessa, Ukraine
Email: 9901020@stud.op.edu.ua

Valentyn Krykun
0000-0002-3764-9255
Odessa Polytechnic National University,
1, Shevchenko Avenue,
Odessa, Ukraine
Email: 9901053@stud.op.edu.ua

Andrii Prokofiev
0000-0002-4520-4248
Odessa Polytechnic National University,
1, Shevchenko Avenue,
Odessa, Ukraine
Email: fallbrick1985@gmail.com

***Abstract*—The work is devoted to the problem of nonlinear modeling of objects based on dynamic neural networks. The aim of the work is to improve the accuracy of modeling dynamic objects with significant nonlinearities using neural network models, and identify the scope of their effective application. This aim is achieved by applying the dynamic nonlinear models in the form of time delay neural networks. The scientific novelty of the work lies in the determination of the dependences between the accuracy of suggested models and the types of model input signals, as well as the amplitudes of model input signals. Practical usefulness of the research lies in the determination of the area of effective use of suggested models of dynamic objects with significantly nonlinear features, such us saturation. Significance of the obtained results: the application of the proposed models for identification dynamic objects with significantly nonlinear characteristics allows to improve the accuracy of the modelling process in comparison with models based on deterministic identification methods, such as integro-power series based on multidimensional weight functions.**

***Index Terms*—Identification, nonlinear dynamic objects, significant nonlinearities, time delay neural networks.**

I. Introduction

TODAY, as a result of the development of technology and science, practical applications increasingly consider dynamic control objects characterized by significantly nonlinear properties. Due to these characteristics, objects can function in more complex modes than objects with characteristics in the form of linear or insignificant nonlinear functions [1]. A non-significantly nonlinear function should be understood as the case when the nonlinear function and its 1st and 2nd order derivatives are continuous over the entire range of input signal changes. Nonlinear links that do not satisfy this condition are considered to be significant nonlinearities [2].

For successful interaction with such objects (solving control, management, and diagnostic tasks), it is first of all necessary to ensure their adequate mathematical support and effective modeling tools. However, the presence of significantly nonlinear characteristics makes the use of existing analytical methods ineffective. Such models don’t reflect the dynamic and nonlinear properties of a real object, so they cannot provide high identification accuracy [2]. An up-to-date approach to modeling nonlinear dynamic objects is the artificial neural network apparatus [3-7].

The aim of the work is to improve the accuracy of modeling dynamic objects with significant nonlinearities using neural network models, and identify the scope of their effective application.

This goal can be achieved by examining the existing architectures of neural networks for modeling nonlinear dynamic objects and identifying their advantages, disadvantages and determining the areas for their effective use. The following tasks are considered within the framework of this work:

1. Study of modeling accuracy of nonlinear objects with smooth nonlinearity.
2. Study of modeling accuracy of nonlinear objects with piecewise linear nonlinearity.

II. Relevant works

Today several methods for modeling nonlinear dynamic objects based on artificial neural networks are known [8, 9]. There are Dynamic Neurospatial Mapping (Dynamic Neuro-SM) [10, 11], Time-Delay Neural Networks (TDNN) [12, 13] and Wiener-type Dynamic Neural Networks (Wiener-type DNN) [14-16].

Dynamic Neuro-SM type models are improvements over the well-known Static Neuro-SM models [10, 11], which aim to map a given approximate model of an object to an exact model. Dynamic Neuro-SM models use neural networks to transform an existing (rough) model into a desired (exact) model using machine learning approach [10]. Such models provide improved accuracy compared to Static Neurospatial Mapping models, but assume some a priori information about the laws of functioning of the object under study [11].

Wiener-type DNN is based on the principle of building a nonlinear dynamic Winner model. This model consists of two parts arranged in series: linear dynamic and nonlinear static models [14, 15]. In this case, the dynamic properties of the object are reproduced by a linear model, and the non-linear properties are reproduced in a static non-linear model. In Wiener-type DNN static non-linear model is implemented as an artificial neural network [14-16]. This structure can significantly increase the reliability of the dynamic neural model, but has a complex (hybrid) structure, which imposes additional requirements on the network learning algorithms and narrows the scope of the model.

Among the considered variants of models TDNN are the most general structures consisting of several layers with direct connection (direct signal propagation) [12]. Such models are capable of learning from the input-output experimental data of nonlinear dynamic objects [12, 13]. These models have good convergence, which is an advantage over the models based on Dynamic Neuro-SM and Wiener-type DNN models, mentioned above. So, using of TDNN for modeling dynamic objects with highly nonlinear characteristics provides unique advantages over other models.

III. Time-Delayed Neural Nets

TDNN models are an effective tool for modeling non-linear dynamic objects with continuous characteristics. The most commonly used structure of TDNN consists of three layers: input, hidden and output.

There are many structures of neural networks: with several hidden layers, different activation functions and topologies. However, using this models gives more complex expression for model output data. This is a significant disadvantage in comparison with three-layer neural networks for modeling nonlinear dynamic objects.

The input layer of TDNN includes M neurons, where M is a length of the object's model memory. The number of neurons M is chosen in such a way as to best reflect the dynamic properties of the object [17].

The number of neurons K is chosen in such a way as to best fit the training set.). It receive input data $\mathbf{x}_n(t)=[x(t_n), x(t_{n-1}), \ldots, x(t_{n-M-1})]$, $t_n=n\Delta t$, $n=1, 2, \ldots$ The hidden layer includes K neurons with a nonlinear activation function. The number of neurons K is chosen in such a way as to best reflect the nonlinear properties of the object.

The output layer includes 1 neuron with a linear activation function. The signal $y_n(t)$ on the output layer at the time t_n depends as on the value of input signal $\mathbf{x}_n(t)$ at the current moment t_n, as on input data $x(t_n), \ldots, x(t_{n-M-1})$ at the times t_n, ... , t_{n-M-1}. So, the output data $y_n(t)$ of TDNN model are determined by the expression:

$$y(t_n) = b_0 + s_0 \sum_{i=1}^{K} w_i\, S_i \left(b_i + \sum_{j=1}^{M} w_{i,j} x(t_{n-j}) \right), \tag{1}$$

where b_0, b_i – biases of the output and hidden layers neurons accordingly; S_0, S_i – activation functions of the output and input layers neurons accordingly; w_i, $w_{i,j}$ – weighing coefficients of the output and hidden layers neurons accordingly.

The activation function can be expressed as a polynomial of degree p. Then the output data $y_n(t)$ of TDNN model are determined by the expression [17, 18]:

$$y(t_n) = b_0 + s_0 \sum_{i=1}^{K} w_i\, S_i \left(b_i + \sum_{j=1}^{M} w_{i,j} x(t_{n-j}) \right)^p . \tag{2}$$

Fig. 1 shows a three-layer TDNN with M inputs, a hidden layer with J neurons, and one output neuron.

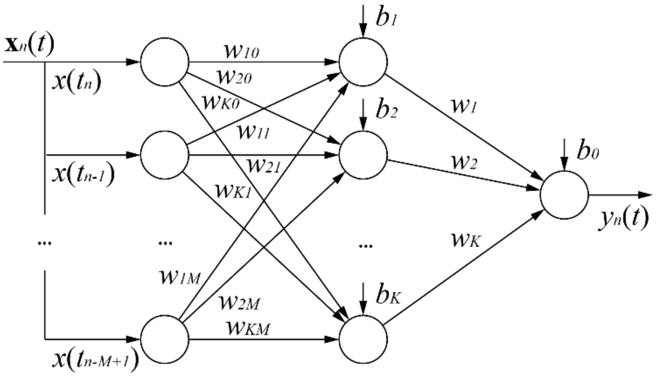


Fig. 1. Three-layer TDNN with M inputs and K hidden neurons

TDNN network can quickly learns dynamic behavior taking into account high-order nonlinear characteristics [18, 19] if they are trained on the data of input-output experiments.

IV. Experimental Setup

The effectiveness of the TDNN models is studied on the example of the test object. Test object simulation model with a first-order dynamic block and nonlinear block in feedback [20] is shown on Fig. 2.

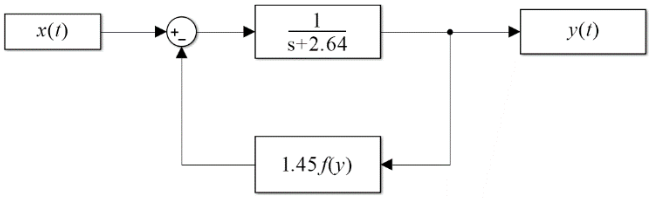


Fig. 2. Block diagram of the test nonlinear dynamical object

The polynomial function and function with saturation are used as a nonlinear feature $f(y)$ in feedback block of the simulation model.

To research the accuracy of modeling dynamic objects with significant nonlinearities using neural network models, and identify the scope of their effective application it is necessary to create training and test datasets from input and output signals.

To form a training and test dataset the test signals $x(t)$ in the form of impulse, step, linear and harmonic functions with different amplitudes a are applied to the input of the simulation model. As a result, a set of output reactions $y(t)$ and sequential segments $\mathbf{x}_n(t)$ of input signal $x(t)$, shifted by one value Δt for each type of nonlinear feature $f(y)$ forms a training and test dataset.

To model objects with different types of nonlinearity, it is necessary to train TDNN on each of the generated datasets [21-23].

To create a neural network, the Keras (keras.io) software tool is used. It is one of the key Python libraries for efficiently organizing APIs when modeling neural networks of any complexity. The library is most effective when building small networks with a sequential structure, where layers follow each other and one input and one output layer. Although it is possible to model more complex neural network structures with multiple inputs and outputs.

To build feedforward networks with Keras we can use an any number of sequential layers of the predefined types: Input, Dense and Activation. The library has a ready-made set of loss functions and optimization algorithms that allow us to quickly train the model and avoid local minima whenever possible.

As a result, a three-layer neural network was created and trained. The input signal $x(t)$ is fed to the M neurons of the input layer. The hidden layer consists of K neurons. The output layer consists of one neuron with a linear activation function. The block diagram of the TDNN is shown on Fig. 3. In this figure, the value *None* in the data dimension vector means a variable number of rows in the dataset.

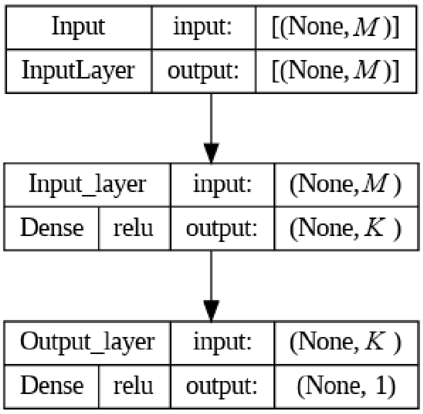


Fig. 3. Structure diagram of the TDNN with M inputs and K hidden neurons

To determine the best values of M and K in the given structure of TDNN a number of neural networks with different numbers of neurons in the input M and hidden K layers are constructed [24]. The result of experiment in the form of the loss as a function of the neurons number in the input M and hidden K layers is presented in Fig. 4.

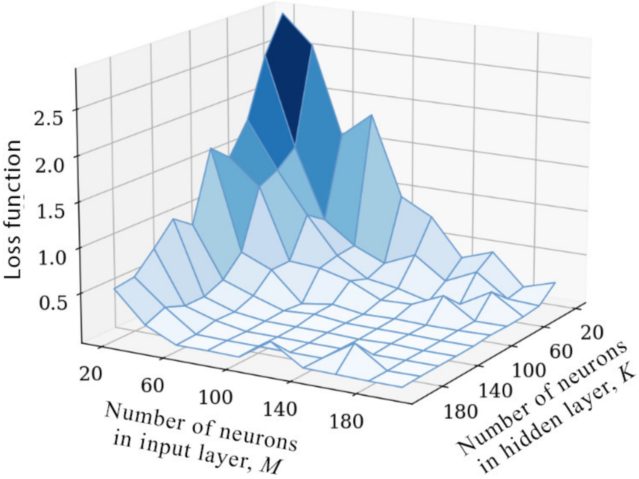


Fig. 4. Loss as a function of the neurons number in the input M and hidden K layers

The result experiment as a dependence of the learning time (epoch) on the neurons number in the input M and hidden K layers are presented in Fig. 5. As a result of TDNN structure experiment, the values M=15 and K=50 were accepted for the number of neurons in the input and hidden layers of the TDNN respectively. The resulting TDNN was used to research the accuracy of models for dynamic objects with significant nonlinearities [25, 26].

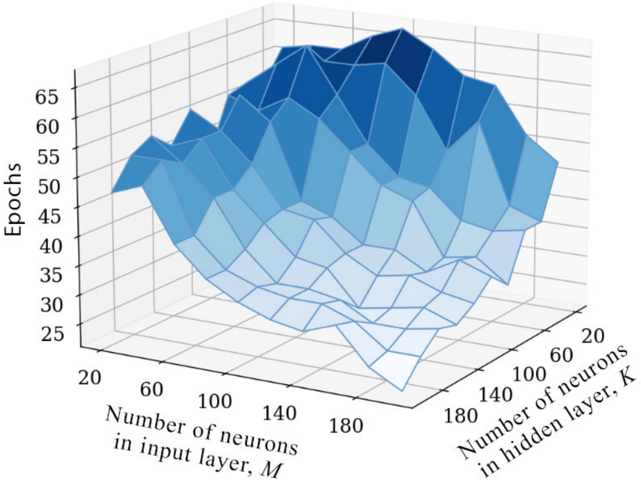


Fig. 5. Epochs as a function of the neurons number in the input M and hidden K layers

For the study the accuracy of modeling dynamic objects with significant nonlinearities using neural network models, and identify the scope of their effective application, two experiments were organized and executed:

1. Study of the scalability of the model to various input signals.
2. Study of extrapolation properties of the model.

The results of both experiments are compared with the results of simulation and identification using deterministic identification methods, such as integro-power series based on multidimensional weight functions.

A. Study of the scalability of the model to various input signals.

The training dataset includes impulse signals $x(t) = a\delta(t)$ various amplitude ($a \in (0, 1]$) at the input of the object and its response $y(t)$ on the output. The test dataset includes step $x(t)=a\Theta(t)$, linear $x(t)=at$ and harmonic $x(t)=asin(t)$ signals various amplitude ($a \in (0, 1]$) at the input of the object and its response $y(t)$ on the output.

A TDNN model builds on the data of the training dataset. The accuracy of the model is tested on the data of the test dataset (signals that are not part of the training sample).

The experiment executes for objects with nonlinear feature $f(y)$ in feedback block in the form of smooth (polynomial) as well as saturation function. Based on the results of the experiment, we make a conclusion about the area of effective use of TDNN models. The model output $y_n(t)$ is compared with the model output $y(t)$ obtained by the simulation and the model output $y_v(t)$ based on deterministic identification methods, such as integro-power series based on multidimensional weight functions [27-29].

Fig. 6 shows a comparison of the output signals $y_n(t)$, $y_v(t)$ and $y(t)$, obtained as a result of the input signal $x(t)=a\delta(t)$ on the TDNN model, integro-power model and simulation the nonlinear dynamical object (fig. 2) for nonlinear feature $f(y)$ in feedback block in the form of polynomial function.

This experiment shows comparable modeling accuracy using TDNN and integro-power models under the action of input signals $x(t) = a\delta(t)$ various amplitude ($a \in (0, 1]$).

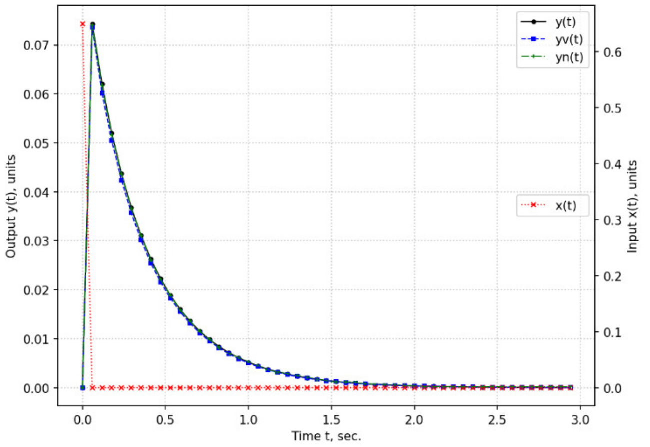


Fig. 6. Comparison of the output signals $y_n(t)$, $y_v(t)$ and $y(t)$, obtained as a result of the input signal $x(t)=a\delta(t)$ (a=0.65) on the TDNN model, integro-power model and simulation the nonlinear dynamical object respectively for nonlinear feature $f(y)$ in feedback block in the form of polynomial function

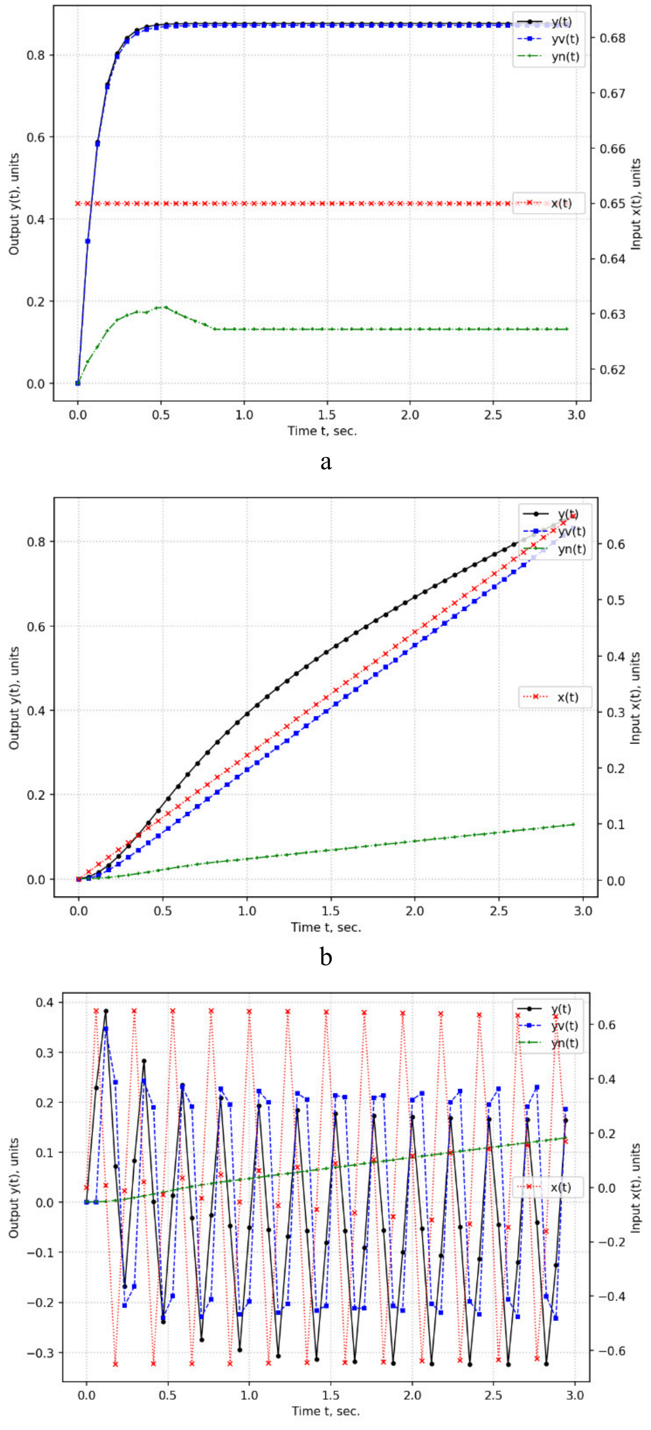
This figure shows a comparison of the output signals $y_n(t)$, $y_v(t)$ and $y(t)$, obtained as a result of the input signals $x(t)=a\Theta(t)$ (fig. 7a), $x(t)=at$ (fig. 7b) and $x(t)=asin(t)$ (fig. 7c) on the TDNN model, integro-power model and simulation the nonlinear dynamical object for nonlinear feature $f(y)$ in feedback block in the form of polynomial function. This experiment shows that the TDNN model is significantly inferior in accuracy to integro-power models under the action of input signals $x(t)$ various amplitude ($a \in (0, 1]$), which were not included in the training dataset.

The conclusion follows from the experiment: TDNN models are not invariant to the form of the input signal. The TDNN model can adequately reflect the properties of the dynamic object in the case of training on a sufficient amount of data. The training dataset must include input signals various amplitude of the same type as in the test dataset. This is a disadvantage of neural network models in comparison with models based on deterministic identification methods, such us integro-power series on the base of multidimensional weight functions [29, 31].

B. *Study of extrapolation properties of the model.*

The training dataset includes impulse $x(t)=a\delta(t)$, step $x(t)=a\Theta(t)$, linear $x(t)=at$ and harmonic $x(t)=asin(t)$ signals various amplitude ($a \in (0, 1]$) at the input of the object and its response $y(t)$ on the output. Simulating data are obtained for objects with nonlinear feature $f(y)$ in feedback block in the form of saturation function.The test dataset includes the same input signals with amplitude in the interval (1, 2] and responses at the output. The accuracy of the model is tested on the data of the test dataset (signals that are not part of the training sample).

The experiment executes for objects with nonlinear feature $f(y)$ in feedback block in the form of saturation function. Based on the results of the experiment, we make a conclusion about the area of effective use of TDNN models. The model output $y_n(t)$ is compared with the model output $y(t)$ obtained by the simulation and the model output $y_v(t)$ based on deterministic identification methods, such as integro-power series based on multidimensional weight functions.



a

b

c

Fig. 7. Comparison of the output signals $y_n(t)$, $y_v(t)$ and $y(t)$, obtained as a result of the input signal $x(t)$ (a=0.65) on the TDNN model, integro-power model and simulation the nonlinear dynamical object respectively for nonlinear feature $f(y)$ in feedback block in the form of polynomial function: a – $x(t)=a\Theta(t)$; b – $x(t)=at$; c – $x(t)=a\sin(t)$; a=0.65

This figure shows a comparison of the output signals $y_n(t)$, $y_v(t)$ and $y(t)$, obtained as a result of the input signal $x(t)=a\Theta(t)$ (a=0.7) on the TDNN model, integro-power model and simulation the nonlinear dynamical object respectively for nonlinear feature $f(y)$ in feedback block in the form of saturation function.

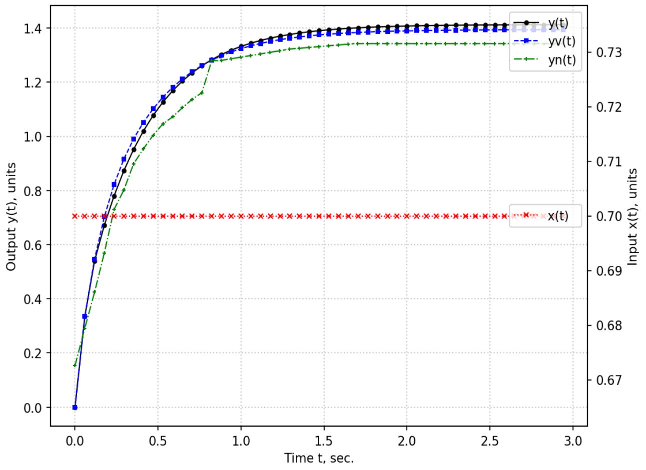


Fig. 8. Comparison of the output signals $y_n(t)$, $y_v(t)$ and $y(t)$, obtained as a result of the input signal $x(t)=a\Theta(t)$ (a=0.7) on the TDNN model, integro-power model and simulation the nonlinear dynamical object respectively for nonlinear feature $f(y)$ in feedback block in the form of saturation function

This figure shows a comparison of the output signals $y_n(t)$, $y_v(t)$ and $y(t)$, obtained as a result of the input signal $x(t)=a\Theta(t)$ (a=1.65) on the TDNN model, integro-power model and simulation the nonlinear dynamical object respectively for nonlinear feature $f(y)$ in feedback block in the form of saturation function.

This experiment shows that the integro-power model lose their accuracy in interpolation and extrapolation tasks when dealing with dynamic objects with significant nonlinear features, for example, saturation function $f(y)$. The obtained simulation results for the dynamic objects with significant nonlinear features allow to conclude that the extrapolation properties of TDNN model are far superior in accuracy to integro-power models under the action of input signals $x(t)$ various amplitude ($a \in (1, 2]$) for all types of signals present in the training dataset.

The obtained results make it possible to determine the area of effective application of TDNN models when modeling dynamic objects with significant nonlinearities.

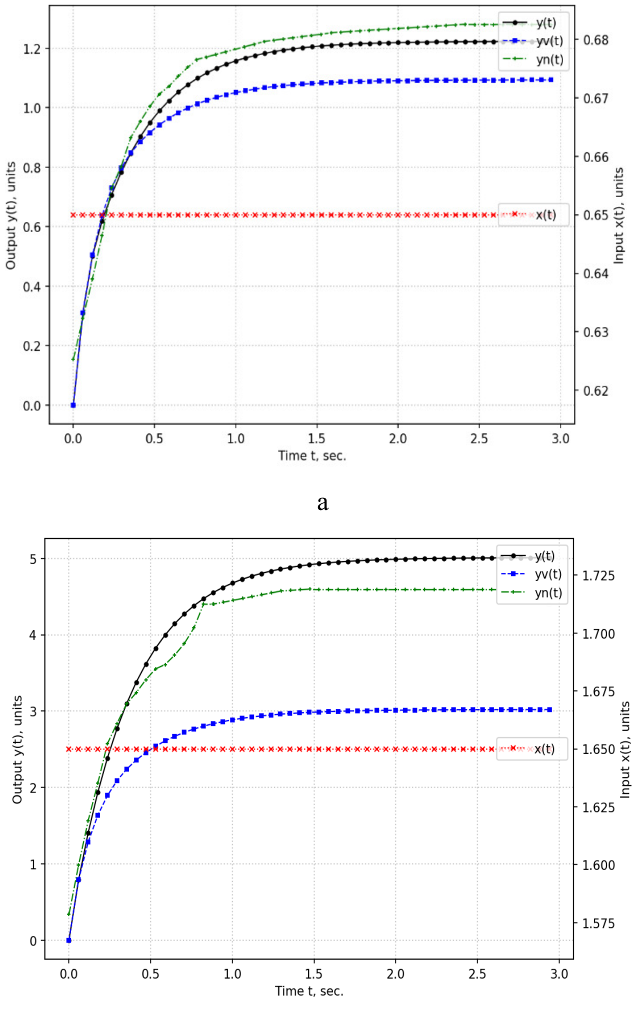
Conclusion

The results of this research are as follows.

The scientific novelty of the work lies in the determination of the dependences between the accuracy of TDNN models and the types of model input signals, as well as the amplitudes of model input signals.

Practical usefulness of the research lies in the determination of the area of effective use of TDNN models – dynamic objects with significantly nonlinear features.

Significance of the obtained results: the application of the proposed models for identification dynamic objects with significantly nonlinear characteristics allows to improve the accuracy of the modelling process in comparison with models based on deterministic identification methods, such as integro-power series based on multidimensional weight functions.



a

b

Fig. 9. Comparison of the output signals $y_n(t)$, $y_v(t)$ and $y(t)$, obtained as a result of the input signal $x(t)=a\Theta(t)$ on the TDNN model, integro-power model and simulation the nonlinear dynamical object respectively for nonlinear feature $f(y)$ in feedback block in the form of saturation function: a – interpolation task (a=0.65); b – extrapolation task (a=1.65)

TDNN models are not invariant to the form of the input signal. The TDNN model can adequately reflect the properties of the dynamic object in the case of training on a sufficient amount of data. The training dataset must include input signals various amplitude of the same type as in the test dataset. This is a disadvantage of neural network models in comparison with models based on deterministic identification methods, such as integro-power series on the base of multidimensional weight functions.

The interpolation and extrapolation properties of TDNN model are far superior in accuracy to integro-power models under the action of input signals $x(t)$ various amplitude ($a \in (1, 2]$) for all types of signals present in the training dataset for dynamic objects with significant nonlinearities.

The obtained results make it possible to determine the area of effective application of TDNN models when modeling dynamic objects with significant nonlinearities.

The proposed models verified using the data of the test dynamical objects with significant nonlinearities such as polynomial and saturation functions.

References

[1] A. Agresti, "Foundations of linear and generalized linear models", Wiley series in probability and statistics, 2017, 480 p.

[2] J. Schoukens and L. Ljung, "Nonlinear System Identification: A User-Oriented Road Map", in IEEE Control Systems Magazine, vol. 39, no. 6, pp. 28-99, Dec. 2019, doi: 10.1109/MCS.2019.2938121.

[3] C. Rudin and J. Radin, "Why are we using black box models in AI when we don't need to? A lesson from an explainable AI competition", Harvard Data Science Review, vol. 2, no. 1, 2019, doi: 10.1162/99608f92.5a8a3a3d.

[4] C. Maszczyk, M. Kozielski and M. Sikora, "Rule-based approximation of black-box classifiers for tabular data to generate global and local explanations", 2022 17th Conference on Computer Science and Intelligence Systems (FedCSIS), Sofia, Bulgaria, 2022, pp. 89-92, doi: 10.15439/2022F258.

[5] Gomolka, Z., Dudek-Dyduch, E., Kondratenko, Y.P. "From homogeneous network to neural nets with fractional derivative mechanism", International Conference on Artificial Intelligence and Soft Computing, ICAISC-2017, Rutkowski, L. et al. (Eds), Part I, Zakopane, Poland, 11-15 June 2017, LNAI 10245, Springer, Cham, 2017, pp. 52-63, doi: 10.1007/978-3-319-59063-9_5.

[6] N. Todorovic and P. Klan, "State of the Art in Nonlinear Dynamical System Identification using Artificial Neural Networks", 2006 8th Seminar on Neural Network Applications in Electrical Engineering, Belgrade, Serbia, 2006, pp. 103-108, doi: 10.1109/NEUREL.2006.341187.

[7] Chi-Hsu Wang, Pin-Cheng Chen, Ping-Zong Lin and Tsu-Tian Lee, "A dynamic neural network model for nonlinear system identification", 2009 IEEE International Conference on Information Reuse & Integration, Las Vegas, NV, USA, 2009, pp. 440-441, doi: 10.1109/IRI.2009.5211647.

[8] W. Liu, W. Na, L. Zhu and Q. -J. Zhang, "A review of neural network based techniques for nonlinear microwave device modeling", 2016 IEEE MTT-S International Conference on Numerical Electromagnetic and Multiphysics Modeling and Optimization (NEMO), Beijing, China, 2016, pp. 1-2, doi: 10.1109/NEMO.2016.7561677.

[9] W. Liu, Y. Su and L. Zhu, "Nonlinear Device Modeling Based on Dynamic Neural Networks: A Review of Methods", 2021 IEEE 4th International Conference on Electronic Information and Communication Technology (ICEICT), Xi'an, China, 2021, pp. 662-665, doi: 10.1109/ICEICT53123.2021.9531270.

[10] L. Zhu, Q. Zhang, K. Liu, Y. Ma, B. Peng and S. Yan, "A Novel Dynamic Neuro-Space Mapping Approach for Nonlinear Microwave Device Modeling", in IEEE Microwave and Wireless Components Letters, vol. 26, no. 2, pp. 131-133, Feb. 2016, doi: 10.1109/LMWC.2016.2516761.

[11] Wenyuan Liu, L. Zhu, Weicong Na and Q. -J. Zhang, "An overview of Neuro-space mapping techniques for microwave device modeling", 2016 IEEE MTT-S Latin America Microwave Conference (LAMC), Puerto Vallarta, Mexico, 2016, pp. 1-3, doi: 10.1109/LAMC.2016.7851276.

[12] M. Sugiyama, H. Sawai and A. H. Waibel, "Review of TDNN (time delay neural network) architectures for speech recognition", 1991 IEEE International Symposium on Circuits and Systems (ISCAS), Singapore, 1991, pp. 582-585 vol. 1, doi: 10.1109/ISCAS.1991.176402.

[13] L. Wenyuan, L. Zhu, F. Feng, W. Zhang, Q.-J. Zhang, L. Qian and G. Liu, "A time delay neural network based technique for nonlinear microwave device modelling, in: Micromachines", Basel, vol. 11, no. 9, 2020, p. 831, doi: 10.3390/mi11090831.

[14] W. Liu, Y. Su, H. Tan, F. Feng and B. Zhang, "A Review of Wiener-Type Dynamic Neural Network for Nonlinear Device Modeling", 2022 IEEE MTT-S International Microwave Workshop Series on Advanced Materials and Processes for RF and THz Applications (IMWS-AMP), Guangzhou, China, 2022, pp. 1-3, doi: 10.1109/IMWS-AMP54652.2022.10106887.

[15] W. Liu, W. Na, F. Feng, L. Zhu and Q. Lin, "A Wiener-Type Dynamic Neural Network Approach to the Modeling of Nonlinear Microwave Devices and Its Applications", 2020 IEEE MTT-S International Conference on Numerical Electromagnetic and Multiphysics Modeling and Optimization (NEMO), Hangzhou, China, 2020, pp. 1-3, doi: 10.1109/NEMO49486.2020.9343530.

[16] A. Balestrino and A. Caiti, "Approximation of Hammerstein/Wiener dynamic models", Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium, Como, Italy, 2000, pp. 70-74 vol.1, doi: 10.1109/IJCNN.2000.857816.

[17] G. Stegmayer, M. Pirola, G. Orengo and O. Chiotti, "Towards a Volterra series representation from a neural network model", WSEAS Transactions on Circuits and Systems, archive 1, 2004, pp. 55–61.

[18] L. Wenyuan, L. Zhu, F. Feng, W. Zhang, Q.-J. Zhang, L. Qian and G. Liu, "A time delay neural network based technique for nonlinear microwave device modelling, in: Micromachines", Basel, vol. 11, no. 9, 2020, p. 831, doi: 10.3390/mi11090831.

[19] F. Alleau, E. Poisson, C. V. Gaudin and P. Le Callet, "TDNN with masked inputs", Fourth International Conference on Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint, Singapore, 2003, pp. 989-993 vol.2, doi: 10.1109/ICICS.2003.1292607.

[20] Fomin, O., Polozhaenko, S., Krykun, V., Orlov, A., Lys, D., "Interpretation of Dynamic Models Based on Neural Networks in the Form of Integral-Power Series", In: Arsenyeva, O., Romanova, T., Sukhonos, M., Tsegelnyk, Y. (eds) Smart Technologies in Urban Engineering. STUE 2022. Lecture Notes in Networks and Systems, vol 536. Springer, 2022, Cham, pp. 258-265, doi: 10.1007/978-3-031-20141-7_24.

[21] J. Sen, "Machine Learning – Algorithms, Models and Applications", London, United Kingdom, IntechOpen, 2021, 154 p., doi: 10.5772/intechopen.94615.

[22] Kondratenko, Y.; Atamanyuk, I.; Sidenko, I.; Kondratenko, G.; Sichevskyi, S. "Machine Learning Techniques for Increasing Efficiency of the Robot's Sensor and Control Information Processing". Sensors 2022, 22(3), p. 1062, doi: 10.3390/s22031062.

[23] A. R. Rao and M. Reimherr, "Non-linear functional modelling using neural networks", 2021, https://arxiv.org/abs/2104.09371, doi: 10.48550/arXiv.2104.09371.

[24] Kondratenko, Y., Gerasin, O., Topalov, A., "A simulation model for robot's slip displacement sensors", International Journal of Computing, Vol.15, Issue 4, 2016, pp. 224-236, doi: 10.47839/ijc.15.4.854.

[25] W. Liu, W. Na, W. Zhang, L. Zhu and M. Wang, "A Review of Recent Neural Network Approaches to the Modeling of Nonlinear Microwave Devices," 2020 13th UK-Europe-China Workshop on Millimetre-Waves and Terahertz Technologies (UCMMT), Tianjin, China, 2020, pp. 1-3, doi: 10.1109/UCMMT49983.2020.9296017.

[26] L. Zhang and Q. -J. Zhang, "Simple and Effective Extrapolation Technique for Neural-Based Microwave Modeling," in IEEE Microwave and Wireless Components Letters, vol. 20, no. 6, pp. 301-303, June 2010, doi: 10.1109/LMWC.2010.2047450.

[27] A. C. Meruelo, D. M. Simpson, S. M. Veres and Ph. L. Newland, "Improved system identification using artificial neural networks and analysis of individual differences in responses of an identified neuron", Neural Networks, vol 75, 2016, pp. 56–65, doi: 10.1016/j.neunet.2015.12.002.

[28] C. A. Mitrea, C. K. M. Lee and Z. Wu, "A comparison between neural networks and traditional forecasting methods: case study", International journal of engineering business management, vol. 1, no. 2, 2009, pp. 19–24, doi: 10.5772/6777.

[29] Pavlenko, V.D. and Pavlenko, S.V., "Deterministic identification methods for nonlinear dynamical systems based on the Volterra Model", Applied Aspects of Information Technology, vol. 1, no. 2, 2018, pp. 11-32, doi: 10.15276/aait.01.2018.1.

[30] V. Z. Marmarelis and X. Zhao, "Volterra models and three-layer perceptrons", in IEEE Transactions on Neural Networks, vol. 8, no. 6, pp. 1421-1433, Nov. 1997, doi: 10.1109/72.641465.

[31] G. Govind and P. A. Ramamoorthy, "Multi-layered neural networks and Volterra series: The missing link", 1990 IEEE International Conference on Systems Engineering, Pittsburgh, PA, USA, 1990, pp. 633-636, doi: 10.1109/ICSYSE.1990.203237.

DOI: 10.15439/2023F4830
ISSN 2300-5963 ACSIS, Vol. 37

Compilation through interpretation

Adam Grabski
0000-0002-6283-8461
Warsaw Univeristy of Technology
ul. Nowowiejska 15/19, 00-665 Warsaw, Poland
Email: adam.gr@outlook.com

Ilona Bluemke
0000-0002-2894-5976
Warsaw Univeristy of Technology
ul. Nowowiejska 15/19, 00-665 Warsaw, Poland
Email: Ilona.Bluemke@pw.edu.pl

***Abstract*—As static metaprogramming is becoming more relevant, compilers must adapt to accommodate them. This requires exposing more information about the code, from the compiler to the programmer as well as more powerful compile-time function execution capabilities. The Interpreter component of a compiler therefore becomes more important.**

In this paper a novel approach to compiler architecture that places the Interpreter as the central component of the compiler is proposed. Translation of user code into the executable form is done by an Interpreter, written in the target language. The data structures of Interpreter are accessible also to the programmer. Such solution significantly improves flexibility and extensibility of the compiler and enables execution of any code at the compile-time.

Compile Time Function Execution (CTFE) First pattern was designed for low-level, non-garbage-collected, reflection-enabled language C-=-1. This is a new programming language, which allows the programmer to execute any code at the compile time, as well as to analyze and modify the program structure. Grace to the extensibility of the designed compiler, programs written in C-=-1 can also generate marshalling bindings for other languages and support a variety of programming paradigms.

The significant flexibility and extensibility of CTFEF caused severe problems in compiler construction. The compiler appeared very complex, its parts, especially Interpreter, were very difficult to debug. Compiler operates on inflexible data structures, accessible also for a programmer. They are therefore a part of the compiled languages standard library and backwards compatibility must be maintained.

I. Introduction

COMPILE-time function execution (CTFE) is an aspect of a programming language, that allows the programmer to execute code at compile-time. It has been gaining popularity with programming languages without virtual machines such as Rust [1] or C++ [2]. The goals and capabilities of CTFE depend on the language and are discussed in details in section II.

In this paper a new approach to compiler construction that places Compile Time Function Execution capabilities as the first priority is proposed. This architecture is called CTFEF: Compile Time Function Execution First. CTFEF builds the compiler around the Interpreter component. The goal of this approach is to shorten the initial stage of compiler bootstrapping [3], [4], better support languages built around static metaprogramming and make the compiler more extensible. The role of the compiler is to construct the semantic model of the program being compiled and pass it as data to the interpreted Compiler-Interface module. Compiler-Interface then transforms the model into the assembly language to be compiled into executable. This approach was created during implementation of the compiler for C-=-1, a new language that prioritizes static metaprogramming. Using CTFEF approach causes many implementation difficulties. Data structures used within the compiler are tightly coupled with the compiled language. These problems and how they were solved in C-=-1 compiler are described in section V.

The design of that language is described in section III. Related work is briefly shown in section II. In section IV the structure of a CTFEF compiler and interactions among its components are presented.

II. Related work

Many currently used programming languages allow the programmer to execute some code at compile time, for example: C# [5], [6], Rust [1], [7] and C++ [8].

Rust language compiler is the most similar in capability, to what was demonstrated with C-=-1, as a feature of CTFEF. It allows the user to write code that performs some transformations of the program, and interact with the environment during the build process. The two relevant features are build scripts and macros. Rust macros, similar to the classic C-style preprocessor macros, generate additional code as text. The major difference between the systems is that Rust macros operate on tokens, rather than text, and use syntax similar to regular Rust code. This mechanism is powerful, allowing the user to rewrite code to avoid repetition, simplify certain tasks and create new syntax. They are however unable to reflect on the program or obtain some information available to the compiler.

Build scripts, on the other hand, are programs that execute before the compilation of the main package. They prepare the environment for building the program, for example, compile external dependencies. The structure of the program being compiled, is not available for build scripts. Build scripts accessing such data, would have to analyze the code by themselves, without any compiler assistance, and that makes automatic generation of bindings for other programming languages very difficult.

C# compiler, called Roslyn [9], has the closest set of capabilities to what CTFEF aims to achieve. Its architecture allows for user-defined analyzers and code generators using a program model generated by the compiler [9]. These compiler extensions can be used and distributed as regular code packages, for example using nuget package manager [10]. C# code analyzers have access to entire analyzed code-base and to semantic analysis functionality of the compiler. Using CTFEF for the purposes of code analysis would not offer any additional benefits.

C# code generators are more limited. They can only add new files, not modify the existing ones. All code generators operate on a read-only snapshot of the code base. This also means that if more than one generator operates during compilation, they are unaware of the files created by other generators[11]. This limitation was introduced for a number of reasons. If source generators could influence each other, the order in which they run would become important and could lead to unpredictable behavior. Compilation performance would also be affected, as parallelization of source generators would become more difficult or even impossible.

Although CTFEF was inspired by the C# compiler, there are significant differences between them. C# compiler extensions are separate, compiled, dynamic libraries loaded by the compiler and executed as a regular code. In CTFEF components that analyze and generate code are part of regular code base and are not compiled separately. The intermediate representation of these modules is interpreted at compile time, together with other parts of compiler, and operate on the same representation of user program. CTFEF is further described in section IV.

III. Design of C-=-1

C-=-1 is a new programming language, designed as low-level and non-garbage-collected and compiled to native machine code, similar to C, C++ or Rust. What differentiates C-=-1 are its two core principles: all code is executable at compile-time and support for metaprogramming. The primary purpose of this language was to investigate how these ideas influence software written in it [12].

C-=-1 is a simple language, built with minimum set of features needed to demonstrate the usefulness of the proposed metaprogramming features. They are discussed further in section III-B. The primary motivation for those mechanisms was to provide to a programmer the ability to create domain specific static analysis and code generation tools, without creating a separate program.

A. Type system

C-=-1 type system is similar to type system in C++. Program may contain user defined classes, with members which may have limited accessibility. Generic programming is achieved by templates, although they are much more limited than the ones present in C++. The user may also use pointers to objects, with arbitrary indirection (for example pointer to pointer to object). Additionally, the language contains the concept of an `interface`, similar to the one found in C# [13].

B. Attributes and metaprogramming

Metaprogramming in C-=-1 is based on attributes. Attributes work in a manner similar to the ones found in C#. They are types, which may contain fields and methods. They can also be used to annotate other elements of the program, such as types, functions or variables.

In C-=-1 these attributes may implement special member functions that react to the use of an annotated program element. For example, an attribute that can be attached to a function, may implement `onCall` special member function. It will be called at compile time, for each invocation of the annotated procedure. Within the special method the attribute will have access to the semantic model of the call site. It may then modify the semantic model or report warnings or errors.

Listing 1 contains an example of a C-=-1 attribute providing static analysis: `noDiscard`. It works in the same manner as the attribute of the same name present in C++17 [14]: the result of invoking the annotated function must be used. To declare an attribute in C-=-1, the `att` keyword is used. After that, attribute targets should be listed in angled brackets, as in line 0 of Listing 1. Valid targets for attributes include a number of language elements, such as types, functions, variables and fields. Attribute from Listing 1 declares two member functions: `attach` in line 4 and `onCall` in line 6. Listing 2 contains an example of using the `noDiscard` attribute. All uses of the `noDiscardFunction`, except for the one on line 3 are valid. Using the function in a statement as opposed to an expression causes a compile-time error.

The `attach` method is called after names of all program elements have been gathered, but before compiler starts to analyze function bodies. It is common among all attribute targets and accepts the descriptor of the attached program element (function, field, type, etc.). Attribute can change aspects of the program that affect function overload resolution, such as whether a function is invokable at run or compile time, only from within the `attach` method.

The `onCall` method is an example of a function reacting to use an annotated program element. These methods are specific to a given attribute target. Within this function, the attribute may analyze and modify the code, as well as raise errors or warnings.

Listing 1: noDiscard attribute in C-=-1

```
public att<function> NoDiscard                              0
{                                                           1
        public fn attach(f: functionDescriptor)             2
        {}                                                  3
        public fn onCall(call:                              4
            functionCallExpression*)
        {                                                   5
                if(call._parentStatment != null<            6
                    IInstruction>())
                        raiseError(                         7
                        &(call._pointerToSource),           8
                        "Return value of a no-              9
                            discard function is not
                            used",
                        123                                 10
                        );                                  11
}                                                           12
}                                                           13
```

Lines 6 to 11 of Listing 1 are an example of a C-=-1 attribute providing static analysis. The `onCall` method checks whether the attached function is invoked in an expression or instruction context. Calling a function as a statement means that the result of that invocation is discarded by the caller. This may indicate an error, when the function has no other side effects. If that is the case, the attribute calls the `raiseError` function, which is provided as a compiler intrinsic, that generates a compilation error. The example presented in Listing 1, although very basic, demonstrates the ability to implement a form of static analysis that typically requires modifying the compiler or creating an external tool.

Listing 2: Example of using noDiscard attribute from Listing 1

```
[noDiscard()]
fn noDiscardFunction() -> usize;
fn main() -> usize {
      noDiscardFunction();
      // error 123: Return value of
      // a no-discard function is not used
      let x = noDiscardFunction();     // ok
      let y = x + noDiscardFunction(); // ok
      return noDiscardFunction();      // ok
}
```

IV. Design of the compiler

CTFEF approach was created during implementation of the first compiler for the C-=-1 language[12]. CTFEF compiler has four major components:

1) Frontend.
2) Interpreter.
3) Compiler Interface.
4) Backend.

Figure 1 contains a diagram with an overview of how these parts interact with each other, during the compilation process. Frontend, described in section IV-A, parses the code in the compiled language and constructs its intermediate representation, using Interpreter's data structures. The structure and form of the intermediate representation is not specified by the CTFEF approach, as long as it contains all semantically relevant information from the source program. It is used to analyze both user code and the Compiler Interface. After the intermediate representation is constructed, it is passed to the Interpreter, which is described in section IV-B. Compiler Interface intermediate representation is then executed, using the user program as data. This step converts the semantic model of the program into the Backends intermediate language. This process is further explained in section IV-C. Finally, the Backend generates the executable file.

A. Frontend

In the CTFEF approach, Frontend serves the same role of constructing the programs intermediate representation, as in conventional compilers [3]. The major difference lays in the data structures used to describe the program. For a CTFEF compiler, they must be accessible to the program running within the Interpreter. This may make Frontend more complex. The additional challenge of representing a user program, using Interpreter data structures, depends on the design of the Interpreter.

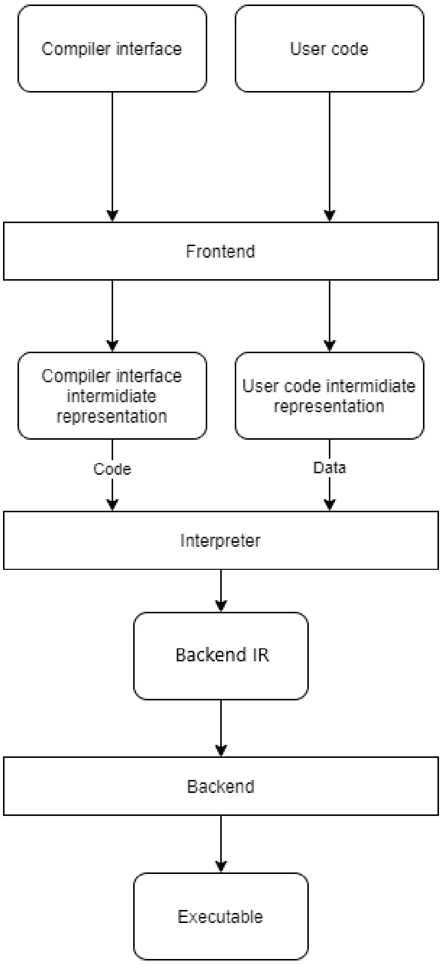

Compiler interface
User code
Frontend
Compiler interface intermidiate representation
User code intermidiate representation
Code
Data
Interpreter
Backend IR
Backend
Executable

Fig. 1: CTFEF compiler structure

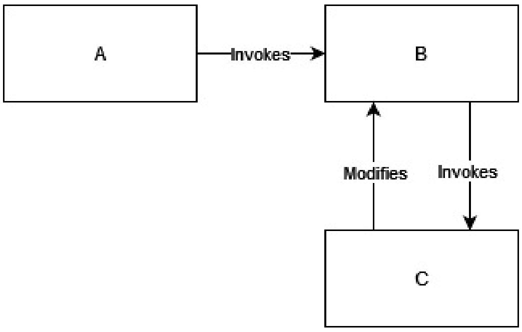


Fig. 2: Example of circular reference in meta code

B. Interpreter

In CTFEF, Interpreter is main component of the compiler. It executes the Compiler-Interface which translates the intermediate representation into the Backend's assembly and serves as what is sometimes called the 'middle-end' of the compiler[15]. To do it, it must be able to treat the program's intermediate representation both as code and data.

An important issue for CTFEF compiler is decision what code can be executed at compile-time. In some languages it may be possible to introduce circular references between the functions that modify the codebase. Figure 2 contains a diagram of such scenario. `A`, `B` and `C` are functions. If function `B` is modified by function `C` and `B` invokes `C`, the behavior of function `A` is unpredictable. This problem will only be magnified by larger program sizes.

One of possible solutions, to the above-mentioned problem, is to restrict which functions can be invoked at compile time. C-=-1 allows code within a compile time context to invoke procedures only from other packages, declared explicitly as dependencies. Circular references between packages, as in most other languages, are forbidden. C-=-1 additionally prohibits modification of dependencies. Therefore, it is impossible for a function to modify a procedure, it depends on.

C. Compiler Interface

Compiler Interface translates the program's intermediate representation into the Backend's assembly language. This component is interpreted during compilation and may be provided by the user. In case of C-=-1, compiler interface could be supplied to the compiler, the same way that the code being compiled is passed, as a collection of source files. Figure IV reflects this decision, treating the Compiler Interface as an input into the compiler, same as user code.

What is unique about CTFEF is that this part of the compiler can be written in the target language, during initial bootstrapping of the compiler bootstrapping process, i.e. initial implementation using another language [3], [4]. In case of the C-=-1 compiler, C++ was used to implement Frontend, Backend and Interpreter, with Compiler Interface written in C-=-1 [12].

Compiler Interface contains a function marked as the Compiler Interface Entry-point. That procedure must accept a set of modules to be compiled and a Compilation Context that is used to generate the Backends assembly. The module descriptors that are passed to the Compiler Interface are built by the Frontend, as can be seen in Figure 1.

After the Compiler Interface finishes generating Backend assembly, the Compiler Backend is invoked to generate the binary executable.

D. Backend

CTFEF does not put any additional requirements on compiler Backend. When using this approach, a generic Backend library can be used. C-=-1 compiler used LLVM[16] as its backend.

The Backend code must be invokable from within the interpreted program in the target language. Depending on how the Interpreter was designed, this may require significant effort. Compiler Backends are large and for the Compiler Interface to take advantage of them, their entire interface must be fully available in the interpreted context. This means exposing each function and type within the library to the interpreted code, by duplicating their signatures. These bindings could feasibly be generated automatically [17], but this technique was not used when implementing C-=-1 compiler.

V. Implementation

The first CTFEF was created for a new language: C-=-1, using generic parser generator and Backend. The most important aspect of implementing a CTFEF compiler is the design of the data structures, described in section V-A, that will be used by the Interpreter.

In order to exploit CTFEF approach in design of a compiler the language should contain a set of data structures to describe user code i.e. a Semantic Model. It will allow the programmer to interact and manipulate the structure of the program at compile-time. The Semantic model designed and implemented for C-=-1 is described in section V-B.

The final part of a CTFEF compiler is the Backend Interface. It is a program, written in the target language, and executed at compile-time that translates the semantic model into the Backends' assembly language. Backend Interface implemented for C-=-1 is relatively small and is described in section V-C.

A. Interpreter data structures

Data structures of the C-=-1 Interpreter have been designed having the ease of development and debugging in mind. They are thus not particularly efficient.

Figure 3 contains a class diagram of most of the types used to represent values within C-=-1. All of them derive from `IRuntimeValue` and are managed via C++ smart pointers. The interface of the base class allows the value to be converted to a human-readable format, serialization, deserialization and copying.

The most primitive types within the hierarchy are `StringValue` and `IntegerValue`. They are simple wrappers for strings and integers, present in host language. Floating point numbers were not implemented as they were not necessary for implementation of a basic compiler.

User-defined types are represented using `ObjectValue`. The contents of an object is kept as a `string - IRuntimeValue` dictionary, with field names as keys and `uniqe_ptr< IRuntimeValue>` as values. C-=-1 object is therefore spread out in memory, even if the fields are directly contained within the class, without any indirection.

There are several types of reference within the C-=-1 Interpreter. The most basic pointer type is a reference to C-=-1 value. It was realized as a pointer to the owning pointer of the value.

B. Program semantic model

A major motivation for creating CTFEF was the ability to support languages with compile-time metaprogramming. This includes reflection and modification of the code being compiled. User program has to be represented as a complete and modifiable object, using the Interpreters' data structures.

C-=-1 language model divides the user program into assemblies. They represent an individual program package: a library or an executable file. The compiler is invoked to compile an assembly, together with its dependencies. Assembly is the root object of C-=-1 program model, it stores a list of assemblies it depends on and the root namespace of the package it

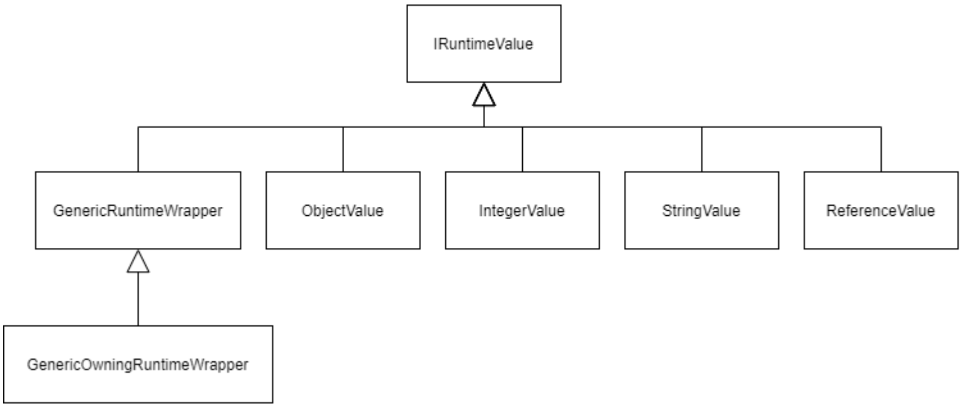


Fig. 3: Class diagram of C-=-1 Interpreter data structures

represents. The remainder of user code is organized into namespaces, types, functions and fields. These parts of the model are represented by native classes of the host language and form the basis for the rest of the model.

The most complex part of the model is the representation of the function body. Like in most programming languages, a C-=-1 function can contain a variety of instruction types. That includes complex statements and blocks of statements that can be arbitrarily nested. Each instruction may also contain expressions of any complexity.

To deal with this complexity, C-=-1 semantic model for functions is build around two interfaces: `IInstruction` and `IExpression` and their concrete implementations. Every category of instruction or expression is represented by its own type. The user may then analyze the structure of the program, using a dynamic type conversion mechanism similar to C++ `dynamic_cast` [2].

All elements of the semantic model, have a `sourcePointer`. It is a simple structure, that contains the filename and the line number of the expression or instruction. This information can be passed to compiler intrinsic functions, such as `raiseError` , to generate error messages for the user. Listing 1 contains an example of this functionality. The `pointerToSource` makes the messages generated by the compiler easier to understand for the programmer.

Component responsible for creating the semantic model is a major part of the compiler. There are two operations that this module performs: building the definition of the types used to describe a program, and creating an instance of the model, given semantic information. C-=-1 compiler has a hard-coded definition of its base library. It contains definitions of primitive types and types used to build the semantic model of a program. The description of this library must be built manually, as it is very closely integrated with the compiler.

C. Backend interface

The C-=-1 Backend interface uses LLVM [18] code generation API that has been exposed by the compiler. The functionality which has been made available to C-=-1 represents a minimal subset of LLVM, that is sufficient to implement a basic compiler.

Besides translating user code, the Backend Interface must also generate the assembly for certain intrinsic operations. Functions such as integer arithmetic operators, array indexers or memory allocators are concepts too low-level to be expressed in C-=-1. They are therefore expressed as functions, without bodies, which are then replaced by appropriate intrinsic operations.

Listing 3 contains a simple function written in C-=-1 (line 2), its ideal LLVMIR (line 5) and LLVMIR generated by C-=-1 compiler (line 21). The current implementation generates a function for each operator, regardless of whether it was defined by the programmer or is a primitive operation. They are merely wrappers around the actual LLVM intrinsic, meant to simplify implementation of the Backend Interface. Future versions, with additional effort, may generate the ideal LLVMIR from line 21 of Listing 3.

Certain other intrinsic operations are defined using external dependencies. C-=-1 memory management library, in the runtime context, uses a simple interface capable of allocating and deleting a continuous buffer. It consists of two functions: `unsafe_new` and `delete`. In the standard library, they are explicitly mapped to `malloc` and `free` functions from the C runtime.

Backend interface must also allow the programmer to influence how the executable code is generated. There are many practical reasons for this capability. Specifying the name of a function in order to link it to an external symbol is one of them. For example, C-=-1 standard library uses `malloc` and `free` to manage memory. These symbols are imported as `unsafe_new` and `delete` in excerpt in Listing 4. This is accomplished using the `mapToExternalSymbol` attribute and specifying the symbol name as a parameter, as was done in lines zero and three.

One of possible ways of achieving this, is to declare an interface for an attribute generating a functions symbol name. Listing 5 contains relevant code of a Backend Interface that uses such an attribute to override mark external symbols. Interface `ISymbolNameOverride` contains only one method:

`createSymbolName` that returns the name of the symbol in the generated assembly.

Listing 3: Representation of average function in C-=-1 and LLVM IR

```
// C-=-1 function in source code
fn average(a: usize, b: usize, c: usize) -> usize {
  return (a + b + c) / 3;
}
// Ideal representation in LLVMIR
define i32 @average(i32 %0, i32 %1, i32 %2){
  %4 = add i32 %1, %0
  %5 = add i32 %4, %2
  %6 = sdiv i32 %5, 3
  ret i32 %6
}
// Generated LLVM IR
define i32 @__operator___plus____usize__usize(i32
    %0, i32 %1){
  %3 = add i32 %1, %2
  ret i32 %3
}
define i32 @__operator___div____usize__usize(i32 %0,
    i32 %1){
  %3 = sdiv i32 %1, %2
  ret i32 %3
}
define i32 @average(i32 %0, i32 %1, i32 %2){
  %4 = call __operator___plus____usize__usize(i32
      %1, i32 %0)
  %5 = call __operator___plus____usize__usize(i32
      %4, i32 %2)
  %6 = call __operator___div____usize__usize (i32
      %5, i32 3)
  ret i32 %6
}
```

Functions `buildFunction` and `getFunctionName`, from Listing 5, are parts of Backend Interface. They are invoked in order to convert a C-=-1 `functionDescriptor` to a LLVMIR function. They were included in the Listing, because they are the only parts of the Backend Interface that need to interact with `ISymbolNameOverride` attributes. Both of these procedures, check whether an attribute implementing this interface is attached to the function they are currently processing. This happens on lines two and eighteen of Listing 5. If that attribute is present, code of that function is ignored, and it is treated as an external symbol: condition on line eighteen omits execution of `build_block` function on line twenty-one. Procedure `getFunctionName` contains a similar condition on line two, that decides how the name should be generated. If an `ISymbolNameOverride` attribute is present, it will be created by `createSymbolName` method of the attribute attached to the function. Otherwise, the `mangleName` function will create name of function's symbol based on its parameters and return type.

Listing 4: C-=-1 memory allocation functions from standard library

```
[mapToExternalSymbol("malloc", "")]
private fn unsafe_new(size: usize) -> char* {}

[mapToExternalSymbol("free", "")]
internal fn delete<typename T>(val: T*) {}
```

Listing 5: Part of a Backend Interface, using `ISymbolNameOverride` interface

```
private fn getFunctionName(f:                        0
    functionDescriptor) -> string {
      let attribute = f.get_attribute<               1
          ISymbolNameOverride>();
      if(attribute != null<                          2
          ISymbolNameOverride>())
        return attribute.createSymbolName();         3
      return mangleName(f);                          4
}                                                    5
private fn buildFunction(                            6
      f: functionDescriptor,                         7
      llvmF: llvmFunction,                           8
      registry: packageRegistry*,                    9
      mod: llvmModule)                               10
{                                                    11
      let variables = dictionary<                    12
          variableDescriptor, llvmValue>();
      let params = f.parameters();                   13
      for(i in enumerate(0, params.length()))        14
        variables.push(params[i], llvmF.             15
            getParameter(i));
      let builder = llvmF.getBuilder();              16
      let attribute = f.get_attribute<               17
          ISymbolNameOverride>();
      if(attribute == null<                          18
          ISymbolNameOverride>())
      {                                              19
        let code = f.code();                         20
        build_block(&code, &builder, &               21
            variables, registry);
      }                                              22
}                                                    23
```

VI. Conclusions

CTFEF is a new approach to compiler construction which offers high degree of compiler extensibility, at the cost of development time and performance, compared with conventional compilers. It places the interpreter as the main component of the compiler and focuses on executing user code at compile time. The user code has access to the same information as the compiler at compile time and may perform analysis or transformation of the compiled program. The thesis that introduced this approach [19] demonstrated numerous practical application: generating bindings for other languages, static analysis and extending semantics of the language. These goals are significantly easier to accomplish, thanks to the access to semantic model of the program, constructed by the compiler. Authors of language tools do not need to analyze the user program.

On the other hand, using CTFEF approach has some drawbacks. Implementing a compiler is more difficult. Since a significant part of the compiler is interpreted, the initial implementation requires working with the target language. The lack of available tools, such as integrated development environments, debuggers and libraries, for this new language, makes this part of the process significantly more difficult. On the other hand also means the work on the compiler in the target language may start at the very beginning of compiler bootstrapping process[3], [4].

The compiler created for C-=-1 has unacceptable performance, as noted by the original C-=-1 paper [12]. Compiling the C-=-1 standard library, containing around 200 lines of code, takes approximately 10 minutes. Majority of that time is spent on interpreting the compiler interface. This is a significant barrier to adopting CTFEF approach. Since the compiler implemented for C-=-1 was made as a research tool with minimal effort, further work is needed to explore the performance issues of CTFEF approach.

References

[1] N. D. Matsakis and F. S. Klock, "The rust language," *ACM SIGAda Ada Letters*, vol. 34, no. 3, pp. 103–104, 2014.

[2] "Programming languages — C++," International Organization for Standardization, Geneva, CH, Standard, Mar. 1998. [Online]. Available: https://www.iso.org/standard/25845.html

[3] A. Puntambekar, *COMPILER DESIGN*. Technical Publications, 2011.

[4] D. Novillo, "Gcc internals," in *International Symposium on Code Generation and Optimization (CGO), San Jose, California*, 2007.

[5] Source generators. [Online]. Available: https://docs.microsoft.com/en-us/dotnet/csharp/roslyn-sdk/source-generators-overview

[6] Dotnet. Roslyn. [Online]. Available: github.com/dotnet/roslyn

[7] S. Klabnik and C. Nichols, *The Rust Programming Language (Covers Rust 2018)*. No Starch Press, 2019.

[8] "Programming languages — C++," International Organization for Standardization, Geneva, CH, Standard, Mar. 2020. [Online]. Available: https://www.iso.org/standard/79358.html

[9] N. Vermeir, ".net compiler platform," in *Introducing .NET 6*. Springer, 2022, pp. 275–295.

[10] M. Balliauw and X. Decoster, "Nuget package manager console powershell reference," in *Pro NuGet*. Springer, 2013, pp. 331–338.

[11] B. O. SLIMÁK and R. J. Pelikán, "Source generators in c#," Master's thesis, Department of Computer Systems and Communications, Masaryk University, 2022.

[12] A. Grabski, "Compilation through interpretation: static metaprogramming in c-=-1," Master's thesis, Warsaw University of Technology, 2022.

[13] A. Hejlsberg, S. Wiltamuth, and P. Golde, *C# language specification*. Addison-Wesley Longman Publishing Co., Inc., 2003.

[14] "Programming languages — C++," International Organization for Standardization, Geneva, CH, Standard, Mar. 2017. [Online]. Available: https://www.iso.org/standard/68564.html

[15] M. Hsu, *LLVM Techniques, Tips, and Best Practices Clang and Middle-End Libraries: Design powerful and reliable compilers using the latest libraries and tools from LLVM*. Packt Publishing, 2021.

[16] C. Lattner, "Llvm and clang: Next generation compiler technology," in *The BSD conference*, vol. 5, 2008.

[17] P. Dietz, T. Weigert, and F. Weil, "Formal techniques for automatically generating marshalling code from high-level specifications," in *Proceedings. 2nd IEEE Workshop on Industrial Strength Formal Specification Techniques*, 1998, pp. 40–47.

[18] J. Zhao, S. Nagarakatte, M. M. Martin, and S. Zdancewic, "Formalizing the llvm intermediate representation for verified program transformations," in *Proceedings of the 39th Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, ser. POPL '12. New York, NY, USA: Association for Computing Machinery, 2012, p. 427–440. [Online]. Available: https://doi.org/10.1145/2103656.2103709

[19] A. Grabski, "Rose parser generator," Bachelor's thesis, Wydział Elektroniki i Technik Informacyjnych, Politechnika Warszawska, 2020.

Communication Papers of the 18th Conference on Computer Science and Intelligence Systems pp. 111–120

DOI: 10.15439/2023F128
ISSN 2300-5963 ACSIS, Vol. 37

Waiting list procedure improvements for master program courses in Information and Computing Sciences

Tomas van Groningen and Sietse Overbeek
Department of Information and Computing Sciences, Faculty of Science, Utrecht University,
Princetonplein 5, 3584 CC Utrecht, the Netherlands
Email: T.J.vanGroningen@students.uu.nl, S.J.Overbeek@uu.nl

***Abstract*—In higher education, at times it happens there are limited places in courses because of, for example, staffing and classroom shortages which can lead to students being wait-listed. Previous research indicates there are numerous waiting list prioritization methods in health care and public housing, whereas research in waiting list prioritization methods for course registration in higher education is very limited. Results of a literature study and interviews with domain experts have been conducted and analyzed to determine how course waiting list procedures can be improved. This has resulted in an improved waiting list procedure including prioritization methods for master program courses in Information and Computing Sciences at Utrecht University, the Netherlands.**

I. Introduction

THE huge increase in jobs in the IT sector that the digital economy has offered, as well as our increasing dependence of computing skills and data, has caused a significant increase in the enrollment of students in undergraduate computer science (CS) related courses and programs across North America [1]. This is partly due to the requirement of at least a bachelor's degree for entry-level jobs within the IT sector [2]. Course enrollment increases are not only caused by the growing number of CS major students, but also due to a substantial rise of non-CS students that want to participate in CS courses [3]. The growth in the number of undergraduate computer science majors has not been matched by an increase in the number of tenure-track or teaching staff [1]. Consequently, teachers have to teach larger classes and more classes are being taught by temporary instructors such as visitors and graduate students. This is partly due to the outflow of graduated CS students into industry rather than staying in academia to teach, as they can earn up to twice as much in industry compared to what professors earn [4]. This causes a snowball effect: the shortage of CS teachers makes it more difficult for CS majors to get into and finish the classes they need to graduate. These teaching and teaching staff shortages, as well as the lack of sufficient classrooms are some of the causes of restricting certain CS courses to a limited number of students. Because the demand for courses sometimes exceeds the course capacity, waiting lists to select which students will be enrolled for the course are used. In this research a case at Utrecht University in the Netherlands will be used and the current handling process of these waiting lists contains multiple registration and prioritization challenges. This research aims to map the current process based on an already available textual description and improvements will be proposed for an optimized waiting list procedure. These improvements will be based on published literature, as well as interviews with domain experts.

This paper is organized into ten sections. In the following section, the design and structure of the research will be explained. In Section III, a motivating scenario will be described, which includes background information as well as the current situation on the procedure of handling waiting lists in the case of Utrecht University. In Section IV, the results of the literature study will be presented. Interviews were held with education coordinators of different study programs within Utrecht University and the results are described in Section V. Section VI triangulates the results from both the literature study and the interviews. Based on these results, an improved waiting list procedure will be presented in Section VII. This improved procedure will be validated with an education coordinator of Utrecht University in Section VIII, to see what final improvements can be made to the procedure. The final results after the validation are discussed in Section IV. The final section includes concluding remarks, research limitations, and avenues for future research.

II. Research design

The research that will be performed in this study is divided into three parts. Firstly, a literature study is performed to find out what is already known about waiting lists, also transcending the domain of higher education. To complement this literature with real life waiting list practices in the domain of higher education, interviews with domain experts are conducted. These experts are education coordinators of different study programs within Utrecht University, who also coordinate the handling of waiting lists for courses. The results are triangulated to find out what the most common and most useful practices are with regard to waiting list handling. Based on these results, an improved waiting list procedure will be created: a textual description complemented with a Business Process Model and Notation (BPMN) model [5]. In the last phase of the research, the proposed improvements will be

Thematic track: Software Engineering for Cyber-Physical Systems

reviewed by another domain expert, to gain more insights and find out strengths and weaknesses of the proposed procedure.

A. Research questions

The aim of this study is to find out if and how waiting list procedures for master program courses in Information and Computing Sciences can be improved over current procedures, so that the chances of participating are as fair as possible for every student. The waiting list procedure for master program courses at Utrecht University are used as a case. The following research question and related sub questions have been conducted:

How can waiting list procedures for Information and Computing Science master program courses be improved to make sure that the chances of participating are as fair as possible for every student?

SRQ 1: How are waiting lists of I&CS master courses currently handled? Before the start of the research, the current situation regarding the waiting lists in I&CS master courses at Utrecht University is investigated in Section III. **SRQ 2:** How are waiting lists handled within other departments at Utrecht University? To answer this question, interviews will be conducted with university staff that handle waiting lists for courses, on which more can be read in Section II-C. **SRQ 3:** How are waiting lists handled at other educational institutes and in other fields? To answer this question, a literature study will be performed, on which more can be read in Section II-B.

B. Literature study design

To find out how other institutes in higher education deal with waiting lists, a literature study is performed. Initially, very limited published literature was found on waiting lists and procedures in education. There is, however, information about waiting lists available university websites, in blogs and news articles. To make sure this information is included in the study, a Multivocal Literature Review (MLR) is performed [6]. A MLR is a type of Systematic Literature Review (SLR) [7] that includes both published literature and grey literature (GL). GL is mostly defined as *"literature that is not formally published in sources such as books or journal articles"*[8] and *"literature that is not controlled by commercial publishers, i.e., where publishing is not the primary activity of the producing body"* [9]. The inclusion of GL in the literature review is supposed to fill in the gaps of published literature by providing other perspectives. For this literature study, multiple search engines have been used to gather scientific as well as GL. For scientific literature, Web of Science and Google Scholar were used. For GL, Google was used. An overview of used search terms and key words per search engine is shown in Table 1.

TABLE I
USED SEARCH TERMS AND KEYWORDS PER SEARCH ENGINE.

Web of Science	**Google Scholar**	**Google**
Waiting lists for courses	Waiting lists for courses	Waiting lists for courses
Waitlisted courses	Waitlisted courses	Waitlisted courses
Waiting list procedure	Waiting list procedure	Information science waitlist course
Waiting list, university	Waiting list, university	
Waiting list prioritization	Waiting list prioritization	

C. Expert opinion interviews

In addition to the literature study, interviews with experts will be conducted. In contrast to mass surveys, experts are typically more informed and motivated than the average participant of a questionnaire [10]. The validity of the information gathered through expert interviews is heavily dependent on the expertise of individuals being interviewed. Because this expertise is so important, the experts for this research are education coordinators of bachelor and master programs within Utrecht University. The interviewed coordinators have been selected because they coordinate programs that have a large number of incoming students every year, which makes waiting list issues likely. A summary of the results of the interviews can be found in Section V, while the entire interview text can be found in Appendix A. In the context of ethics and privacy, an Ethics and Privacy Quick Scan was conducted (see Appendix B). Based on the scan, this research has been classified as low risk with no further ethics or privacy assessment required. All participants were asked to fill out a consent form that asks their permission to participate in the research and informs them about it.

D. Validation

To validate the effectiveness of a treatment, it must be demonstrated that it would achieve the desired outcomes for stakeholders when applied to the specific problem at hand [11]. The requirements for the treatment must be clearly defined and justified, and the treatment must meet these requirements to be considered validated. When both the literature study and expert opinion interviews have been finalized, an improved waiting list procedure will be created based on several requirements, that will be drawn up based on the results from the literature study and interviews as well as aspects of the current procedure. In Section VII, this improved procedure, including requirements for the treatment, will be presented. The procedure will then be presented to the education coordinator of the master programs within the I&CS department of Utrecht University. This coordinator reviews the improved procedure and provides feedback, based on which the final waiting list procedure will be presented in Section VIII.

III. MOTIVATING SCENARIO

As mentioned earlier, the research is based on the case of the I&CS department at Utrecht University. The problem with waiting lists in this situation has to do with different enrollment periods for master courses and incoming students. In this section, the context of having a maximum number of places for a course and different course registration periods will be explained.

A. Broader context

Waiting lists are sometimes necessary for students before they can participate in a course. This limitation can be there because of teaching staff and room capacity. Schedules for an academic year that start in September are made in January of that same year. Program directors have to provide schedulers with the number of students they expect to take part in courses for that particular year. When the enrollment for a course starts, it is possible that the demand is higher than the supply, meaning that all students that have signed up beyond the maximum capacity of the course will be placed on a waiting list. The course coordinator can ask for lecture rooms with bigger capacity, but because schedules have been made already, it is not likely that this is possible. Alternatives are teaching the course in hybrid mode or dividing the students over multiple rooms on campus, but then there is still the problem of hall availability, as well as needing more staff to guide the streaming of the lecture to multiple rooms. Another possibility is changing the design of the course to allow more students. This can range from adjusting group projects to complex implementations like using IT to automate parts of a course. One of these methods is blended learning: ways of enhancing traditional teaching with digital methods like additional online material or even building a completely new course with a blended learning concept from scratch [12]. This method could facilitate more students to take part in courses, as it allows for hybrid teaching and learning. A reason why there could be a maximum capacity of a course is because of a limited number of teachers and teaching assistants (TAs). At the I&CS department, courses in master programs generally do not have TAs, because a TA for a master course must have followed the course already and a master program is only one or two years. On top of that, many of those are also TAs for advanced-level bachelor courses. The result is that remaining TAs may have limited effect on the overall teaching staff workload reduction for master programs and may have an impact on the number of students that can be accommodated.

B. Types of courses

A master course can be of a different level for each student. There are three types of courses, the first one being a **mandatory course**: a course that the student of a certain master's program must follow. This means that when a student signs up for this course, they will *always* be enrolled. Secondly, there are **primary electives**. These are courses of which a student has to follow a certain amount. Finally, there are **secondary electives**: several courses that you can freely choose from any master program. Students on a waiting list for such a course will only be enrolled if there are seats left after having handled all other waitlisted students.

C. Course registration periods

The first course registration period for a course is the first time the enrollment for a course is opened, which means that the maximum capacity of the course is available for registration. In an academic year, there are four teaching periods (or blocks). For the 2022/2023 academic year, the course registrations are opened for the dates that can be seen in Figure 1 in Appendix D [13].

After the initial course registration period, there is another possibility to enroll for a course: the post (course) registration. This registration period is only opened for two days, usually two weeks before the start of the new period for which this post registration is meant. However, chances are high that the course is already full by then.

D. As-is situation

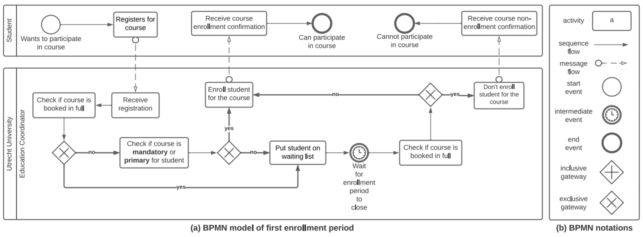
New students can start a master program in either September or February. These students are unable to sign up for courses that take place in the first period when all places are still available (initial course registration), as they do not have access to an online environment where students can sign up for courses yet. They can only sign up for courses when the deadline of the first course enrollment period has passed, which means that there is a chance that a course is already booked in full. In that case a student will be put on a waiting list. A workaround for such a student is following the course in the second year as they can sign up for the course during the first enrollment period. Giving new students priority over already enrolled students could ensure them a spot in the course they want to follow. However, if this is always done, other students will have a higher chance of not being able to follow their preferred primary elective courses. Following that, students that would like to follow a certain course as a secondary elective would be denied participation in that course, as they are then last in line in terms of priority. This is not ideal either, as there should be chances for any student to participate in a course, regardless of the course type.

The current situation regarding waiting list handling is as follows: students are placed on a waiting list if the course is full or if the course is not listed in their exam program as mandatory or primary. For the first enrollment period, only waiting lists for courses that have been booked in full are looked into. For the late enrollment period, all waiting lists are looked into. If there are waiting list candidates for whom the course is mandatory, the student will *always* be enrolled. If the course is definitely booked in full, waitlisted students that still need another course to follow in the upcoming period are offered participation in courses that have places left. The full as-is situation can be found in Appendix C.

E. Business Process Model and Notation (BPMN) model

The Business Process Model and Notation (BPMN) is a process modeling language that is used to present business processes graphically. This allows users to comprehend the activity flows, role assignments and usage of data and information [5]. In Figure 1a, the waiting list handling of the first enrollment period, as described in Section III-B, is modeled. Figure 1b shows a legend of what the different symbols and elements of the BPMN represent. Two pools, 'student' and

‘Utrecht University’, are used in Figure 1 and 2, with each pool representing a participant in a process [14]. These are used when the model includes multiple participants that are not physically connected to each other, which is the case here. As activities within a pool are treated as a separate and independent processes, sequence flow cannot be used to connect activities between pools. Therefore, message flow is be used for the communication between two pools. The procedure for the late enrollment period is modelled in the same way in Figure 2.


Student
Wants to participate in course
Registers for course
Receive course enrollment confirmation
Can participate in course
Cannot participate in course
Receive course non-enrollment confirmation
Utrecht University
Education Coordinator
Check if course is booked in full
Receive registration
Enroll student for the course
no
yes
Don't enroll student for the course
Check if course is **mandatory** or **primary** for student
yes
no
Put student on waiting list
Wait for enrollment period to close
Check if course is booked in full
yes
activity
a
sequence flow
message flow
start event
intermediate event
end event
inclusive gateway
exclusive gateway

(a) BPMN model of first enrollment period

(b) BPMN notations

Fig. 1. BPMN model of first enrollment period waiting list procedure.

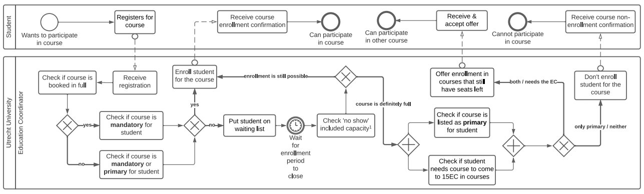

Student
Wants to participate in course
Registers for course
Receive course enrollment confirmation
Can participate in course
Can participate in other course
Receive & accept offer
Cannot participate in course
Receive course non-enrollment confirmation
Utrecht University
Education Coordinator
Check if course is booked in full
Receive registration
Enroll student for the course
enrollment is still possible
Offer enrollment in courses that still have seats left
both / needs the EC
Don't enroll student for the course
yes
yes
Check if course is **mandatory** for student
no
Put student on waiting list
Wait for enrollment period to close
Check 'no show' included capacity¹
course is definitely full
Check if course is listed as **primary** for student
only primary / neither
no
Check if course is **mandatory** or **primary** for student
Check if student needs course to come to 15EC in courses

Fig. 2. BPMN model of late enrollment period waiting list procedure.

IV. Literature study results

As explained in before, a MLR will be performed in this research. To gain insights in what has been formally published, sectors where waiting lists research has appeared a lot, like health care and public housing, will be included in this study.

A. Health care

Waiting for healthcare refers to the period of time between when a person becomes aware of a health issue and when they receive a diagnosis and treatment for it [15]. Managing and handling waiting lists is a complex problem that affects physicians, patients, healthcare systems, and governments from clinical and ethical perspectives [16]. Waiting lists have found to be the secondary most significant ethical challenge that is encountered by patients and their families in Canada [17].

The main reasons to worry about how waiting lists are handled are fairness or equity: in principle, patients who are in the greatest need for treatment should be treated first, if all else is equal. In addition, patients with the same level of urgency should have to wait the same amount of time before getting treated [18], [19]. Many waiting lists lack the implementation of this fairness, equity, and organization, as they have been constructed under time pressure with little forethought [20]. Formal policies and procedures for waiting lists should be developed and defined by health care organizations to maximize their organization and fairness. Five key concepts that should be considered during the assessment of patients on waiting lists are considered to be severity, urgency, relative priority, need, and expected benefit [19]. Severity refers to three aspects: the level of suffering, limitations in performing daily activities and the risk of passing away early. Urgency indicates the need for immediate clinical intervention or surgery. Urgency and priority seem strongly related: when a patient’s situation is very urgent, he or she should be prioritized over a patient whose situation is not as urgent. The term need is often a point of discussion, as sometimes it is seen as equal to severity while others define it more like urgency. According to guidelines from the Victorian Government Department of Human Services in Australia, patients should be placed in different urgency categories based on their clinical need [21]. The patients that are in the most urgent need for surgery are placed in a higher category than others, giving them higher priority. Expected benefit refers to the degree of benefit a patient would experience from surgery, and the probability of that benefit occurring. These two factors can differ from patient to patient, e.g., the benefit of surgery can be large for a certain patient while the likelihood of occurrence is low, and vice versa. There are two types of benefits: the elongation of a patient’s lifespan and improvements in a patient’s life quality, where the latter is often seen as most important. To make waiting lists fair for every patient, several prioritization methods have been created. A commonly used prioritization method is ranking patients that need elective surgery by the urgency of the needed treatment based on clinical and social criteria [22]. Patients are put in a group that represents their medical situation, which is assigned to a maximum time for the surgery those patients should have to wait [23]. These medical conditions vary from emergency (needs immediate revascularization) to marked delay (3 to 6 months waiting time). Another method to rank patients on a waiting list is a scoring system that assigns a ‘priority score’ to each patient [24]. Patients will be ordered on a waiting list based on their priority score, which is based on a patient’s urgency of need. This priority score will be added up to an accumulating waiting list score each week, which allow waitlisted patients to climb up.

Within the literature, many articles explicitly define certain priority criteria. These criteria are based on the severity of the disease and consequently the urgency of treatment, while social factors are often considered as well [25]. Priority criteria are criteria on which a patient is ranked before being put on a waiting list. This ranking will then define their level of urgency and thus their place on the waiting list. They have two essential functions: they are used to prioritize and schedule surgeries for patients on the waiting list, as well as to create profiles of the patients on the waiting list based on their specific needs and characteristics. A priority criteria form for hip and knee replacement can be seen in Appendix D, Table I [26]. The higher level a patient scores for each criterion, the higher their score will be for the level and thus for their overall score. The panel members that tested this method concurred that the criteria accurately reflected the way surgeons perceive priority and urgency of patients that are in need for hip or knee replacement. Similar criteria and urgency levels have been developed for general surgery in Western Canada [25]. After a series of testing, a set of final priority criteria was drawn up (see: Table II in Appendix D). These criteria were deemed logically valid and easy to use. The use of clinical priority criteria is supported by doctors from western Canadian provinces, and they think they are aligned with worldwide expert opinions on the urgency of surgery. In Australia, at least up until 2010, patients that were in need for surgery were placed in one of three clinical urgency categories which have been established nationally [21]. These priority criteria categories can be found in Table III in Appendix D. This system makes for a simple form of prioritization, but lacks specific guidelines for surgeons to make decisions on a patient's clinical need for treatment as it does not consider many other (social) factors when deciding on the urgency. In Italian National Health Service, recommended maximum waiting times for patients based on the urgency level of their needed treatment have been formalized [22]. Five urgency-related groups have been created, all with a maximum number of days a patient should have to wait before receiving surgery (see: Table IV in Appendix D).

B. Public housing

Most states in Australia use a waiting list with several categories of housing need to prioritize applicants on waiting lists [27]. To even be eligible to sign up for public housing, an applicant must fulfill certain criteria, like living in the state or area of applying, owning an Australian passport and having a significantly low income. When a participant fulfills all of these criteria, they will be put on a waiting list based on the category they are put in. Each state operates their own form of ranking applications, which can be seen in Table V in Appendix D [28]. These criteria consider both medical and social factors. More general priority criteria for Australian public housing have been drawn up as well. Some common reasons for someone to be prioritized over others are domestic violence, disability, and homelessness [29]. Another method to rank patients on a waiting list is the priority points system. With this system, applicants receive points based on their level of need for housing. Both the level of need and the number of points can be reconsidered as long as an applicant is on the waiting list [29]. Points are often awarded for current housing conditions, disability, family size, and medical need.

C. Educational institutions

The Universites of North Georgia and University of Denver state that their waiting lists for (most) courses are handled with a first come, first served function [30], [31]. The only criteria that students must meet to get on a waiting list, in most cases, are not having schedule conflicts, fulfilling course entry requirements and not already taking the maximum number of hours. The University of Auckland may select the order of students on waiting lists for courses on academic merits, for example a GPA requirement [32], while the Washington College of Law waiting lists take priority to make the waitlist process as fair as possible [33]. Due to high demand for courses, the Information Science (IS) department from Cornell University has set up priority criteria for course enrollment: IS majors come first, followed by IS major applicants, then IS minors, fourth come seniors in others fields and finally all other students, where seniors are prioritized over freshmen [34]. Boston University also gives preference to seniors over undergraduate students, but only after giving priority to all College of Arts & Sciences and Graduate School of Arts & Sciences students [35].

V. Expert opinions

For the expert opinions, multiple semi-structured interviews were performed with education coordinators of different departments and study programs within Utrecht University. The interviews were conducted with the education coordinators of the Biology and Pharmacy bachelor programs and the education coordinators of the Graduate School of Life Sciences.

All of the interviewed experts pointed out that they use a different approach to the 'regular' waiting list procedure. Rather than filling up the course capacity before placing students on a waiting list, the course capacity is set to zero in the course registration system, which means that every student that signs up for a course is placed on a waiting list automatically. After the end of the initial course registration period, the waiting list is closed and a spreadsheet with data about students is generated. When the course capacity allows all students that registered for the course to participate, all students are placed for the course. When the number of registrations exceeds the course capacity, the education coordinator or teacher has to decide which students will be enrolled for the course. Criteria have been set up to make this process as fair as possible. A frequently occurring problem all of the interviewees mentioned is that students often register themselves for multiple courses in the same timeslot, hoping they will be registered for at least one of those courses. By doing that, they essentially take an additional spot on the waiting list for a course. The Biology bachelor program came up with an additional method to overcome this problem. A student can sign up for only one course per timeslot but can name a second and third choice course including motivation. When they cannot be placed for their first choice course, the education coordinator will check if the student can be registered for a course of their second or third choice, which is almost always possible. In the rare case that the student cannot be registered for either of the three preferred choices, the student will have to choose another course during the post registration period. All interviewed education coordinators also mentioned that they do not offer elective courses in the first teaching period in the first period of the first year of the program. Only compulsory courses are offered in this period, for which (new) students are signed up by the university.

For both the Biology and Pharmacy bachelor programs, a list of priority criteria is used to decide which students will be enrolled for a course when it happens to be overbooked [36].

The criteria for Pharmacy bachelor courses can be found in Table VI in Appendix D. Priority criteria 7 and 8 are only used in case of great urgency. When a course within the Biology bachelor program happens to be overbooked, all registered students will be filtered through several criteria to narrow the registrations down to the maximum capacity the course allows. The so-called hard criteria are handled first as seen in Table VII in Appendix D, and if after that the number of students still exceeds the course capacity, the remaining criteria from Table VIII will be looked into [37]. Students that fulfill the most remaining criteria will in this case be placed for the course. This means that there is a possibility for a student that they will not be enrolled for the course, regardless of fulfilling the hard criteria, because of the maximum course capacity.

VI. Triangulation of results

A very common factor across all of the published literature is the importance of the urgency, severity or need when prioritizing someone, rather than waiting time. Generally, the time spent, position on a waiting list and even order of registration for a waiting list are not often used. Prioritizing people based on several factors that have nothing to do with time is considered to be a fair method to deal with the allocation of people to waiting lists. Many universities in the US of which data was available handle their waiting lists for courses on a first come, first served basis. This is a rather simple method compared to the methods found in the other literature and the extensive procedures and filtering criteria to filter out students from a waiting list. This importance of the urgency, severity or need when prioritizing someone for medical elective surgery or for public housing is captured in priority criteria. These are criteria, often ranked from most important to least important, on which someone is prioritized according to what they score for each criterion. For good procedures, these criteria should be clearly defined, ranked and sometimes even categorized. When this is done, procedures are transparent which allows precise insights for everyone on the waiting list on which place they have been put and on the basis of what criteria this place has been decided. In some waiting list prioritization methods, the importance of time is considered. In some cases, patients that have been ranked at the hand of several priority criteria will be placed in one of multiple categories, which are connected to a recommended maximum waiting time before surgery for a patient. In a few cases, time spent on a waiting list is considered as important, and was taken into account while creating certain priority formulae. This means that someone who has spent a long time waiting already will get a slightly higher priority score than someone who has not been waiting for surgery or treatment as long. Priority criteria in the literature on waiting lists in the medical domain go further than just deciding who should be prioritized according to the severity and urgency for needed surgery. The criteria often also consider the impact the disease or untreated condition has on the patient's social factors such as their ability to perform their daily activities, work and live independently. The same goes for the literature on public housing waiting list prioritization. Here, social factors like a person's medical condition and their age are considered when deciding someone's place on a waiting list. This inclusion of social factors are not part of the priority criteria for courses as mentioned by the interviewees.

VII. Optimized waiting list procedure

To construct an improved waiting list procedure for the I&CS department, a list of requirements for this improved procedure will be set up. A requirement is defined as a goal for the treatment that is going to be designed [11].

A. *Defining requirements*

A requirement is a desired characteristic or objective for a treatment that is being developed [11]. The treatment that is being designed here is the optimized waiting list procedure for master program courses in I&CS in the case of Utrecht University. Based on the problems that are faced currently, and all of the information that was gathered from the literature study and expert opinion interviews, the following two requirements have been drawn up to realize this optimized procedure: 1) The new procedure should provide fairer chances for all new and active students to participate in a course; 2) a list of ranked prioritization criteria should be created, to make the prioritization of students on waiting lists fair and transparent.

B. *Defining priority criteria*

The literature and interview results show that there are certain eligibility criteria to even get on a waiting list for public housing or for waitlisted courses respectively. Of course, if entry requirements are set for a course, students cannot participate in the course if they do not meet these requirements. Therefore, this should be considered during the course enrollment periods, like the Biology and Pharmacy bachelor programs do: if a student does not meet the entry requirements of the course, the student will not be eligible and thus not be registered for the course. Rather than using this as a priority criterion, it will be an eligibility criterion, like found in the literature. Before a student is put on a waiting list for a course to have a chance of participating in the course, he or she must fulfill the entry requirement(s), if any apply (E1 in Table II). A rule that was already in use is that students for which the course is mandatory should always be enrolled for the course. It is related to the priority criteria from Cornell University, where IS majors are prioritized over others. Therefore, this rule will be maintained for the improved procedure in the form of a priority criterion (P1). Some of the I&CS master programs offer certain study paths that a student can follow. A student that wants to follow a course for the path should get priority over students for which the course is not included in their path. This is defined as priority criterion 2 (P2) in Table II. The next priority criterion concerns primary elective courses. These are courses that are less important compared to mandatory courses (P1) and courses that are part of a path (P2), and thus, are given a lower ranking (P3) in Table II. Another important factor regarding prioritization is the need for at least two courses per period. Every student should be able to obtain 15 EC (7.5 EC per course) per period. If a student is still on a waiting list for a course after the late course enrollment period and needs to follow that course to come to 15 EC in courses for the upcoming period, the student will be prioritized. This will only apply for the waiting list handling of the late course enrollment period, as after that period, students will not get another chance to sign themselves up for a course. This rule is maintained from the existing waiting list handling procedure and formulated as a priority criterion (P4). The final factor that will be used is considering students that signed up for a course in a previous year but were not enrolled. Both the Biology and Pharmacy bachelor programs have formalized this into a priority criterion, which is what will be done for the optimized I&CS procedure as well. It is listed as priority criterion 5 (P5) in Table II. Finally, if there are a few spots left for a course where there are too many waitlisted students, a draw will take place to decide which remaining students will be enrolled for the course.

TABLE II
ELIGIBILITY AND PRIORITY CRITERIA FOR OPTIMIZED WAITING LIST PROCEDURE.

Rank	Criterion
E1	Student must fulfill entry requirements of the course
P1	The course is labeled as mandatory for student: the student must be enrolled for the course
P2	The course is part of the track or path that the student is following[1]
P3	The course is labeled as primary elective for student
P4	Student needs course to come to 15 EC in courses for upcoming period[2]
P5	Student signed up for course before but was not enrolled

[1] Only applies for courses in master programs that offer tracks or paths.

[2] Only applies for waiting list handling of late course enrollment period.

C. To-be situation

The proposed waiting list procedure for I&CS master program courses is divided into two segments: an improved procedure for both the first/initial and the late/post course enrollment period. The improved procedures are based on the current procedure, but also assess the problem described in Section VII-A and include the priority criteria that have been defined in Section VII-C. For the (first/initial) course enrollment period for period 1 (P1) of a new academic year, all active students can sign up for courses before the summer break. For the 2022/2023 academic year, this enrollment period was opened from May 30, 2022, to June 24, 2022 (see Table III). For the improved procedure, all students that sign up for courses within this period will be put on waiting lists for the courses. This is possible by putting the course capacity for all courses at zero, so that all students are placed on the waiting list automatically. As seen in Table III, there is a period of more than three months between the final application deadline for new students, June 1, and the start of the teaching period, on September 12. The university will review any application that is completed before the deadline and strives to inform the applicant within 20 working days [38]. This means that there is enough time left between the review of all applicants and the start of the first period of the new academic year starting in September. Thus, these students can still be added to the waiting lists for the courses they would like to follow.

TABLE III
DEADLINES & DATES REGARDING NEW STUDENTS STARTING IN SEPTEMBER & FEBRUARY [13], [39].

Event	September 2022	February 2023
Application deadline (non-EU)	01/04/2022	01/09/2022
Application deadline (EU)	01/06/2022	15/10/2022
Start course enrollment period	30/05/2022	31/10/2022
Deadline course enrollment	24/06/2022	25/11/2022
Post course enrollment period	22/08/2022- 23/08/2022	23/01/2023- 24/01/2023
Start of teaching period	12/09/2022	06/02/2022

The procedure will work identical for the enrollment for courses that start in period 3 (P3) in February. For this period, there are new students as well as already enrolled students that have to sign up for courses. The final application deadline for new students starting in February 2023 is October 15, 2022. The start of the teaching period is February 6, 2023, almost 4 months later than the final application deadline (see Table 12 for dates). Again, the applications are aimed to be reviewed within 20 working days, leaving enough time between the review of all applicants and the start of the teaching period. This allows all new students to put themselves on waiting lists for courses. When the number of registrations for a course does not exceed the course capacity all registered students will be enrolled. Otherwise, certain students will be selected to be enrolled for the course based on the priority criteria defined in Section VII-C. For teaching periods 2 and 4 (P2 & P4), there is no inflow of new students. All students sign up for courses during the course enrollment periods as seen in Table 12 and are automatically put on a waiting list. After the enrollment period has closed, the selection process will be performed based on the priority criteria defined in Section VII-C. When the course capacity is exceeded, students will be filtered through the priority criteria. The procedure can be seen in the ‘selection process’ sub process in Figure 4 in Section VII-D. The handling of the waiting lists of the first enrollment period has to be done before the post course enrollment period opens. Students that were not enrolled for a course during the waiting list selection process of the first enrollment period will be informed before the post course enrollment period opens. That means that they have time to look for another course to follow, for which they can register themselves during the late enrollment period. For the post enrollment period, only courses that have not been booked in full during the first enrollment period will be shown to students. Students that have not been enrolled for one or more course(s) can then only choose from these courses. The waiting list handling procedure will work similarly to the process of the first course enrollment period. The only difference is that priority criterion P4 will also be used during this process, which does not apply for the first course enrollment period waiting list handling. To make sure students understand how the procedure works and how the waiting lists are handled, the procedure should be written down clearly. As the procedures for the first and late enrollment period are now almost identical, the redesigned BPMN model found in Figures 5 and 6 below applies for both the course enrollment periods. The process maintains many elements from the models that were created based on the as-is situation in Section III-D. One new element that has been added is the sub-process 'Selection process', denoted as an activity with a small square at the bottom with a plus sign. It explains the selection of students on the waiting list by using the defined priority criteria. This collapsed sub process can be found in Figure 4.

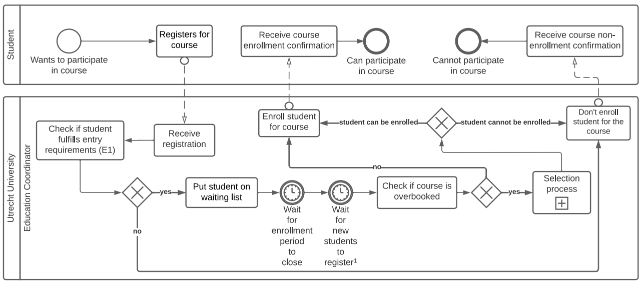

Student
Wants to participate in course
Registers for course
Receive course enrollment confirmation
Can participate in course
Cannot participate in course
Receive course non-enrollment confirmation
Utrecht University
Education Coordinator
Check if student fulfills entry requirements (E1)
Receive registration
Enroll student for course
student can be enrolled
student cannot be enrolled
Don't enroll student for the course
no
yes
Put student on waiting list
Wait for enrollment period to close
Wait for new students to register¹
Check if course is overbooked
yes
Selection process
no

Fig. 3. BPMN model of proposed improved procedure.

The sub-process 'Selection process' in Figure 6 below shows the handling of a waiting list when all students are placed on the waiting list and the course is overbooked. The list of students will then be filtered through the priority criteria that have been defined in Section VII-C until the number of students is sufficiently reduced so that the course capacity is not exceeded. The full process is described in Figure 6. As steps 2-5 are recurrent, there are no separate activities for these steps and a loop has been created to represent these cyclic steps. The procedure continues in Figure 5 at the outflowing sequence flow from activity 'Selection process'.

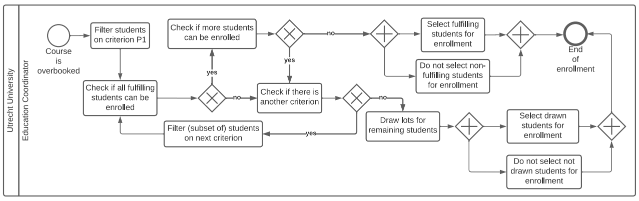

Utrecht University
Education Coordinator
Course is overbooked
Filter students on criterion P1
Check if more students can be enrolled
no
yes
Select fulfilling students for enrollment
Do not select non-fulfilling students for enrollment
End of enrollment
yes
Check if all fulfilling students can be enrolled
no
Check if there is another criterion
no
Filter (subset of) students on next criterion
yes
Draw lots for remaining students
Select drawn students for enrollment
Do not select not drawn students for enrollment

Fig. 4. BPMN model of 'Selection process'.

VIII. Validation

To validate the optimized waiting list procedure that has been created in Section VII above, another expert opinion will be used. The optimized procedure will be proposed to the education coordinator of the I&CS master program courses, as that person possesses all of the necessary knowledge about the current process and the problems that are faced. This knowledge is needed for the validation to work, as only an expert on the domain can imagine realistic problem contexts and make assumptions on how the proposed procedure could work in practice [11]. If the proposed waiting list procedure does not satisfy the expert, the artifact will be revisited and adjusted or even redesigned according to expert feedback.

A. I&CS education coordinator interview results

The following questions, with the answers that were given written out directly below each question, were asked based on the proposed procedure. **What aspects of the proposed improved procedure are improvements over the current procedure?** The improved procedure does indeed solve the problem of late incoming new students and will provide these students with a fairer chance to participate in elective courses in the periods in which they start (periods 1 and 3) compared to the current procedure. **What drawbacks are there to the proposed improved procedure?** The handling of waiting lists that contain all students for a course will require significantly more time, a higher budget and staffing. It will take more time and staffing to sort out waiting lists by hand, as for the improved procedure, all courses use waiting lists. The proposed improved procedure is also not in line with current back-office procedures. Active students will feel less motivated to sign up for courses timely, as they will not be registered straight away if they do. It could also lead to them signing up for multiple courses, as with a waiting list, they cannot immediately see if they will be signed up for a course. **What aspects of the proposed improved procedure could be done differently or revised?** All of the mentioned drawbacks can be improved upon, e.g., the extra time, effort and resources the improved procedure requires. **Could the proposed improved procedure be used in the future; is it realistic and feasible?** Yes, but it will require alignment with the back-office as it will cost time and money to follow an enrollment procedure that is different from the usual procedure.

As the aim for this research was to make chances of participating in courses fairer for every student, the procedure with placing all students on a waiting list for a course will be maintained. Of course, handling all waiting lists for all courses there are by hand requires a lot of time from support staff, as also mentioned during the expert interviews in Section V. A solution to this problem could be automating this process. As the set of priority criteria is defined clearly, the enrollment of students could be automated. An example of how such a filtering system could work is shown in Figure 5, that is revised based on the feedback from the expert. The sorting out of the waiting lists is now done by the 'Student Filtering Program'. This lane has been added in Figure 5, and 'Education Coordinator' has been changed to 'Student Filtering Program' in Figure 6. The 'Student Affairs' lane has been added as well, as they officially enroll students for courses, not the education coordinators.

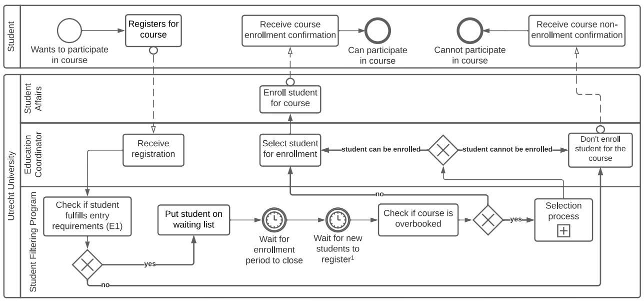

Student
Wants to participate in course
Registers for course
Receive course enrollment confirmation
Can participate in course
Cannot participate in course
Receive course non-enrollment confirmation
Utrecht University
Student Affairs
Enroll student for course
Education Coordinator
Receive registration
Select student for enrollment
student can be enrolled
student cannot be enrolled
Don't enroll student for the course
Student Filtering Program
Check if student fulfills entry requirements (E1)
Put student on waiting list
Wait for enrollment period to close
Wait for new students to register1
Check if course is overbooked
yes
no
Selection process
yes
no

Fig. 5. Redesigned BPMN model.

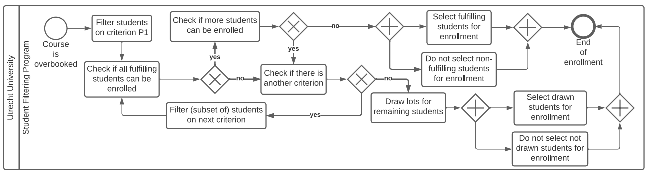

Utrecht University
Student Filtering Program
Course is overbooked
Filter students on criterion P1
Check if all fulfilling students can be enrolled
Check if more students can be enrolled
yes
no
no
yes
Check if there is another criterion
no
yes
Filter (subset of) students on next criterion
Select fulfilling students for enrollment
Do not select non-fulfilling students for enrollment
Draw lots for remaining students
Select drawn students for enrollment
Do not select not drawn students for enrollment
End of enrollment

Fig. 6. Redesigned sub-process ’Selection process’.

IX. Discussion

The handling of waiting lists is a complex process. The main theme across literature and the conducted interviews is the importance of urgency and need when allocating applicants for waiting lists. In contrast, available grey literature showed a predominance of rather simpler first come, first served methods. As these methods would not contribute to solving problems faced in the situation at Utrecht University, this method was not considered for the proposed improved procedure.

The studied literature provided waiting list management and handling in other domains, while the interviews added an educational perspective. The combination of the most important findings from those two methods has provided a new perspective on waiting list handling, namely in the domain of education. The proposed improved procedure was presented to the education coordinator of the I&CS department at Utrecht University as part of the validation of the new process. According to the feedback, the new procedure does indeed tackle the problem that was aimed to solve, but also has drawbacks. The biggest drawback is that the procedure would require significantly more time and resources to operate. This problem can be solved by automating the process, as proposed in Section VIII-B. Other issues, like more uncertainty for students if and when they will be placed for a course were not solved for now, but could be investigated further. Even though many studies have been conducted in the field of waiting list procedures, prioritization and management, no formal literature was found on the handling of waiting lists for courses in (higher) education. This means that the proposed improved waiting list procedure is not based on research in the exact same application domain. The mentioned interviews were conducted with domain experts from one university in one country. Moreover, interviews were only performed with education coordinators within the Science Faculty. Due to time limitations and because programs from other faculties that were approached said they do not use waiting lists for courses, no others were interviewed. This may have an impact on the generalizability of the results.

X. Conclusions & Future Research

The problems that were aimed to tackle in this research were related to waiting list problems for master program courses within the I&CS department at Utrecht University. The main research question was the following: “How can waiting list procedures for master program courses in I&CS be improved to make sure that the chances of participating are as fair as possible for every student?”. The methods mentioned in Section VIII-B after refining the proposed procedure from Section VII based on received feedback provide an answer. Based on literature and interview results, waiting list procedures can be improved by integrating three aspects into the process. The first one is changing the use of waiting lists. Instead of using a waiting list only as soon as the course is booked in full, all students will be put on a waiting list upon registration for a course automatically. This allows for more students to get onto the waiting list for a longer period of time, so incoming students can register themselves for courses as well. The handling of the waiting list can then be postponed until all students have signed up for a course. The second aspect is formalizing the handling of the waiting list by defining priority criteria. In the current procedure, there are some rules about who to prioritize over others, but there is no (ranked) list of the exact criteria on which waiting lists are handled. Explicit information about waiting list handling is also currently not provided to students. The defined priority criteria for the improved procedure allow for a fair process of selecting which students will be enrolled for a course. When a student is not selected for a course, the student can easily be notified by mail that explains why. For transparency reasons, these criteria should also be published on the department’s web page. The final optimization that has been made over the current procedure is the automation of the process. Based on that filtering process, an information system or education coordinator decides which students will be enrolled.

Future research could gain more insights on the handling of waiting lists for courses in higher education. Research on a larger scale could be conducted to investigate how other colleges and universities handle waiting lists for courses. This will lead to an even better understanding of how these processes work in a wider variety of educational institutes in multiple countries. Based on a larger data set that contains several perspectives on waiting list handling for courses, an even more optimal procedure could be realized. The importance of including factors such as student age, bad habits, or disabilities could be investigated in future work. Finally, as a result of the procedure that was created after the validation session a proposal for automation is made which is part of future research.

Appendices

- For Appendix A, see: https://osf.io/h3g5x
- For Appendix B, see: https://osf.io/pe43y
- For Appendix C, see: https://osf.io/eprbw
- For Appendix D, see: https://osf.io/agn3h

References

[1] Tracy Camp, W. Richards Adrion, Betsy Bizot, et al. “Generation CS”. In: *ACM Inroads* 8.2 (May 2017), pp. 44–50. DOI: 10.1145/3084362. URL: http://dx.doi.org/10.1145/3084362.

[2] E Krutsch. *Computer Science Education Week: Explore In-Demand IT Jobs*. Dec. 2022. URL: https://blog.dol.gov/2022/12/01/computer-science-education-week-explore-in-demand-it-jobs.

[3] S Zweben and B Bizot. “2021 Taulbee Survey”. In: *Computing Research News* 34.5 (May 2022). URL: https://cra.org/wp-content/uploads/2022/05/2021-Taulbee-Survey.pdf.

[4] Esther Shein. “The CS teacher shortage”. In: *Communications of the ACM* 62.10 (Sept. 2019), pp. 17–18. DOI: 10.1145/3355375. URL: http://dx.doi.org/10.1145/3355375.

[5] Andre L. N. Campos and Toacy Oliveira. “Software processes with BPMN: an empirical analysis”. In: *Product-Focused Software Process Improvement* (2013), pp. 338–341. DOI: 10.1007/978-3-642-39259-7_29. URL: http://dx.doi.org/10.1007/978-3-642-39259-7_29.

[6] Vahid Garousi, Michael Felderer, and Mika V. Mäntylä. “Guidelines for including grey literature and conducting multivocal literature reviews in software engineering”. In: *Information and Software Technology* 106 (Feb. 2019), pp. 101–121. DOI: 10.1016/j.infsof.2018.09.006. URL: http://dx.doi.org/10.1016/j.infsof.2018.09.006.

[7] B Kitchenham and S Charters. “Guidelines for performing Systematic Literature Reviews in software engineering”. In: *EBSE Technical Report* EBSE-2007-01 (July 2007). URL: https://www.elsevier.com/__data/promis_misc/525444systematicreviewsguide.pdf.

[8] Carol Lefebvre, Eric Manheimer, and Julie Glanville. “Searching for studies”. In: *Cochrane Handbook for Systematic Reviews of Interventions* (2008), pp. 95–150. DOI: 10.1002/9780470712184.ch6. URL: http://dx.doi.org/10.1002/9780470712184.ch6.

[9] D.J. Farace and J Schöpfel. *Introduction grey literature*. Grey Literature in Library and Information Studies, 2010, pp. 1–7. URL: https://library.oapen.org/bitstream/handle/20.500.12657/45661/626361.pdf?sequence=3&isAllowed=y.

[10] Han Dorussen, Hartmut Lenz, and Spyros Blavoukos. “Assessing the reliability and validity of expert interviews”. In: *European Union Politics* 6.3 (July 2005), pp. 315–337. DOI: 10.1177/1465116505054835. URL: http://dx.doi.org/10.1177/1465116505054835.

[11] Roel J. Wieringa. “Treatment validation”. In: *Design Science Methodology for Information Systems and Software Engineering* (2014), pp. 59–69. DOI: 10.1007/978-3-662-43839-8\{_}7. URL: http://dx.doi.org/10.1007/978-3-662-43839-8_7.

[12] Ali Alammary, Judy Sheard, and Angela Carbone. “Blended learning in higher education: three different design approaches”. In: *Australasian Journal of Educational Technology* 30.4 (Sept. 2014). DOI: 10.14742/ajet.693. URL: http://dx.doi.org/10.14742/ajet.693.

[13] Utrecht University. *Academic Year Calendar Science Faculty 2022-2023*. 2022. URL: https://students.uu.nl/sites/default/files/roosterplanning21_22_1.pdf.

[14] S.A. White and IBM Corporation. *Introduction to BPMN*. July 2004. URL: https://www.bptrends.com/bpt/wp-content/publicationfiles/07-04%20WP%20Intro%20to%20BPMN%20-%20White.pdf.

[15] Caroline Fogarty and Patricia Cronin. “Waiting for healthcare: a concept analysis”. In: *Journal of Advanced Nursing* 61.4 (Jan. 2008), pp. 463–471. DOI: 10.1111/j.1365-2648.2007.04507.x. URL: http://dx.doi.org/10.1111/j.1365-2648.2007.04507.x.

[16] Marianna Karamanou, Dimitrios Vrachatis, and Dimitrios Tousoulis. “The ethics of waiting lists for TAVR procedures”. In: *European Heart Journal* 41.6 (Feb. 2020), pp. 735–736. DOI: 10.1093/eurheartj/ehaa043. URL: http://dx.doi.org/10.1093/eurheartj/ehaa043.

[17] Jonathan M Breslin, Susan K MacRae, Jennifer Bell, and Peter A Singer. “Top 10 health care ethics challenges facing the public: views of Toronto bioethicists”. In: *BMC Medical Ethics* 6.1 (June 2005). DOI: 10.1186/1472-6939-6-5. URL: http://dx.doi.org/10.1186/1472-6939-6-5.

[18] S Lewis, M.L. Barer, C Sanmartin, S Sheps, S.E.D. Shortt, and P.W. McDonald. “Ending waiting-list mismanagement: principles and practice”. In: *Canadian Medical Association Journal* 162.9 (May 2000), pp. 1297–1300. URL: https://www.cmaj.ca/content/cmaj/162/9/1297.full.pdf.

[19] D.C. Hadorn and Steering Committee of the Western Canada Waiting List Project. “Setting priorities for waiting lists: defining our terms”. In: *Canadian Medical Association Journal* 163.7 (Oct. 2000), pp. 857–860. URL: https://www.cmaj.ca/content/163/7/857.full.

[20] Seth A. Brown, Jefferson D. Parker, and Phillip R. Godding. “Administrative, clinical, and ethical issues surrounding the use of waiting lists in the delivery of mental health services”. In: *The Journal of Behavioral Health Services &; Research* 29.2 (May 2002), pp. 217–228. DOI: 10.1007/bf02287708. URL: http://dx.doi.org/10.1007/bf02287708.

[21] Andrea J Curtis, Colin O H Russell, Johannes U Stoelwinder, and John J McNeil. “Waiting lists and elective surgery: ordering the queue”. In: *Medical Journal of Australia* 192.4 (Feb. 2010), pp. 217–220. DOI: 10.5694/j.1326-5377.2010.tb03482.x. URL: http://dx.doi.org/10.5694/j.1326-5377.2010.tb03482.x.

[22] A. Testi, E. Tanfani, R. Valente, G. L. Ansaldo, and G. C. Torre. “Prioritizing surgical waiting lists”. In: *Journal of Evaluation in Clinical Practice* 14.1 (Jan. 2008), pp. 59–64. DOI: 10.1111/j.1365-2753.2007.00794.x. URL: http://dx.doi.org/10.1111/j.1365-2753.2007.00794.x.

[23] M E Seddon, J K French, D J Amos, K Ramanathan, S C McLaughlin, and H D White. “Waiting times and prioritisation for coronary artery bypass surgery in New Zealand”. In: *Heart* 81.6 (June 1999), pp. 586–592. DOI: 10.1136/hrt.81.6.586. URL: http://dx.doi.org/10.1136/hrt.81.6.586.

[24] B Davis and S R Johnson. “Real-time priority scoring system must be used for prioritisation on waiting lists”. In: *BMJ* 318.7199 (June 1999), pp. 1699–1699. DOI: 10.1136/bmj.318.7199.1699. URL: http://dx.doi.org/10.1136/bmj.318.7199.1699.

[25] Mark Taylor and David C. Hadorn. “Developing priority criteria for general surgery: results from the Western Canada Waiting List Project.” In: *Canadian Journal of Surgery* 45.5 (Oct. 2002), pp. 351–7.

[26] Gordon Arnett and David Hadorn. “Developing priority criteria for hip and knee replacement: results from the Western Canada Waiting List Project.” In: *Canadian Journal of Surgery* 46.4 (Aug. 2003), pp. 290–6.

[27] Alfred Michael Dockery, Rachel Ong, Stephen Whelan, and Gavin Wood. “The relationship between public housing wait lists, public housing tenure and labour market outcomes”. In: *AHURI Research Paper* (Sept. 2008). URL: https://apo.org.au/sites/default/files/resource-files/2008-09/apo-nid8026.pdf.

[28] Gavin Wood, Rachel Ong, and Alfred Michael Dockery. “What has determined longer run trends in public housing tenants’ employment participation 1982-2002?” In: *AHURI Research Paper* 5 (June 2007), pp. 1–96. URL: https://apo.org.au/sites/default/files/resource-files/2007-06/apo-nid7188.pdf.

[29] Terry Burke and Kath Hulse. “Allocating social housing”. In: *AHURI Positioning Paper* 47 (Feb. 2003). URL: https://apo.org.au/sites/default/files/resource-files/2003-12/apo-nid8297.pdf.

[30] University of North Georgia. *Waitlisted Ccurses*. URL: https://ung.edu/registrar/waitlisted-courses.php.

[31] University of Denver. *Closed seats & waitlists*. URL: https://www.du.edu/registrar/registration/how-register/waitlists.

[32] University of Auckland. *Course enrolment waitlist*. URL: https://uoa.custhelp.com/app/answers/detail/a_id/239/%7E/course-enrolment-waitlist.

[33] American University Washington College of Law. *Waitlist process - Courses & registration - Current students - Office of the registrar*. URL: https://www.wcl.american.edu/academics/academicservices/registrar/current-students/courses-registration/waitlist/.

[34] Cornell University. *Enrollment/waitlist information*. Oct. 2022. URL: https://infosci.cornell.edu/courses/enrollmentwaitlist.

[35] Boston University. *Waitlist | Boston University Department of Computer Science | Computer Science*. URL: https://www.bu.edu/cs/undergraduate/undergraduate-life/courses/cs-waitlists/.

[36] Utrecht University. *In- en uitschrijven cursus*. URL: https://students.uu.nl/beta/farmacie-b/praktische-zaken/in-en-uitschrijving/in-en-uitschrijven-cursus.

[37] Utrecht University. *Studiegids Bacheloropleiding Biologie 2022-2023*. Tech. rep. Nov. 2022. URL: https://students.uu.nl/sites/default/files/v5_Studiegids_BachelorBiologie_UU_22-23_3nov2022_binder.pdf.

[38] Utrecht University. *Admission and application - Degree from a Dutch research university*. URL: https://www.uu.nl/en/masters/business-informatics/admission-and-application/degree-from-a-dutch-research-university.

[39] Universiteit Utrecht. *Online modules*. URL: https://www.uu.nl/onderwijs/educate-it/docentontwikkeling/online-modules.

DOI: 10.15439/2023F5720
ISSN 2300-5963 ACSIS, Vol. 37

Conceptualizing sustainability in the context of ICT. A literature review analysis.

Tobias Hassmann
Munich Business School
80687 Munich, Germany
Email: tobias.hassmann@munich-business-school.de

Markus Westner
OTH Regensburg
93053 Regensburg, Germany
Email: markus.westner@oth-regensburg.de

***Abstract*—This paper examines the conceptualization of sustainability in the context of information and communication technology (ICT) research. Through an inductive text analysis of sixteen literature reviews spanning from 2014 to 2023, key themes and concepts are identified, highlighting the complex relationship between ICT and sustainability. ICT is perceived both as an enabler and a problem for sustainability. Furthermore, the terminology and concept of sustainability in the context of ICT remain unclear. The emergence of digitalization as a novel socio-technical phenomenon poses additional challenges for conceptual alignment. While a holistic view of sustainability in ICT is desired, business and social implications receive less attention. The paper summarizes and discusses the developments in research on this topic over the past decade.**

***Index Terms*—Sustainability in ICT, sustainability by ICT, digital sustainability, digital transformation and sustainability.**

I. Introduction

ENVIRONMENTAL challenges have become one of the most pressing contemporary issues for humankind. They are paired with social and economic transformations. For that matter, sustainability research has become a topic of interest as it promises practical solutions for these challenges [1–3]. It is postulated that fundamental sustainability transformations at the macro-, meso- and micro-level are required to address the manifold and complex challenges on the social, economic, and environmental levels involving multiple actors [4]. At the same time, digitalization has become a global and ubiquitous [5] phenomenon, which now is quasi-irreversible after COVID-19 [6]. This raises the question if and how these two megatrends – digitalization and sustainability – are intertwined or apt to drive deep transformations [7–9]. Digitalization is a socio-technical phenomenon [10] that enables the utilization of novel "*technologies, communication methods, business functions, and models*" [6, p. 15] to achieve different objectives. ICT is the backbone of digitalization. It provides hardware and software solutions that enable digitalization. Thus, ICT deeply impacts the economy [11] and enables socio-technical megatrends commonly referred to as SMACIT, that is, social platform, mobile, analytics, cloud computing, and internet of things [8, 10].

The term and concept of sustainability originate from forestry and originally describe the approach to harvest just the amount of wood that regrows [12, 13]. The concept became popular in 1987, when the World Commission on Environment and Development published the Brundtland Report, which provided a seminal description of the sustainability concept [12] by distinguishing three pillars [13]: environmental, economic, and social sustainability, also referred to as planet, profit, people [3, 12]. Additionally, the Brundtland report defined sustainable development, whose aim is to satisfy "*the needs of the present without compromising the ability of future generations to meet their own needs*" [13, p. 684].

Today, the sustainability concept is applied to a variety of other – complex, novel, and broad – problem spaces [14] such as ICT. Consequently, the term has become a multi-layered concept [15]. In the context of ICT, the concept of sustainability remains opaque [16].

This paper aims to provide guidance, clarity, and an overview for both researchers and practitioners in the field on how the concept of sustainability is related to ICT. Such conceptual alignment will support them operationalizing actions and strategies to achieve sustainability goals. Given the fundamental conceptual challenges of applying sustainability to the domain of ICT, this paper aims to address three research questions:

RQ 1: *How is the concept of sustainability discussed in the context of ICT?*
RQ 2: *What are the key concepts and themes that characterize the relationship between sustainability and ICT?*
RQ 3: *How has this relationship evolved over the past ten years?*

In the context of this paper, a *concept* is a mutually exclusive, well defined, and known insight from extant literature on that topic. A concept can become a subcategory and can be mapped to a well-known and well discussed research *theme* [17]. For example, the terminological misalignment regarding sustainability in the context of ICT is a *theme* but is discussed at two different *conceptual* levels: sustainability in the context of ICT, as well as sustainability in the context of digitalization or digital transformations. While both strands of discussion are mutually exclusive on a *conceptual* level, they form a *theme* inasmuch they inherently are terminological debates. For the analysis, an inductive text analysis was applied to identify *concepts* and *themes* from literature reviews on sustainability in ICT.

A five-step process was applied to identify themes and concepts [17]. First, it was decided to focus on literature reviews from the past ten years that treat the topic of sustainability and ICT on a holistic and conceptual level. Second, searches for relevant literature reviews on Scopus, Web of Science, and EBSCO were performed. Inclusion and exclusion criteria were applied, and search queries were reformulated to retrieve pertinent content. Third, the retrieved literature reviews were assessed as to whether they match the initially defined search criteria. Literature reviews that met the inclusion criteria were added to the final sample for further analysis. Hence, a selective strategy was chosen for creating the sample for analysis. Fourth, the sampled literature reviews were carefully examined and scrutinized using open, axial, and selective coding, and comparative analysis to synthesize research findings and gaps. From these findings, conceptual commonalities were induced and mapped to *concepts*. The *concepts* were eventually also subsumed to overarching *themes*. Fifth, the findings are eventually presented in the following chapters.

II. Methodological approach

A. Search strategy

Scopus, Web of Science, and EBSCO were used to find relevant literature reviews. Only articles, reviews, or conference proceedings in academic journals dating from between the years 2014 through 2023 and written in English language were considered. Additional restriction criteria, such as selecting a set of subject areas (Scopus), categories (Web of Science), and databases (EBSCO), were applied. The search query was constructed such that it retrieves matches in the title of documents to limit the number of matches. The following keywords were applied and concatenated using the OR- and AND- operators: ((*digitalization* OR *digital* OR *ict* OR (*information system*) OR (*information science*)) AND ((*sustainability* OR *sustainable* OR (*problem* OR *solution*)) AND (*literature* OR *review* OR *concept* OR *research* OR *definition*). The results that were found in the three different databases were then merged and duplicates were eliminated. A first temporary sample was chosen based on carefully studying the abstracts. Only literature reviews that cover sustainability in the context of ICT holistically were considered. That is, they must refer to all sustainability pillars, treat the topic of sustainability in the context of ICT in general, and not in context of a specific industry, sector, technology, or application scenario. Table I shows the applied search approach for retrieving relevant literature reviews.

To refine the result set, another round of reading and assessment was performed. As a result, some literature reviews were discarded because they did not meet the inclusion criteria: for example, literature reviews focusing on sustainability for smart cities, efficient manufacturing, industry 4.0, or business models were neglected. Reviews that address a specific industry or sector, for instance, textile industry, or fishery, were discarded, too. Two reviews were incorporated in the analysis despite showing a clear tendency towards a specific sustainability pillar [8, 18]. However, those literature reviews actively tried to integrate their focus area into the greater context of sustainability and ICT. Additionally, forward- and backward citations were used to identify pertinent research articles.

An additional "sanity-check" on Google Scholar was performed to identify potentially overlooked literature reviews. For that, the same keywords were applied that were used to retrieve relevant literature from Scopus, Web of Science, and EBSO databases. Finally, sixteen literature reviews remained for the in-depth analysis. Table V in the appendix lists the literature reviews that were sampled for the analysis.

Table I.
Search approach (matches on *title*-property) for literature selection from Scopus, Web of science, and EBSCO

Property	**Inclusion criteria**	**Scopus**	**Web of Science**	**EBSCO**
Subject area	Computer science OR business, management, and accounting OR social sciences OR engineering OR environmental science	407		
	Computer science OR environmental sciences OR green sustainable technology OR environmental studies OR management OR business OR engineering		204	
	Business source premier OR green file			81
Document types	Articles, conference papers, reviews, and proceeding papers in academic journals	361	192	69
Publication Years	2014 to 2023	297	147	55
Language	English	270	154	50
Thematic focus	Literature reviews that treat sustainability in ICT holistically	15	7	2
# Total after merging and deduplication		**#16**		

B. Data analysis

The sampling and data synthesis processes, including the conceptualization and thematization steps, are illustrated in Fig. 1.

First, open coding was performed by carefully reading the sampled literature reviews and coding findings and research gaps. As a result, a total of 97 findings and gaps were synthesized: 59 findings (61%) and 38 gaps (39%). Next, axial coding was applied by inducing *concepts* and *themes* from the findings and gaps. Thirteen different *concepts* and five *themes* were induced, and the 97 individual findings and gaps were classified accordingly. Selective coding was then applied to identify connections, interdependencies, and validity of the concepts and themes.

Themes and concepts were then compared to each other to avoid overlaps or fuzziness with regards to their content. The concepts and themes eventually enabled addressing all three research questions. The concepts and themes that were identified address RQ1, that is, understanding *how the concept of sustainability is discussed in the context of ICT*. They also directly address RQ2, whose aim is to understand *what the key concepts and themes are that characterize the relationship between sustainability and ICT*. To answer RQ3, which is to understand *how the relationship between sustainability and ICT has evolved over the past ten years*, an additional step was performed: the identified themes and concepts were analyzed as to which dimensions they are composed of, how the meaning and importance of those dimensions have evolved, and which additional dimensions have emerged by validating them against a conceptualization proposal from 2014 [16].

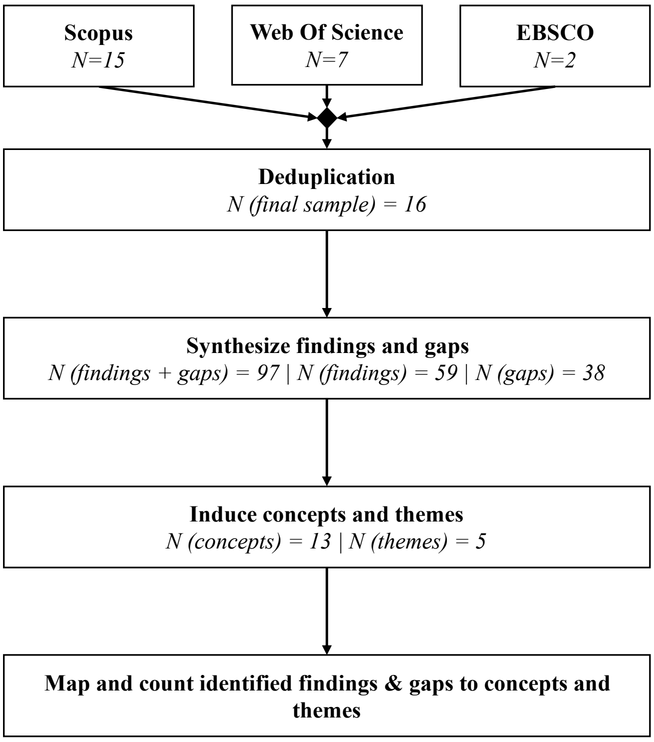

Scopus
N=15
Web Of Science
N=7
EBSCO
N=2
Deduplication
N (final sample) = 16
Synthesize findings and gaps
N (findings + gaps) = 97 | N (findings) = 59 | N (gaps) = 38
Induce concepts and themes
N (concepts) = 13 | N (themes) = 5
Map and count identified findings & gaps to concepts and themes

Fig 1. Literature sample and data synthesis approach

III. Findings

A. Concepts and themes

Our analysis of the findings and gaps in the analyzed literature reviews identified thirteen concepts, which were then grouped and mapped into five themes: (A) application, (B) sustainability concept, (C) impact, (D) mitigation, and (E) stakeholders. The concepts (1) application scenarios, (2) application sectors, (3) application technology, and (4) geographic perspectives were mapped to the theme *application*. The theme *sustainability concept* is divided into the concepts: (5) need to align concepts of digitalization, ICT, digital sustainability, and digital transformations and (6) terminological misalignment of sustainability in the context of ICT. Further, the *impact*-theme is composed of the categories (7) ICT as enabler, (8) ICT as problem, (9) ICT as problem and enabler and (10) the measurement of impacts of ICT on sustainability. The theme of *mitigation* consists of the concepts (11) mitigation strategies to address sustainability challenges and the (12) need for an interdisciplinary and holistic approach to sustainability in ICT. The theme *stakeholders* is defined by one concept, which is the (13) role of stakeholders and governance.

Table II shows the concepts and themes that were identified and are used to directly address RQ1 and RQ2. The following paragraphs explain the detected themes and concepts in detail.

Table II.
Themes and concepts identified addressing RQ1 and RQ2

Theme	Concept	ID
Application	Application scenarios	A#1
	Application sectors	A#2
	Application technology	A#3
	Geographic perspectives	A#4
Sustainability concept	Need to align concepts of digitalization, ICT, digital sustainability, and digital transformations	B#5
	Terminological misalignment sustainability in the context of ICT	B#6
Impact	ICT as enabler	C#7
	ICT as problem	C#8
	ICT as problem and enabler	C#9
	Measurement of impacts of ICT on sustainability	C#10
Mitigation	Mitigation strategies to address sustainability challenges	D#11
	Need for an interdisciplinary and holistic approach to sustainability in ICT	D#12
Stakeholders	Role of stakeholders and governance	E#13

B. Aspects of discussing sustainability in the context of ICT

The famous Brundtland report assigned ICT a leading role in achieving sustainability goals [19]. Today, the concept of sustainability in ICT is discussed under two competing yet intertwined concepts (C#9): *ICT as enabler* (C#7) or *problem for sustainability* (C#8). Research lists different application areas where ICT functions as an enabler for efficiency, for example, by reducing the overall energy consumption [6, 20], and is thus considered a historic opportunity [8, 11]. On the other hand, ICT directly impacts the environment, for example, via the hardware value chain for which raw materials need to be extracted, and energy is required for producing, using, refurbishing or reusing, and finally for disposing hardware as e-waste [5]. Also, operating large data centers, the backbone for supporting an increased usage of cloud-based services [21], requires significant energy [22]. Additionally, there are negative indirect impacts such as obsolescence, induction, and rebound effects [19, 23, 24]. Consequently, the theme of *mitigation* (D#11 and D#12) is the prevailing strand of discussion in literature. Also, the presentation of concrete proposals on how to *measure the impacts of ICT on sustainability* (C#10) is of importance in that context [6, 23, 25]. Another negative consequence is the social divide between those participating in digitalization and those being left out or behind due to missing skills or not keeping up with using the benefits of using digital technologies [9, 26, 27]. Such negative spillover effects of technology adoption beyond the environmental dimension are widely neglected [3, 6, 27]. Also, adverse environmental consequences of novel technologies, such as bit coin mining, have been insufficiently considered [3, 6, 27].

However, the analyzed literature reviews show a tendency towards describing *ICT as enabler* (C#7). We identified that seven, i.e., 7% of all findings and gaps, referred to the consideration of *ICT as both a problem and enabler* (C#9) for sustainability. This finding aligns with what is reported in other literature reviews. For example, one study reports that 58% of their examined studies focus exclusively on the positive effects, while only 15% analyzed focused on negative effects [9]; another review found 52% of papers in their sample that describe ICT solutions as enabler [28]. Different enablement scenarios are mentioned in the analyzed reviews, for example, promoting renewable energy, driving energy efficiency [7, 8, 28], ensuring pollution and waste control that help create smart cities [8], managing energy demand and supply [8, 28], platforms for helping establish a circular economy [29], promoting sustainable mobility [27], or enabling education for sustainability [19]. However, there is little research on how ICT-based solutions can enable deep sustainability transformations, such as supporting the move to more sustainable agriculture [28]. There is also criticism that the belief that digital solutions will consistently result in positive sustainability outcomes represents an inherent risk (*digital solutionism*) [19].

Table III shows the distribution of the thirteen induced categories based on the findings and gaps identified in the sampled literature reviews. ICT as enabler for sustainability is the most mentioned finding – and ranks low in terms of being a research gap.

C. Concepts and themes characterizing the relationship between sustainability and ICT

Twenty-one, that is 22% of the identified findings and gaps, emphasize the need for an *interdisciplinary and holistic approach to sustainability in ICT* (D#12) [3, 6, 7, 25, 26, 28]. There is agreement that all three pillars are intrinsically interconnected and cannot be considered in isolation [9, 27]. Particularly, social aspects such as the social divide caused by digitalization are thematized [18, 19]. Three papers mention that there is comparably little attention in management literature on the topic [6, 7, 26]. Another challenge pointed out is that the economic and environmental sustainability pillars can overlap. For example, topics such as the circular economy are considered both a social [28] and environmental aspect of sustainability [8]. Within the economic pillar, key topics of discussion are how ICT-supported digital solutions drive open innovation [29], or how digital technology – in combination with technology savvy human capital – is crucial for businesses to improve their products and services [30]. It is further pointed out that adopting a more interdisciplinary and systemic approach is essential to overcome the biases, limitations, and gaps identified in the current research landscape. Collaboration and engagement with multiple disciplines are required to understand the systemic nature of digital transformation and the link between digitalization and sustainable development [6, 9, 31].

TABLE III.
FREQUENCY OF CONCEPTS AND THEMES IDENTIFIED IN THE SAMPLED LITERATURE REVIEWS

Theme	Concept	[#] Finding	[#] Gap	[#] Total
Mitigation	Need for an interdisciplinary and holistic approach to sustainability in ICT	9	12	21
Impact	ICT as enabler	14	0	14
Mitigation	Mitigation strategies to address sustainability challenges	3	8	11
Sustainability concept	Need to align concepts of digitalization, ICT, digital sustainability, and digital transformations	3	7	10
Impact	Measurement of impacts of ICT on sustainability	5	4	9
Stakeholders	Role of stakeholders and governance	6	3	9
Impact	ICT as problem and enabler	5	2	7
Application	Application scenarios	4	0	4
Application	Geographic perspectives	2	2	4
Application	Application technology	4	0	4
Impact	ICT as problem	2	0	2
Sustainability concept	Terminological misalignment of sustainability in the context of ICT	1	0	1
Application	Application sectors	1	0	1
Grand Total		59	38	97

Mitigation strategies to address sustainability challenges (D#11) are referred to in eleven, that is 11% of all findings and gaps. Though, the discovered and proposed strategies to mitigate sustainability impact do vary. The need for incorporating sustainability into strategic decision-making in organizations with the goal to establish sustainable business models is underlined [8]. The importance of operationalizing and providing guidance for sustainability is also mentioned to be important, as the actual acting stakeholders need guidance on how to act sustainably [16]. Business innovation is brought forward, too, as a possible strategy [30]. Other strategies mentioned are the focus on digital education [19] as well as the implementation of policies [19, 28]. Policies are crucial to establish sufficiency-oriented strategies and transformations at the structural level, which are underrepresented in research [28]. The sampled literature reviews show differences regarding the focus on where these policies ought to be applied. While it is acknowledged that individual users drive sustainability outcomes through their choices [6, 27], it is also proposed that only regulations and policies can ensure sustainable behavior [28]. This applies to all actors: individuals, and business organizations. However, it is conceded that it is comparably feasible to establish regulations for hard- or software providers, but employing policies for organizations or individuals is complex and difficult to monitor [28]. In this context, it is pointed out that it will be challenging to lower service standards without impacting an ICT provider's business by affecting customer expectations [27]. Overall, while the sources agree on the importance of sustainability strategies, they differ in their focus and approach to achieving sustainability goals. One the one hand, the emphasis on consistency and sufficiency strategies driven by policies are highlighted [28], while the need to examine how incentive systems or broad sustainability goals impact individuals' behaviors and beliefs is stressed out, too [6, 27].

The broad semantic and conceptual scope of the term sustainability is subject to academic discussions: the term is described to be implicitly normative [1, 32–35] and polysemous [36, 37]. More recently and importantly, the *need to align the concepts of digitalization and ICT, digital sustainability, and digital transformations* (B#5) emerged as a new important aspect [3, 7, 8, 25, 29, 31], which further contributes to the *terminological misalignment of sustainability in the context of ICT* (B#6). Ten, that is 10% of all identified findings and gaps, stressed out the lack of a description for the relationship and terminology for how sustainability and digitalization are connected. Hence, a paradigm-shift is suggested [31] to understand the connection between the two interdependent concepts [3]. Also, a clear delineation of the concepts of green IS, which focuses on the sustainable use of technology, and green IT, which aims to achieve sustainability goals by leveraging technology, is recommended [25]. The terminological discussion is further characterized by its focus on integrating environmental sustainability principles into business models and organizational strategies, as well as the alignment between organizational strategy and digitalization. It is highlighted that a deeper understanding of the environmental and social implications [18] of digitalization is needed. Therefore, a public goods approach is suggested to consider deep social, economic, and environmental impacts in the context of digitalization and the ubiquitous use of technology [27]. To conclude, the topic of how digitalization and digital transformation, sustainability, and sustainability transformations are connected is considered important, but there is not yet a terminological alignment for the scientific discourse on that topic.

Also, the *role of stakeholders and governance* (E#13) can be synthesized as a theme within sustainability in ICT – nine, that is 9% of all gaps and findings, refer to it. Stakeholders need to address challenges in operationalizing sustainability in ICT research [16]. In business organizations they are responsible for incorporating and driving for strategic sustainability goals [8, 25]. There is a need to better understand how the different organizational departments and functions can collaborate to achieve sustainability goals or implement sustainability initiatives, and what the role the IT department can play in that context [25]. Furthermore, the importance of practitioners and researchers for collaboration across disciplines to conduct comprehensive sustainability research is stressed [6]. Governmental actors are expected to design and implement policies that can enforce and encourage sustainable behaviors for individuals, individuals in organizations and ICT manufacturers and providers [28]. One study also mentions trading-blocs or countries as actors [3]. However, a sharp distinction between business and governmental actors is noticed, whereby the former is associated with sustainable business model creation, and the latter with policy development [7]. Finally, political participation and activism, for example via grassroot movements, and public goods approaches [27], are described to be required for fostering broad consensus on sustainability-specific matters [28]. To conclude, stakeholders in sustainability for ICT range from individuals to supra-national organizations; consequently, the range of ownership and responsibilities to drive sustainability outcomes is broad. Another emerging theme is the variety of *application* (A#1-4) areas to which sustainability is applied. Sustainability is applied to different *application sectors* (A#2) and *scenarios* (A#1) in conjunction with different ICT-enabled *technologies* (A#3) [6, 7, 25, 28]. This comprises sectors such as agriculture, rural communities, manufacturing, and logistics, libraries, digital learning, smart cities, healthcare, tourism, digital learning, production, or the energy sector [7]. Within each sector, ICT-supported sustainability solutions are applied, for example, for e-waste management, pollution control or efficient manufacturing [3] different technologies such as 3D-printing, IoT, automation or big data are used [6]. Finally, three reviews highlight the limited *geographical perspective* (A#4) and the lack of comparative research in understanding the relationship between digital transformation and sustainability. It is concluded that studying different countries and contexts is required to achieve a more robust and generalizable understanding [7, 30, 31] of sustainability in the context of ICT.

IV. Discussion

One of the earliest and a widely cited literature review in the sample dates from 2014. It is used to gauge the evolvements in the research field over the past ten years [16]. Table IV summarizes the sustainability dimensions presented as in [16], and specifies the thematic and conceptual evolvements that have occurred. New dimensions that have emerged over the past ten years are marked in grey and added to the row '*Additions*', which covers the novel aspects '*Application areas*' and '*Geo(graphic) perspective*'. The thematic and conceptual evolvements are summarized in the column '*Thematic and conceptual evolvements*' and are highlighted in grey, too. These additions address RQ3 on *how the relationship between sustainability and ICT has evolved over the past ten years*.

The conceptualization proposal provided in [16] first describes the conceptual misalignment of sustainability in ICT. It is found that academic literature on sustainability in the context of ICT implicitly assumes that the conceptual dimensions of the sustainability concept are common knowledge. As a result, no further specification or definition of the concept is provided, and only references to other sources, which attempt to clarify the concept, are provided [16]. This conceptual and terminological under-specification can also be confirmed in the literature reviews analyzed for this paper. However, the theme that has emerged in the context of terminological ambiguity is the need to align *the concepts of digitalization and ICT, digital sustainability, and digital transformations* (B#5) [7, 8, 18, 25, 29, 31].

It is also observed that the three sustainability pillars are widely used to describe sustainability in ICT. It is pointed out that the economic and ecological perspectives, referred to as eco-effectiveness and eco-efficiency goals, overlap [16]. This view is valid in the more recent literature reviews; however, the importance of a holistic approach to sustainability in ICT is emphasized, recognizing the interconnectedness of economic, social [18], and environmental pillars. It is widely acknowledged that these pillars cannot be considered in isolation and that social aspects, such as the social divide caused by digitalization [18, 19], are of utmost importance, although it is conceded that social implications of digitalization need further research [3, 6, 7, 25, 27].

The second aspect of the definition, as provided in [16], is that sustainability is attributed in the literature to four categories of *reference objects*: first, individual and organization stakeholders, who drive sustainable development. Second, enablers, which allow stakeholders to act in a sustainable manner. *Stakeholders* are differentiated into *individuals*, *individuals in organizations* and *organizations*. Third, *consequences*, which are the result of sustainable activities. Finally, *sustainable activities* are tied to all entities, that is stakeholders, actors, and consequences [16]. All these four categories of reference objects hold true – but the reference objects and their roles can be updated and augmented.

Table IV.
Conceptualization of sustainability in ICT and its evolvements over the past ten years to address RQ3

Dimensions		Findings from [16]	Thematic and conceptual evolvements
Dimensions of the concept of sustainability in context of ICT as in [16]	Definition	Implicit assumptions on sustainability dimensions prevail in analyzed literature.	Focus on how digitalization, ICT, digital sustainability, and digital transformations are linked and can be conceptualized.
	Pillar	Environmental and economic pillars overlap. Holistic approach is considered important.	Increased focus on interdisciplinary and holistic approaches. Topics such as social impacts, e.g., social divide, receive increased attention.
	Enablers	Allows stakeholder to act sustainably (ICT artifacts, sustainability goals, strategies, etc.)	The concept of ICT as enabler prevails, but there is increased attention on negative side-effects and consequences.
	Stakeholder	Individuals and individuals in organizations	Increased attention on role of civil society, organizational, and governmental stakeholders, and their importance for mitigation strategies. Acknowledging gap on how individual behaviors are impacted by beliefs or social opinions.
	Activity	Activities links stakeholders, and enablers to sustainable consequences.	Evolved conceptualization of mitigation strategies (sufficiency, consistency, efficiency). Relationship between sustainability and digitalization, and digital transformations poses new areas for research.
	Consequences	Result of a sustainable activity. No differentiation between positive or negative consequences.	Consensus that there is a need for a more balanced view, in which negative social consequences receive more attention. Novel measurement and assessment approaches (Life-cycle-, enabling-, structural effects).
Additions	Application areas	Sector or industry (e.g., agriculture, industry 4.0, healthcare, etc.)	
		Technologies (e.g., big data, machine learning, etc.)	
		Scenarios (e.g., e-waste management, pollution control, etc.)	
	Geo-perspective	Local, regional, national, transnational, global perspectives	

Due to the socio-technological development in the context of digitalization, the fundamental role of ICT in sustainability – ICT as enabler or part of the problem – remains a seminal strand of discussion and hence a key part of the sustainability concept in the context of ICT. The range of stakeholders is actively discussed. It is acknowledged that the roles of individuals and how their beliefs or social norms impact sustainability behaviors [27], the roles of policy makers in enforcing sustainability regulations [28], or the role of organizational stakeholders in initiating sustainability programs [25], are areas for further exploration. There are also advancements with regards to how different stakeholders can be mapped to different mitigation strategies [24].

In terms of consequences, the research field has provided advanced models and categorizations to better assess and understand impacts on the different sustainability pillars holistically. One example is the seminal LES model that differentiates life cycle, enabling- and structural effects [38].

The role of activities to achieve sustainability goals remains uncontested, but research has also delivered frameworks such as the sustainability mitigation strategies [24] – sufficiency, consistency, and efficiency – and impact measurement approaches [23] that are suited to guide and classify sustainability-oriented actions.

The role of ICT as an enabler is also further specified by analyzing concrete digital solutions enabled by novel digital technologies such as big data, or artificial intelligence in the context of a specific sector and specific scenarios are explicitly mentioned [6, 7]. Geographical aspects have gained more attention, which is also expressed in publication addressing specific national or regional aspects. Hence, the application scenarios, including sector-, scenario-, and technology-, and geographic-specific views, can be added as an additional dimension in a conceptualization matrix for sustainability in the context of ICT.

V. Conclusion

This paper aimed to address three research questions with regards to the concept of sustainability in the context of ICT.

For RQ1 it was shown that the role of ICT as either an enabler or problem for sustainability remains an important aspect in the discussion. ICT is predominantly seen as an enabler for sustainability that creates opportunities for efficiency-gains, energy reductions, or facilitates smart cities. However, the negative impacts such as the environmental footprint caused throughout the hardware lifecycle or the social divide created by dividing society into those participating in digitalization, and those who do not, attract increasing attention. Hence, practitioners and researchers should actively look for and transparently point out potential negative side-effects and consequences when planning to apply or investigate ICT-enabled solutions to achieve sustainability goals.

Regarding RQ2, the analysis confirmed that there is continuous terminological misalignment of sustainability in ICT, but with the important modification that the focus now is on integrating the concepts of digitalization and digital transformation, which is expressed by the term of *digital sustainability* [7, 25]. The overarching importance of an interdisciplinary and holistic approach to sustainability in the context of ICT and digitalization is stressed out as a requirement to address all three sustainability pillars. Therefore, it is important for both practitioners and researchers to look at sustainability in the context of ICT holistically and to tackle sustainability-related initiatives and projects in an interdisciplinary manner.

By addressing RQ3, it was revealed that important evolvements of the sustainability concept over the past ten years occurred: recent technologies, and new digital solutions, which are often global and international phenomena, broadened the range of application scenarios for the concept of sustainability. Also, the role of stakeholders and the mitigation strategies associated with or applied by them have seen further amendments and specifications.

As a result, future research should consider an interdisciplinary approach to better understand the complex connections and interdependencies between stakeholders, ICT, digitalization, and sustainability transformations to explore innovative solutions for sustainable digital transformation [28]. But to manage the scope and complexity of the topic, research should focus on specific application areas or technologies. More recent technologies such as artificial intelligence [7, 39–41] or digital twins are already being discussed in the context of sustainability [7, 42]. More targeted research can help narrow the research scope to an applicable and practical level. Another opportunity for further research is how stakeholders and practitioners in organizations can implement sustainability initiatives in their respective organizations and then measure the benefits that those initiatives yield [25] .

The limitations of this study are its selective approach and its focus on the most recent literature in the field: articles published between 2021 and 2023 represent 81% of the examined corpus. The sampling step can result in the exclusion of relevant content and hence it must be acknowledged that synthesizing more papers could have attributed additional insights. Also, the terminological misalignment and the absence of well-established keywords for research on sustainability in context of ICT and digitalization can result in excluding pertinent literature reviews. These shortcomings and additional potential gaps are an opportunity for further research.

Appendix

Figure 2 summarizes the total count of findings and gaps, mapped to the five synthesized themes.

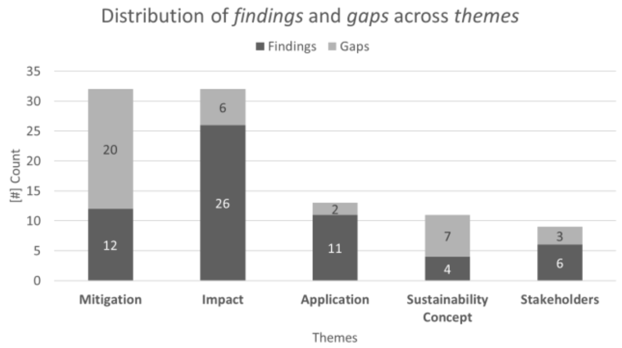


Fig 2. Distribution of themes

Table V lists the sampled literature reviews including coding results, differentiated by themes, concepts, as well as findings and gaps.

TABLE V.
LITERATURE CODING RESULTS

Theme		Application					Sustainability concept			Impact						Mitigation				Stakeholder	
Concepts		A#1	A#2	A#3	A#4		B#5		B#6	C#7	C#8	C#9		#C10		D#11		D#12		E#13	
Finding (F) / Gap (G)		F	F	F	F	G	F	G	F	F	F	F	G	F	G	F	G	F	G	F	G
Author	Published																				
J. Bieser and L. Hilty	2018													2		1					
F. Chasin	2014								1								1		1	2	
A. K. Feroz, H. Zo, and A. Chiravuri	2021	1						1									2				
I. Guandalini	2022		1	1	1			2											4	1	1
Y. Karki and D. Thapa	2021					1	1	3									2	1			
J. Kotlarsky, I. Oshri, and N. Sekulic	2023			1			1			1					1				1	1	1
A. Kuntsman and I. Rattle	2019	1								2	1	3		1							
S. Lertpiromsuk and P. Ueasangkomsate	2022					1				3											
M. Mouthaan, K. Frenken, L. Piscicelli, and T. Vaskelainen	2023									1				1				3	1	1	
P. Perera, S. Selvanathan, J. Bandaralage, and J.-J. Su	2023	1			1		1				1										
G. Robertsone and I. Lapiņa	2023			1				1		3			1								
A. Rosário and J. Dias	2023									4						1		2			
P. Sacco, E. R. Gargano, and A. Cornella	2021			1								1			2			1		1	
T. Santarius and J. Wagner	2023	1										1	1			1	1		3		1
G. D. Sharma, D. Reppas, G. Muschert, and V. Pereira	2021																	2	1		
D. J. Veit and J. B. Thatcher	2023													1	1		2		1		

[A] Application – (1) *Application scenarios*, (2) *Application sectors*, (3) *Application technology*, (4) *Geographic perspectives*

[B] Sustainability concept – (5) *Need to align concepts of digitalization, ICT, digital sustainability, and digital transformations*, (6) *Terminological misalignment sustainability in the context of ICT*.

[C] Impact – (7) *ICT as enabler*, (8) *ICT as problem*, (9) *ICT as problem and enabler*, (10) *Measurement of impacts of ICT on sustainability*

[D] Mitigation – (11) *Mitigation strategies to address sustainability challenges*, (12) *Need for an interdisciplinary and holistic approach to sustainability in ICT*

[E] Stakeholders: (13) Role of stakeholders and governance

REFERENCES

[1] S. Lenz, "Is digitalization a problem solver or a fire accelerator? Situating digital technologies in sustainability discourses," *Social Science Information*, vol. 60, no. 2, pp. 188–208, 2021, doi: 10.1177/05390184211012179.

[2] M. Stojanovic, "Philosophy of sustainability experimentation _ experimental legacy, normativity and transfer of evidence," *European Journal for Philosophy of Science*, vol. 11, no. 3, pp. 1–22, 2021, doi: 10.1007/s13194-021-00383-4.

[3] A. Rosário and J. Dias, "Sustainability and the digital transition: A literature review," *Sustainability*, vol. 14, no. 7, p. 4072, 2022, doi: 10.3390/su14074072.

[4] J. Köhler et al., "An agenda for sustainability transitions research: State of the art and future directions," *Environmental Innovation and Societal Transitions*, vol. 31, pp. 1–32, 2019, doi: 10.1016/j.eist.2019.01.004.

[5] T. T. Lennerfors, P. Fors, and J. van Rooijen, "ICT and environmental sustainability in a changing society," *Information Technology & People*, vol. 28, no. 4, pp. 758–774, 2015, doi: 10.1108/ITP-09-2014-0219.

[6] P. Sacco, E. R. Gargano, and A. Cornella, "Sustainable digitalization: A systematic literature review to identify how to make digitalization more sustainable," in *Creative Solutions for a Sustainable Development*, pp. 14–29, 2021, doi: https://doi.org/10.1007/978-3-030-86614-3_2.

[7] I. Guandalini, "Sustainability through digital transformation: A systematic literature review for research guidance," *Journal of Business Research*, vol. 148, pp. 456–471, 2022, doi: 10.1016/j.jbusres.2022.05.003.

[8] A. K. Feroz, H. Zo, and A. Chiravuri, "Digital transformation and environmental sustainability: A review and research agenda," *Sustainability*, vol. 13, no. 3, pp. 1–20, 2021, doi: 10.3390/su13031530.

[9] M. Mouthaan, K. Frenken, L. Piscicelli, and T. Vaskelainen, "Systemic sustainability effects of contemporary digitalization: A scoping review and research agenda," *Futures*, vol. 149, p. 103142, 2023, doi: 10.1016/j.futures.2023.103142.

[10] C. Legner et al., "Digitalization: Opportunity and challenge for the business and information systems engineering community," *Business & Information Systems Engineering*, vol. 59, no. 4, pp. 301–308, 2017, doi: 10.1007/s12599-017-0484-2.

[11] Y. K. Dwivedi et al., "Climate change and COP26: Are digital technologies and information management part of the problem or the solution? An editorial re-flection and call to action," *International Journal of Information Management*, vol. 63, p. 102456, 2022, doi: 10.1016/j.ijinfomgt.2021.102456.

[12] T. Kuhlman and J. Farrington, "What is sustainability?," *Sustainability*, vol. 2, no. 11, pp. 3436–3448, 2010, doi: 10.3390/su2113436.

[13] B. Purvis, Y. Mao, and D. Robinson, "Three pillars of sustainability: in search of conceptual origins," *Sustainability Science*, vol. 14, no. 3, pp. 681–695, 2019, doi: 10.1007/s11625-018-0627-5.

[14] F. Popa, M. Guillermin, and T. Dedeurwaerdere, "A pragmatist approach to transdisciplinarity in sustainability research: From complex systems theory to reflexive science," *Futures*, vol. 65, pp. 45–56, 2015, doi: 10.1016/j.futures.2014.02.002.

[15] S. S. Vildåsen, M. Keitsch, and A. M. Fet, "Clarifying the epistemology of corporate sustainability," *Ecological Economics*, vol. 138, pp. 40–46, 2017, doi: 10.1016/j.ecolecon.2017.03.029.

[16] F. Chasin, "Sustainability: Are we all talking about the same thing? State-of-the-art and proposals for an integrative definition of sustainability in information systems," in *Proceedings of the 2014 conference ICT for Sustainability*, Stockholm, Sweden, 2014, pp. 342–351.

[17] J. F. Wolfswinkel, E. Furtmueller, and C. P. M. Wilderom, "Using grounded theory as a method for rigorously reviewing literature," *European Journal of Information Systems*, vol. 22, no. 1, pp. 45–55, 2013, doi: 10.1057/ejis.2011.51.

[18] P. Perera, S. Selvanathan, J. Bandaralage, and J.-J. Su, "The impact of digital inequality in achieving sustainable development: a systematic literature review," *EDI*, 2023, doi: 10.1108/EDI-08-2022-0224.

[19] A. Kuntsman and I. Rattle, "Towards a paradigmatic shift in sustainability studies: A systematic review of peer reviewed literature and future agenda setting to consider environmental (un)sustainability of digital communication," *Environmental Communication*, vol. 13, no. 5, pp. 567–581, 2019, doi: 10.1080/17524032.2019.1596144.

[20] J. A. S. Laitner, "The energy efficiency benefits and the economic imperative of ICT-enabled systems," in *ICT Innovation for Sustainability. Advances in Intelligent Systems and Computing*, pp. 37–48, Jan. 2015, doi: https://doi.org/10.1007/978-3-319-09228-7_2.

[21] M. Omurgonulsen, M. Ibis, Y. Kazancoglu, and P. Singla, "Cloud computing," *Journal of Global Information Management*, vol. 29, no. 6, pp. 1–25, 2021, doi: 10.4018/JGIM.20211101.oa40.

[22] L.-D. Radu, "Green cloud computing: A literature survey," *Symmetry*, vol. 9, no. 12, p. 295, 2017, doi: 10.3390/sym9120295.

[23] J. Bieser and L. Hilty, "Assessing indirect environmental effects of information and communication technology (ICT): a systematic literature review," *Sustainability*, vol. 10, no. 8, p. 2662, 2018, doi: 10.3390/su10082662.

[24] T. Santarius et al., "Digital sufficiency: conceptual considerations for ICTs on a finite planet," *Annals of Telecommunications*, pp. 1–19, 2022, doi: 10.1007/s12243-022-00914-x.

[25] J. Kotlarsky, I. Oshri, and N. Sekulic, "Digital Sustainability in Information Systems Research: Conceptual Foundations and Future Directions," *Journal of the Association for Information Systems*, vol. 24, no. 4, pp. 936–952, 2023, doi: 10.17705/1jais.00825.

[26] G. D. Sharma, D. Reppas, G. Muschert, and V. Pereira, "Investigating digital sustainability: A retrospective bibliometric analysis of literature leading to future research directions," *First Monday*, 2021, doi: 10.5210/fm.v26i11.12355.

[27] D. J. Veit and J. B. Thatcher, "Digitalization as a problem or solution? Charting the path for research on sustainable information systems," *Journal of Business Economics*, pp. 1–23, 2023, doi: 10.1007/s11573-023-01143-x.

[28] T. Santarius and J. Wagner, "Digitalization and sustainability: A systematic literature analysis of ICT for Sustainability research," *GAIA - Ecological Perspectives for Science and Society*, vol. 32, S1, pp. 21–32, 2023, doi: 10.14512/gaia.32.S1.5.

[29] G. Robertsone and I. Lapiņa, "Digital transformation as a catalyst for sustainability and open innovation," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 9, no. 1, p. 100017, 2023, doi: 10.1016/j.joitmc.2023.100017.

[30] S. Lertpiromsuk and P. Ueasangkomsate, "Digitalization, sustainability and innovation: A systematic literature review," in *2022 IEEE Technology & Engineering Management Conference - Asia Pacific (TEMSCON-ASPAC)*, Bangkok, Thailand, 2022, pp. 84–88.

[31] Y. Karki and D. Thapa, "Exploring the link between digitalization and sustainable development: Research agendas," *Responsible AI and Analytics for an Ethical and Inclusive Digitized Society*, pp. 330–341, 2021, doi: https://doi.org/10.1007/978-3-030-85447-8_29.

[32] S. Baumgärtner and M. Quaas, "What is sustainability economics?," *Ecological Economics*, vol. 69, no. 3, pp. 445–450, 2010, doi: 10.1016/j.ecolecon.2009.11.019.

[33] E. Meyer and D. Peukert, "Designing a transformative epistemology of the problematic: A perspective for transdisciplinary sustainability research," *Social Epistemology*, vol. 34, no. 4, pp. 346–356, 2020, doi: 10.1080/02691728.2019.1706119.

[34] A. S. Mitchell, M. Lemon, and W. Lambrechts, "Learning from the anthropocene: Adaptive epistemology and complexity in strategic managerial Thinking," *Sustainability*, vol. 12, no. 11, p. 4427, 2020, doi: 10.3390/su12114427.

[35] M. Nagatsu et al., "Philosophy of science for sustainability science," Sustainability Science, vol. 15, no. 6, pp. 1807–1817, 2020, doi: 10.1007/s11625-020-00832-8.

[36] W. A. Salas-Zapata and S. M. Ortiz-Muñoz, "Analysis of meanings of the concept of sustainability," *Sustainable Development*, vol. 27, no. 1, pp. 153–161, 2019, doi: 10.1002/sd.1885.

[37] W. A. Salas-Zapata, L. A. Ríos-Osorio, and J. A. Cardona-Arias, "Methodological characteristics of sustainability science: a systematic review," *Environment, Development and Sustainability*, vol. 19, no. 4, pp. 1127–1140, 2017, doi: 10.1007/s10668-016-9801-z.

[38] L. M. Hilty and B. Aebischer, "ICT for sustainability: An emerging research field," in *ICT Innovation for Sustainability. Advances in Intelligent Systems and Computing*, pp. 3–36, 2015, doi: https://doi.org/10.1007/978-3-319-09228-7_1.

[39] T. Schoormann, G. Strobel, F. Möller, D. Petrik, and P. Zschech, "Artificial Intelligence for sustainability—a systematic review of information systems literature*," Communications of the Association for Information Systems*, vol. 52, pp. 199–237, 2023, doi: 10.17705/1CAIS.05209.

[40] S. L. Pan and R. Nishant, "Artificial intelligence for digital sustainability: An insight into domain-specific research and future directions," *International Journal of Information Management*, vol. 72, p. 102668, 2023, doi: 10.1016/j.ijinfomgt.2023.102668.

[41] R. Nishant, M. Kennedy, and J. Corbett, "Artificial intelligence for sustainability: Challenges, opportunities, and a research agenda," *International Journal of Information Management*, vol. 53, p. 102104, 2020, doi: 10.1016/j.ijinfomgt.2020.102104.

[42] S. S. Kamble, A. Gunasekaran, H. Parekh, V. Mani, A. Belhadi, and R. Sharma, "Digital twin for sustainable manufacturing supply chains: Current trends, future perspectives, and an implementation framework," *Technological Forecasting and Social Change*, vol. 176, p. 121448, 2022, doi: 10.1016/j.techfore.2021.121448.

DOI: 10.15439/2023F7674
ISSN 2300-5963 ACSIS, Vol. 37

FAS-CT: FPGA-Based Acceleration System with Continuous Training

1st Manuel L. González
ITCL
Burgos, SPAIN
manuel.gonzalez@itcl.es

2st Jorge Ruiz
ITCL
Burgos, SPAIN
jorge.ruiz@itcl.es

3st Randy Lozada
ITCL
Burgos, SPAIN
randy.lozada@itcl.es

4st E.S. Skibinsky-Gitlin
ITCL
Burgos, SPAIN
erik.skibinsky@itcl.es

5th Ángel M. García-Vico
Dept. of Communication and Control Systems
UNED, Spain
amgarcia@scc.uned.es

6st Javier Sedano
ITCL
Burgos, SPAIN
javier.sedano@itcl.es

7st José R. Villar
University of Oviedo
Oviedo, SPAIN
villarjose@uniovi.es

***Abstract*—This paper presents FAS-CT a novel approach to a distributed low-latency Deep Learning inference system based on a Field Programmable Gate Array (FPGA). The system incorporates continuous training capabilities based on Concept Drift Detection, where each model prediction is compared with the ground truth to detect a change in the data patterns that the model requires to adapt to. FAS-CT is formed by two main execution pipelines. Firstly, the prediction pipeline is powered with Xilinx® Zynq® UltraScale+™ MPSoC FPGA and where low latency is the target. Secondly the Retraining pipeline aims adapting the model the model when Concept Drift is detected. A complete characterization of FAS-CT is provided in this article using a neural network model and an experimental setup. The latency of the Prediction pipeline achieved was 5.79 ms. The total degradation of the model when continuous training is activated is 57% in contrast to when is deactivated which is 1609%. These results demonstrate that FAS-CT is suited for real-time Deep Learning inference and can be automatically adapted to evolving data environments.**

I. Introduction

IN RECENT times, Deep Learning (DL) has rapidly emerged as a powerful tool, demonstrating unparalleled potential in domains such as computer vision, natural language processing, and predictive analytics, surpassing traditional machine learning techniques [1]. Conventionally, cloud computing has been the favored approach for deploying DL models for such applications, harnessing the vast processing power and storage capabilities of data centers [2]. However, the growth in data traffic coupled with the stringent low-latency requirements of various DL services has begun to challenge this centralized computing approach [3].

The Edge Computing (EC) paradigm has gained prominence, offering a solution to these challenges by processing data closer to its source. EC is a decentralized paradigm that places computational resources, memory, and services closer to the data origination point, thereby accelerating response times and reducing the burden on communication bandwidth. Despite its potential, EC poses its own challenges, particularly in terms of limited computational power and memory resources when compared to cloud-based systems [4]. These EC limitations have a direct impact on DL solutions. Real-time inference or model adaptation to new data could be compromised.

Executing DL models on EC devices in order to achieve real-time inference is a non-trivial task due to the inherent resource constraints. Despite optimization for edge deployment [5], these models continue to demand substantial computational resources. Model adaptation represents another significant challenge within the EC paradigm. DL models necessitate continuous updates to maintain relevance in rapidly evolving data environments or in the presence of Concept Drift (CD) [6]. These widely recognized problems of DL models on EC have been investigated in literature [5] and are typically addressed through an orchestration system architecture or continuous training techniques.

Orchestration architecture, a key theme across several articles, is primarily used to manage and optimize distributed resources. The research in [7] uses it for distributing processing between EC devices and the cloud in a real-time image recognition system. In [8], the authors use an orchestration architecture for continuous training by integrating new data into existing models, whereas in [9], it enables continuous updates to the seizure prediction model. However, it is noteworthy to mention that these approaches typically focus either on orchestration or on continuous training, but rarely integrate both elements effectively. This highlights the novelty of the current work, which aims to bridge this gap by developing a mechanism capable of detecting CD and performing continuous training while managing and optimizing resources through an orchestration system architecture. These studies emphasize the importance of orchestration architecture in managing complex distributed systems and enabling continuous training. This topic is particularly interesting, as highlighted in [8], [9], and [10]. [8] and [9] demonstrate its importance in medical applications, where models are continually updated with new data to improve accuracy. [10] extends this idea to IoT devices, introducing a loss compensation mechanism to improve Federated Incremental Learning, highlighting the applicability of continuous training across various fields.

On DL there are different types of models each one with a target application [1]. Those DL architectures have been implemented successfully in different EC devices. Implementations of Fully Connected Networks (FCN) have been explored in [11], [12]. Convolutional Neural Networks (CNN) on [7], [13], [8] and Recurrent Neural Networks (RNN) particularly Gated Recurrent Units (GRU) and Long Short Term Memory (LSTM), are used in [14], [15], [16]. Typically used EC devices for DL implementation have coupled processing technologies like CPU + GPU [17], [18], CPU + FPGA [13], [19] or CPU + TPU [18], [20]. CPU's mean purpose is to manage data and connections with orchestration architecture, meanwhile, GPU, FPGA, or TPU are used to accelerate DL model inference.

The use of FPGA is explicitly discussed in [11], which presents ZyNet to automate FCN implementation on low-cost FPGA platforms. This tool facilitates the deployment of FCN in edge computing devices and is a promising approach for making FPGA-based DL computing more accessible. In [13] an FPGA with LeNet-5 is studied. This article implements an autonomous architecture with continuous training of the model in a Xilinx® Multi-Processing System on Chip (MPSoC) device. The training is performed in MPSoC CPU and the inference is executed on FPGA. This solution leads to inference times of 2.2 ms and training times of 286s. Also in [12], [16] a Xilinx® FPGA is used to implement inferences of FCN and LSTM, their performance is analyzed. These implementations have maximum inference times of 1.09ms for FCN and 2.6ms for LSTM.

In the present work we introduce the FPGA-Based Acceleration System with Continuous Training (FAS-CT). This novel EC architecture is designed for the orchestration of DL model inference on FPGA, explicit CD detection, and continuous training. As is exposed, DL model inference on FPGA can yield low-latency responses. The CD is explicitly identified to monitor any degradation in the model's performance. If CD is detected, the retraining stage of the DL model is launched. This process allows continuous training of the model and its automatic update in FPGA. The main contributions of this article are listed as follows:

1) We put forth an architecture that efficiently coordinates various technologies best suited for different tasks. FPGA for DL model inference, GPU for DL model training, and CPU for preprocessing, postprocessing, CD detection, and data communication.
2) A complete description and characterization of FAS-CT is presented. The description of each component and the interaction between them is detailed. The characterization is examined with real-world data concerning response time, model performance, and model updates.
3) CD detection is included in the architecture to perform model retraining only when needed. This stage in the orchestration scheme allows for saving energy because a power-hungry GPU is used only when the model performance is worsening.

The rest of the article is organized as follows. Section II details a complete description of FAS-CT architecture, focusing on different technologies for each component. Section III explains CD detection and its implementation in a module on FAS-CT. Section IV explains the implementation of DL model inference on FPGA and its communication with FAS-CT. Section V focuses on the setup and the experiments performed on FAS-CT to get a complete characterization of the architecture. Section VI exposes the results of metrics defined in the previous section. The article finishs in section VII with conclusions and future research work.

II. Continuous Training System Architecture

FAS-CT architecture is designed around a central orchestration framework that maximizes the benefits of each technology it incorporates. It leverages FPGA for real-time neural network inference, GPU for model training, CPUs for pre and post-processing of data, CD detection, and management of data communication. This collaborative design facilitates high performance and ensures seamless integration of these key processes. FAS-CT has been designed to enable easy adaptation to diverse hardware, software, and data configurations. Each step can be deployed on separate hardware, thus satisfying most latency, performance, or throughput requirements with ease. For instance, feature preprocessing, inference of the neural network, and result postprocessing can be performed near the data source or prediction consumer, while feature storage and model retraining can be executed on more powerful devices like computer servers with GPU.

FAS-CT is composed of different stages or modules. Each of the stages shown in Fig.1 is handled by a different service. Modules are grouped in two pipelines. The first pipeline is responsible for inference in FPGA. The second pipeline is responsible for retraining the model when CD is detected.

To facilitate the service deployment, management, and monitoring, Docker [21] has been used for each module. A description of the task and purpose for each stage is given below.

A. Data Propagation

To communicate the different stages in FAS-CT the Message Queuing Telemetry Transport (MQTT) [22] protocol has been used. MQTT is a lightweight messaging protocol based on the publish-subscribe pattern. The protocol operates on top of the TCP/IP network stack and has support for multiple Quality of Service (QoS) levels to ensure reliable message delivery.

To manage the message queues the open-source Eclipse Mosquitto [23] MQTT Broker has been used. Mosquitto is licensed under EPL2, and it is one of the most suitable MQTT brokers due to its high performance [24], being multi-platform MQTT 5 compliant and having Transport Layer Security (TLS) support. The data has been serialized using Google Protocol Buffers [25].

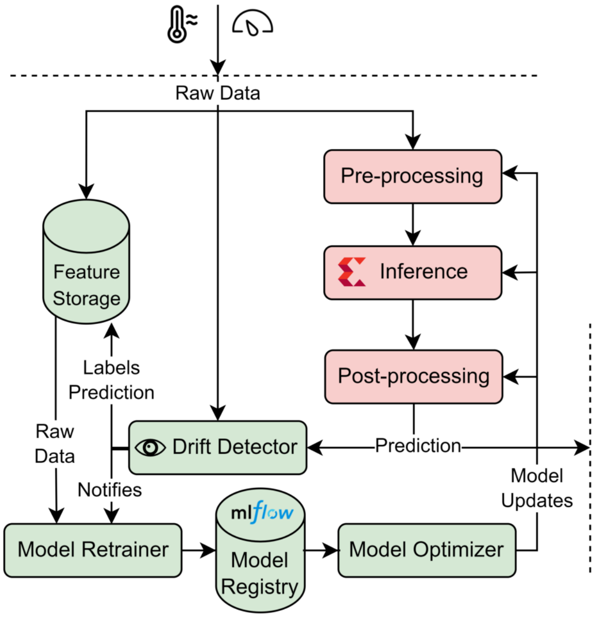


Fig. 1. FAS-CT architecture diagram. There are eight modules grouped in two pipelines. Modules of the Prediction pipeline are highlighted in red and the Retraining pipeline in green. The inference module is executed in FPGA, he Model Retrainer module in GPU and the rest of modules in CPU.

B. Feature Store

A database records each set of model input features with a unique sequential ID, along with the corresponding ground truth, model inference result, and drift level of the model for that prediction. The database can be queried using a set of remote procedure calls (RPC) [25] to retrieve the various features required for the model retraining process.

C. Feature Inference Preprocessing

The feature inference preprocessing stage involves a series of transformations that are applied to the raw data to make it compatible with the machine learning models. This stage may include data cleaning, where inconsistencies, errors, and outliers in the data are identified and corrected or removed. Another common step is data normalization or standardization, which is crucial for algorithms that are sensitive to the scale of the features. This process adjusts the values of numeric features so that they share a common scale, without distorting the differences in the ranges of values or losing information. Feature engineering is another integral part of preprocessing, creating new features based on existing ones, which can enhance the predictive power of the machine learning model.

D. FPGA Edge Inference

An FPGA receives a set of tensors from the preprocessing stage which serves as inputs for the neural network. These tensors are essentially multi-dimensional arrays of data, prepared and structured to be ingested by the model for making predictions.

Simultaneously, the device also listens for model updates from the continuous training pipeline. Upon receiving these model updates, the FPGA substitutes the current model with the new model. Essentially the model is evolving its ability to make accurate predictions in line with the most recent trends in the data. This process of continuous listening and updating ensures that the model deployed on the FPGA is always synchronized with the most recent version and maintains accuracy even in the face of changing data landscapes.

E. Result Postprocessing

Certain models may need a postprocessing step to enable an accurate comparison between the ground truth and the prediction. For instance, if the preprocessing stage involved scaling or standardization of features, an inverse transformation might be necessary for the postprocessing stage to convert predictions back to the original scale. In this way, the predictions can be compared with the ground truth. Another common postprocessing step involves the treatment of probability outputs. Many machine learning models, especially in classification tasks, output probabilities of each class. A thresholding operation might be necessary to convert these probabilities into discrete class labels. The choice of threshold can significantly affect the model's performance metrics and can be fine-tuned based on the requirements of the specific task.

F. Drift Detector

This module is continuously monitoring the error that the neural network is generating. If the error is between some limits or thresholds, the model is considered to be providing a correct prediction. If the error increases, CD is detected and the Retraining pipeline is executed.

There are several Drift Detectors that can be placed at this stage. For FAS-CT we choose Drift Detector Method. This is an algorithm developed by J. Gama et al. [26]. It is computationally lightweight and has low memory requirements, in line with the two main constraints in EC. A description of this algorithm is detailed in Section III.

G. Model Retraining, Validation and Registry

Upon a Drift Detector notification, the Model Retraining stage updates the scalers and the neural network to fit the newest data. The data has been stored properly in the Feature Storage module and is served to perform new training on the model.

The validation process of the updated model entails a comparison between the Drift Detector error metric of both old and new models. If the error metric of the new model is less than the current drifted mean, the updated model is stored in the model registry and publishes a model update on FAS-CT. However, if the error metric of the new model is assessed as an improvement, the model is stored.

The experiments, models, and scalers are tracked by MLFlow [27]. MLFlow is an open-source platform for machine learning workflows that includes features such as experiment and model registry, allowing for efficient management of models. Furthermore, MLFlow's experiment tracker stores and organizes all the models, data, and metrics of a retraining process.

H. Model Optimization and Update

Upon completion of the model's retraining and validation process, the subsequent step involves its conversion and optimization into a specific format that can be consumed by edge devices like FPGA. The new model is serialized in JSON format and it is sent to the FPGA as detailed above.

The model update could also involve changes in Preprocessing and Post-processing stages. In this case, some functions like scalers must be updated. Finally, if the model is retrained, the Drift Detector is also reset with new parameters. This process is explained in the following section.

III. Concept Drift Detection

CD refers to the phenomenon where the statistical properties of data, on which the model has been trained, change over time in unforeseen ways, causing the model's performance to degrade. This happens because most predictive models are designed and trained under the assumption that future patterns will remain consistent with historical ones, which is often not the case in real-world scenarios. Real-world data is continually evolving and changing, and so too is the context in which DL models operate. Changes can occur in various forms e.g. gradual, abrupt, incremental, or recurring changes [6]. Different types of CD exist, such as real, virtual, and dual CD [6], [28]. In this work, only real CD will be considered and will be referred as CD for the sake of simplicity.

DL models, though powerful and highly accurate, have a significant weakness when it comes to CD. Detecting CD and subsequently retraining the models can be a solution to mitigate this impact. The process typically involves monitoring the model's performance metrics over time, and if a significant decline is detected, it's an indication that CD might be occurring. Once identified, the model can be retrained with the latest data, which reflects the new patterns. By applying this continuous training approach, DL models can become more adaptable to the evolving nature of real-world data.

The CD detection method implemented in the Drift Detector module, shown in Fig. 1, is similar to the Drift Detection Method (DDM) proposed by [26]. This method is based on the error signal produced by a binary classifier. The error signal is the probability of misclassifying an instance plus the standard deviation. The error signal fits a Bernoulli distribution because a binary classifier is assumed in [26]. DDM can be extended to forecasting or regression models by monitoring the error in prediction. On these models, DDM studies error signal mean and standard deviation based on a Gaussian distribution:

$$e_n = y_n^{true} - y_n^{pred} \tag{1}$$

$$\mu_n = \frac{n-1}{n} \cdot \mu_{n-1} + \frac{1}{n} e_n \tag{2}$$

$$\sigma_n = \sqrt{\frac{n-1}{n}\sigma_{n-1}^2 + \frac{1}{n-1}\left(e_n - \mu_n\right)^2} \tag{3}$$

$$ddm\_e_n = \mu_n + \sigma_n \tag{4}$$

Where $\mu_0(e) = \sigma_0(e) = 0$ and n is the number of monitored predictions. Recurrent formulas for μ_n and σ_n are used to avoid storing previous values of the error and satisfy the memory requirements of the processing system. These formulas are derived in detail in Appendix A. The value ddm_e_n is known as the DDM error metric and is used to determine when CD is detected. Notice that for DDM it is necessary to have the y^{true} value. Particularly, this is possible for a forecasting task on time-series data because y^{true} will be available at a certain time. Two configuration parameters are needed μ_{min} and σ_{min}. These parameters are the minimum mean and the minimum standard deviation calculated during the training process. After that, DDM is configured and starts monitoring the model. The warning level is triggered if:

$$ddm\_e_n >= \mu_{min} + 2 \cdot \sigma_{min} \tag{5}$$

At the warning level, the performance of the DL model is starting to worsen and the CD may arise. To adapt the model input and target data are stored to retrain the model. The drift level is triggered if:

$$ddm\_e_n >= \mu_{min} + 3 \cdot \sigma_{min} \tag{6}$$

At the drift level, CD is detected, retrain is performed with stored data, DL model adaptation is executed and DDM parameters are restored. In FAS-CT the new model is changed on FPGA and the inference is executed with the new adapted model. This is an endless loop of inferring, monitoring, retraining, and adapting the DL model that could generate an updated response in an evolving environment.

IV. FPGA Inference

The edge inference module in the prediction pipeline shown in Fig. 1 is implemented using an FPGA device. These devices are highly versatile integrated circuits that can be reconfigured and programmed to perform specific tasks, making them ideal for application acceleration, including neural network inference [11]. FPGA devices offer several advantages for such tasks. First, their parallel processing architecture allows multiple operations to be performed efficiently and simultaneously, resulting in high throughput and low latency [13], [16]. This is particularly beneficial for neural network inference, which involves intensive matrix calculations. In addition, FPGAs offer the flexibility to customize hardware designs, enabling the implementation of highly optimized neural network architectures tailored to specific application requirements. The ability to fine-tune hardware resources at the circuit level enables efficient utilization of FPGA resources, resulting in improved power efficiency [12]. In addition, FPGAs can be integrated with existing systems, including CPUs and GPUs, to leverage their respective strengths in a heterogeneous computing environment. Overall, the programmability, parallelism, customization, and integration capabilities of FPGAs make them a compelling choice for accelerating neural network

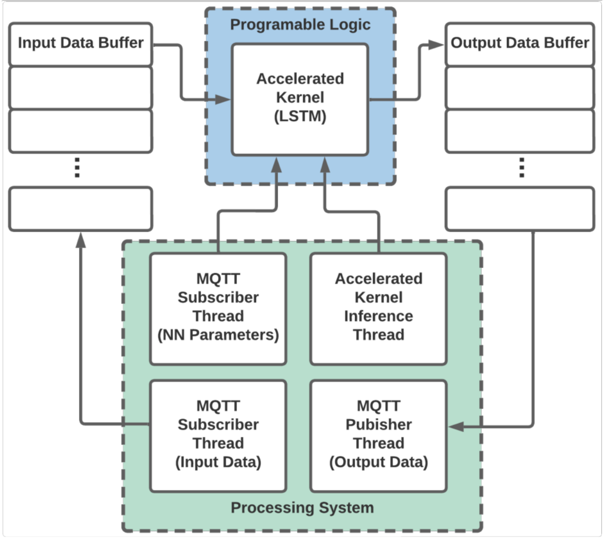

Input Data Buffer
Programable Logic
Accelerated Kernel (LSTM)
Output Data Buffer
MQTT Subscriber Thread (NN Parameters)
Accelerated Kernel Inference Thread
MQTT Subscriber Thread (Input Data)
MQTT Pubisher Thread (Output Data)
Processing System

Fig. 2. FPGA inference diagram

inference, offering significant performance gains and energy efficiency for a wide range of applications.

The inference stage must be able to process various data sent by the FAST-CT using the MQTT protocol. The transmitted data are in JSON format. The FPGA processing system must be in charge of extracting the necessary information to carry out the inference on the input data and the configuration of the corresponding neural network parameters. Likewise, the FPGA device transmits the result of the inference of each set of data received. The neural network is implemented on an acceleration kernel in the programmable logic side of the FPGA. This kernel communicates with the processing system and accelerates the inference task.

For the management of all these tasks, the architecture shown in Fig. 2 has been implemented. Four processing threads are used for data processing, control of the inference process, and configuration of the acceleration kernel. Three threads manage the data received and sent by MQTT. The other one is in charge of the inference execution. As can be seen in Fig. 2, two buffers are used for the synchronization between the tasks of reading input data, inference, and writing output data.

The inference process in the prediction pipeline on FAS-CT works as follows. The MQTT subscriber thread for input data writes the inference data to the input buffer. The accelerated kernel inference thread orders the execution of the acceleration kernel and stores the result in the output data buffer. Finally, the MQTT publisher thread for output data sends the results of the inference to continue the prediction pipeline. The FPGA device is also integrated into the retraining pipeline. In the fourth thread, the MQTT subscriber for neural network parameters is listening for updates in neural network weights. These updates are codified in JSON format. This thread manages the configuration of the acceleration kernel parameters and sets it for the next execution.

V. Materials and Methods

A. Setup

In this study, we employed FAS-CT with an LSTM neural network for time series forecasting. Due to client confidentiality, the data and results of the experiment have been anonymized. The purpose of the network is to forecast the value of a single sensor with a prediction horizon of 10 minutes. The model has been trained with a 32-minute sliding window of 4 different correlated sensors with a sample rate of 1 minute. All data is stored in a database so it is available for any experiment.

The LSTM network is a sequential model with an input LSTM layer of 32 units followed by a fully connected layer of 16 neurons with $tanh$ activation function and a single neuron as output with a $linear$ activation function. The neural network training uses the mean squared error loss function and Adam as the optimizer. The training process is limited to 100 epochs, with an early stopping of 10 epochs of patience.

The implementation of all the modules specified in Section II, with the exception of the inference kernel, are implemented in Python using common libraries like Paho-MQTT, Scikit Learn, GRPC, SQL Alchemy, and Tensorflow. The Preprocessing module receives data from 4 different sensors and appends them into a First-in-First-out queue of size 32 working as a sliding window. The data is then scaled into the interval $[0,1]$ using the MinMaxScaler algorithm. The Post-Processing module receives the network result and implements the inverse scale transformation.

The description of the acceleration kernel has been performed using HLS in the Xilinx® Vitis™ HLS development environment. This acceleration kernel describes the LSTM neural network with a 32-cell LSTM layer, 4 input features, and a 32-sample time window. This LSTM network also has two dense layers, the first of 16 neurons and the second of one neuron. For the implementation of the MQTT communication protocol, the MQTT-C [29] library has been used.

All the modules, with the exception of the inference, are running on a local host PC inside Docker containers on top of a Linux OS on a CPU Intel Core i7-13700k, a GPU Nvidia RTX 3060-12GB, and a memory RAM of 32 GB. The Model Retrainer uses the power of the GPU to accelerate the training of the neural network. The FPGA is connected to the local LAN network via Ethernet. The FPGA used for the LSTM implementation is the Ultra96v2 evaluation board which contains a Xilinx® Zynq® UltraScale+™ MPSoC device.

B. Experiments and Characterization

To evaluate the performance of the FAS-CT, backtesting experiments have been executed. Different metrics have been monitored during experiments. To characterize the prediction pipeline, the latency of each process, and communications are measured. To characterize the retraining pipeline, the error of the model and the number of retrains are monitored. On the retraining pipeline, the focus is on studying the DDM Drift Detector because it is the module in charge of executing the retraining.

A base model was trained offline with the first 5% of the available data whereas the rest was used for backtesting experiments. All the experiments have been performed using a configuration of 1000 input samples per minute. Three distinct experiments were conducted. The initial test use static scalers, which were initially fitted with the feasible range of values that each sensor can detect to prevent any bias from the training set. The second test involved employing dynamic scalers, during the retrain step the scalers are fitted with the updated train dataset. The final test examined the behavior of the system without any retraining, serving as a baseline. The three tests were conducted using the same base model and initial scalers.

VI. Results and Discussion

A. Latency

The latency of different modules and communications has been measured in the prediction pipeline. The method was to calculate the difference between input and output message timestamps for each module. As a control for communication, the latency between the FAS-CT host and the FPGA was measured using a ping command. The ping package yielded an average latency of 1.253 ± 0.614 milliseconds.

TABLE I
Latency of each process for prediction pipeline in FAS-CT

	Process Name	Process Type	Technology	Latency (ms)
1	Pre-processing	Execution	CPU	0.856 ± 0.457
2	Pre-P → Infer	Communication	CPU	1.442 ± 0.834
3	Inference	Execution	FPGA	1.894 ± 0.085
4	Infer → Post-P	Communication	CPU	1.372 ± 0.890
5	Post-processing	Execution	CPU	0.336 ± 0.150
6	Prediction	Orchestration	FAS-CT	5.792 ± 1.396

The latency between Pre-processing and inference measures the time that it takes for a tensor to arrive at the FPGA (Table I row 2). Similarly, the latency between inference and Post-processing measures the time required for a prediction to reach the Post-processing module (Table I row 4). None of both measurements include the run-time of the involved stages. The execution latency of Pre-processing, Inference, and Post-processing modules has also been measured in Table I rows 1, 3, and 5. Finally, latency measurements have been conducted to determine the time required for generating a prediction from the moment the sensors are polled (Table I row 6). This measurement includes communication and execution of all modules.

B. DDM Backtesting Results

This section focuses on the behavior of the DDM algorithm on the backtest dataset and the consequent start of the retraining pipeline. In table II the experiment results are shown. DDM error metric is calculated in backtests using Eq. 4. The optimal continuous training configuration involves having the least mean DDM error and maximizing the number of samples in the No Drift region.

TABLE II
Backtesting results on retraining pipeline

	Retrain		No Retrain
Scaler Type	**Static**	**Dynamic**	**Static**
Level No Drift %	**65,23%**	35,77%	1,11%
Level Warning %	23,50%	**21,22%**	0,00%
Level Drift %	**11,26%**	43,01%	98,89%
N Retrains	8	24	0
Initial DDM Error	0,1053		
Last DDM Error	0,1653	**0,1475**	1,8
Mean DDM Error	**0,391 ± 1,36**	0,464 ± 1,37	1,335 ± 2,03

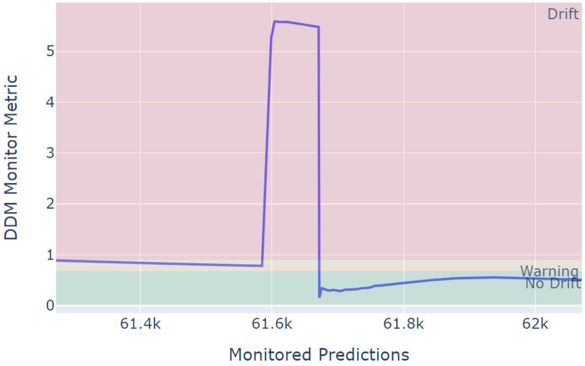


Fig. 3. Outlier in data that causes a sudden increment on DDM error metric, plotted in blue. The green area is the no drift region, the yellow area is the warning region and the red area is the drift region.

Among the three different test configurations, the configuration that yields the best results is the one that retrains the base model and keeps the scalers static. This configuration labels 66.2% of the predictions as No Drift with an average error of 0.391 ± 1.36. These statistics have been achieved with 8 retrains during backtesting.

The configuration with dynamic scalers updates them after each retrain stage. On this configuration, there are 43.01% predictions labelled as Drift. This is 281% more than the previous configuration with static scalers. In contrast with the previous experiment, the final DDM error is lower but with a higher mean DDM error of 0.464 ± 1.37. This is an increase of 18.6% in the DDM error metric with also an increase in the number of retrains. This behaviour is due to the dynamic scalers altering the data distribution after each retraining, worsening the model generalization.

Lastly, the no retraining configuration results in 99% of drifted predictions with an average error of 1.335. Furthermore, in this particular scenario, no prediction was within the warning region as the model encounters an outlier among the first 1.1% of the data that, drastically increments the DDM error metric, as seen in Fig.3. This is the worst configuration meaning that continuous training is needed for this neural network to operate with real-world data.

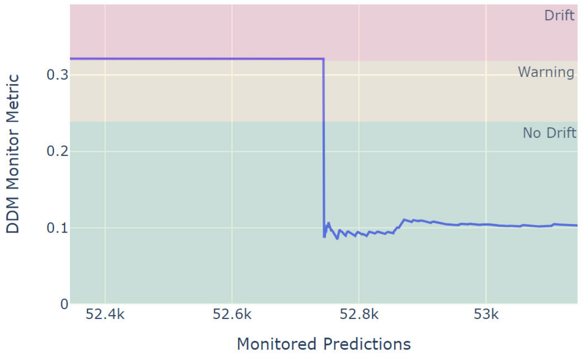


Fig. 4. Retrain number 5 in Table III where after 23333 predictions reach drift level, the retrain pipeline is executed and DDM error metric (plotted in blue) goes again to No Drift region. The green area is the no drift region, the yellow area is the warning region and the red area is the drift region.

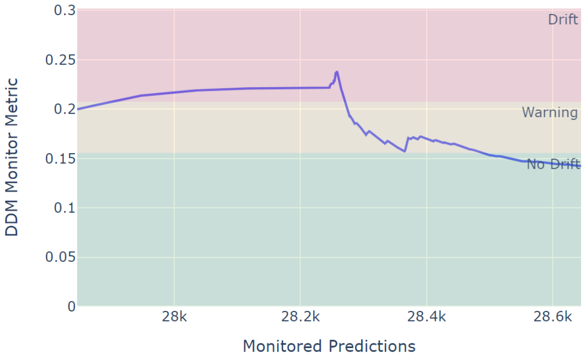


Fig. 5. Retrain number 3 in Table III where the new model could only make 1166 predictions before being substituted. After the retrain number 3 DDM error metric (plotted in blue) does not reach immediately the no drift region. The green area is the no drift region, the yellow area is the warning region and the red area is the drift region.

C. Retraining

After establishing that static scalers are the optimal configuration for FAS-CT with this data, we will now delve closer. We will focus on this experiment, studying how retraining improves the model. The self-retraining process using the static scalers configuration has been detailed in table III. The table includes the model DDM error improvements between each retraining step, as well as the count of predicted samples by the model prior to it being updated.

A successful retrain of the model greatly improves the error metric over its predecessor and generalizes enough that it can make reliable predictions for a substantial portion of samples without triggering another retrain. An example of a successful model retrain would be the fourth retrain, which predicts 23333 samples and presents minimal drift, as seen in the left part of Fig. 4.

TABLE III
DETAILED IMPROVEMENT ON EACH RETRAIN WITH STATIC SCALERS.

Retrain	Previous DDM Error	New DDM Error	Improvement in DDM Error	New Model Predictions
1	12,532	12,085	3,57%	876
2	1,5295	0,1783	88,34%	28248
3	0,2421	0,2217	8,43%	1166
4	0,4559	0,2832	37,89%	23333
5	0,3212	0,0883	72,50%	6035
6	0,2032	0,1368	32,66%	1557
7	0,4283	0,4187	2,24%	1335
8	0,9455	0,1491	84,23%	8053

Furthermore, other retrain attempts are not as successful, such as the first or the third retrain. These retrains happen far from the drift level Eq. 6. Fig. 5 shows the third retrain that only has a slight improvement. This model can only predict 1166 samples before being replaced with a more accurate model.

Lastly, the speed of the retraining process affects the number of predictions beyond the drift level. By reducing the retraining time of the model, the number of predictions in the drift region can be reduced. A comparison between the training performance on the CPU and GPU of the system using Tensorflow Keras is presented in Table IV. The training time depends on various factors such as the number of retrain samples or early stopping configuration. Because of that milliseconds per training batch of 64 samples has been used as a comparative metric.

TABLE IV
TRAINING PERFORMANCE COMPARISON IN DIFFERENT DEVICES

	RTX 3060 12GB		i7 13700k	
Batch	**Mean (ms)**	**Std (ms)**	**Mean (ms)**	**Std (ms)**
32	2.81	8.05	3.59	4.40
64	2.86	5.22	3.96	5.07
128	2.94	6.95	4.93	7.55
256	3.07	10.02	10.006	11.04
512	3.46	12.97	14.99	14.04

GPU is faster per training batch than CPU as expected. Also, training on GPU leverages CPU that can perform better in other modules like in the prediction pipeline, where low latency is required.

VII. CONCLUSION AND FUTURE WORK

This article introduces FAS-CT, a distributed DL inference architecture with FPGA acceleration and continuous training based on CD detection. This architecture is focused on enhancing the performance and reliability of deep learning predictions in changing or difficult-to-predict environments. To achieve that purpose, FAS-CT is composed of two execution pipelines. First is the prediction pipeline that orchestrates model inference in FPGA. Second is the retraining pipeline which monitors the error metric of the model and manages the actualization of the model.

One of the components of FAS-CT is the CD Detector, an algorithm that labels the model predictions with three possible values, No Drift, Warning, and Drift. Once a prediction is labelled as Drift, FAS-CT retraining pipeline is launched using two possible configurations, static or dynamic scalers.

The reliability of FAS-CT has been backtested using an LSTM neural network trained for a forecasting task. The results in table II showed that both retraining configurations outperform the default behaviour of a simple prediction pipeline without continuous training. In addition, the implementation with static scalers stands out by labelling 75% fewer predictions as drift and having a lower mean DDM error metric than the implementation using dynamic scalers.

Additionally, the article also studies the latency of each module involved in the prediction pipeline. FPGA has an inference latency of 1.9ms whereas the complete pipeline has an average latency of 5.8ms, with communication between different components accounting for over 2.5ms of the total latency.

Overall, the article demonstrates that FAS-CT is a reliable low-latency DL inference system that adapts over time. This system is suitable for real-time complex tasks that must be executed on the edge. Also, this article demonstrates that is completely feasible the coordination between a drift detector and an FPGA as an accelerator.

A. Future Work

Regression Outlier Resilience: The presence of outliers can significantly influence the efficacy of CD detection. In scenarios where a model fails to accurately regress an outlier value, the DDM update process may erroneously identify CD, triggering the retraining of a stable model.

Synchronous Model Update: Updating a model while ensuring consistent distributions across different components can be a complex task as communication is asynchronous. Updating the model, the DDM parameters or scalers can happen in different timestamps, resulting in incoherent distributions until all the models are updated.

Monolithic Scaling and Inference block: The communication between the Pre-processing, Inference, and Post-processing modules introduces latency to the inference task, as highlighted in Table I. It is possible to consolidate the three blocks into a single monolithic block executed on the FPGA.

Enhancing Dynamic Scalers Configuration: Not all scenarios can be deployed using the static scalers configuration, as the working data interval might be unknown or can drastically change over time. An algorithm that detects when the scalers are outdated so they can be dynamically updated could be developed.

Early CD detection and model adaptation: Since the DDM error metric reaches the drift level until the model is updated, the system makes some predictions in the drift region. These predictions have been minimized using GPU for training. However, it is necessary to reduce them as much as possible. To achieve this, the proposal is to use methods that detect CD early, such as the Early Drift Detector Method [30]. Further research is needed in this area.

Acknowledgment

This work has been founded by: the Ministry of Science and Innovation under CERVERA Excellence Network CER-20211003 (IBERUS) and CER-20211022 (CEL.IA), Missions Science and Innovation MIG-20211008 (INMERBOT), CDTI (Centro para el Desarrollo Tecnológico Industrial) under project CER-20211022, ICE (Junta de Castilla y León) under project CCTT3/20/BU/0002 and CCTT3/20/0003, the SEDIA of Spanish Ministry of Economic and Digital Transformation under the program R&D Missions in Artificial Intelligence MIA.2021.M01.0004, the Spanish Ministry of Economics and Industry under the grant PID2020-112726RB-I00, the Spanish Ministry of Science, Innovation and Universities with code PID2019-107793GB-I00 and TED2021-131983B-I00, the Principado de Asturias under the grant SV-PA-21-AYUD/2021/50994 and the FEDER funds, ref. 1380734.

Appendix A: Mathematical derivation of mean and standard deviation recurrent formulas

Definitions of mean and standard deviation over a set of items $\{e_i\}_{i=1,\ldots,n}$ are:

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} e_i; \qquad \sigma_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(e_i - \mu_n\right)^2}$$

These definitions imply that all items of e_i must be available in order to calculate μ_n and σ_n for any n. This could be a problem for a limited memory process system if the set is too big or infinite. Due to this limitation, it is necessary to rewrite the mean and the standard deviation definitions as recurrent formulas where only a few values must be stored. Starting with the mean:

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} e_i = \frac{n-1}{n}\mu_{n-1} + \frac{1}{n}e_n$$

For the calculation of the mean, it is only necessary to store three values: the previous mean μ_{i-1}, the number of items n, and the last item e_i. Now deriving the same recurrent formula for standard deviation:

$$\sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(e_i - \mu_n\right)^2 = \frac{1}{n}\sum_{i=1}^{n-1}\left(e_i - \mu_n\right)^2 + \frac{1}{n}\left(e_n - \mu_n\right)^2$$

Notice that the first term is not σ_{n-1}^2 because the mean is the updated mean μ_n and not μ_{n-1}. It is necessary to substitute μ_n with the recurrent formula:

$$\sigma_n^2 = \frac{1}{n^3}\sum_{i=1}^{n-1}\left[n\left(e_i - \mu_{n-1}\right) - \left(e_n - \mu_{n-1}\right)\right]^2 + \frac{1}{n}\left(e_n - \mu_n\right)^2$$

Developing the square of the binomial, applying the definition of μ_{n-1} and σ_{n-1}^2 and arranging all the terms:

$$\sigma_n^2 = \frac{n-1}{n}\sigma_{n-1}^2 + \frac{n-1}{n^3}\left(e_n - \mu_{n-1}\right)^2 + \frac{1}{n}\left(e_n - \mu_n\right)^2$$

Now it is possible to derive the second or the third term depending if the final formula is μ_n or μ_{n-1} dependent.

Developing the second term by substituting the expression of μ_{n-1} in terms of μ_n:

$$\frac{n-1}{n^3}\left(e_n-\mu_{n-1}\right)^2=\frac{1}{n\left(n-1\right)}\left(e_n-\mu_n\right)^2$$

Now the second term has the same dependence as the third term. Summing those terms, the recurrent formula for standard deviation is:

$$\sigma_n=\sqrt{\frac{n-1}{n}\sigma_{n-1}^2+\frac{1}{n-1}\left(e_n-\mu_n\right)^2}$$

For the calculus of the standard deviation, it is only necessary to store four values: the previous standard deviation σ_{n-1}, the current mean μ_n, the number of items n, and the last item e_n.

These recurrent formulas for mean and standard deviation can satisfy the memory requirements in a process system where data is continuously arriving like in FAS-CT or any other system that deals with data streams.

References

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015. doi: 10.1038/nature14539

[2] B. Varghese and R. Buyya, "Next generation cloud computing: New trends and research directions," *Future Generation Computer Systems*, vol. 79, pp. 849–861, Feb. 2018. doi: 10.1016/j.future.2017.09.020

[3] A. Arpteg, B. Brinne, L. Crnkovic-Friis, and J. Bosch, "Software Engineering Challenges of Deep Learning," in *2018 44th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, Aug. 2018. doi: 10.1109/SEAA.2018.00018 pp. 50–59.

[4] M. P. Véstias, R. P. Duarte, J. T. de Sousa, and H. C. Neto, "Moving Deep Learning to the Edge," *Algorithms*, vol. 13, no. 5, p. 125, May 2020. doi: 10.3390/a13050125

[5] A. Gholami, S. Kim, Z. Dong, Z. Yao, M. W. Mahoney, and K. Keutzer, "A Survey of Quantization Methods for Efficient Neural Network Inference," Jun. 2021, arXiv:2103.13630 [cs].

[6] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, and G. Zhang, "Learning under Concept Drift: A Review," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 12, pp. 2346–2363, Dec. 2019. doi: 10.1109/TKDE.2018.2876857

[7] E. A. Castillo and A. Ahmadinia, "A Distributed Smart Camera System Based on an Edge Orchestration Architecture," *Journal of Circuits, Systems and Computers*, vol. 30, no. 04, p. 2150059, Mar. 2021. doi: 10.1142/S0218126621500596

[8] L. D. Biasi, A. A. Citarella, M. Risi, and G. Tortora, "A Cloud Approach for Melanoma Detection Based on Deep Learning Networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 3, pp. 962–972, Mar. 2022. doi: 10.1109/JBHI.2021.3113609

[9] H. K. Fatlawi and A. Kiss, "An Adaptive Classification Model for Predicting Epileptic Seizures Using Cloud Computing Service Architecture," *Applied Sciences*, vol. 12, no. 7, p. 3408, Mar. 2022. doi: 10.3390/app12073408

[10] B. Cao, W. Wu, and J. Zhou, "LCFIL: A Loss Compensation Mechanism for Latest Data in Federated Incremental Learning," in *2022 IEEE 19th International Conference on Mobile Ad Hoc and Smart Systems (MASS)*. Denver, CO, USA: IEEE, Oct. 2022. doi: 10.1109/MASS56207.2022.00055. ISBN 978-1-66547-180-0 pp. 332–338.

[11] K. Vipin, "ZyNet: Automating Deep Neural Network Implementation on Low-Cost Reconfigurable Edge Computing Platforms," in *2019 International Conference on Field-Programmable Technology (ICFPT)*. Tianjin, China: IEEE, Dec. 2019. doi: 10.1109/ICFPT47387.2019.00058. ISBN 978-1-72812-943-3 pp. 323–326.

[12] R. Lozada, J. Ruiz, M. L. González, J. Sedano, J. R. Villar, A. M. García-Vico, and E. S. Skibinsky-Gitlin, "Performance/Resources Comparison of Hardware Implementations on Fully Connected Network Inference," in *Intelligent Data Engineering and Automated Learning – IDEAL 2022*, ser. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2022. doi: 10.1007/978-3-031-21753-1_34. ISBN 978-3-031-21753-1 pp. 348–358.

[13] Y. Zheng, B. He, and T. Li, "Research on the Lightweight Deployment Method of Integration of Training and Inference in Artificial Intelligence," *Applied Sciences*, vol. 12, no. 13, p. 6616, Jun. 2022. doi: 10.3390/app12136616

[14] J. Violos, S. Tsanakas, T. Theodoropoulos, A. Leivadeas, K. Tserpes, and T. Varvarigou, "Hypertuning GRU Neural Networks for Edge Resource Usage Prediction," in *2021 IEEE Symposium on Computers and Communications (ISCC)*. Athens, Greece: IEEE, Sep. 2021. doi: 10.1109/ISCC53001.2021.9631548. ISBN 978-1-66542-744-9 pp. 1–8.

[15] X. Wang, A. Khan, J. Wang, A. Gangopadhyay, C. Busart, and J. Freeman, "An edge–cloud integrated framework for flexible and dynamic stream analytics," *Future Generation Computer Systems*, vol. 137, pp. 323–335, Dec. 2022. doi: 10.1016/j.future.2022.07.023

[16] M. L. González, R. Lozada, J. Ruiz, E. S. Skibinsky-Gitlin, A. M. García-Vico, J. Sedano, and J. R. Villar, "Exploring the implementation of LSTM inference on FPGA [Manuscript submitted for publication]," *Proc. of the International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME 2023)*, 2023.

[17] S. An and U. Y. Ogras, "MARS: mmWave-based Assistive Rehabilitation System for Smart Healthcare," *ACM Transactions on Embedded Computing Systems*, vol. 20, no. 5s, pp. 72:1–72:22, Sep. 2021. doi: 10.1145/3477003

[18] A. Pardos, A. Menychtas, and I. Maglogiannis, "On unifying deep learning and edge computing for human motion analysis in exergames development," *Neural Computing and Applications*, vol. 34, no. 2, pp. 951–967, Jan. 2022. doi: 10.1007/s00521-021-06181-6

[19] W. Jiang, X. Ye, R. Chen, F. Su, M. Lin, Y. Ma, Y. Zhu, and S. Huang, "Wearable on-device deep learning system for hand gesture recognition based on FPGA accelerator," *Mathematical Biosciences and Engineering*, vol. 18, no. 1, pp. 132–153, 2021. doi: 10.3934/mbe.2021007

[20] B. C. Dos Santos Melício, G. Baranyi, Z. Gaál, S. Zidan, and A. Lorincz, "DeepRehab: Real Time Pose Estimation on the Edge for Knee Injury Rehabilitation," in *Artificial Neural Networks and Machine Learning – ICANN 2021*, ser. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2021. doi: 10.1007/978-3-030-86365-4_31. ISBN 978-3-030-86365-4 pp. 380–391.

[21] D. Merkel, "Docker: lightweight linux containers for consistent development and deployment," *Linux journal*, vol. 2014, no. 239, p. 2, 2014.

[22] "Mqtt specification," https://mqtt.org/mqtt-specification/.

[23] R. A. Light, "Mosquitto: server and client implementation of the mqtt protocol," *Journal of Open Source Software*, vol. 2, no. 13, p. 265, 2017. doi: 10.21105/joss.00265

[24] B. Mishra, B. Mishra, and A. Kertesz, "Stress-testing mqtt brokers: A comparative analysis of performance measurements," *Energies*, vol. 14, no. 18, 2021. doi: 10.3390/en14185817

[25] "grpc," https://grpc.io/.

[26] J. Gama, P. Medas, G. Castillo, and P. Rodrigues, "Learning with Drift Detection," in *Advances in Artificial Intelligence – SBIA 2004*, ser. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2004. doi: 10.1007/978-3-540-28645-5_29. ISBN 978-3-540-28645-5 pp. 286–295.

[27] "Mlflow - a platform for the machine learning lifecycle | mlflow," https://mlflow.org/.

[28] M. L. González, J. Sedano, A. M. García-Vico, and J. R. Víllar, "A Comparison of Techniques for Virtual Concept Drift Detection," in *16th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2021)*, ser. Advances in Intelligent Systems and Computing. Cham: Springer International Publishing, 2022. doi: 10.1007/978-3-030-87869-6_1. ISBN 978-3-030-87869-6 pp. 3–13.

[29] "Liambindle/mqtt-c: A portable mqtt c client for embedded systems and pcs alike." https://github.com/LiamBindle/MQTT-C.

[30] M. Baena-Garcıa, R. Gavalda, and R. Morales-Bueno, "Early Drift Detection Method."

Communication Papers of the 18th Conference on Computer Science and Intelligence Systems pp. 141–146
DOI: 10.15439/2023F6100
ISSN 2300-5963 ACSIS, Vol. 37

Automatic Construction of Knowledge Graph of Tea Diseases and Pests

1st Qiang Huang
0000-0002-2552-2687
College of Information engineering
Sichuan Agricultural University
Yaan City, China
13499@sicau.edu.cn

2nd YouZhi Tao
0009-0009-5429-4888
College of Information engineering
Sichuan Agricultural University
Yaan City, China
taoyouzhi@stu.sicau.edu.cn

3nd Shitao Ding
0009-0000-9886-5077
College of Information engineering
Sichuan Agricultural University
Yaan City, China
2020319014@stu.sicau.edu.cn

4th Yongbo Liu
0009-0000-4009-680X
Agricultural Information and Rural Economy Institute
Sichuan Academy of Agricultural Sciences
Chengdu City, China
182382602@qq.com

corresponding: Francesco Marinello
0000-0002-3283-5665
University of Padova
Padua City, Italy
francesco.marinello@unipd.it

***Abstract*—Tea production involves several stages, usually pests and diseases can negatively impact the quality of tea and reduce the harvest, limiting the industry's development. Therefore, it is important to unify the knowledge on tea pests and diseases. Unfortunately, the current knowledge graph construction for tea pests and diseases relies mainly on semi-automated and manual methods, resulting in inefficiency and failing to meet production demands. This research combines three model of Bidirectional Encoder Representation from Transformers, Bi-directional Long Short-Term Memory, Conditional Random Fields for joint extraction of data. The model abbreviated as BERT-BiLSTM-CRF, using the model can automatically generates the triplets, and then store them in the Neo4j database. The study shows that this model has improved accuracy compared to previous methods, and provides effective support for scientific management and production services of tea pests and diseases. The research offers a reference for quickly constructing knowledge graphs in the crop domain.**

***Index Terms*—tea pests and diseases, knowledge graph, BERT-BiLSTM-CRF, Neo4j**

I. Introduction

THE knowledge graph is a structured semantic knowledge base, which is essentially a semantic network that describes the relationships between entities, and can now be used to refer to various large-scale knowledge bases. The knowledge graph represents the relationships between entities in the form of an entity-relationship-entity triplet [1-2] and has been widely used in the fields of research, internet and artificial intelligence [3-4]. With the continuous development of artificial intelligence, machine learning, big data and other disciplines, knowledge graph has achieved better results in domain knowledge management, and the construction of knowledge graph in specific areas of agriculture has gradually become the focus of research by researchers at home and abroad. Yongbo Liu constructed a tea knowledge graph based on the BERT-WWM and attention mechanism approach [5]. Haussmann constructed a knowledge graph of agricultural information [6], which enables users to select the available agricultural products to make food. Dandan Wang combined 2 methods, bottom-up and top-down, to construct a knowledge graph of rice [7], Xu Xin used Neo4j and NLP technology to construct a knowledge graph of wheat varieties [8], which solved the problem of high knowledge repetition rate and unclear knowledge association in variety data. In the process of tea production and marketing, we will face several aspects such as planting, management and processing, each of which requires scientific technical guidance. In actual production, tea yield reduction caused by pests and diseases is generally 15%-20%, which can lead to no tea harvesting in serious cases [9]. Pests and diseases are important factors limiting the development of the tea industry. However, the currently existing knowledge graph in the field of tea pests and diseases are mainly constructed in a semi-automatic and manual way, with low construction efficiency, which cannot meet the actual production needs.

In this paper, we constructed a domain text dataset based on ME+R+BIESO annotation, and used the BERT-BiLSTM-CRF model for joint triplet extraction of entities and relationships from unstructured data to realize the automated construction of knowledge graph, which provides a reference basis for the rapid construction of knowledge graph in crop domains.

II. Tea Pest And Disease Knowledge Graph Construction Process

The knowledge graph can be divided into general domain knowledge graph and vertical domain knowledge graph according to different application directions [10-11]. General domain knowledge graph has the characteristics of being oriented to the whole domain, having a wide range of audiences and involving shallow industry knowledge, and are mostly used in business scenarios such as Internet search engine and content recommendation, e.g., Google search en-

Funded by: Sichuan Provincial Financial Independent Innovation Special Plan Project - Application Research of Spatiotemporal Big Data Analysis in Agricultural Production Services (2022ZZCX034), China

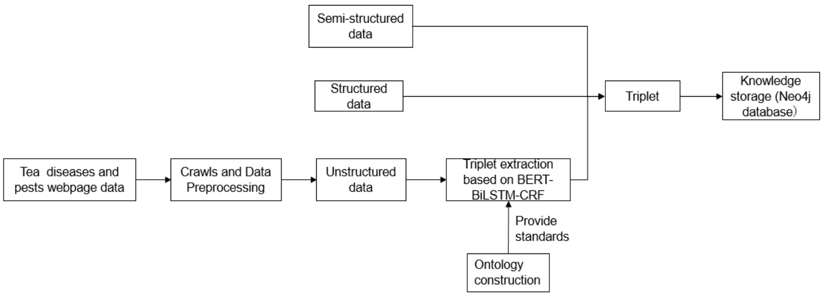

Semi-structured data
Structured data
Triplet
Knowledge storage (Neo4j database)
Tea diseases and pests webpage data
Crawls and Data Preprocessing
Unstructured data
Triplet extraction based on BERT-BiLSTM-CRF
Provide standards
Ontology construction

Figure 1. Process for constructing a knowledge graph of tea pests and diseases

gine, FreeBase [12], DBpedia [13], etc. In this study, the knowledge graph of tea pests and diseases belongs to the vertical domain, and the top-down approach is used to build the graph ontology. This approach requires defining the ontology and data schema first, and then populating the entities and their relationships into the knowledge graph. As shown in Figure 1, it specifically includes the following five stages: (1) Data acquisition and processing. Data is obtained from a variety of sources, which include structured, semi-structured and unstructured data. Structured data can be obtained from third party databases, while semi-structured data typically includes HTML web pages and JSON data. In this paper, unstructured data is obtained from the tea pest knowledge website and data cleaning and pre-processing operations are carried out to obtain the raw text data. (2) Ontology construction. Construct tea pest and disease ontology based on domain corpus, define classes, relations and attributes, and set corresponding constraints to clarify the boundary of knowledge extraction; (3) Data annotation. Use Brat data annotation tool to annotate the text set and obtain the relevant training set and test set after processing by Python code; (4) Knowledge extraction. The BERT-BiLSTM-CRF model is used for training, and the trained model is used to do triplet extraction of the data set. (5) Knowledge storage. The extracted tea pests and diseases triplet data are stored in the Neo4j graph database [14] and visualized.

A. *Data Acquisition and Pre-processing*

The data related to tea pests and diseases in this paper are mainly from the website China Crop Germplasm Information Network, which comes from the Institute of Crop Science, Chinese Academy of Agricultural Sciences. The page contains information on pests and diseases of various crops, and the data on tea pests and diseases mainly contains information on symptoms, alias, pathogen categories and other attributes. The website contains data of 71 tea pests and 21 tea diseases. The Scrapy crawler framework is used for data crawling, and the data pre-processing is combined with rules and manual review to obtain a noise-free plain text corpus.

B. *Ontology Construction*

The architecture of the knowledge graph is generally divided into two layers: the schema layer and the data layer. The schema layer is the core of the knowledge graph structure and is built on top of the data layer. Designing the schema layer of tea pest and disease knowledge graph before data extraction is beneficial to reduce data redundancy. According to the data characteristics of tea pests and diseases, the tea disease knowledge graph concept and tea pest knowledge graph concept are designed respectively, as shown in Figure 2.

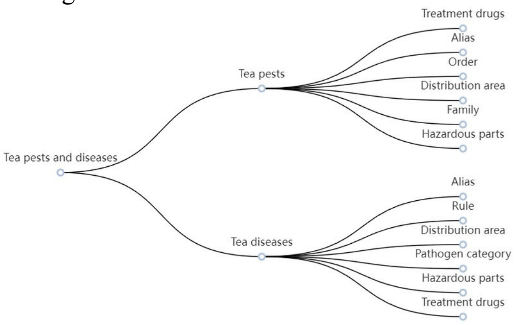


Figure 2. Conceptual level of tea pest and disease knowledge graph

C. *Data Annotation*

ME+R+BIESO annotation method is used to annotate the main entity and the relationship between the main entity and other entities. First, the main entity is labeled as “ME”, and when there is a relationship between an entity and the main entity in a piece of data, the entity Ei is labeled as relationship Ri. The information of each character in the entity is indicated by using the Begin-Inside-End-Single-Other, BIESO) flag to indicate the information of each character in this entity. When the complete set of BE, BIE or S of the main entity ME and a certain relation Ri is matched, the main entity and Ei corresponding to this tag set are taken out and the (ME, Ri, Ei) triplet is formed by data parsing.

Take the data of tea pest “Artaxa flava” as an example, as shown in Figure 3. First of all, “Artaxa flava” is labeled as ME (Main Entity), and “yellow poisonous moth” is an alias of “Artaxa flava”, so “yellow poisonous moth” is labeled as alias. After the data labeling task is completed, the generated a file and the original txt text file are used to label each character in the text with a corresponding label using Python code, and other irrelevant characters are labeled as “O”. When matching the main entity ME and the set of BIE or BE tags with the relationship “Alias”, the mapping of tags can generate a triplet (Artaxa flava, Alias, yellow poisonous moth).

The ME+R+BIESO annotation method focuses on the annotation of the relationship type Ri between the main entity and other entities, without focusing on the entity type to which the entity itself belongs. This method only annotates and extracts on a predefined set of relationships to reduce redundancy and error propagation of irrelevant entity pairs. For the case of overlapping relationships between the main entity ME and multiple entities Ei, multiple corresponding triplets can be obtained by label matching and mapping.

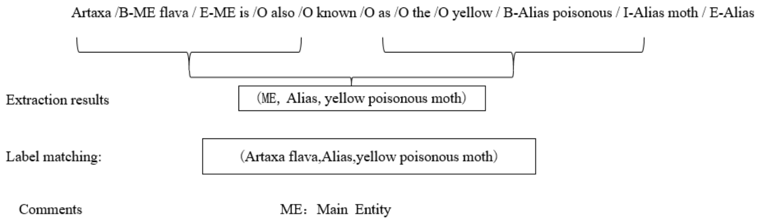

Artaxa /B-ME flava / E-ME is /O also /O known /O as /O the /O yellow / B-Alias poisonous / I-Alias moth / E-Alias
Extraction results
（ME, Alias, yellow poisonous moth）
Label matching:
（Artaxa flava,Alias,yellow poisonous moth）
Comments
ME：Main Entity

Figure 3. Example of ME+R+BIESO labeling method

Compared with the traditional entity relationship extraction methods, the ME+R+BIESO method can synchronize the labeling of bodies and relationships, which reduces the labeling cost and improves the efficiency.

D. BERT-BiLSTM-CRF Model

For the upstream task of entity recognition, traditional corpus learning models such as Word2Vec [15], Glove [16] and other single-layer neural networks cannot characterize the multi-sense of words well in Chinese language environment, so this study chooses the Bidirectional Encoder Representations from Transformers as the linguistic pre-processing model for graph construction. Representations from Transformers as the language preprocessing model for graph construction, in order to obtain high-quality word vectors for entity extraction and classification of downstream tasks. In 2015, the BiLSTM-CRF [17] model proposed by Baidu Research Institute was used for named entity recognition.

A BiLSTM-CRF end-to-end model based on BERT word embedding is used to train and predict tags based on the ME+R+BIESO annotation model. The model consists of three components: namely the BERT layer, the bi-directional LSTM layer and the CRF layer [18], and the overall framework of the model is shown in Figure 4. Firstly, the previously annotated corpus is encoded by the BERT pre-training model to extract the tea pest text corpus features and generate word vectors corresponding to words based on the contextual features of the current words. The key part of the BERT pre training model is in the Transformer layer. The core of the Transformer layer is to calculate the correlation between words through the self-attention function Attention, in order to allocate the weight of words [19].

$$\text{Attention}\ (Q,K,V)=\text{softmax}\left(\frac{Qk^{t}}{\sqrt{d_k}}\right)V \qquad (1)$$

In the equation, headi represents single headed Attention, and MultiHead represents multiple

head attention, W is the weight matrix, through multiple different linear variables change the projection of Q, K and V, and then use the concatenation function concat to concatenate the results of the self-attention mechanism by multiplying them by weights, and calculate the position information of different spatial dimensions.

$$\text{head}_i=\text{Attention}\ \left(QW_i^Q,KW_i^K,VW_i^V\right) \qquad (2)$$

$$\begin{gathered}\text{MultiHead}(\text{Q},\text{K},\text{V})=\\ \text{Concat}(head_1\ ,head_2\ \dots,head_n\ )\ W^0\end{gathered} \qquad (3)$$

In the formula, Q, K, and V are all word vector matrices, and dk represents the input dimension,

WiQ, WiK, WiV represent the weight matrix, and W0 represents the additional weight matrix.

Obtain the word vector corresponding to the input sequence through the BERT model, and then input the word vector into the BiLSTM module for bidirectional encoding. The BiLSTM model overcomes the dependency limitations

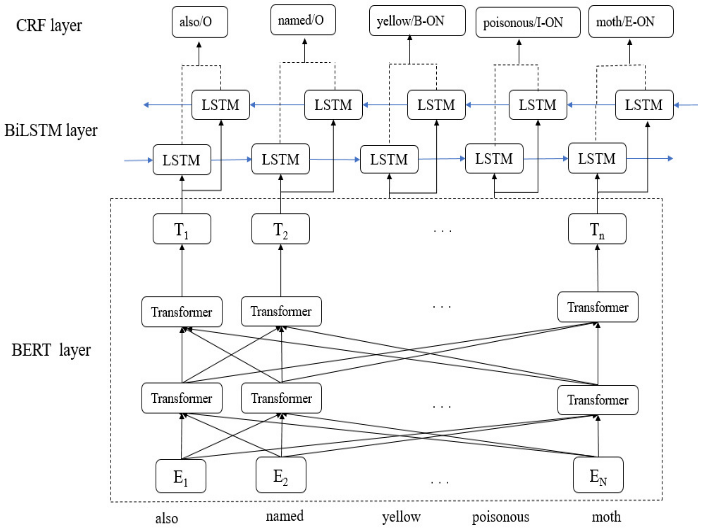


Figure 4. Overall framework of BERT-BiLSTM-CRF model

of traditional machine learning, and bidirectional encoding allows contextual information to be read into the model to achieve effective prediction of tag sequences, and the model outputs a predicted score value for each tag. Finally, the output of the BiLSTM model is decoded using the CRF model to obtain the final predicted annotated sequence. Compared with general deep learning named entity recognition models, the most important feature of this model is the incorporation of a BERT pre-training model, which does not require pre-training of word vectors, and the rich word-level features, syntactic structure features and semantic features of the sequences can be extracted by directly feeding the sequences into the BERT model.

E. Triplet Extraction

The trained model was saved and automated triplet extraction of unstructured text was performed. The automated extraction process is shown in Figure 5, using data from pest-related websites and using the Scrapy crawler framework to automatically obtain unstructured text. In the web data, a page is described for the same tea pest content. To improve the accuracy of the triplet extraction, the main entity, the tea pest name entity, can be identified using a split word method. Now only the corresponding relationships and tail entities need to be predicted using the model. The corresponding label for each word is obtained from the saved model predictions, and the predicted labels "B-Ri", "I-Ri" and "E-Ri" are then used according to the predicted labels "B-Ri", "I-Ri" and "E-Ri" are combined in the order of "BIE" or "BE" to form the corresponding tail entity Ei, which forms a ternary data set of the form (ME, Ri, Ei). Combined with the mapping of the relationship type corresponding to the relationship label R and the custom entity label dictionary, it is converted into the final (main entity, relationship, tail entity) form to complete the automated extraction of the triplet.

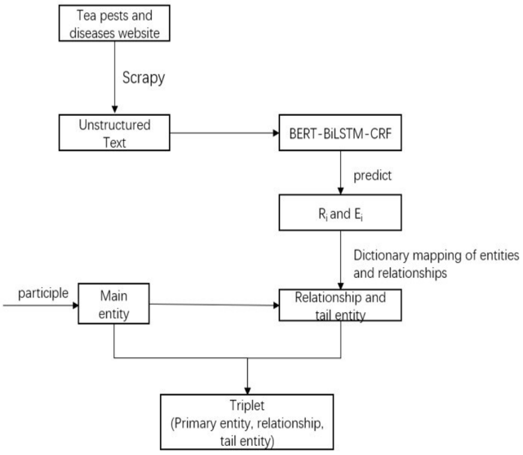

Tea pests and diseases website
Scrapy
Unstructured Text
BERT-BiLSTM-CRF
predict
R_i and E_i
Dictionary mapping of entities and relationships
participle
Main entity
Relationship and tail entity
Triplet (Primary entity, relationship, tail entity)

Figure 5 Automatic extraction of triplets

F. Knowledge Storage

Neo4j is a popular graph database that can store entities, attributes, and relationships in the knowledge graph using graphical representation. This storage mode makes visualization and query of knowledge graph very convenient. Use the Neo4j graph database for knowledge storage, as shown in Figure 6, which is a visual knowledge graph of tea pests "Aleuro10bus marlatti Quaintance", "Aonidiella aurantia" and "Artaxa flav". The data statistics of the triplets are shown in Table I.

III. Experimental Results And Analysis

The experimental environment is Python 3.8 and Pytorch 1.8, the model uses precision, recall and F1 values as evaluation metrics.

A. Comparison of different models

In order to validate the BERT-BiLSTM-CRF model of superiority, four groups of experiments were set up in this paper. The LSTM, LSTM-CRF and BiLSTM-CRF models were chosen as control experiments respectively; the results of the model comparison experiments are shown in Figure 7.

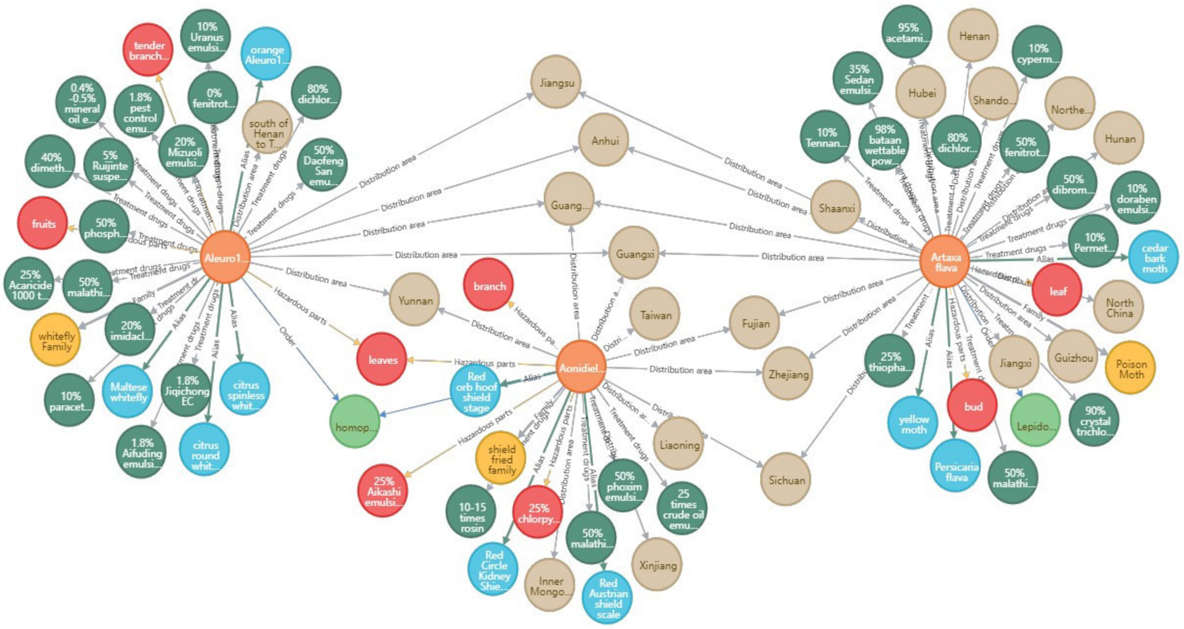


Figure 6. Example of tea pest data visualization

Table I.
Triplet data statistics

Name	Quantity	Meaning
Alias	200	Alias information for tea disease or pest entities
Order	70	Biological classification of tea pest entities "Order"
Family	71	Biological classification of tea pest entities "Family"
Treatment drugs	572	Information on drugs for the control of tea diseases or pest entities
Distribution area	662	Information on the regional distribution of tea disease or pest entities
Hazardous parts	176	Information on the damage sites of tea diseases or pest entities
Pathogen category	20	Information on the pathogenic categories of tea disease entities
Rule	36	Information on the occurrence pattern of tea disease entities
Total	1807	Total number of triplets

Compared with BiLSTM-CRF and LSTM-CRF, the BERT-BiLSTM-CRF model improves the accuracy by 1.76%~5.48%, the recall by 4.25%~8.61%, the F1 score by 3%~7.05%, and the F1 score by 90.53%. The BERT-BiLSTM-CRF model improved the F1 score by 3% after adding the BERT pre-trained language model to the BiLSTM-CRF, indicating that BERT can assist in improving the model's semantic representation of the text and capture the interrelated entity relationships in the tea pest text to a greater extent, thus optimizing the entity effect of the relationship extraction "task".

B. Prediction results of different relationships

The prediction results of the BERT-BiLSTM-CRF model for the relationship between the main entity and each entity are shown in Figure 6, which shows that the overall effect is good, with an F1 score of 90.53%. The "Order", "Family", "Alias", "Distribution area", "Treatment drugs" and "Pathogen category" of tea pests and diseases are shown in Figure 8, The six types of relationships, "Order", "Family", "Alias", "Distribution area", "Treatment drugs" and "Pathogen category", were identified well. In particular, the prediction accuracy of "Order" and "Family" was close to 100%, because the data characteristics of these two types of relationships were very obvious. The BERT-BiLSTM-CRF model can effectively learn the textual information. However, the prediction results of the relationship between "Hazardous parts" and "Rule" were significantly lower than the average. By analyzing the text of the corresponding corpus and the final prediction results of the relationship between "Hazardous parts" and "Rule", we can see that the damage sites of different pests and diseases in this paper are different, such as "leaf" , "stems", "branches", "tea tree root system" and other words are describing the damage site; the same as "relative humidity 85%-87%", "poor ventilation and light penetration", "temperature 25-28℃", "high temperature and low humidity", etc. are all describing the occurrence pattern of the disease. The inconsistency of description methods makes it difficult for the model to fully learn the characteristics of the damage site, which makes the recall rate of "Hazardous parts" and "Rule" is low.

In this paper, the F1 values of the named entity recognition model basically reached more than 90% for entities other than "Hazardous parts" and "Rule". In summary, the BERT-BiLSTM-CRF model in this paper has a relatively good recognition effect in the named entity recognition task of tea pests and diseases.

IV. Conclusion

The BERT-BiLSTM-CRF model used in this paper extracts the triplet data of tea pests and diseases and automates the construction to generate the knowledge graph of tea pests and diseases. The experimental results show that the accuracy value reaches 90.10% and the recall rate

is 90.53%. It provides effective support for the scientific management and production services of tea pests and diseases, and the study also provides a reference basis for the rapid construction of knowledge graph in crop fields.

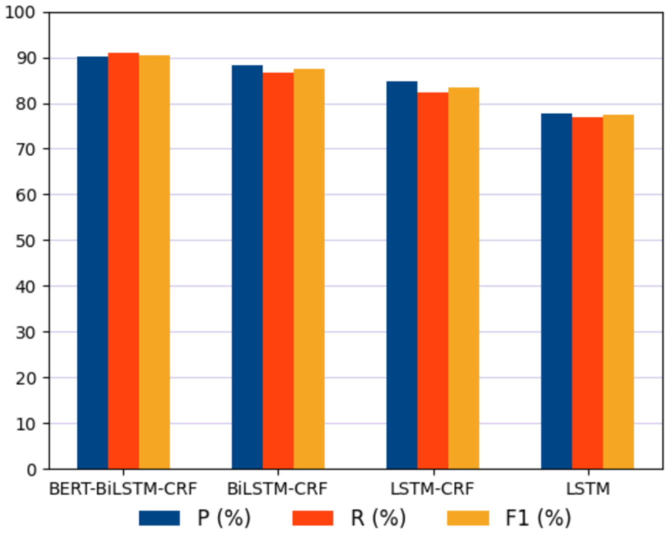


Figure 7. Entity extraction model performance comparison

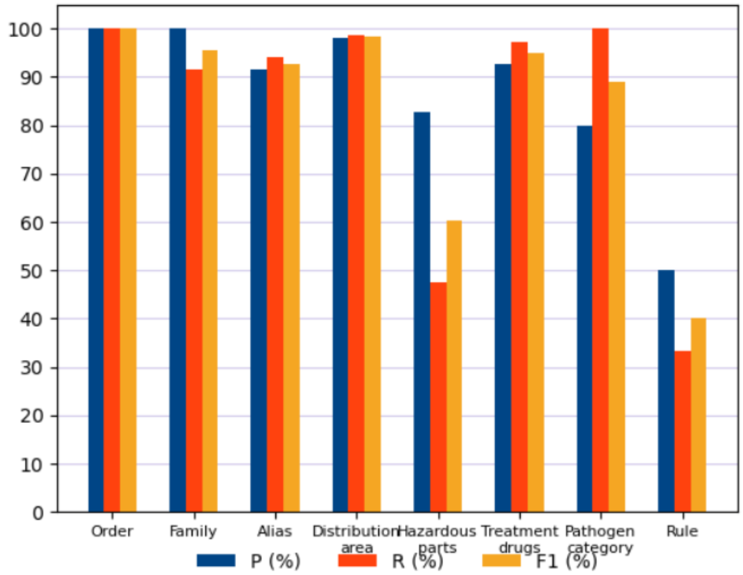


Figure 8. Prediction results of each relationship in the BERT-BiLSTM-CRF model

References

[1] Xu Zenglin, Sheng Yongpan, He Lirong, et al. "A review of knowledge graph techniques," Journal of University of Electronic Science and Technology of China, 2016, 45(4):18.

[2] Paulheim H, "Knowledge graph refinement: A survey of approaches and evaluation methods," Semantic web, 2017, 8(3): 489-508, to be published.

[3] Pujara J, Hui M, Getoor L, "Large-Scale Knowledge Graph Identification using PSL" Aaai Fall Symposium, 2013, to be published.

[4] Zhang Qingling, Li Xianzheng, Li Hangyu, et al. "Application of knowledge graph in agriculture," Electronic Technology & Software Engineering, 2019(7):3.

[5] Liu YB, Huang Q, Gao WB, et al. "Construction of tea knowledge graph by integrating BERT-WWM and attention mechanism," Southwest Journal of Agriculture,2022,35(12):2912-2921.

[6] Haussmann S, Seneviratne O, Chen Y. "Food KG: a semantics-driven knowledge graph for food recommendation," International Semantic Web Conference. Springer, Cham, 2019: 146-162, to be published.

[7] Wang Dandan, "Research and application of knowledge graph construction method for Ningxia rice," Northern University for Nationalities, 2020.

[8] Xu Xin, Yue Jinzhao, Zhao Jinpeng, et al. "Research on the construction and visualization of knowledge graph of wheat varieties," Computer System Applications,2021,30(06):286-292.

[9] Tan Rongrong, Liu Mingyan, Gong Ziming, et al. "Analysis of the types and occurrence patterns of major pests and diseases in tea areas of Hubei Province," Tea Newsletter, 2013, 40(04):36-38.

[10] MA Mohamed, Pillutla S, "Cloud computing: a collaborative green platform for the knowledge society," Vine, 2014, 44(3): 357-374, to be published.

[11] Hu Fanghuai, "Research on Chinese knowledge graph construction method based on multiple data sources," Shanghai: East China University of Science and Technology, 2015.

[12] Yue B, Gui M, Guo J. "An effective framework for question answering over freebase via reconstructing natural sequences," Proceedings of the 26th International Conference on World Wide Web Companion. 2017: 865-866, to be published.

[13] Ritze D, Bizer C. "Matching web tables to dbpedia-a feature utility study," context, 2017, 42(41): 19-31, to be published.

[14] Webber J. "A programmatic introduction to neo4j," Proceedings of the 3rd annual conference on Systems, programming, and applications: software for humanity. 2012: 217-218, to be published.

[15] Goldberg Y, Levy O. "word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method," arXiv preprint arXiv:1402.3722, 2014.

[16] Pennington J, Socher R, Manning C D. "Glove: Global vectors for word representation," Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014: 1532-1543, to be published.

[17] Huang Z, Xu W, Yu K. "Bidirectional LSTM-CRF models for sequence tagging" arXiv preprint arXiv:1508.01991, 2015.

[18] Sutton C, Mccallum A. "An Introduction to Conditional Random Fields," Foundations and Trends in Machine Learning, 2010, 4(4):267-373, to be published.

[19] Wu Z，Jiang D，Wang J，et al. "Knowledge-based BERT: a method to extract molecular features like computational chemists," Briefings in Bioinformatics，2022,23(3):bbac131, to be published.

Communication Papers of the 18th Conference on Computer Science and Intelligence Systems pp. 147–152
DOI: 10.15439/2023F1961
ISSN 2300-5963 ACSIS, Vol. 37

Harness Old Media: a cross-disciplinary approach to utilizing television data for media content analysis.

Piotr Jabłoński
0000-0001-9360-3502
Faculty of Mathematics and Computer Science
Adam Mickiewicz University, Poznan, Poland
e-mail: piotr.jablonski@amu.edu.pl

***Abstract*—The phenomenon of disinformation has become a common theme in studies across various fields. Both qualitative and quantitative methodologies are typically used, focusing primarily on content sourced from the internet. This article introduces a method to extend this focus to include content from 'Old Media'[1] specifically from Television which as an unstructured medium, presents a combination of textual and visual layers. Despite this complexity, the integration of these elements allows for the design of algorithms capable of analyzing video streams and extracting individual news from main news programs of nationwide broadcasters. The proposed solution facilitates the extraction of transcriptions generated by the research tool. The aim of this research is to allow access to the content of television to enable its inclusion in research, performed in a manner analogous to Internet content. This research is part of a project that deals with the development of algorithms for combining, classifying and comparing content from different media in order to design an imprecise classifier of disinformation content.**

I. Introduction

THE spread of the term 'Fake News' worldwide is frequently associated with pivotal political events of 2016, including the US presidential election and the Brexit referendum in the UK, during which the internet and social media were flooded with fabricated content. As outlined in [2], the concept of fake news itself emerged much earlier, dating back to the 19th century in tandem with the rapid development of yellow journalism in the United States. The phenomenon of disinformation, has also reached Poland, causing the emergence of numerous educational campaigns targeting a broad audience in Poland. Among many different initiatives, there was a high rise of activities included the core curriculum of primary and secondary schools, implemented through educational projects. A variety of outreach programs geared toward the general public have also been inaugurated, coordinated by a group of NGO institutions, media consortia and governmental agencies. Post-2016, the phenomenon of Fake News has been a focal point in a multitude of scientific research conducted by different institutions not only in Poland, but also worldwide. These studies span across numerous disciplines, including but not limited to social communication and media studies, linguistics and computer science. It is noteworthy to mention that this phenomenon is also incorporated into the specific objectives of the Infostrateg program, launched by the National Center for Research and Development in 2020 [3].

II. Current state of research

So far, the research conducted by scientists is mostly focused on the analysis of one medium - print press / television / Internet. In current research, one can notice a tendency to choose the latter medium more often, while as Naturel stated: "Querying and retrieving information from a large television (or video) corpus is still a challenge, for both professional archivers and simple TV users as well." [4]. That's why the justifications for the choice of internet medium points to its universality and great dynamism, unavailable in the old media. Vasoughi also pointed out that on the Internet, news classified as falsehood, was able to reach first 1500 people six times faster than the true news [5].

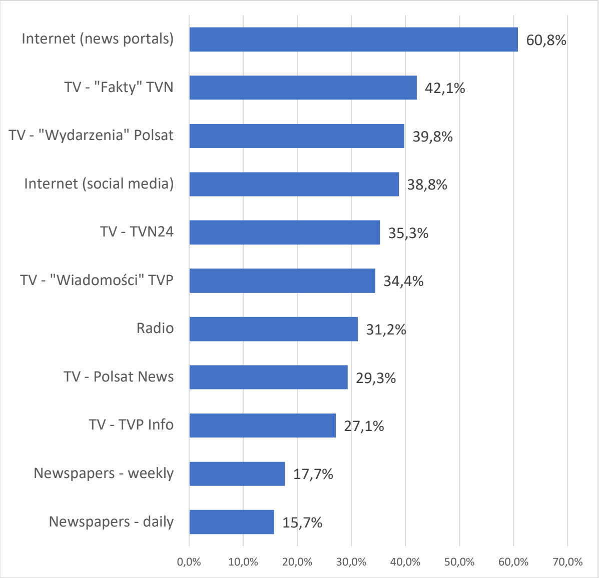


Figure 1. Results of the survey about the preferences of Poles regarding sources of information about Poland and the world. Source: [6]

Undoubtedly, studying the content of a single medium has a major impact on the consistency of research methodology. The

[1] The terms "Old Media" and "New Media" began to be used by scholars and academics studying the changes in communication caused by the growth of digital technologies in the 1990s. New media includes forms of communication in online form such as electronic books, email, informational web portals, and social media [1].

Thematic track: International Workshop on AI in Digital Humanities, Computational Social Sciences and Economics Research

Internet as a medium, as mentioned earlier, not only enables a high rate of news dissemination, but is also a frequent source of information for the Polish public. As shown in the "Survey of Poles' preferences for sources of information about Poland and the world" [6] presented in the Figure 1, 60.8 percent of Poles cling to information from online portals. However, this group does not include social media, which is the main choice of information source for 38.8% of respondents. In the same survey, TV programs such as Fakty TVN, Wydarzenia Polsat, and Wiadomości TVP received 42.1%, 39.8%, and 34.4% respectively.

In the same survey, it was also shown that social media is a more frequent choice as a source of information than online portals only in the 18-29 age group, although their advantage is quite small (75.8% to 71.8%). Taking into account the above statistics, as well as a report showing that 93.3% of households in 2022 had access to the Internet [7], television, as a news medium, should be analyzed just as often as internet in media content studies. As stated in [8] "Even in many developed and technologically advanced countries (...), among the people who say they use the Internet daily, a large percentage also say they use television daily for information purposes".

Despite the majority of studies based on material from the Internet, there are studies conducted simultaneously analyzing the content of different media. One of the largest studies conducted to date is the work of Claudia Melladio's international team "Journalistic Role Performance - second wave". This team consists of researchers from 37 countries, studies the content of television, radio, press and Internet portals. The study of Polish team, taking part in this research included 14 outlets. In this group three of them were television programs broadcasted by main nationwide television stations. Uniformity of sampling in this international study consisted of examining two constructed weeks from the entire year 2020. As a result, 541 news cases from Polish news programs broadcast on television were analyzed. This accounted for 8.6% of all news stories analyzed in Poland in the aforementioned study. One of the reasons for conducting research on such a limited research sample is the extremely time-consuming process of coding the video, which involves reviewing and analyzing the content in each message according to a designed codebook.

From the above analysis arose the need to enable access to TV content in a manner equivalent to Internet content - in text form. This form of data makes it possible to use the tools and statistical methods available in natural language processing techniques.

An analysis of the scientific literature has revealed a prevailing shortage in the area of both the discipline of social communication and media sciences and computer science. The research conducted in the field of social sciences on media content analysis using quantitative data analysis methods implemented on a sample of just over half a thousand records is the largest studies conducted in the world to date. The data analysis methods used in the referenced study do not involve any statistical language analysis processes and rely entirely on human input. In contrast, in the discipline of computer science, where such methods would be expected, research is not conducted at all, due to the lack of access to research material. The research carried out by the author, instead, focuses on creating the development of methods for automatic verification and analysis of media content.

III. Context of the study

The purpose of the project is to develop algorithms to fuse, classify and compare content from different news media for the purpose of designing an imprecise classification of disinformation content. The problems of verifying the authenticity and reliability of published information are a direct result of the oversupply of content and information noise, determined by intertwining elements such as truth, rationality, objectivity, but also rumor, hearsay or conspiracy theories. For this reason, not only the recipients, but also the editors who are the intermediary of the media message have a problem with their verification, and instead of stopping further publications, they amplify the process of spreading unverified information in the media. The oversupply of information is noticeable in social media, where in 2020 Twitter published about 380,000 twitts a day in Polish only, which is more than 260 new messages per minute[2].

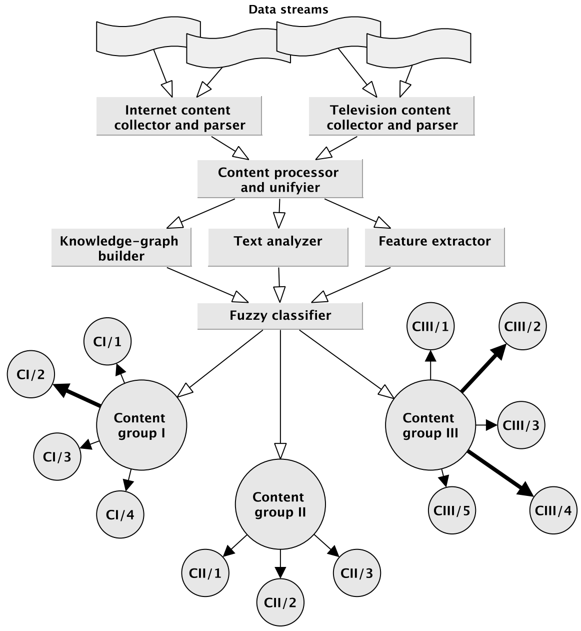


Figure 2. Diagram of the disinformation detection system.

At present, there is a lack of solutions capable of verifying in an automated manner, on the basis of television broadcasts,

[2] The data comes from research conducted by the author in 2021/2022. The research involved an analysis of the volume of information published on web news portals and twitter platform. Their effect was used in the work of the team implementing the study "From urban legend to fake news. A global detector of contemporary falsehood" funded by NCBiR.

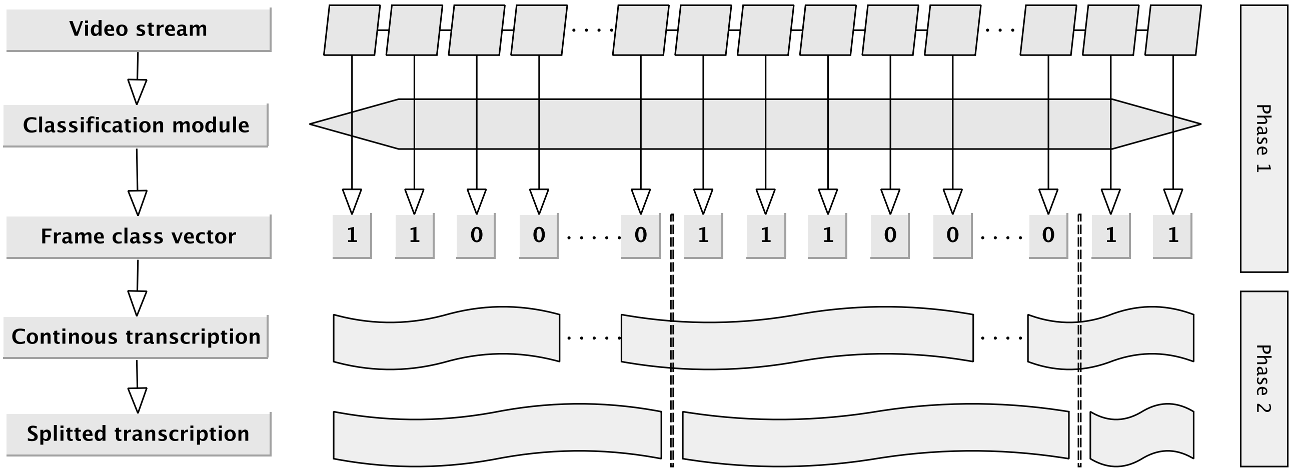


Figure 3. Scheme of the designed system splitting news outlet to separate news.

whether a specific content of a message originating from the Internet media has the hallmarks of disinformation content. Developing a solution to the above problem, will help create a mechanism for detection of disinformation content, which can stop the spread of disinformation (fake news) by professional media broadcasters and Internet users who are unaware of the threat posed by disinformation. The research presented in this article is part of a system presented at Figure 2, which in its next steps uses knowledge-graphs and fuzzy logic to fuse content from different media.

IV. Developed algorithm

Such a large amount of information delivered in a limited amount of time can affect the creation of information chaos that media audiences have to deal with. In this worldwide confusion, the spread of false information (so-called "fake news") is becoming more frequent, not only on social media, but also by professional news outlets.

In order to collect the research material in the form of transcript text files, an algorithm was designed for news programs broadcasted by major national TV stations (Polsat, TVN, TVP). Developed algorithm detects individual news stories in an outlet and then extracts their transcript, covering detected news which can be later treated as single text documents. The database in the form of video footage and transcription file, for current and archived program outlets, is the CAST (Content Analysis System for Television) system, operating at the Faculty of Political Science and Journalism at Adam Mickiewicz University in Poznan [9] [10]. *The system records 6 channels: TVP 1, TVP 2, Polsat, TVN, TVN24, TVP Info continuously, (24 hours a day), since mid-2014. (...) The broadcasts are stored in a database, described with metadata generated from the EPG (...) Another useful feature present in the system is speech-to-text transcription. Each Polish-language broadcast found in the database contains a text transcription of all the issues spoken in it."* [11].

Thanks to the data collected by the CAST system, it became possible to design a two-phase algorithm. The first one analyzes the video stream, assigning to each frame one of two classes - a frame with a studio image (1), and a frame with a non-studio image (0). Each of broadcasted program have a set of dedicated model and mask in size of frames to determine its class. For this purpose, the OpenCV library is used, with the 'matchTemplate' method additionally using 'masks' that exclude variable parts of the studio image. For each program, minimum of three key frames are defined with a mask applied to it, designed to make the analysis independent of variable elements of the scene. This is an extremely important part of the overall analysis, as current studio arrangements provide not only for multiple presenters, but also for variable camera settings and dynamic backgrounds, often occupying as much as 65,5% percent of the scene area. Figure (4) presents an example set of templates and masks for the TVN Fakty program. The process designed in this phase accepts a video stream as input, which is then analyzed. The video can be of any size but must be encoded with a codec that can be parsed by the OpenCV library. During testing different size of video source[3] was used but the algorithm showed an insensitivity to the resolution of the video analyzed. Currently, it covers a wide range of codecs like MPEG1/2/4, H.264, HEVC, VP8/9, VC1, but also JPEG and uncompressed video. At the beginning of the process, algorithm detects the number of frames per second in the input stream, which is then set as the fixed number of frames to be skipped between successive analyzed key-frames. In the proposed solution, the number of frames to be skipped is equal to the number of frames per second, resulting in the analysis of exactly one frame every one second. The implemented approach optimizes the performance of the image analysis process, which was confirmed empirically

[3] During test following resolutions was checked 1080x1920px, 1280x720px and 640x360px. None of these resolutions have shown greater effectiveness of identifying studio scenes.

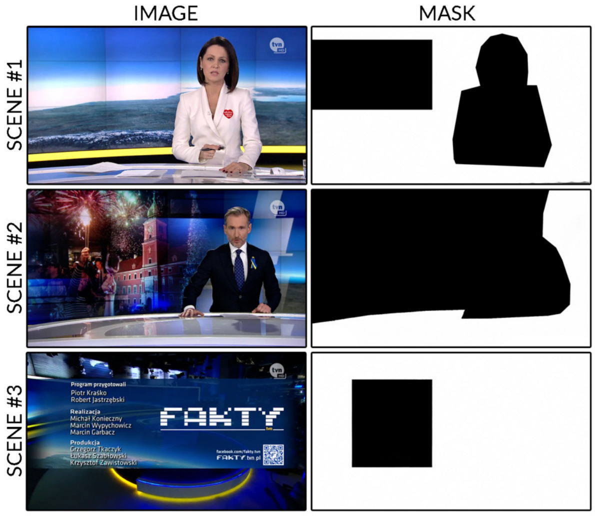


Figure 4. Figure shows examples of different templates and corresponding masks used in algorithm to detect studio frame between news.

Table I
METADATA AVAILABLE ON ARCHIVED OUTLETS OF NEWS WITH EXAMPLE CONTENT.

Field	Description	Content
UUID	Unique object identifier in the CAST system	b61ca13f-b0f2-47f0-ad49-ecedf6107de1
Title	EPG program name	Wiadomości
Description	EPG description	News service presenting (...) economy, culture and social life.
Channel	Channel	TVP 1
Category	EPG Category	news/current affairs (general)
Start date	Scheduled recording start date	2022-03-30
Start time	Scheduled recording start time	19:30:00
End date	Scheduled recording end date	2022-03-30
End time	Scheduled recording end time	20:05:00

during the study. Decreasing the step, did not result in an increase in efficiency in detecting the class of the frame, but increased the number of frame comparison operations. In the CAST system, video is recorded at 50 frames per second. This means that a video stream, lasting 30 minutes, contains 90,000 video frames. As a result of optimization involving frame bypassing, only 1,800 frames are analyzed, which is 2% of all video frames. If a higher step value (equal to two, three or five times the number of frames per second) was adopted, a noticeable delayed detection of the studio's frame was created. In effect the beginning of presenter speech could be cut off from the next material. Then each key-frame is analyzed using the OpenCV library, which determines the class of the frame (studio/non-studio). Comparison of frame with template and mask is made using the function `matchTemplate()` (`cv.matchTemplate(image, templ, method[, result[, mask]])`, where `image` is the analyzed frame, `templ` is the prepared template file and `mask` is the matching mask file. Currently OpenCV uses one of two methods, which support mask usage: `TM_SQDIFF` and `TM_CCORR_NORMED`. In this research was used the last one, which stands for *Correlation Coefficient*. Function this return a matrix of values, which is then search as global maximum with usage of `minMaxLoc()` function. As the effect of first phase of the algorithm, a vector of binary values is generated, containing a values representing the classification of each analyzed key-frame.

In phase two of the algorithm, the previously generated vector is converted into time code values according to the parameters of the analyzed stream. This vector is then used to indicate the locations of studio and news boundaries in the transcription file into parts corresponding to the detected image sequences. Each of isolated transcription part is then supplemented with a metadata metric in the form of parameters presented in Table I for easy transcription file identification. The gathering and saving of metadata on each of the broadcast programs is a key to the ability of determining the publish location and exact time of the video's broadcast. The sole recording of a TV program or analysis of available video footage (especially in the context of archival material, available on the Internet) does not allow for precise temporal placement, which can be crucial for identifying the source of the disinformation. With precise metadata, including the name of the program, it becomes possible not only to develop the route of the spread of content in the media, but also the appearance of actors in a specific time frames.

V. CONCLUSION

As a result of the research, a set of files was created, representing the content of the main news channels of Polish national TV stations. The result of the program for a period of 26 weeks, is depicted by 6989 text files, representing the extracted news. Each of program outlet consist of average of 12.78 news items per program.

Algorithm, tested on small human annotated test set, consisting of 30 episodes, presents an outstanding performance of scene identification. Table II presents exact results, in which it is shown performance of detecting studio scenes in every news program. The developed algorithm achieves the following results: Precision 97,95%, Recall 97,46% and F1-score: 97,70%.

The data set built with the presented algorithm will contribute positively to the ability to analyze the content of TV

Table II
THE EFFICIENCY OF DETECTING STUDIO SCENES IN THE TEST SET

Detecting method	Total	TVP	Polsat	TVN
Human annotated programs	30	10	10	10
Human decision	383	124	132	127
Scene correctly detected by Algorithm (TP)	366	117	128	121
Scene not detected by Algorithm (FN)	10	4	2	4
Scene incorrectly detected by Algorithm (FP)	8	2	2	4

programs. Undoubtedly, this is an important step in making the content of this medium available for quantitative research, particularly when comparing it with Internet content.

Each document retains the appropriate structure and detailed metadata, enabling extensive quantitative content analysis. It may be possible to apply methods of automatic text summarization [12] and NLP Tasks with the use of Transformer Models [13]. It can be also used to analyze the affect [14] in news content. The data can also be used to more accurately analyze the content of the messages according to 5W Lasswell's model of communication (who?", "says what?", "in what channel?", "to whom?", "with what effect?") [15] over a broader timeframe. Proposed content distribution can also be another step in the growth of data journalism where big data plays an important role [16].

In the near future, the development of the designed algorithm should also include the possibility of identifying experts and speakers in the broadcast. Currently, the CAST system implements a module for reading the content of information contained in lower third[4], which is in the fine-tuning stage. Thanks to this functionality, another layer of information is being added, which can help in the detection of actors appearing in the media message. All unstructured data collected from television and structured data from the Internet should be organized into a Knowledge Graph, enabling the creation of an efficient connection among all instances of knowledge. According to Zhang[17], it is valuable to employ embedding and clustering algorithms to implement a topic hierarchy for enhancing the Knowledge Graph's performance.

Effects of this research will also positively contribute to the author's project covering methods of imprecise classification of disinformation content. The work is intended to test the possibility of content analysis on the basis of media content from both the Internet and the television broadcasts. Another task is to measure the effectiveness of content clustering and classification using fuzzy logic methods. One of the elements developed in the research will be a system for analyzing various media messages on a selected topic, taking into account similarities between messages. These similarities will be developed, among other things, on the basis of the results of summarizing modules, sentiment analysis and the combination of identified named entities.

[4] Lower third is a graphical overlay, placed in lower part of screen, containing information about current story or appearing person, like name, surname and affiliation.

References

[1] W. J. Dizard, *Old Media New Media: Mass Communications in the Information Age*, Second edition. New York: Longman, 1996, ISBN: 9780801317439.

[2] D. Halagiera, "Fake news jako nowe (stare) wyzwanie dla świata mediów – portal YouTube w walce z nieprawdziwymi informacjami," in *Kryzysy współczesnego świata. Różne ujęcia problemów globalnych i regionalnych*, 2019, pp. 91–105.

[3] Narodowe Centrum Badań i Rozwoju, "Program Strategiczny INFOSTRATEG „Zaawansowane technologie informacyjne, telekomunikacyjne i mechatroniczne”," Narodowe Centrum Badań i Rozwoju, Warszawa, Tech. Rep., Apr. 2020. [Online]. Available: https://archiwum.ncbr.gov.pl/fileadmin/Programy_Strategiczne/Opis_Programu_INFOSTRATEG.pdf.

[4] X. Naturel and P. Gros, "Detecting repeats for video structuring," *Multimedia Tools and Applications*, vol. 38, no. 2, 2008, ISSN: 13807501. DOI: 10.1007/s11042-007-0180-1.

[5] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, 2018, ISSN: 10959203. DOI: 10.1126/science.aap9559.

[6] Instytut Badań Internetu i Mediów Społecznościowych and Instytut Badań Rynkowych i Społecznych. "Badanie preferencji Polaków dot. źródeł informacji o Polsce i świecie." (Jan. 2021), [Online]. Available: https://ibims.pl/skad-polacy-czerpia-informacje-o-polsce-i-swiecie-raport-ibims-i-ibris/.

[7] Eurostat, *Digital economy and society statistics*, Dec. 2022. [Online]. Available: https://ec.europa.eu/eurostat/databrowser/view/isoc_ci_in_h/default/table?lang=en.

[8] S. J. Shaikh, "Television versus the internet for information seeking: Lessons from global survey research," *International Journal of Communication*, vol. 11, 2017, ISSN: 19328036.

[9] Faculty of Political Science na Journalist. "CAST." (2020), [Online]. Available: https://wnpid.amu.edu.pl/en/home/cast.

[10] J. Wyszyński, *Content Analysis System for Television*, 2017. [Online]. Available: http://cast.info.pl/.

[11] A. Stępińska and J. Wyszyński, "Ilościowa analiza zawartości przekazów w badaniach nad dyskursem populistycznym," in *Badania nad dyskursem populistycznym: wybrane podejścia*, 2020, ch. VII, pp. 107–129.

[12] N. Andhale and L. A. Bewoor, "An overview of text summarization techniques," *Proceedings - 2nd International Conference on Computing, Communication,*

Control and Automation, ICCUBEA 2016, 2017. DOI: 10.1109/ICCUBEA.2016.7860024.

[13] A. Gillioz, J. Casas, E. Mugellini, and O. A. Khaled, "Overview of the transformer-based models for nlp tasks," *Proceedings of the 2020 Federated Conference on Computer Science and Information Systems, FedCSIS 2020*, 2020. DOI: 10.15439/2020F20.

[14] P. Subasic and A. Huettner, "Affect analysis of text using fuzzy semantic typing," *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 4, pp. 483–496, Aug. 2001, ISSN: 10636706. DOI: 10.1109/91.940962.

[15] H. D. Lasswell, "The structure and function of communication in society," *The Communication of Ideas*, no. 1948, 1948.

[16] A. Veglis and T. A. Maniou, "The mediated data model of communication flow: Big data and data journalism," *KOME*, vol. 6, no. 2, 2018, ISSN: 20637330. DOI: 10. 17646/KOME.2018.23.

[17] Y. Zhang, M. Pietrasik, W. Xu, and M. Reformat, "Hierarchical topic modelling for knowledge graphs," pp. 270–286, 2022. DOI: 10.1007/978-3-031-06981-9_16.

Communication Papers of the 18th Conference on Computer Science and Intelligence Systems pp. 153–156

DOI: 10.15439/2023F9900
ISSN 2300-5963 ACSIS, Vol. 37

Enhancing Image Quality through Automated Projector Stacking

Franciszek Jełowicki
0009-0001-7676-6884
Warsaw University of Technology, Faculty of Mathematics and Information Science
ul. Koszykowa 75, 00-662 Warszawa, Poland
franciszek.jelowicki@pw.edu.pl

***Abstract*—This paper describes a system that combines several projectors to display a common image. Human visual perception of an image is largely dependent on contrast. When external light sources are present, the contrast of the projected image decreases. Increasing the brightness of the projector is limited by technology. By combining several projectors into one system it is possible to increase brightness, and thus contrast, without using more expensive projectors. The method of calibrating the system involves displaying the ChArUco board and taking pictures of them with a smartphone camera. Based on the detected markers, homographies are found. Then the image is modified so that each projector displays the same pixel of the input image at each point of the common projection area. Compared to existing commercial systems this one does not require a dedicated projector or camera model. Nevertheless, the results show an improvement in image quality.**

I. Introduction

DIGITAL projectors are a popular choice since they allow to display big images - bigger than monitors with comparable price. When the room where the projection takes place is well darkened the image is well visible. Conversely, when there are additional light sources, the image visibility decreases. This happens because human vision depends on contrast rather than absolute light intensity [1], [2]. By contrast we are referring to the brightness ratio of two adjacent pixels (white and black). If absolute intensity of a white pixel is I_w and of a black pixel - I_b the contrast is $\frac{I_w}{I_b}$. With additional ambient light with intensity I_a contrast is $\frac{I_w+I_a}{I_b+I_a}$. When I_a tends to infinity, contrast tends to 1, which means no contrast at all. Decreased contrast especially reduces the visibility of text, which is important in presentations.

Increasing a projector's maximal brightness (I_w) is a technological challenge and generally comes with an increased device price. The idea presented in this work involves using multiple projectors set up in such a way to project a common image. This way light intensities from each projector may be added to improve image quality in bright environments.

Another gain from this method may be increased resolution, understood as the number of pixels on a given area. When, e.g., two identical projectors project image on the same area, it is possible to achieve two times bigger pixel density. A similar solution used in some projectors is called XPR [3]. It involves projecting the image four times, moving it half a pixel up/down or right/left. Pixel size is the same as always, but since there are more of them, the image is perceived as having better quality. Increasing resolution by using multiple projectors was described in [4].

When using a few projectors, the main problem is to calibrate them so that each projector displays the same pixel of the input image at each point of the surface on which the projection takes place. There is a need to designate a new projection area contained in the common part of projection areas with the same aspect ratio as the input image. Inside that area brightness will be increased. On the other hand, if the calibration was inaccurate, the images from individual projectors would not overlap. As a result, image perception could be worse.

In this work, we have described a simple system that meets the objectives outlined above. It should work independently of the used projectors and only requires a smartphone with a camera as additional hardware. The performed tests suggest that the calibration achieved by this method on ordinary equipment increases image quality.

II. Related Work

There exist a few commercial solutions similar to the one proposed there, but to our best knowledge, no results of these systems have been published. This section describes these methods and alternative methods that can be used to combine a few projectors.

A. Using projector keystone correction

When the projector is not perpendicular to the projection plane, the projection area is distorted from a rectangle into an irregular quadrilateral. To compensate for that effect, most projectors contain an option called a *keystone correction*. It allows modifying the corners of the quadrilateral back into a rectangle.

Theoretically, with two or more projectors, it is possible to align all of them to a common projection area using that feature. That solution was proposed, e.g., by Epson, who also created a tool to assist in that process [5].

The main advantage of this method is its versatility – even without additional tools, most projectors have a keystone correction option that can be used independently of the system. On the other hand, this process has many disadvantages. Manual calibration is hard, takes much time, and is inaccurate.

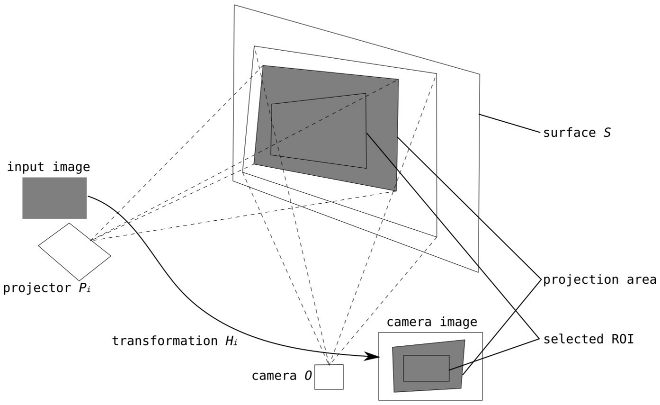


Fig. 1. Schematic system overview.

Since keystone correction has a constant step, the shift on the projection surface depends on the projector's distance from the projection surface.

B. Expanding projection area using projection mapping

Projection mapping is a method of projecting an image onto various, often very irregular, surfaces and objects [6]. It often involves using a few projectors to illuminate one object. These methods may be used to display common image by multiple projectors, one next to the other in an expanded projection area. As the projection area is expanded, projectors may be placed closer to the surface so that light intensity will be bigger.

Expanding the projection area has a few drawbacks for that use case. Firstly it requires placing projectors closer to the surface, which may be inconvenient or impossible. Secondly, when using, e.g., two projectors aspect ratio of projection will be non-standard - since we increased either width or height by two times. Thirdly, projection mapping methods are typically complex to perform advanced tasks like geometry correction. In our case, simpler methods may be used.

Nevertheless, the method presented here and also aforementioned commercial solutions use similar methods to projection mapping. The objective of this work was to create a simple and accessible solution.

C. Epson's automatic solutions

One solution for automatic projector stacking was developed by Epson [7]. This technology is a bit similar to the one presented in this work - the user connects projectors, sets calibration settings in the program, then a sequence of SL patterns is projected. Nevertheless, it still has many disadvantages. It is intended only for a small subset of Epson's high-end devices. Moreover, it needs to display many images to calibrate the system - on a shared promotional video, the calibration of two projectors takes around 2 minutes.

D. domeprojection.com solution

Company *domeprojection.com* developed its own solution [8]. It is most similar to the one presented in this work. Projectors are calibrated using a smartphone as a camera. As a calibration pattern ArUco markers are used, which allows the use of only one image per projector, so calibration is fast. However, it is working with only few Barco projectors.

III. Method

A. Method overview

The method assumes that N projectors $(P_1, P_2 ... \ P_N)$ are projecting to mostly common projection area on flat surface S. There is also one camera (smartphone) O set up to observe the projection area. In that case, the camera observes images displayed by each projector. Input image sent to projector P_i is seen by the camera as warped by some homography transformation H_i as seen in Fig. 1.

If we know all homographies H_i, we can assume a certain parametric surface $S(s,t)$ we use as a new, common projection area for all projectors (*Region of interest* - ROI). In practice, we want to use a rectangle with commonly used aspect ratio (like 16×9, 16×10, or 4×3). Our goal is to calculate for every projector a map $M_i(u,v)$ which transforms input image $I(u,v)$ so it will be seen by the camera in selected projection area $S(s,t)$.

To achieve that for all projector coordinates (u,v), we can calculate their position in camera space by applying homography and then find coordinates of that point in new projection area space $S(s,t)$.

Therefore the method is split into three main steps: acquiring data, calculating homographies, and calculating transformation map.

B. Acquiring data

As introduced above, the camera sees the input image of each projector P_i as transformed by homography H_i into camera space. It follows that if we have a set of corresponding points in the camera and input image, we can calculate homography. Precisely speaking, homography has 8 degrees of freedom ([9], p. 33), so it can be calculated using only 4 points (e.g., projection area corners). However, to achieve better accuracy it is best to have more points.

Finding correspondence between the projector and camera is a well-known problem. These methods consist of capturing by a camera set of patterns displayed by a projector. There are many of these methods, e.g. work [10] lists more than 40 algorithms. The goal of these algorithms is to:

- detect as many points as possible - preferably to get a dense map - for each camera pixel observing the projecting area,
- achieve good accuracy of detected points,
- achieve good robustness - resistance to uncontrolled parameters like ambient light,
- use as few images as possible.

In our use case, we need a pattern that:

1) uses as few images as possible to speed up calibration - preferably only one,
2) features good accuracy of detected points,

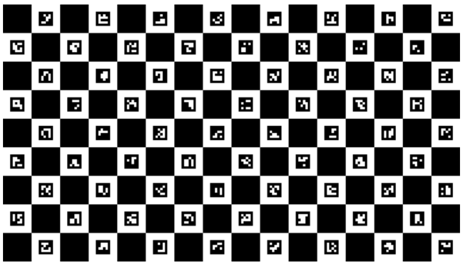


Fig. 2. ChArUco board used in tests

3) features good resistance to ambient light.

The number of detected points is not that important since, as already mentioned, homography can be calculated using only 4 points. Also, the assumption that projection is onto a flat surface allows us to choose methods that work well on planes but have problems on non-plane surfaces.

Taking all of this into account, we used the ChArUco board as the calibration method. ChArUco board combines chessboard and ArUco markers - ArUco markers are placed into white fields of a chessboard. Chessboard is the pattern used commonly for camera calibration since chessboard field edges and corners can be detected with high, subpixel accuracy even when they are blurred [11]. ArUco markers are binary (white-black) square markers. Combining both techniques in one pattern makes it possible to keep high accuracy detection of chessboard corners with unique identification provided by ArUco. Since this is still a white-black pattern, it is very resistant to ambient light. There is a popular and good implementation of ChArUco board detection in openCV [12], but new methods are also being developed [13].

The ChArUco board may be adjusted by modifying the size of squares. Since to calculate homography, only a few points are needed, in tests we used the board presented in Fig. 2. It is a 16×9 board, so there are 144 fields. Therefore for ArUco we can use 4×4 markers. That way pattern is very robust (big markers, small ArUco dictionary) while providing around 100 detected points.

So in the first phase of calibration ChArUco board is displayed and captured by the camera. Then points (chessboard corners) are detected. As a result, we got a list of pairs of points in input image space and camera space.

C. Calculating homographies from detected points

Pairs of points acquired in the previous step may now be used to calculate homography. Unfortunately, homography is a transformation in homogeneous coordinates, and therefore it cannot be found as a solution to a system of linear equations. Nevertheless, although it is a non-linear system, it is still very simple.

D. Calculating transformation map

The last step is calculating the transformation map, which specifies how the input image should be warped. For each image coordinates (u, v) we can find camera coordinates using found homography. Then if a point is outside of ROI, we want to display black here - so we insert a special value e.g. -1, as a coordinate. If the point is inside ROI, we must find coordinates in that ROI space (s, t). The exact method depends on the ROI used. For rectangular ROI, we can calculate the inverse of bilinear interpolation to get coordinates. Then the value of the transformation map on point (u, v) is (s, t).

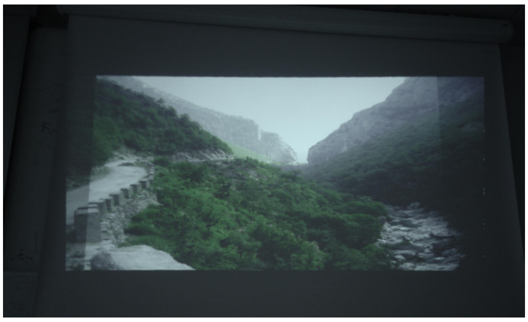


Fig. 3. System at work. In this case, we selected ROI bigger than common projector area so the brightness gain clearly is visible.

An important aspect is the method of selection of ROI. We used two algorithms. The first selects a maximal rectangle with a given aspect ratio within the common part of all projection areas. That method assumes that the observer position is similar to the camera position so that the observer will see a rectangular projection. The second method assumes one main projector. In that method, ROI is based on quadrilateral being the projection area of that main projector scaled to be contained within the common part of all projection areas. That method assumes that the main projector has good geometry from the observer's point of view.

IV. Tests and results

Measuring image quality improvement is a hard task since it depends on many factors. As already mentioned in our case improved image quality is a result of increased brightness and thus contrast and potential increased effective resolution as described in [4]. On the other hand, bad calibration quality will make images not overlap and thus degrade quality. Considering that, tests were designed to measure image quality changes with one or more projectors in different light conditions.

The test procedure was based on LogMAR charts [14]. These patterns were designed to measure visual acuity. Normally one chart (which is printed, not displayed, so it has a very high contrast ratio) is used to measure the visual acuity of different people. In our case, we assumed that different imperfect charts (displayed by projectors in different conditions, so with various contrast) seen by one person in one position could be used to measure image quality.

Normally LogMAR chart consists of a few rows of letters. Each row is assigned points equal to logarithm of letter size in that row measured in minutes of visual angle. A person with normal sight sees details as big as 1 minute of visual angle corresponding to 0.0 LogMAR points. Higher values mean worse eyesight and smaller better eyesight.

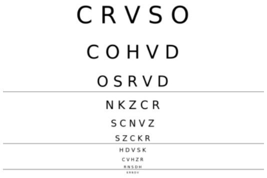


Fig. 4. Example of LogMAR type pattern used for tests.

In our case, we used a chart consisting of 10 rows with 5 letters each. The first row used the biggest font size, and each successive row used font size $\sqrt[3]{2}$ smaller than the previous. That means that in the LogMAR system, each row was worth 0.1 points more than the previous. Because letter sizes were not normalized and we do not need (and can) measure real visual acuity we decided to simplify these calculations. Each fully recognized row is worth 1 point, and each correctly recognized letter in the first not fully recognized row is worth 0.2 points.

We assumed that we could simultaneously perform tests on multiple people and averaged the results. Each person has different eyesight and sits in a different position relative to the projection area, but if the usage of the system changes image quality, we should observe a change in average LogMAR results.

The test consisted of displaying charts described above to 60 people in four consecutive scenarios: (a) using one projector, (b) using two projectors, (c) using one projector but with an additional illumination from an additional projector displaying a plain white pattern, (d) like (c) but using two projectors. That means that tests show results for 1 or 2 projectors in the presence of more or less ambient light. Each participant was asked to write down displayed letters and also subjectively rate visibility in each scenario.

The tested projectors were two Benq MX660P. This model utilizes DLP, lamp, and has a native resolution of 1024×768. Smartphone Lenovo K6 Note was used as the camera - it has a 16MP, 4632×3474 sensor.

Results are presented in Fig. 5. It shows that in both scenarios using two projectors improves image quality, and improvement is bigger in the presence of brighter ambient light - the difference is 0.4 points between (b) and (a) scenario and 0.46 between (d) and (c). Also, 82% of the people stated that visibility is better when using two projectors, 13% that differences are negligible, and only 5% that using two projectors decrease image quality. Since we cannot measure all factors exactly, no conclusions can be drawn from these data about the exact quality of improvement, but only that system really improves quality in a bright environment.

V. Summary

Obtained results show that the proposed approach significantly improve projection quality. Therefore this means that other commercial solutions based on similar ideas could be useful, which was not proved before. Above all, however, these results mean that it is possible to create a similar solution without relying on specific, high-end hardware and to develop an open solution to be commonly used.

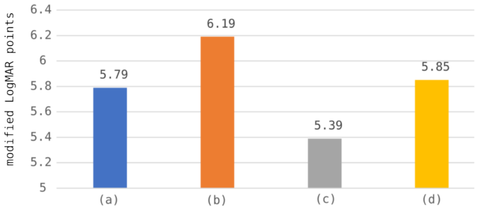


Fig. 5. Number of points scored in each test: (a) one projector, (b) two projectors, (c) one projector with additional illumination, (d) two projectors with additional illumination.

Another aspect is further development of the described method. As for now it has many advantages to similar technologies, but it may be further improved by researching faster, higher-precision correspondence finding methods in bright environments. Another field of further research is a combination of the described method with other projection mapping applications like geometry correction.

References

[1] F. W. Campbell and J. G. Robson, “Application of fourier analysis to the visibility of gratings,” *The Journal of Physiology*, vol. 197, no. 3, pp. 551–566, 1968. doi: 10.1113/jphysiol.1968.sp008574

[2] B. GJ, “Contrast discrimination by the human visual system,” vol. 40, 1981. doi: 10.1007/BF00326678

[3] S. E. Smith, “Multi-axis gimbal extended pixel resolution actuator,” Patent US20 190 227 261A1, 2018. [Online]. Available: https://patents.google.com/patent/US20190227261A1

[4] D. G. Aliaga, Y. H. Yeung, A. Law, B. Sajadi, and A. Majumder, “Fast high-resolution appearance editing using superimposed projections,” *ACM Trans. Graph.*, vol. 31, no. 2, 4 2012. doi: 10.1145/2159516.2159518. [Online]. Available: https://doi.org/10.1145/2159516.2159518

[5] Epson. Stacking multiple projectors for increased brightness and 3-d projection. Accessed 18.04.2023. [Online]. Available: www.youtube.com/watch?v=favBGq9iLRk

[6] A. Grundhöfer and D. Iwai, “Recent advances in projection mapping algorithms, hardware and applications,” *Computer Graphics Forum*, vol. 37, no. 2, pp. 653–675, 2018. doi: https://doi.org/10.1111/cgf.13387

[7] Epson. Epson projector professional tool - geometry correction assist for stacking. Accessed 18.04.2023. [Online]. Available: www.youtube.com/watch?v=xHpkC4YYe5I

[8] Stacker app. Accessed 18.04.2023. [Online]. Available: www.domeprojection.com/products/stacker-app/

[9] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004.

[10] J. Salvi, S. Fernandez, T. Pribanić, and X. Lladó, “A state of the art in structured light patterns for surface profilometry,” *Pattern Recognit.*, vol. 43, pp. 2666–2680, 2010.

[11] A. Fabijańska, “Gaussian-based approach to subpixel detection of blurred and unsharp edges,” in *Proceedings of the 2014 Federated Conference on Computer Science and Information Systems*, vol. 2. IEEE, 2014. doi: 10.15439/2014F136 pp. pages 641–650.

[12] G. Bradski, “The OpenCV Library,” *Dr. Dobb’s Journal of Software Tools*, 2000.

[13] D. Hu, D. DeTone, and T. Malisiewicz, “Deep charuco: Dark charuco marker pose estimation,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. doi: 10.1109/CVPR.2019.00863 pp. 8428–8436.

[14] I. L. Bailey and J. Lovie, “New design principles for visual acuity letter charts,” *Optometry and Vision Science*, vol. 53, p. 740–745, 1976.

Communication Papers of the 18th Conference on Computer Science and Intelligence Systems pp. 157–164

DOI: 10.15439/2023F9294
ISSN 2300-5963 ACSIS, Vol. 37

Mushroom Picking Framework with Cache Memories for Solving Job Shop Scheduling Problem

Piotr Jedrzejowicz, Izabela Wierzbowska
0000-0001-6104-1381
0000-0003-4818-4841
Department of Information Systems,
Gdynia Maritime University,
ul. Morska 81-87, 81-225 Gdynia, Poland;
Email: {p.jedrzejowicz, i.wierzbowska}@wznj.umg.edu.pl

***Abstract*—Applying population-based metaheuristics is a known method of solving difficult optimization problems. In this paper the search for the best solution is conducted by decentralized, self-organized agents, working in parallel threads, in the so called mushroom-picking method. The search is enhanced by remembering in which part of the recently improved solution the last successful change took place and intensifying the search in this part. A computational experiment shows that introducing the component for remembering the most recent changes may improve the results obtained by the model in the case of JSSP problems.**

I. Introduction

POPULATION-BASED methods belong to the most effective approaches when dealing with computationally difficult optimization problems, including combinatorial optimization ones. A population, in a broad term, represents here solutions of the problem, potential solutions, parts of solutions, or some constructs that can be somehow transformed into solutions.

The main focus while using population-based methods is to find a proper mechanism controlling their three fundamental components - intensification, diversification, and learning [1]. Diversification is understood as the method of identifying diverse promising regions over the whole search space, while intensification is the method of finding a solution by exploring some promising regions. Learning is gaining and using the knowledge of where and how applying intensification and diversification operations. Population-based methods belong to a wider class of metaheuristics. Metaheuristics differ between themselves by using different intensification, diversification and learning rules.

Pioneering population-based methods included genetic programming – (GP) [2], genetic algorithms (GA) [3], evolutionary computations - EC [4], [5], ant colony optimization (ACO) [6], particle swarm optimization (PSO) [7] and bee colony algorithms (BCA) [8]. Since the nineties a massive number of metaheuristics using the population-based paradigm have been proposed (see reviews [9], [10]), some of them named after biological, physical, chemical, and other natural sciences phenomena. Generally, the discussed metaheuristics have achieved a certain level of perfection in finding solutions to many computationally difficult problems. None of them, however, could be considered a champion and their performances differ depending on the problem characteristics and specifications. Moreover, a vast majority of well-performing population-based metaheuristics require fine-tuning or even adaptation to produce high-quality solutions to difficult computational problems.

An important step towards increasing the effectiveness of the population-based methods was the emergence, during the last few decades, of commonly accessible technologies enabling parallelization of search over the solution space. Among the successful attempts to parallelize searching for the best solution among the population of solutions was parallel GA on MapReduce framework using the Hadoop cluster [11]. The method was designed for solving instances of the travelling salesman problem (TSP). A similar approach for implementing a parallel genetic algorithm with the Hadoop MapReduce for TSP can be found in [12]. Spark-based ant colony optimization algorithm for solving the TSP was proposed in [13]. A Spark-based version of the population learning metaheuristic applied, among others, to job-shop scheduling problem (JSSP) can be found in [14]. Some extensive reviews of developments in the field of parallel metaheuristics can be found in [15], [16], and [17].

Job-shop scheduling problem (JSSP) is one of the “classic” computationally hard problems. In recent years several approaches to solving the JSSP using population-based metaheuristics have been proposed. Some examples of such approaches include:

- Specialized cuckoo search algorithm [18].
- A hybrid particle swarm optimization (PSO) and neural network algorithm [19].
- An improved whale optimization algorithm [20].
- Genetic Algorithm for JSSP [21].
- Wolf pack algorithm for JSSP [22].

- Biomimicry hybrid bacterial foraging optimization algorithm for JSSP [23].
- Hybrid harmony search algorithm for JSSP [24].
- Discrete particle swarm optimization (PSO) algorithm for JSSP [25].
- Hybrid PSO optimization algorithm with nonlinear inertial weight and Gaussian mutation for JSSP. [26]

To take advantage of the expected speed-up several attempts of using parallel computational environments for solving the JSSP have been also recently reported. MapReduce coral reef algorithm for solving JSSP instances was proposed by [27]. The distributed Evolutionary Algorithm for scheduling large-scale problems was suggested by [28]. An efficient parallel tabu search for the blocking job shop scheduling problem was suggested by [29].

Another metaheuristic designed for parallel environments named Mushroom Picking Algorithm (MPA) was proposed by the authors in [30]. The approach proved successful in solving the JSSP instances. Motivated by the good performance of the MPA when solving the JSSP we proposed in [31] an extension of the MPA in the form of the software framework named Mushroom Picking Framework (MPF). The MPF provides a generic functionality of parallel searching for the best solution among population members and is implemented as a multiple-agent system. MPF can be adopted by a user to fit particular combinatorial problem requirements.

In this paper, we further extend the MPF by equipping each solution with a cache memory where recent changes that occur during the search process are stored. The extended MPF is denoted as MPF+. The rest of the paper is organized as follows. Section 2 contains a description of the Mushroom Picking Framework with cache memories. Section 3 recalls briefly the Job Shop Scheduling Problem. Section 4 explains our implementation of the proposed MPF with cache memories for solving JSSP instances. Section five presents the results of the computational experiment held to validate the approach. The final section contains conclusions and suggestions for future research.

II. Mushroom Picking Framework with cache memories

A. Mushroom Picking Algorithm (MPA)

Mushroom Picking Algorithm was introduced in [30] for solving difficult optimization problems. It is characterized by the following features:

- It operates on a population of individuals representing solutions to the given combinatorial optimization problem.
- It uses a set of agents, that may improve a solution or solutions from the population
- Randomly selected agents try to improve randomly selected solutions. If successful, the resulting solution may replace a solution from the population.

Each agent reads a solution or solutions, depending on its improvement method and requirements as to the number of the input solutions (here one or two). Each agent then runs its internal improvement algorithm creating a new solution. If the fitness of a newly generated solution is better than the fitness value of the solution drawn from the population, or is better than the worse fitness of the two initially drawn solutions, it replaces the worse solution.

In order to avoid obtaining a population with a very low diversity of solutions, whenever two solutions drawn from the population are similar to one another, one of them is replaced by a new, random solution. The similarity of solutions is decided through a comparison of the respective fitness values. If these values are identical or differ by a predefined value, solutions are considered identical. The procedure allows for the maintenance of the required diversity of solutions in the population.

B. Mushroom Picking Framework (MPF)

Mushroom Picking Framework works in the following manner:

- The initial population of solutions is generated.
- The search for the best solution is performed in cycles.
- In each cycle, the population of solutions is divided into several subpopulations of equal size.
- Using the Apache Spark functionality subpopulations are independently and in parallel explored by improvement agents. Each subpopulation is processed in a separate thread.
- In each subpopulation the Mushroom Picking Algorithm is used to improve solutions.
- After each cycle, all the solutions from the subpopulations are drawn back into the common memory and shuffled.
- A predefined number of the worst solutions may be replaced by the currently best one. Then the next cycle begins.

The process of searching for the best solution is iterative and runs as described by Algorithm 1. The stopping criterion is defined by the maximum number of consecutive cycles in which the best solution in the population does not improve (the process ends when the best makespan of the solutions has not changed for predefined number of consecutive cycles).

What happens within the method $optimize$ is shown as Algorithm 2. In each subpopulation, the process of applying improvement agents to solutions - MPA - is represented by the $p.applyOptimizations$ in Algorithm 2. Attempted improvements are executed by improvement agents. The set of agents in each subpopulation is identical and consists of an equal number of agents of the same type.

Algorithm 1 Mushroom Picking Framework operation

```
$solutions \leftarrow$ set of random solutions;
while ! stoppingCriterion do
    $solutions \leftarrow solutions.optimize$;
    $bestSolution \leftarrow$ the best solution chosen from $solutions$;
end while
return $bestSolution$;
```

Algorithm 2 Method *optimize*

```
Require:
    $solutions$ = set of solutions;
    $k$ = number of parallel threads in which the solutions will be processed;
Ensure:
    $populations \leftarrow solutions$ divided to a list of $k$-element subpopulations;
    $populationsRDD \leftarrow populations$ parallelized in Apache Spark;
    $populationsRDD \leftarrow populationsRDD.map(p => p.applyOptimizations)$;
    $solutions \leftarrow$ solutions collected from $populationsRDD$;
    $solutions \leftarrow solutions$ with $l$ worst solutions replaced with the best solution;
```

The framework may be used for all problems, for which a task, a solution with a method that returns the fitness value, and a set of agents working as in MPA is defined.

C. Cache memories

For the JSSP instances, we use the MPF implementation proposed in [31] extended by adding the cache memory to each population member. The extended framework will be denoted as MPF+.

The general assumption is that solutions are encoded in the form of lists. In our framework improvement agents try to improve current solutions by moving, swapping, or modifying parts of the lists that encode solutions. The proposed cache memory is used to record and store for each solution the position (index in the list representing the solution), in which the last successful improvement move or change took place. The above feature helps to intensify the search for successful moves in the vicinity of the recent change. The idea of using information stored in the cache memory is to improve the synergistic effects of agent interactions, by providing the current information on which part of the solution they should focus it the next step.

Since we consider solutions that are represented as a list, changing an element in such a list (for example moving or swapping it with another element), results in saving its position together with the respective weight which is allocated by the user. In the next iteration of improvement in $ApplyOptimizations$, the new starting position for a change is drawn at random from the close neighborhood of the element at the saved position, and either the weight's value is reduced by one or - if the next change successfully led to a solution with better makespan - a new position with the maximum weight is remembered. When after several iterations the weight reaches value 0, the saved position receives a random value.

The process described above is shown as Algorithm 3. For simplicity's sake, Algorithm 3 covers the case of a single argument agent. The cache memory is represented by $solution.position$ and $solution.weight$ values. When creating a random solution, its $position$ and $weight$ are set to random position and maximum weight. The radius and the maximum weight are both predefined as the algorithm parameters.

III. Job Shop Scheduling Problem (JSSP)

Job shop scheduling problem (JSSP) is a well-known NP-hard optimization problem in which n jobs $(J_1, \ldots J_n)$ must be processed on m machines $(m_1, \ldots m_m)$.

In JSSP each job consists of a list of operations, the operations must be processed in the exact order as in the given list, and only after all preceding operations have been completed. Also every operation has to be processed on a specific machine in the given time. The operations cannot be interrupted and each machine can process only one operation at a time.

The makespan is defined as the length of the schedule, or the time in which all operations of all jobs will be processed. The JSSP objective is to find such schedule, that its makespan is minimal.

In JSSP a solution may be represented as sequence of jobs' numbers of the length $\leq n \times m$. In this sequence each job j appears at most m times, and the i-th occurence of the job corresponds to the i-th operation of this job. The algorithm in this paper uses this representation to find the solution with the smallest makespan.

Fig. 1 presents a solution of JSSP problem that may be represented by, for example, list (1,2,0,1,1,2,0,0) or (2,1,0,1,2,1,0,0).

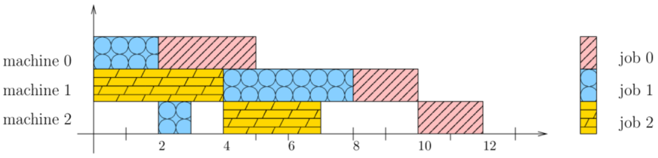


Fig. 1: Solution of JSSP task with makespan 12

IV. MPF+ Implementation for solving JSSP instances

In the Mushroom Picking Framework, one has to define the task, the solution and the agents. We implement the solutions as lists of job numbers - as it has been described in the previous section. All solutions in the initial population are randomly generated. The agents that try to improve solutions transform the lists by changing the order of elements or moving the elements to different positions. In the MPF+ for JSSP, the following agents are used:

Algorithm 3 *ApplyOptimizations* with cache memory of recent changes

```
Require:
  $W$ =maximum weight;
  $R$ =radius;
Ensure:
  for $iteration$ in the given range do
    $agent$ ← agent drawn at random according to some probability set by the user;
    $solution$ ← random solution;
    if $solution.weight > 0$ then
      $newWeight \leftarrow solution.weight - 1$
      $newPosition \leftarrow random(solution.position - R, solution.position + R)$;
    else
      $newWeight \leftarrow solution.weight$;
      $newPosition \leftarrow random(0, solution.size)$;
    end if
    $optimizedSolution$ with $optPosition$ and $optWeight \leftarrow agent(solution, start = newPosition)$;
    if $optimizedSolution$ is better than $solution$ then
      return $optimizedSolution$ with $optPosition$ and $optWeight$;
    else
      return $solution$ with $newPosition$ and $newWeight$;
    end if
  end for
```

- RandomSwap - replaces pairs representing jobs on two random positions in the list of pairs. If successful, the position of the first swapped element is remembered as the base for future exploring.
- RandomMove - moves one element representing the job and moves it to another, random position. If successful, the original position of the element is remembered as the base for future exploration.
- RandomOrder - takes a random slice of the list and shuffles the elements in this slice (the order of the slice' elements changes at random). If successful, the middle element of the slice is remembered as the base for future exploration.
- RandomCrossover - requires two randomly drawn solutions. A slice from the first solution is extended with the missing elements in the order as in the second solution. If successful, the middle element of the slice is remembered as the base for future exploration.

Each agent stores in the solution's memory index of one element of the list representing solution. When the solution is again sent to an agent (in the next iteration in the *ApplyOptimizations* method), the agent will draw the starting point for the transformation from the part of the solution given by the range of indices: $(solution.position - R, solution.position + R)$, where R is given as the algorithm parameter.

Each iteration of the *ApplyOptimizations* method starts with drawing at random an agent. The agents are drawn with the following probabilities: 0.28 for each one-argument agent and 0.14 for RandomCrossover agent, to maintain the empirically identified required frequency of calling a single and double argument agents.

V. Results

A. Computational experiment

To validate the proposed approach, we have carried out several computational experiments. Experiments were run on a benchmark dataset for the JSSP problem: the set of 40 instances proposed by Lawrence [32], that have sizes from 5x10 to 15x15. All computations have been run on the Spark cluster at the Centre of Informatics Tricity Academic Supercomputer and Network (CI TASK) in Gdansk. In all experiments 240 subpopulations have been used, each consisting of 3 solutions. These subpopulations have not been processed literally in parallel due to a varying temporary constraint on the number of available nodes. Using a cluster with more allocated nodes would lead to shorter computation times, as demonstrated in [31].

The use of our cache memory has been controlled by two parameters R and W. R is used to define the range of solution elements from which the next starting point for an agent will be drawn. The starting point is drawn from $(position - R, position + R)$. R parameter has been set to 5 in all experiments. W is the weight assigned to a solution after an agent performs a change. Its value was set to 0, 10, or 15, where 0 value results in not using the cache memory at all. For tasks from la01 to la15 and task la31, if the solution was calculated using the cache memory, the weight of 10 was used. For the remaining solutions, their initial weight was set as 15.

The time of computations mainly depends on the number of iterations in one cycle, and the stopping criterion, which is the maximum number of cycles in which the best solution does not change, denoted mwc. The number of iterations was set to 1000, 3000 or 6000 iterations in one cycle. The mwc was set to 2 or 5. The values of parameters that were used in the experiment are described in the Table I.

TABLE I: Parameters used at the experiments

Task	R and W if cache used	Number of iterations	
la01	5, 10	1000	2
la02	5, 10	3000	2
la03	5, 10	3000	5
la04	5, 10	3000	2
la05-la15	5, 10	3000	2
la16-la27	5, 15	3000	5
la28-la30	5, 15	6000	5
la31	5, 10	3000	2
la32	5, 15	3000	5
la33	5, 15	3000	2
la34-la35	5, 15	3000	5
la36-la40	5, 15	6000	5

Average computation times and average errors are shown in Table II. BKS stands for the best-known solution (in terms of the solution makespan) and the errors have been calculated for BKS values. The average errors and times have been calculated from at least 30 results.

Tables III and IV contain a comparison of results obtained by MPF+ with results obtained by other recently published algorithms. Table III contains Q-Learning Algorithm (QL, [33]) and a hybrid EOSMA algorithm [34] that mixes the strategies of Equilibrium Optimizer (EO) and Slime Mould Algorithm (SMA). Table IV shows results for the Coral Reef Optimization (CROLS, [35]). The average errors for CROLS, QL and EOSMA algorithms have been calculated based on average results given in the original papers. In the case of the Coral Reef Optimization results reported in [35] were given for only chosen instances of the problem. The Coral Reef algorithm was run for three different reef sizes. For each task in the table, the best Coral Reef Optimization result was chosen from among the three available results in [35]. In the case of QL0 and QL1 [33] running times of algorithms were not given. Algorithm EOSMA needed from over 10 seconds to 10^3 seconds of running time.

Figures 2 and 3 present convergence for six different runs of the algorithm for tasks ls03 and la26 respectively. For both figures such runs of the algorithm have been chosen for which solutions with the best known makespan were found. In three of the runs the cache memories were used (red lines with triangles), and three runs did not use the cache memories (blue lines with circles).

VI. Discussion

From Table II it can be seen, that in many cases intensifying the search using data stored in the proposed cache memory leads to better results obtained in comparable and even occasionally shorter times. To gain better knowledge of the performance of the proposed MPF+ implementation as compared with its earlier version (MPF) we have carried out a pairwise comparison using the Wilcoxon matched pairs tests.

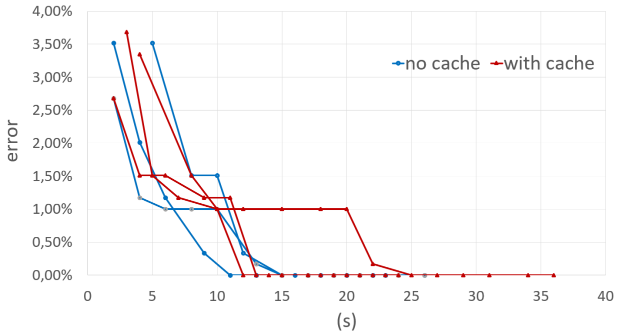


Fig. 2: Convergence for runs with and without cache on la03

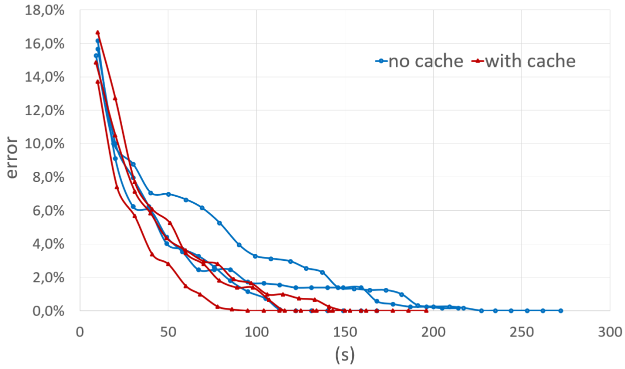


Fig. 3: Convergence for runs with and without cache on la26

The null hypothesis in such a case states that results produced by two different methods are drawn from samples with the same distribution. With T statistics equal to 53.00, Z statistics equal to 2.771429, and a p-value equal to 0.005581, the null hypothesis has to be rejected at the significance level of 0.05.

Analysis of results from Tables III and IV allows observing that MPF+ implementation for solving the JSSP instances performs well as compared with several other approaches offering for numerous instances better performance or shorter computation time.

From Fig. 2 and Fig. 3 it can be noticed that most runs in which the cache was used required less time to finish. The markers show the error of the best makespan found at the time when a cycle ends, and after most of the cycles the best makespan value found so far was better in cases when the cache was used.

VII. Conclusion

The main contribution of the paper is extending the earlier proposed Mushroom Picking Framework by incorporating the, so-called, cache memory. It serves to store recent changes to solutions effected by improvement agents. A novel version of the framework referred to as MPF+ can be used for solving a variety of computationally hard combinatorial optimization problems. The approach takes advantage of better controlling the intensification part of searching for the best solution. The

TABLE II: Average computation times and average errors obtained in the experiment

		MPF		MPF+	
Task	**BKS**	**Avg err**	**Avg time (s)**	**Avg err**	**Avg time (s)**
la01	666	0.00%	3.53	0.00%	**3,07**
la02	655	0.00%	11.33	0.00%	**9,93**
la03	597	0.45%	23.67	**0.25%**	**22,37**
la04	590	0.07%	10.97	**0.02%**	13.33
la05	593	0.00%	8.23	0.00%	8.40
la06	926	0.00%	11.93	0.00%	13.10
la07	890	0.00%	14.33	0.00%	**13.20**
la08	863	0.00%	10.23	0.00%	10.40
la09	951	0.00%	10.03	0.00%	10.10
la10	958	0.00%	12.23	0.00%	**10.63**
la11	1222	0.00%	13.20	0.00%	13.43
la12	1039	0.00%	13.50	0.00%	**13.27**
la13	1150	0.00%	13.57	0.00%	**13.20**
la14	1292	0.00%	13.23	0.00%	13.60
la15	1207	0.00%	18.23	0.00%	20.27
la16	945	0.47%	60.23	**0.34%**	**43.07**
la17	784	0.39%	42.83	**0.27%**	46.43
la18	848	0.44%	45.85	**0.33%**	48.63
la19	842	1.15%	54.13	1.34%	**47.23**
la20	901	0.65%	38.60	**0.63%**	**37.43**
la21	1046	3.94%	121.70	**3.57%**	125.93
la22	927	2.64%	152.43	2.80%	**113.73**
la23	1032	0.05%	96.23	0.07%	**87.07**
la24	935	4.14%	103.90	**4.12%**	111.63
la25	977	4.13%	126.53	**3.75%**	**122.73**
la26	1218	2.18%	199.20	**1.96%**	**197.83**
la27	1235	5.60%	171.40	**5.40%**	200.10
la28	1216	3.23%	321.67	3.31%	419.83
la29	1152	7.95%	363.13	**7.43%**	428.90
la30	1355	0.92%	375.63	**0.60%**	**347.87**
la31	1784	0.00%	146.73	0.00%	155.53
la32	1850	0.00%	227.83	0.00%	235.50
la33	1719	0.00%	222.00	0.00%	**176.17**
la34	1721	0.58%	366.90	0.62%	388.87
la35	1888	0.11%	274.83	**0.06%**	298.97
la36	1268	4.46%	290.15	4.51%	311.85
la37	1397	4.95%	372.18	**4.83%**	380.53
la38	1196	6.42%	434.65	**6.11%**	**411.63**
la39	1233	4.25%	375.90	**3.15%**	436.53
la40	1222	4.37%	350.20	**4.36%**	390.07
avg		1.59%	138.08	1.50%	143.81

mechanism helps enhance the synergetic effects of interactions between agents by providing constantly updated information pointing directly at a part of the solution in which the recent change caused some improvement of the fitness function value. As a test-bed for validation purposes, we have selected one of the classic computationally hard combinatorial optimization problems – the job shop scheduling problem. While incorporating the proposed cache memory has not caused a dramatic improvement in the quality of results, it helped nevertheless to improve some of them, and in many cases has led to a shortening of the computation time. For the JSSP results obtained in the experiments are competitive, even if the cluster environment that has served as the platform to run the programs has not been able to provide fully parallel computations for all threads.

We believe that the MPF+ could be further improved. Future research should focus on finding mechanisms for the automatic setting of weights values depending on the scale of improvements. Another possibility is to take advantage of reinforcement learning techniques for controlling and managing the course of computations.

Acknowledgment

This research was funded by Gdynia Maritime University, grant number WZNJ/2023/PZ/03

References

[1] F. Glover and M. Samorani, "Intensification, diversification and learning in metaheuristic optimization," *Journal of Heuristics*, vol. 25, 03 2019. doi: 10.1007/s10732-019-09409-w

[2] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, MA, USA: MIT Press, 1992. ISBN 0-262-11170-5

[3] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading, MA: Addison-Wesley, 1989.

[4] D. B. Fogel, "On the relationship between the duration of an encounter and the evaluation of cooperation in the iterated prisoner's dilemma," *Evol. Comput.*, vol. 3, no. 3, pp. 349–363, 1995. doi: 10.1162/evco.1995.3.3.349. [Online]. Available: https://doi.org/10.1162/evco.1995.3.3.349

TABLE III: Comparison of results obtained by MPF+ with other recently published results (average error and average running time)

Task	**MPF+**		**QL01** [33]	**QL02** [33]	**EOSMA** [34]
	err	**time (s)**	**err**	**err**	**err**
la01	0.00%	3.07	0.00%	0.15%	0.00%
la02	0.00%	9.93	4.58%	3.21%	0.00%
la03	0.25%	22.37	3.69%	5.03%	2.35%
la04	0.02%	13.33	5.08%	3.05%	0.00%
la05	0.00%	8.40	0.00%	0.00%	0.00%
la06	0.00%	13,10	0.00%	0.00%	0.00%
la07	0.00%	13,20	8.65%	0.11%	0.00%
la08	0.00%	10.40	1.51%	0.00%	0.00%
la09	0.00%	10.10	0.00%	0.00%	0.00%
la10	0.00%	10.63	0.00%	0.00%	0.00%
la11	0.00%	13.43	0.00%	0.00%	0.00%
la12	0.00%	13.27	0.00%	0.00%	0.00%
la13	0.00%	13.20	0.00%	0.00%	0.00%
la14	0.00%	13.60	0.00%	0.00%	0.00%
la15	0.00%	20.27	7.87%	4.06%	0.00%
la16	0.34%	43.07	4.02%	5.08%	3.24%
la17	0.27%	46.43	2.04%	2.55%	1.13%
la18	0.33%	48.63	2.95%	3.07%	2.22%
la19	1.34%	47.23	3.92%	6.29%	4.17%
la20	0.63%	37.43	4.22%	4.55%	1.67%
la21	3.57%	125.93	5.83%	12.81%	7.07%
la22	2.80%	113.73	10.25%	18.99%	5.79%
la23	0.07%	87.07	0.58%	6.59%	1.45%
la24	4.12%	111.63	6.95%	15.19%	7.91%
la25	3.75%	122,73	9.93%	14.23%	7.11%
la26	1.96%	197.83	6.90%	17.65%	4.82%
la27	5.40%	200.10	10.45%	18.95%	9.02%
la28	3.31%	419.83	11.68%	15.79%	7.26%
la29	7.43%	428.90	18.32%	24.91%	11.83%
la30	0.60%	347.87	2.58%	14.10%	3.88%
la31	0.00%	155.53	4.09%	6.84%	0.08%
la32	0.00%	235.50	3.46%	8.22%	0.19%
la33	0.00%	176.17	5.70%	6.92%	0.17%
la34	0.62%	388.87	6.22%	12.38%	2.16%
la35	0.06%	298.97	5.14%	11.55%	0.42%
la36	4.51%	311.85	11.59%	16.72%	6.38%
la37	4.83%	380.53	9.74%	14.96%	9.14%
la38	6.11%	411.63	11.54%	19.98%	11.19%
la39	3.15%	436.53	10.14%	18.09%	8.43%
la40	4.36%	390.07	7.28%	22.09%	8.93%
avg	1.37%	126.96	5.17%	8.35%	3.20%

[5] Z. Michalewicz, "Genetic algorithms + data structures = evolution programs," in *Springer Berlin Heidelberg*, 1996. doi: 10.1007/978-3-662-03315-9

[6] M. Dorigo, V. Maniezzo, and A. Colorni, "Ant system: optimization by a colony of cooperating agents," *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society*, vol. 26 1, pp. 29–41, 1996. doi: 10.1109/3477.484436

[7] R. Poli, J. Kennedy, and T. M. Blackwell, "Particle swarm optimization," *Swarm Intelligence*, vol. 1, pp. 33–57, 1995. doi: 10.1109/icnn.1995.488968

[8] T. Sato and M. Hagiwara, "Bee system: finding solution by a concentrated search," *1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation*, vol. 4, pp. 3954–3959 vol.4, 1997.

[9] H. Ma, S. Shen, M. Yu, Z. Yang, M. Fei, and H. Zhou, "Multi-population techniques in nature inspired optimization algorithms: A comprehensive survey," *Swarm Evol. Comput.*, vol. 44, pp. 365–387, 2019. doi: 10.1016/j.swevo.2018.04.011

[10] P. Jedrzejowicz, "Current trends in the population-based optimization," in *Computational Collective Intelligence*, N. T. Nguyen, R. Chbeir, E. Exposito, P. Aniorté, and B. Trawiński, Eds. Cham: Springer International Publishing, 2019. doi: 10.1007/978-3-030-28377-3_4343. ISBN 978-3-030-28377-3 pp. 523–534.

[11] H. R. Er and N. Erdogan, "Parallel genetic algorithm to solve traveling salesman problem on mapreduce framework using hadoop cluster," *JSCSE*, 01 2014. doi: 10.7321/jscse.v3.n3.57

[12] E. Alanzi and H. Bennaceur, "Hadoop mapreduce for parallel genetic algorithm to solve traveling salesman problem," *International Journal of Advanced Computer Science and Applications*, vol. 10, 01 2019. doi: 10.14569/IJACSA.2019.0100814

[13] Y. Karouani and Z. Elhoussaine, "Efficient spark-based framework for solving the traveling salesman problem using a distributed swarm intelligence method," in *2018 International Conference on Intelligent Systems and Computer Vision (ISCV)*, 2018. doi: 10.1109/ISACV.2018.8354075 pp. 1–6.

[14] P. Jedrzejowicz and I. Wierzbowska, "Apache spark as a tool for parallel population-based optimization," in *Intelligent Decision Technologies 2019*, I. Czarnowski, R. J. Howlett, and L. C. Jain, Eds. Singapore: Springer Singapore, 2020. ISBN 978-981-13-8311-3 pp. 181–190.

[15] E. Alba, G. Luque, and S. Nesmachnow, "Parallel metaheuristics: Recent advances and new trends," *International Transactions in Operational Research*, vol. 20, pp. 1–48, 08 2012. doi: 10.1111/j.1475-3995.2012.00862.x

[16] P. González, X. Pardo, R. Doallo, and J. Banga, "Implementing cloud-based parallel metaheuristics: an overview," *Journal of Computer Science and Technology*, vol. 18, p. e26, 12 2018. doi: 10.24215/16666038.18.e26

TABLE IV: Comparison of results obtained by MPF+ with other recently published results (average error and average running time for chosen la instances)

Task	**MPF+**		**CROLS1** [35]		**CROLS2** [35]	
	err	**time (s)**	**err**	**time (s)**	**err**	**time (s)**
la01	0.00%	3.07	0.00%	39,91	0.00%	15.64
la02	0.00%	9.93	0.00%	40.91	0.00%	15.27
la06	0.00%	13,10	0.00%	151.82	0.00%	92.45
la07	0.00%	13,20	0.08%	148.18	0.00%	88.73
la11	0.00%	13.43	0.00%	224.09	0.00%	136.55
la12	0.00%	13.27	0.03%	228.55	0.00%	149.91
la16	0.34%	43.07	0.29%	125.91	0.39%	132.55
la17	0.27%	46.43	0.17%	198.55	0.39%	130.09
la21	3.57%	125.93	0.27%	269.36	0.63%	165.55
la22	2.80%	113.73	0.62%	185.18	0.56%	265.55
la26	1.96%	197.83	0.98%	281.73	1.01%	440.36
la27	5.40%	200.10	0.33%	260.18	0.34%	447.36
la32	0.00%	235.50	0.16%	453.45	0.12%	418.45
la33	0.00%	176.17	0.08%	643.09	0.29%	617.27
la39	3.15%	436.53	0.70%	675.45	0.37%	502.82
la40	4.36%	390.07	1.33%	585.45	2.16%	495.55
avg	1.37%	126.96	0.32%	281.99	0.39%	257.13

[17] M. Essaid, L. Idoumghar, J. Lepagnot, and M. Brévilliers, "GPU parallelization strategies for metaheuristics: a survey," *International Journal of Parallel, Emergent and Distributed Systems*, vol. 34, pp. 1–26, 01 2018. doi: 10.1080/17445760.2018.1428969

[18] H. Hu, W. Lei, X. Gao, and Y. Zhang, "Job-shop scheduling problem based on improved cuckoo search algorithm," *International Journal of Simulation Modelling*, vol. 17, pp. 337–346, 06 2018. doi: 10.2507/IJSIMM17(2)CO8

[19] Z. Zhang, Z. Guan, J. Zhang, and X. Xie, "A novel job-shop scheduling strategy based on particle swarm optimization and neural network," *International Journal of Simulation Modelling*, vol. 18, pp. 699–707, 12 2019. doi: 10.2507/IJSIMM18(4)CO18

[20] J. Zhu, Z. Shao, and C. Chen, "An improved whale optimization algorithm for job-shop scheduling based on quantum computing," *International Journal of Simulation Modelling*, vol. 18, pp. 521–530, 09 2019. doi: 10.2507/IJSIMM18(3)CO13

[21] X. Chen, B. Zhang, and D. Gao, "Algorithm based on improved genetic algorithm for job shop scheduling problem," in *2019 IEEE International Conference on Mechatronics and Automation (ICMA)*. IEEE Press, 2019. doi: 10.1109/ICMA.2019.8816334. ISBN 978-1-7281-1698-3 p. 951–956. [Online]. Available: https://doi.org/10.1109/ICMA.2019.8816334

[22] F. Wang, Y. Tian, and X. Wang, "A discrete wolf pack algorithm for job shop scheduling problem," in *2019 5th International Conference on Control, Automation and Robotics (ICCAR)*, 2019. doi: 10.1109/ICCAR.2019.8813444 pp. 581–585.

[23] A. Vital-Soto, A. Azab, and M. Baki, "Mathematical modeling and a hybridized bacterial foraging optimization algorithm for the flexible job-shop scheduling problem with sequencing flexibility," *Journal of Manufacturing Systems*, vol. 54, pp. 74–93, 01 2020. doi: 10.1016/j.jmsy.2019.11.010

[24] H. Piroozfard, K. Y. Wong, and A. D. Asl, "A hybrid harmony search algorithm for the job shop scheduling problems," in *2015 8th International Conference on Advanced Software Engineering & Its Applications (ASEA)*, 2015. doi: 10.1109/ASEA.2015.23 pp. 48–52.

[25] R. Krishnaswamy and C. Rajendran, "A novel discrete PSO algorithm for solving job shop scheduling problem to minimize makespan," *IOP Conference Series: Materials Science and Engineering*, vol. 310, p. 012143, 02 2018. doi: 10.1088/1757-899X/310/1/012143

[26] H. Yu, Y. Gao, L. Wang, and J. Meng, "A hybrid particle swarm optimization algorithm enhanced with nonlinear inertial weight and gaussian mutation for job shop scheduling problems," *Mathematics*, vol. 8, no. 8, p. 1355, Aug 2020. doi: 10.3390/math8081355. [Online]. Available: http://dx.doi.org/10.3390/math8081355

[27] C.-W. Tsai, H.-C. Chang, K.-C. Hu, and M.-C. Chiang, "Parallel coral reef algorithm for solving JSP on spark," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2016. doi: 10.1109/SMC.2016.7844511 pp. 001 872–001 877.

[28] L. Sun, L. Lin, H. Li, and M. Gen, "Large scale flexible scheduling optimization by a distributed evolutionary algorithm," *Computers & Industrial Engineering*, vol. 128, pp. 894–904, 2019. doi: https://doi.org/10.1016/j.cie.2018.09.025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S036083521830442X

[29] A. Dabah, A. Bendjoudi, A. AitZai, and N. Nouali-Taboudjemat, "Efficient parallel tabu search for the blocking job shop scheduling problem," *Soft Computing*, vol. 23, p. 13283–13295, 12 2019. doi: 10.1007/s00500-019-03871-1

[30] P. Jedrzejowicz and I. Wierzbowska, "Parallelized swarm intelligence approach for solving TSP and JSSP problems," *Algorithms*, vol. 13, no. 6, p. 142, Jun 2020. doi: 10.3390/a13060142. [Online]. Available: http://dx.doi.org/10.3390/a13060142

[31] P. Jedrzejowicz, E. Ratajczak-Ropel, and I. Wierzbowska, "A population-based framework for solving the job shop scheduling problem," in *Computational Collective Intelligence: 13th International Conference, ICCCI 2021, Rhodes, Greece, September 29 – October 1, 2021, Proceedings*. Berlin, Heidelberg: Springer-Verlag, 2021. doi: 10.1007/978-3-030-88081-1_26. ISBN 978-3-030-88080-4 p. 347–359. [Online]. Available: https://doi.org/10.1007/978-3-030-88081-1_26

[32] S. Lawrence, "Resource constrained project scheduling - technical report," Carnegie-Mellon University: Pittsburgh, PA, USA, Tech. Rep., 1984.

[33] M. A. Belmamoune, L. Ghomri, and Z. Yahouni, "Solving a job shop scheduling problem using Q-learning algorithm," in *Service Oriented, Holonic and Multi-Agent Manufacturing Systems for Industry of the Future*, T. Borangiu, D. Trentesaux, and P. Leitão, Eds. Cham: Springer International Publishing, 2023. doi: 10.1007/978-3-031-24291-5_16. ISBN 978-3-031-24291-5 pp. 196–209.

[34] Y. Wei, Z. Othman, K. Mohd Daud, S. Yin, and Q. Luo, "Equilibrium optimizer and slime mould algorithm with variable neighborhood search for job shop scheduling problem," *Mathematics*, vol. 10, p. 4063, 11 2022. doi: 10.3390/math10214063

[35] C.-S. Shieh, T.-T. Nguyen, W.-W. Lin, D.-C. Nguyen, and M.-F. Horng, "Modified coral reef optimization methods for job shop scheduling problems," *Applied Sciences*, vol. 12, no. 19, p. 9867, Sep 2022. doi: 10.3390/app12199867. [Online]. Available: http://dx.doi.org/10.3390/app12199867

DOI: 10.15439/2023F8078
ISSN 2300-5963 ACSIS, Vol. 37

Ausklasser - a classifier for German apprenticeship advertisements

Kai Krüger
German Federal Institute for Vocational Education and Training
Email: kai.krueger@bibb.de

***Abstract*—The German labor market system heavily relies on apprenticeships. Online Job Advertisements (OJAs) become an increasingly important data source to monitor labor market. Commonly, researchers use Information Extraction (IE) methods from Natural Language Processing (NLP) to extract entities such as skills and tasks from OJAs and draw conclusions about the labor market by aggregating them based on relevant variables such as occupations. Depending on the research question, it may be valuable to be able to exclude apprenticeships from these analyses, because apprentices will not be expected to have a specialized skill-set yet. As a result, Apprentice OJAs (AOJAs) may not reflect the dynamics in occupations and labor market as much as Regular OJAs (ROJAs). Furthermore, certain analyses may benefit from examining apprenticeships exclusively. This paper provides an efficient distilBERT based text classification model for this task and discusses findings from an experiment pipeline designed to identify the best possible implementation strategy of this task given the current NLP toolkit.**

I. Introduction

ONLINE Job Advertisements (OJAs) have been used to monitor labor market [3] with regard to dimensions such as skills and tasks [12], working tools [11], education [2], the impact of the covid pandemic [20, 13], specific industry sectors [16] and others. It is safe to say that OJAs are a valuable data source for monitoring labour market for years to come. Methodologically, researchers usually use Natural Language Processing (NLP) and Information Extraction (IE) to gain insights into the contents of the OJAs to aggregate information about entities such as skills or tasks based on dimensions they are interested in such as occupations or industry sectors.

Being able to exclude Apprenticeship OJAs (AOJAs) or consider them exclusively is an important variable for these analyses, especially in Germany. The goal of this paper is therefore to develop a text classification model, that can predict whether an input OJA is an AOJA or a Regular OJA (ROJA), and make it publicly available[1]. It also contributes by conducting experiments to find best way to implement such a model within the current NLP landscape. Other researchers can use the findings to build their own models for other languages or similar tasks. Specifically, it tests for eight parameters that concern composing the training data, model choice, hyper-parameter choice and task modeling. For these parameters, hypotheses are formulated and tested in a random search with 100 trials. The final parameter setup is then reported for another 13 runs to reduce the effect of randomness and the model is published publicly. The structure of the remaining paper is as follows. In Chapter II we briefly explain the need for an AOJA classifier. In Chapter III we frame building this classifier as a NLP problem and introduce the different choices we had to make when constructing such a classifier. Corresponding to these choices we formulate hypotheses. The choices can be represented by parameters in our experiment pipeline, which is explained in Chapter IV. Chapter V presents the results and Chapter VI discusses them and mentions major limitations of the experiment. Finally, Chapter VII concludes this paper and gives suggestions for further research.

II. Building an AOJA classifier for labor market research

The apprenticeship system is a key factor for the German labor market. Although higher education gets increasingly important, vocational education and training (VET) makes up about half of the German post-secondary education system [4, 5]. The VET system provides the labour market with skilled workers and is an established process for the transition from school to work. Since VET is still part of people's education, apprentices are not expected to have a major skill-set yet that companies could demand in their ads. Furthermore, due to the formalization of VET in Germany, tasks listed in AOJAs are likely to be more generic. Therefore, AOJAs do not reflect labour market dynamics as much as ROJAs. Inversely, AOJAs isolated are a valuable source to ask research questions specifically with regard to apprenticeships. Research questions could include finding out how employers try to attract apprentices or predicting trends in the popularity of certain occupations based on AOJA number development. To the best of our knowledge, across languages there is no model currently available to classify OJAs as to whether an apprentice is being sought or not. Papers analysing labor market on the basis of OJAs have so far not made any distinction between AOJAs and ROJAs. Few publications use OJAs to specifically conduct research about apprenticeships and those that do use a smaller size of hand selected ads [9, 8]. The primary goal of this paper is therefore to build and publish an AOJA classifier. In the next chapter we frame this task as a binary text classification problem and discuss a variety of aspects that ought to be considered in the process of constructing the model given the current NLP landscape.

[1] Model: https://huggingface.co/KKrueger/ausklasser
Code: https://github.com/KruegerETRF/ausklasser

III. AOJA CLASSIFIER AS A NLP PROBLEM

Text classification is a well explored NLP problem. With the transformer architecture and Hugging Face infrastructure powerful off the shelve solutions are available with minimal time invest and coding efforts. At the same time, the particular task of classifying AOJAs and ROJAs has not been explored and no datasets or benchmarks are available. In an explanatory data analysis [19] it was found that the texts show distinguishable characteristics for the task at hand. In most cases an AOJA explicitly states that an apprentice is being sought. This not only makes it plausible to build an AOJA classifier, but also makes it a comparably easy task. So, then the question is, whether it is sufficient to simply take any pretrained model on Hugging Face and finetune it on some hundred samples and be done?

We argue that there is still a variety of decisions to be made to tackle this problem in the most optimal way given the current state of NLP. We include different dimensions into the evaluation of our model including robustness, epistemological validity, efficiency, ethical concerns, flexibility and generalizability. We derive research questions from these decisions that we formulate as hypotheses that are being empirically tested via the experiment pipeline described in section IV. In that sense, this paper serves as a reference point to other researchers facing similar problems. The structure of this chapter is to formulate the hypotheses and then discuss their background and relevance.

Hypothesis H_1: There will be no difference between multi- and monolingual pretrained models.

Hypothesis H_2: Domain adapted models will perform better than generic models.

Hypothesis H_3: Lighter models will perform equally well to bigger models.

The first three hypotheses deal with the choice of the pretrained model, which is the first decision to be made. The most obvious goal is to choose the best performing model for the given task. A first reference point could be the standard BERT model [7], which has been trained multilingually[2], including German texts. Then, there are monolingual models for the German language like [6]. Now, given the task at hand, which one performs better? Hypothesis H_2 is concerned with a BERT model domain adapted to German OJAs developed by [10]. It is plausible to expect that this model performs better, because it has already internalized patterns of OJA and might be able to generalize quicker, for example by knowing relevant synonyms for the German word for apprentice, *Auszubildender*, (e.g. *Azubi*) or other keywords.

Both of these research questions have the primary goal to find a very robust and well performing model in model construction. Beyond this, however, there is another relevant aspect for model choice: computational cost. Higher computational cost means that model training is more expensive financially and has an increased environmental impact [18]. Even though in [18] models are being trained from scratch, fine-tuning models is also costly. Additionally, the efficiency of the trained model at inference is a relevant factor for research institutes and companies with a smaller budget. Generally, the more efficient a model is the better as long as its performance does not suffer. Since the task at hand is rather simple, it is plausible that lighter models such as [17] perform equally well, which is tested by Hypothesis H_3.

Hypothesis H_4: Hyperparameter search can be neglected

Also intertwined with the points about training cost is the search for optimal hyperparameters. The more different setups are tried, the more resources need to be used. Therefore, the authors in [18] suggest to use hyperparameter optimization algorithms. The study in [14], however, shows that these techniques can fail given an insufficient time budget and are prone to overfitting. Given the simplicity of the task, the question is, if it is even necessary to perform extensive hyperparameter search or if using default or common configurations is sufficient. Hypothesis H_4 therefore tests whether a single model hyperparameter consistently affects model's performance. As we will see in section IV the pipeline chooses hyperparameters so that the search space only affects commonly used values (including default values). Given the simplicity of the task the hypothesis is that it does not matter significantly, which learning rate, for example, is chosen. Certainly, choosing absurd values for the learning rate would affect models performance significantly, but this is not relevant to the hypothesis. With the regard to learning rate it has to be mentioned that all models (and configurations) explored in this experiment pipeline use the adaptive gradient algrotihm [15] AdamW, which means the importance of the initial learning rate hyperparameter decreases over time.

Hypothesis H_5: Given two datasets D1 and D2 from different sources and substantial textual differences, models trained primarily on D1 data will perform better when testing on D1 data and vice versa.

Hypothesis H_6: Using more than two labels will increase the robustness of the model on downstream binary predictions

The datasets used are described in Appendix A in detail. For Hypothesis H_5 it is important to know that there are two different labeled datasets available that each are different structurally. Specifically, one of the datasets (D1) has a lot of boilerplate remains from the scraping process, whereas the other one (D2) has only cleaned text without boilerplate, containing only the actual ad. Cleaning D1 is not an option. Also, D1 comes from various online sources, whereas D2 comes from the same source website, which might lead to other biases/differences, such as D2 being more homogeneous linguistically or in terms of labor market specific factors (ie. certain industry sectors are more likely to appear in D2 than

[2] Of course, there is also the monolingual English version, but that is irrelevant in this case.

others compared to D1). The goal for our model is to not be influenced by such factors. Ideally, it generalizes over any German OJAs fed into it. Preliminary analysis showed, however, that simpler models trained on D2 perform worse on D1 data. Therefore, Hypothesis H_5 tests if such effects are also true for transformer based models. This aspect is also relevant in the context of publishing our model publicly. Researchers might be able to trust its performance on their data more, if we can falsify Hypothesis H_5.

Another aspect of the two datasets is that they are labeled more fine grained than only AOJA and ROJA. There are further categories like internship or leading position. These categories, however, are not common between both datasets. Also, the most important goal of the classifier is to distinguish between AOJAs and ROJAs. It would, however, be a potential future advantage, if some of the other classes could also be identified by our classifier. With regard to Hypothesis H_6 another factor has to be considered: given the high amount of OJAs that seek "regular" workers in both datasets, binary set ups will often end up without many other "special" categories such as internships. This might lead to the model not learning to differentiate between AOJA and non-AOJA, but between ROJA and non-ROJA positions. This would be prevented by the more sophisticated categories. Note, that in case a multiclass model was trained its predictions were still aggregated to do binary classification during test phase (see section IV for further details).

Hypothesis H_7: Balanced datasets will perform better than unbalanced datasets

Hypothesis H_8: Models will perform well with limited training data

The amount of AOJAs compared to ROJAs is much smaller (rougly 14 percent). There is no sophisticated balancing in place such as data augmentation methods, but we will compare simple over- and undersampling to not balancing data and see, how this influences the performance on a balanced testset.

Another question when building a model for a new NLP dataset where no benchmarks exist yet is how much data labeled data is required. Since we have a lot of labeled data available, we can test different sizes up to 10.000 ads, but we hypothesize that given the simplicity of the task models should be able to achieve good results with limited training data. See section IV for further details on how much data is used exactly.

IV. Experiment Pipeline

In this section we describe the experiment pipeline. Based on the hypothesis described in section III there are eight parameters introduced to the pipeline, three of which are common hyperparameters fed into model training. Table 1 shows the parameters and their search space. The pipeline consists of three steps explained below. The parameters are inserted in the first two steps (compose data and training), whereas the last step (testing) reports the final metrics. It is important to mention that these metrics are always tested against the same testset, regardless of how the training (and evaluation) data ended up after the first step. Fig. 1 gives an overview of the pipeline. In the initial experiment a random search with 100 trials was performed.

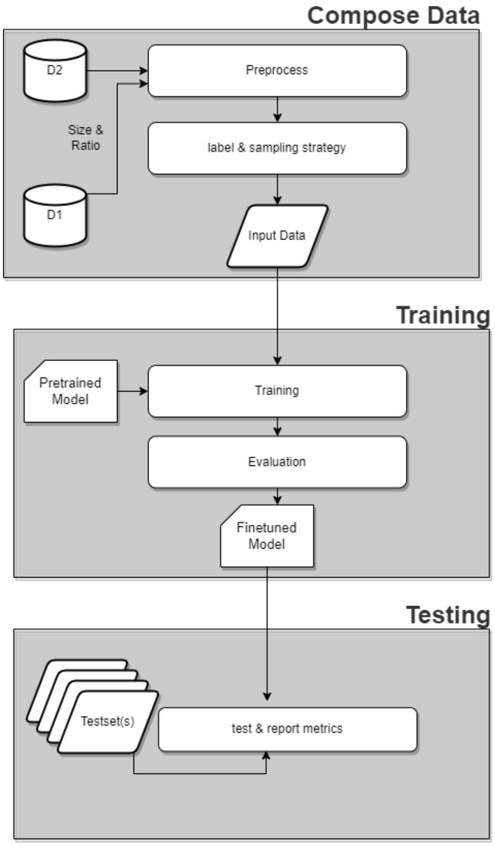


Fig. 1. Simple visualization of the experiment pipeline

A. Step 1: Compose Data

The compose data step consists of accessing the data from the two datasets with regard to the size and ratio parameters. It then harmonizes both datasets into one and performs a check to prevent any ads that are used in the testset later to be included in the training data. Then it aggregates labels based on the label strategy chosen. For the binary option all non-AOJAs are being aggregated into one category (ROJA). For the multiclass option two additional label classes are added. They are described in more detail in the appendix. Finally, it considers the sampling strategy, either leaving the data as is (no balance) or performing over- or downsampling. Both work similar. The highest/lowest number of instances for a class is located and then data is either randomly duplicated (oversampling) or removed (downsampling) for all other classes until that number is reached. The final dataset is then being forwarded

TABLE I Parameters and search space

Parameter	Options
model	multilingualBERT, germanBERT, jobBERT, distilBERT
size	100, 500, 1000, 5000, 10000
ratio	1, 0.7, 0.5, 0.3, 0
label strategy	binary, multiclass
balance strategy	oversample, downsample, no balance
learning rate	0.0001, 0.00001, 0.000001
epochs	3, 5, 7
warmup	0, 500

to the training step. [3]

B. Step 2: Training

The training step loads in the pretrained tokenizers and models from the options listed in Table 1 as well as the input data from step 1. Then, the standard fine-tune process is being initiated, where the above described hyperparameters are being varied. A statement about hardware used and the energy consumption can be found in the appendix. Most notable is that the batch size had to be kept relatively small (eight) due to hardware limitations, which potentially hinders performance.

The dataset is split 0.7/0.3 for evaluation during training. For evaluation accuracy, precision, recall and f1 are being measured and logged along with the training loss. When there are four labels precision, recall and f1 are being averaged via the macro average method. Once the training is done, the model with the fine-tuned weights is saved and forwarded to the testing step.

C. Step 3: Testing

In this step, the saved model from the training steps is loaded and tested against the independent testset as described above. The testset contains 80 job ads split 40/40 between D1 and D2, and then split 20/20 between both classes. To ensure validity of the testset, all ads have been reevaluated blindly by two annotators each. In case the model was trained on multiple classes its predictions were aggregated to the binary labels.

Like in training the metrics accuracy, precision, recall and f1 were being used. Since the testset is balanced, accuracy can well indicate model performance. For the other metrics AOJAs have been labeled the positive category, because it is more important. Each metric was measured four times:

- For the entire testset
- For testset data only from D1
- For testset data only from D2
- For testset data whose texts surpass the 512 max token length that the models pose

Building sub-datasets to include only data from D1 or D2 respectively was to study the effect on input data specifics on the model's ability to generalize and test Hypothesis H_5. To make the model more robust, a dataset with only those texts that contained more than 512 tokens has been build to ensure that the models performance is not influenced by truncation.

V. Results

The above described experiment pipeline has been used multiple times with different purposes to test the hypothesis or to build the final model. This chapter is split into two subsections. First, the initial search with 100 runs to test the influence of the different parameters is described and reported. Then, the configuration of the final model is described and the results of 13 runs are reported.

A. Initial search

The initial search was a random search with 100 experiment runs randomly selected from the 10.800 possible parameter combinations possible. The goal was to get an overview over the general performance and different parameters. Of these 100 experiment runs, 8 ended unfinished due to hardware issues. The exact parameter configurations and corresponding results can be accessed in the repository. Fig. 2 shows distribution of experiments for different metrics. We can observe a very strong fluctuation in experiment results, ranging from 0 F1 in the testset to 1.0. Further analysis shows that models with 0 F1 tend to have 0.5 accuracy, which leads to the conclusion that the model has overfitted into always predicting one category. Generally, the three parameters size, label strategy and balance strategy can lead to datasets where there either not much data left at all (because of downsampling) or data is heavily imbalanced so that the model performs well in training just by predicting the largest category (normal OJAs). However, this is not true for all cases. For example, run *66c90a54* [4] had 1.250 samples for both categories (downsampled from 10.000 overall), which should be more than sufficient to learn a meaningful representation of the two classes. Also, accessing training metrics showed that these models do have reasonable performance in the evaluation. Further analysis, however, showed that exploding gradients were likely a problem in these cases. Based on this we can already falsify hypothesis H_4. Choosing the correct hyperparameters based on the setup **is** important as it can prevent overfitting. Especially the number

[3] Note, that for this step the code is only partially published, because the data is not published and the access to the database works with internal procedures. In the published repository this part is replaced with pseudo code

[4] individual runs can accessed from the repository

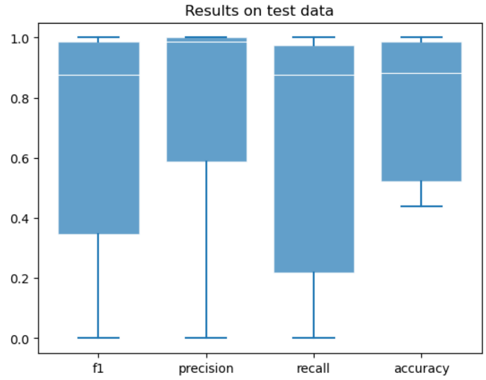


Fig. 2. Results of the initial search on testset performance by different metrics. The white line indicates the median.

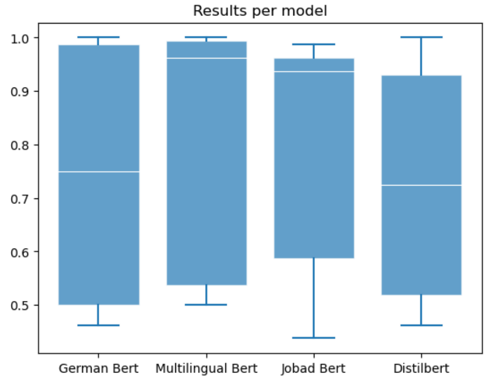


Fig. 3. Accuracy of all runs per model on testset performance. The white line indicates the median.

of epochs mattered significantly. Overall, the amount of models not working downstream while showing reasonable metrics during training was very high, which proves the importance of using a separate testset. Due to the high amount of outliers, further analysis will prefer to analyse median over mean values.

Overall no single parameter consistently performs bad. Each single parameter achieves accuracy scores > 0.9. The median for accuracy is 0.88, which confirms that good performing models can be build in a variety of ways. Also, several runs achieved accuracy on the testset of 1. Generally, fig. 2 also shows that precision is higher than recall most of the times. This might again be due to imbalanced datasets where the model learns to prefer predicting ROJAs. Another general observation is that overall models did not perform worse on longer truncated texts.

The first three hypotheses all deal with the choice of the pretrained model. The results show that the multilingual BERT model outperforms the monolingual model. The domain adapted model outperforms the monolingual model it was adopted from, but performs slightly worse than the multilingual model overall while, however, showing less deviation. The lighter distilBERT model also performs worse, on par

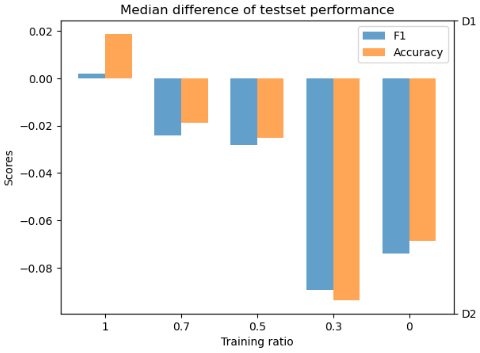


Fig. 4. Median difference of all runs per ratio on testset performance on D1 data versus D2 data. Positive values mean models performed better on D1 data, negative values mean models performed better on D2 data

with the German BERT model. In that sense, none of the first three hypotheses can be verified. With regard to hypothesis H_5 the results falsify the hypothesis. Fig. 4 shows more data from one dataset does not necessarily lead to better performance on the same data downstream. Specifically, the D1:D2 70:30 ratio performs better overall on D2 data in the test phase. However, also the balanced data performs better on D2 data. This might likely be due to D1 data containing boilerplate that confuses any model. With regard to the actual performance, 0 and 0.7 ratios perform best, which indicates that differences must also be attributed to other parameters. Therefore, a balanced 50:50 split still seems like the most reasonable option. hypothesis H_6 was verified based on these runs: multiclass models usually performed better, achieving 0.91 median accuracy, whereas binary models only got to 0.81 accuracy. Potentially this is because the additional classes helped prevent the models from overfitting.

In terms of balancing the data (hypothesis H_7) oversampling (0.94 median accuracy) and not balancing the data (0.93 median accuracy) performs much better than downsampling (0.58 median accuracy). The low performance of downsampling is however explained by the low amount of total data left when small datasets are being accessed to begin with. If considering only the experiments with at least 1.000 ads, for example, downsampling achieves 0.9 median accuracy. Considering only larger training sets for oversampling, however, also increases median performance to 0.97 accuracy. In any case we cannot confirm hypothesis H_7, because not balancing the data did not hurt performance significantly. With regard to hypothesis H_8 it shows that actually increasing data to several thousand samples still significantly increases performance and is the only major influence parameter. Although in some cases 100 or 500 sample setups showed good results, the overall performance increases with data size. This means that even for seemingly simple tasks researchers can likely increase performance by gathering more data.

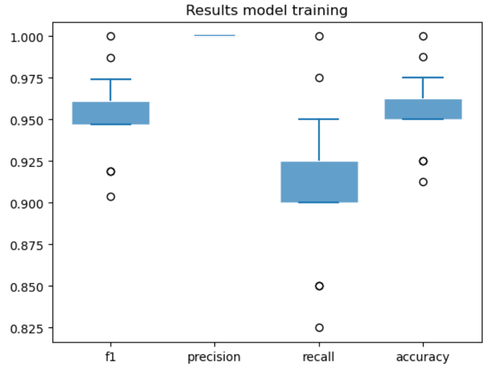


Fig. 5. Results of 13 experiments with fixed parameters

B. Training the model

Given the results from the initial search, how to construct the final model? Despite worse performance, it was decided to first try to use a pretrained distilBERT model, because given enough data those would still perform very well and have the benefit of having reduced cost. Since multilabels worked better initially, it was chosen here as well. For balancing, no balancing was decided upon, because it worked almost as well as oversampling and due to the lower amount of training samples is more energy and time efficient to train. The amount of data chose was 10.000, because a greater sample improved results in the initial search and the ratio was 0.5, which proved to be a robust choice. A learning rate of 0.0001 and no warmup steps were also chosen. For the epochs, a new value of 4 was introduced, because the analysis of training curve showed that models often needed more than three epochs, but sometimes started to overfit at five already. With these fixed parameters 13 runs have been performed to reduce the influence of randomness when reporting the final model. The results (Fig. 5) show a consistently good performance, but outliers still having a variance of roughly nine percent. Precision was higher than recall again, indicating a slight bias towards the ROJA category. Of the experiments a robust model was chosen and uploaded publicly. It achieved .98 accuracy on the testset and .9 accuracy in training evaluation (on four classes). All metrics can be accessed

VI. Discussion and Limitations

The primary goal of this paper was to publish a well performing model to classify German AOJAs and ROJAs. The model published achieved high results in training and testing and is able to classify German OJAs into four categories and AOJAs are one of them. Using the distilBERT architecture, our model is also small and efficient.

In terms of gaining insight into the decisions to be made when building such a model, there were a number of interesting findings. First, while the overall performance was good for many models, there were heavy outliers, which shows that experiments need to be conducted thoroughly and final training metrics cannot be taken for granted. Also, researchers need carefully select hyperparameters like the number of training epochs. Learning rate and warmup steps seemed to play less of an important role. Another key finding was that using more categories, if available, might boost performance for downstream tasks with less categories, because the model learns differntiations more explicitly. Also, the model was able to perform better on clean data, even if boilerplate data was favoured in training (70:30 split). If models trained on mostly clean text on the other hand get exposed to data containing boilerplate the performance drops significantly. Longer texts that had to be truncated did not pose an extra challenge for any of the models. This is not unexpected, because the information required to distinguish the texts for the task at hand is usually found in the opening section of the ads. Our findings suggest that unbalanced training data can also lead to good generalization, if there is enough data overall.

In terms of the models, every pre-trained model was able to produce good fine-tuned models. Some, however, did so more consistently than others. Given the factors described in Chapter III, it might be advisable to use smaller and more efficient models, despite them perform worse in some setups. As shown in Chapter IV B, they perform just as well with the correct setup and offer additional advantages. The most important parameter for any setup as to our findings was the size of the dataset. It showed that even for comparably simple tasks performance can be increased by increasing the data size.

The major limitation of the experiments in the initial search was that the hypotheses were not tested individually. When a certain parameter setting performed worse than another, for example a pretrained model performed worse overall than another, this might not have been the effect of that setting, but because randomly it had less favourable settings in other parameters such as size. Regarding size it also has to be mentioned that for low data settings no methods designed specifically for such cases [1] have been tested. Furthermore, the size of the testset was relatively low given the amount of labeled data that was used during training. This is because the quality of the labels was questionable and testset labels have been reevaluated manually, see Appendix A. A larger testset would have been beneficial, but was not feasible given the resources available.

VII. Conclusion and Outlook

Despite the limitations mentioned above, the published model can be regarded as robust and usable for researchers analyzing German labor market with OJAs. Other researchers can council the findings of the experiment pipeline to make more profound decisions in model construction. Some of the findings here are also worth investigating further. Especially the effect of different datasets on the ability of models to generalize and the effect boilerplate generally has on NLP models is worth looking into. It is also important to keep this aspect in mind when using public models directly. How does the data from that model differ from the data it is supposed to be used on? Scraping boilerplate might be a significant pitfall here. Another aspect worth looking into is the strategy of using more labels than required for the downstream task at hand. Because the categories are more fine-grained, the models will be more robust downstream. The obvious downside of this approach of course is that it requires more data to begin with. But if a lot of data or resources to label additional data are available, adding additional categories might be a fruitful strategy to increase performance. Further investigation into these topics might include setting up more controlled experiments testing single parameters only across different tasks, datasets, models, etc.

Appendix A
Data

For this task, two labeled datasets are available to the authors. Both datasets are protected and cannot be published. The description here serves for transparency and reproducability. The two datasets are:

- Scraped-Data (D1): This dataset comes from a commercial provider of scraped OJAs from 2015 to 2022. In a project, roughly 15.000 OJAs have been labeled according to the type of worker being sought (including apprentices). The annotation process, however, was only a single blind annotation without further quality control. Also, this dataset suffers from unclean full texts, meaning that boiler plate and texts fields are often not separated from the actual ad and stored together as the full text in the data base.
- BA-Data (D2): This dataset is being provided by the Bundesagentur für Arbeit (BA). It consists of roughly 10 Million OJAs from 2011 to 2022 and comes with labeled metadata. Usually this metadata is of good quality, because it is hand labeled by the expert employees of the BA. However, it was not originally intended for scientific use and there are no further control mechanisms for label quality in place. Also, the metadata is not always consistent and label schemes may change over time. Another relevant information is that the full texts come manually cleaned and are free from any boiler plates or text fields that are often found in scraped data.

As shown, neither dataset is of undisputed validity. Therefore, it was decided to construct a test dataset of 80 OJAs that are equally split between the two datasets and among these splits are also equally split between the two categories. In other words there are 20 OJAs for each category for each dataset. This dataset was cross validated by two independent annotators (One male, one female, both German, one student in economics, one researcher in NLP). In some cases OJAs were too short to contain sufficient information or consisted of only boilerplate. These cases were removed. In all remaining cases both annotators agreed on the labels. This testset is being kept entirely out of model training, but each experiment run tests against it in the end.

Appendix B
Categories

Both datasets do not actually come in a binary labeled form, but have further categories of different job positions like internship or leading role. These categories differ between datasets and it is not possible to establish an unambiguous mapping. Of course, apprenticeships are a category in both datasets and the other labels can be aggregated to the ROJA category. However it was decided to include a set up with four different categories into the experiment pipeline. The categories then are as follows:

1) Apprenticeships
2) Other minor positions
3) Leading position
4) Regular workers

To see the exact mapping of other minor positions, please access the mapping dictionary in the utils.py file in the repository.

Appendix C
Hardware & Parameters

All experiments were run on a NVIDIA GeForece RTX 3080. Training time varied between roughly five and twenty minutes per run depending on dataset size and number of epochs. The batch size for training and evaluation were eight. All other hyperparameteres that potentially influence the models performance (and are not includeded in the experiment pipeline) were the defaults of the huggingface training arguments from the trainer class.

References

[1] Iz Beltagy et al. “Zero- and Few-Shot NLP with Pretrained Language Models”. In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*. Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 32–37. DOI: 10.18653/v1/2022.acl-tutorials.6. URL: https://aclanthology.org/2022.acl-tutorials.6.

[2] Phillip Brown and Manuel Souto-Otero. “The end of the credential society? An analysis of the relationship between education and the labour market using big data”. In: *Journal of Education Policy* 35.1 (2020), pp. 95–118. ISSN: 0268-0939. DOI: 10.1080/02680939.2018.1549752.

[3] Marlis Buchmann et al. “Swiss Job Market Monitor: A Rich Source of Demand-Side Micro Data of the Labour Market”. In: *European Sociological Review* (2022). ISSN: 0266-7215. DOI: 10.1093/esr/jcac002.

[4] Statistsiches Bundesamt. *Berufsbildungsstatistik*. Accessed on August 17, 2023. URL: https://www-genesis.destatis.de/genesis/online?operation=previous&levelindex=2&step=2&titel=Ergebnis&levelid=1690804374122&acceptscookies=false#abreadcrumb.

[5] Statistsiches Bundesamt. *Statistik der Studenten*. Accessed on August 17, 2023. URL: https://www-genesis.destatis.de/genesis/online?sequenz=tabelleErgebnis&selectionname=21311-0010#abreadcrumb.

[6] Branden Chan, Stefan Schweter, and Timo Möller. "German's Next Language Model". In: *Proceedings of the 28th International Conference on Computational Linguistics*. Barcelona, Spain (Online): International Committee on Computational Linguistics, Dec. 2020, pp. 6788–6796. DOI: 10.18653/v1/2020.coling-main.598. URL: https://aclanthology.org/2020.coling-main.598.

[7] Jacob Devlin et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". In: *CoRR* abs/1810.04805 (2018). arXiv: 1810.04805. URL: http://arxiv.org/abs/1810.04805.

[8] Khristin Fabian and Ella Taylor-Smith. *How are we positioning apprenticeships? A critical analysis of job adverts for degree apprentices*. English. United Kingdom: Society for Research in Higher Education, 2021.

[9] Khristin Fabian et al. "Signalling new opportunities? An analysis of UK job adverts for degree apprenticeships". In: *Higher Education, Skills and Work-Based Learning* ahead-of-print.ahead-of-print (2023). ISSN: 2042-3896. DOI: 10.1108/HESWBL-02-2022-0037.

[10] Ann-Sophie Gnehm, Eva Bühlmann, and Simon Clematide. "Evaluation of Transfer Learning and Domain Adaptation for Analyzing German-Speaking Job Advertisements". In: *Proceedings of the 13th Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, 2022.

[11] Betül Güntürk-Kuhl, Philipp Martin, and Anna Cristin Lewalder. *Die Taxonomie der Arbeitsmittel des BIBB: Revision 2018*. 2018.

[12] Robert Helmrich et al. *Berufsbildung 4.0 – Fachkräftequalifikationen und Kompetenzen für die digitalisierte Arbeit von morgen: Säule 3: Monitoring- und Projektionssystem zu Qualifizierungsnotwendigkeiten für die Berufsbildung 4.0*. 1. Auflage. Vol. 214. Wissenschaftliche Diskussionspapiere. Leverkusen: Verlag Barbara Budrich, 2020. ISBN: 9783962082024. URL: https://www.bibb.de/dienst/veroeffentlichungen/de/publication/show/16688.

[13] Jakob de Lazzer and Martina Rengers. "Auswirkungen der Coronakrise auf den Arbeitsmarkt: Experimentelle Statistiken aus Daten von Online-Jobportalen". In: (2021).

[14] Xueqing Liu and Chi Wang. *An Empirical Study on Hyperparameter Optimization for Fine-Tuning Pre-trained Language Models*. 2021. arXiv: 2106.09204 [cs.CL].

[15] Ilya Loshchilov and Frank Hutter. *Decoupled Weight Decay Regularization*. 2019. arXiv: 1711.05101 [cs.LG].

[16] Mirjana Pejic-Bach et al. "Text mining of industry 4.0 job advertisements". In: *International Journal of Information Management* 50 (2020), pp. 416–431. ISSN: 02684012. DOI: 10.1016/j.ijinfomgt.2019.07.014.

[17] Victor Sanh et al. *DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter*. 2020. arXiv: 1910.01108 [cs.CL].

[18] Emma Strubell, Ananya Ganesh, and Andrew McCallum. *Energy and Policy Considerations for Deep Learning in NLP*. 2019. arXiv: 1906.02243 [cs.CL].

[19] Dennis Ulmer et al. "Experimental Standards for Deep Learning in Natural Language Processing Research". In: *Findings of the Association for Computational Linguistics: EMNLP 2022*. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, Dec. 2022, pp. 2673–2692. URL: https://aclanthology.org/2022.findings-emnlp.196.

[20] Stefan Winnige and Alexandra Mergener. "Homeoffice-Boom im Zuge der Corona-Pandemie: Welche Potenziale zeichnen sich langfristig für akademisch und beruflich Qualifizierte ab?" In: *Berufsbildung in Wissenschaft und Praxis* 50.2 (2021), pp. 27–31.

Communication Papers of the 18th Conference on Computer Science and Intelligence Systems pp. 173–180
DOI: 10.15439/2023F865
ISSN 2300-5963 ACSIS, Vol. 37

Developing an interval method for training denoising autoencoders by bounding the noise

Bartłomiej Jacek Kubica

0000-0002-5547-3759
Institute of Information Technology,
Warsaw University of Life Sciences – SGGW,
ul. Nowoursynowska 159, 02-776 Warsaw, Poland
E-mail: bartlomiej_kubica@sggw.edu.pl

***Abstract*—This paper discusses prospects of using interval methods to training denoising autoencoders. Advantages and disadvantages of using the interval approach are discussed. It is proposed to formulate the problem of training the proper neural network as a constraint-satisfaction, and not optimization, problem. Pros and cons of this approach are considered. Preliminary numerical experiments are also presented.**

I. Introduction

Autoencoders (AE) are commonly used, nowadays, and they found applications in various branches related to machine learning (ML). They can be used, i.a., for feature extraction, dimensionality reduction, denoising, and even as some kind of generative models (so-called variational autoencoders).

Interval methods, while used by several authors for neural network training (see, e.g., [1], [2], [3], [4], [5], [6], [7]), have rarely been used in conjunction with AE, so far. The only known exception is the paper [8]. This is surprising, because interval methods have – as we shall see – several natural advantages, when applied to training AEs. This paper intends to fill this gap, at least to some extent.

The paper is organized as follows. Section II introduces the idea of autoencoders. It (briefly) discusses various types of the AEs, their features, and applications. Section III introduces the interval calculus, and basics of the algorithms that use it. In Section IV, we discuss the interval approach to training AEs, and discuss the possible solutions.

II. Autoencoders and their types

Autoencoders (also called autoassociotors, at least in the early papers) are a specific kind of unsupervised (or semi-supervised) feed-forward neural networks [9], [10]. Their use dates back to the eighties; cf. [11], [12], [13].

The AE consists of at least three layers: the input layer (x), the hidden layer (h), and the output layer (y). Its essence is to reproduce the input on the output, but not in a trivial manner: $y = x$, but approximately. Depending on the structure and dimensionality of the hidden layer, the input can be reconstructed more or less precisely, and – as we shall see – the reconstruction process will capture various features of the data.

An AE can be logically decomposed into two parts:

- the *encoder*, transforming the input x to the representation of data, in the hidden layer: $h = f(x)$,
- the *decoder*, transforming the representation to the data from the original space: $y = f^*(h)$.

The f^* function in the above description is some sort of an ‘approximate inverse’ of f.

We train the AE to have:

$$y = f^*(f(x)) \approx x \ , \tag{1}$$

but how close y can get to x depends on the restrictions on the representation h: what is its dimensionality, etc.

Training is usually performed by minimizing some loss function (least-squares or the Kullback-Leibler divergence [14]), possibly plus some regularization term(s) forcing the satisfaction of some additional conditions, e.g., the presence of some features.

The general structure of the AE is presented in Fig. 1.

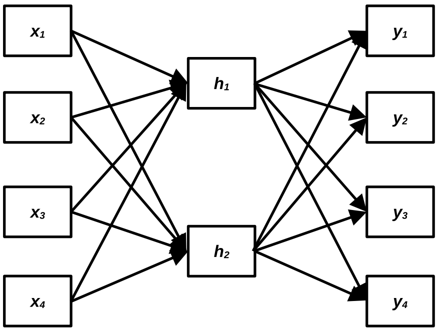


Fig. 1. General structure of an AE

This figure suggests that there are fewer neurons in the hidden layer than there are in the input and output layers. This is often the case: such an AE is called *undercomplete*, and the limitation of the representation in the hidden layer forces the network to learn the most important features only.

Topical area: Advanced Artificial Intelligence in Applications

Yet, another type of an AE is used as well: an *overcomplete autoencoder*, that has more neurons in the hidden layer than in the input or output ones (Fig. 2).

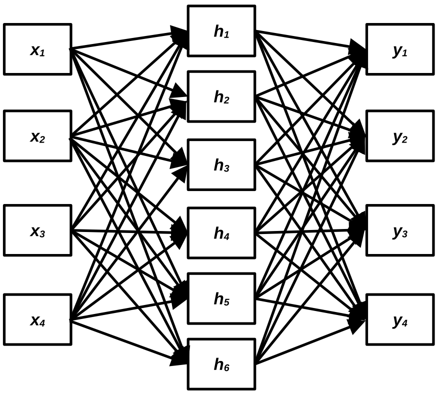


Fig. 2. An overcomplete autoencoder

How does such an AE work? Will it simply learn to approximate the identity function $y = x$, or will it find a useful representation? As it has already been mentioned, the essence is to provide the proper regularization in the learning process: the minimized loss function should contain a regularization term, being the 'penalty' for using too many neurons.

An example of such a regularization is used in the so-called *sparse autoencoder* (SAE). In this case, it is penalized to activate too many neurons in the hidden layer. This can be obtained, in particular, by using as the regularization term the Kullback-Leibler divergence of the hidden layer – see, e.g., [15].

Overcomplete AEs are usually considered to be more difficult to train than undercomplete ones, but more powerful. The source of their power is the ability to omit local optima, during the learning process.

Another kind of the AE, particularly important in this paper is a *denoising autoencoder* (DAE). In this case, the training process is subtly modified to obtain a representation that is robust in the presence of noise. Hence, instead of feeding the input layer with training vectors x, we use perturbed training data (let us denote it by $x + \varepsilon$), yet expecting the output layer to return $y \approx x$, and not $y \approx x + \varepsilon$. How can it be achieved?

Firstly, we have to assume some specific distribution of the noise; usually Gaussian distribution is used, but it it not the only possibility. Secondly, the training set has to be increased, as for each input vector x, we must now have a few its perturbed counterparts. Thirdly, reconstructing the values is usually done basing on the interdependency of various components of x, but this is not the only possibility. Sometimes, the values of x belong to a discrete set, and small perturbations can easilly be corrected.

One of the drawbacks of the traditional approach to training DAEs is the enlargement of the training set. Actually, later in the paper we propose an approach to mittigate it, by using some interval methods for training the DAE. Please compare also the latter discussion in Section VI.

In general, the autoencoder does not have to be limited to three layers. The number of intermediate layers can be increased; such an AE is called a *deep autoencoder* (or a *stacked autoencoder*). An example structure of a deep AE is presented in Fig. 3.

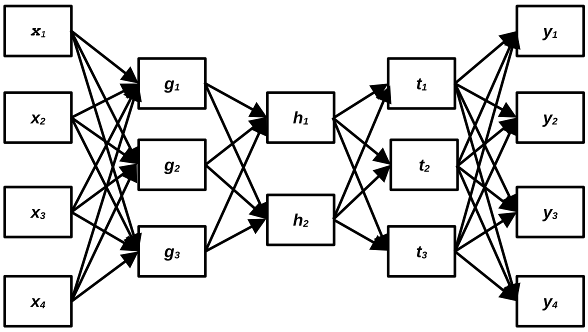


Fig. 3. A deep autoencoder

The virtue of a deep AE is that it can be trained iteratively: first we find the weights for a single hidden layer, such that $g = f_1(x)$, and $y = f_1^*(g) \approx x$, then $h = f_2(g)$, and $t = f_2^*(h) \approx g$, thus obtaining:

$$\begin{aligned} h &= f_2(f_1(x)) \ , \\ y &= f_1^*(f_2^*(h)) \ . \end{aligned}$$

Yet another kind of an AE is a contracting autoencoder (CAE). The essence is to train the AE so that we had $h = f(x)$, and the derivative of f was close to zero at the training points. Usually, it is obtained by adding the the loss function a regularization term, penalizing a norm of the Jacobi matrix of f. An analog for a deep CAE is straightforward.

But why would an AE satisfy such a condition? What do derivatives close to zero imply?

Actually, the idea is pretty similar (but not mathematically equivalent!) to a DAE: when the derivative of f is close to zero, adding a noise to x does not change its representation significantly. There are two key differences between DAE and CAE:

- DAE enforces some conditions on the output layer, i.e., on $y = f^*(f(x))$, while CAE enforces some conditions on the hidden layer, i.e., $h = f(x)$,
- CAE does not make any assumptions about the distribution of the noise, while DAE uses the noise generated using a specific distribution.

There are some approaches to combine DAE and CAE, e.e., the marginalized DAE (mDAE), proposed in [16].

Let us conclude this survey with a *variational autoencoder* (VAE). This kind of an AE tends to learn rather a probabilistic distribution of the data than a representation of a single element of the training set. The idea has been proposed by [17].

Usually, it is assumed that the distribution is Gaussian, and we have to fit its mean value and standard deviation, but other distributions can be used as well [10]. To find the optimal values of the parameters, the Kullback-Leibler divergence is minimized, so that the distribution was as close to the given one, as possible.

A VAE is a neural network very different from the ones described up to now. In all previous architectures, the most important part of the network was the encoder; the role of the decoder was reduced to validating the encoder. For a VAE, the decoder is the most important part. It is worth noting that the role of such a network is to *generate* the data resembling (but not identical to) the input ones. We shall not use VAEs in this paper, but the topic is worth further consideration.

Activation functions

Let us mention one more detail: what activation functions are used in AEs? Several such functions happen to be used in artificial neural networks [9], [10], but for AEs two of them are the most common. We shall stick to using them, as well.

Two such activation functions – ReLU (Rectified Linear Unit):

$$ReLU(t) = \max(t, 0) \ . \tag{2}$$

and the sigmoid function:

$$\sigma(t) = \frac{1}{1 + \exp(-\beta \cdot t)} \ , \tag{3}$$

III. Interval methods

The interval calculus is the tool of choice for us, in this paper. What is it?

It can be defined a branch of numerical analysis and mathematics that operates on intervals rather than precise numbers. A good introduction can be found in many classical textbooks, including, i.a., [18], [19], [20], [21], [22], [23], or a most recent one [24].

Arithmetic operations (as well as other operations and functions) on intervals are designed, so that the following condition was fulfilled:

$$\odot \in \{+, -, \cdot, /\}, \ a \in \mathbf{a}, \ b \in \mathbf{b} \ \text{ implies } \ a \odot b \in \mathbf{a} \odot \mathbf{b} \ . \tag{4}$$

In other words, the result of an operation on numbers will should contained in the result of an analogous operation on intervals, containing these numbers.

This results in the following formulae for arithmetic operations (cf., e.g., the aforementioned textbooks):

$$\begin{aligned}
[\underline{a}, \overline{a}] + [\underline{b}, \overline{b}] &= [\underline{a} + \underline{b}, \overline{a} + \overline{b}] \ , \\
[\underline{a}, \overline{a}] - [\underline{b}, \overline{b}] &= [\underline{a} - \overline{b}, \overline{a} - \underline{b}] \ , \\
[\underline{a}, \overline{a}] \cdot [\underline{b}, \overline{b}] &= [\min(\underline{a}\underline{b}, \underline{a}\overline{b}, \overline{a}\underline{b}, \overline{a}\overline{b}), \max(\underline{a}\underline{b}, \underline{a}\overline{b}, \overline{a}\underline{b}, \overline{a}\overline{b})] \ , \\
[\underline{a}, \overline{a}] \ / \ [\underline{b}, \overline{b}] &= [\underline{a}, \overline{a}] \cdot \left[1 / \overline{b}, 1 / \underline{b}\right] \ , \qquad 0 \notin [\underline{b}, \overline{b}] \ .
\end{aligned} \tag{5}$$

It is worth noting that the above formulae are not the only possible ones. Alternative (and even more general) formulations are possible as well. Details can be found, i.a., in Chapter 2 of [24]. Also, let us mention that the division by an interval containing zero is also possible. This is done in the so-called extended Kahan-Novoa-Ratz arithmetic; cf., e.g., [20] for details.

Similarly to the arithmetic operations, we can define the power of an interval:

$$[\underline{a}, \overline{a}]^n = \begin{cases} [\underline{a}^n, \overline{a}^n] \text{ for odd } n \\ [\min\{\underline{a}^n, \overline{a}^n\}, \max\{\underline{a}^n, \overline{a}^n\}] \text{ for} \\ \qquad \text{even } n \text{ and } 0 \notin [\underline{a}, \overline{a}] \\ [0, \max\{\underline{a}^n, \overline{a}^n\}] \text{ for even } n \text{ and} \\ \qquad 0 \in [\underline{a}, \overline{a}] \end{cases} \ , \tag{6}$$

and other transcendental functions, like:

$$\begin{aligned}
\exp\big([\underline{a}, \overline{a}]\big) &= [\exp(\underline{a}), \exp(\overline{a})], \\
\log\big([\underline{a}, \overline{a}]\big) &= [\log(\underline{a}), \log(\overline{a})], \text{ for } \underline{a} > 0. \\
\dots
\end{aligned}$$

For details, please consult, e.g., Section 2.3 of [24].

A. The problem under solution for interval algorithms

The approach described in the preamble of this section finds several applications. Not to repeat the whole discussion from Chapter 4 of [24], let us state that they can be used to seek the solutions of problems of the following form:

$$\text{Find } \textit{all } x \in X \text{ such that } P(x) \text{ is fulfilled.} \tag{7}$$

Here, $P(x)$ is a formula with a free variable x and $X \subseteq \mathbb{R}^n$. In particular, constraint satisfaction problems (CSP), and optimization problems (unconstrained and constrained ones) are specific instances of Problem (7).

It should be noted that Formula P can contain, in addition to the variable x, also some parameters; let us denote them by $a \in \mathbb{R}^k$. We have two main possibilities here:

$$\text{Find } \textit{all } x \in X \text{ such that } (\forall a \in \mathbf{a}) \ \ P(x, a) \text{ is fulfilled.}$$

or:

$$\text{Find } \textit{all } x \in X \text{ such that } (\exists a \in \mathbf{a}) \ \ P(x, a) \text{ is fulfilled.}$$

Other variants use various quantifiers for various components of the vector a, e.g., $(\forall a_1 \in \mathbf{a}_1)(\exists a_2 \in \mathbf{a}_2)(\forall a_3 \in \mathbf{a}_3)$, etc.

In the above manner, the intervals give us a natural tool to bound several kinds of uncertainty. While, in the opinion of the author, interval calculus should *not* be understood as a tool of uncertainty description, but rather as a general approach to seek points satisfying a certain logical condition, the uncertainty description remains an important family of its applications.

B. Interval branch-and-bound type methods

How to solve problems of type (7)? The generic algorithm proposed in [24] is called the branch-and-bound type method (B&BT). Instances of the B&BT algorithm are, among others, these popular ones:

- the branch-and-bound methods for optimization problems,

- the branch-and-prune method for CSPs.

The essence in both cases is to subdivide the boxes subsequently (starting from the initial box or the list of initial boxes), discarding the boxes that do not contain solutions, and possibly verifying some boxes to contain the solution(s). Details can be found in [24].

It is a specific instance of the so-called divide-and-conquer strategy (but the author himself dislikes this term).

Let us focus on the problem of solving a CSP, i.e., trying to compute the set:

$$S = \{x \in X \mid g_i(x) \leq 0, i = 1, \ldots, m\} \ ,$$

or succinctly: $S = \{x \in X \mid g(x) \leq 0\}$, where $g = (g_1, \ldots, g_m)$.

Using interval methods, we compute two lists of boxes: *verified* and *possible* solutions.

In case of a system of inequalities, the interior of the solution set S is nonempty and the verified solutions are boxes contained in this interior (boxes that contain solutions only). Possible boxes lie usually on the boundaries of S, and they contain some points both from the set S and from its complement $\mathbb{R}^n \setminus S$. Typically there are several possible boxes, unless S is a box itself (which would be highly unlikely).

The branch-and-prune algorithm for a CSP can be formulated as follows:

The 'rejection/reduction tests', mentioned in the algorithm have been described in the author's previous papers; specifically, please consult [25], [26], [27], [28] and the references therein.

In our version of the solver, the most important tool is hull-consistency (HC).

Definition 3.1: A box $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$ is hull-consistent with respect to a constraint $c(x_1, \ldots, x_n)$, iff:

$$\forall i \ \mathbf{x}_i = \square\{s \in \mathbf{x}_i \mid \exists x_1 \in \mathbf{x}_1, \cdots \exists x_{i-1} \in \mathbf{x}_{i-1}, \\ \exists x_{i+1} \in \mathbf{x}_{i+1} \cdots \exists x_n \in \mathbf{x}_n \\ c(x_1, \ldots, x_{i-1}, s, x_{i+1}, \ldots, x_n)\} \ .$$

A popular algorithm to enforce HC is called HC4. Details can be found, i.a., in [28] or Subsect. 5.5. of [24].

IV. Interval algorithms for training AEs

A. Interval methods and neural networks

As it has already been mentioned, interval algorithms have been extensively used for training various kinds of neural networks. Two approaches have been formulated; in [3], they are simply called 'type 1' and 'type 2' interval neural network problems. 'Type 1' problems can be solved, by obtaining the solution of an equations system:

$$f(x_i, w) = y_i, \text{ for } i = 1, \ldots, N \ , \tag{8}$$

while 'type 2' problems require solving an optimization problem:

$$\min_w \Big(||f(x_i, w) - y_i|| \Big) \ . \tag{9}$$

Algorithm 1 Interval branch-and-prune algorithm for a system of inequalities

```
Require: $\mathbf{x}^{(0)}, \mathrm{g}, \varepsilon$
{$\mathbf{x}^{(0)}$ is the initial box, $\mathrm{g}(\cdot)$ is the interval extension of the function $g\colon \mathbb{R}^n \to \mathbb{R}^m$}
{$L_{ver}$ – verified solution boxes, $L_{pos}$ – possible solution boxes}
$L_{ver} = \emptyset$
$L_{pos} = \emptyset$
$\mathbf{x} = \mathbf{x}^{(0)}$
loop
  compute $\mathbf{y} = \mathrm{g}(\mathbf{x})$
  optionally, process the box $\mathbf{x}$, using additional rejection/reduction tests
  if ($\underline{y} > 0$) then
    discard $\mathbf{x}$
  else if ($\overline{y} \leq 0]$) then
    push ($L_{ver}$, $\mathbf{x}$)
  else if ($\mathrm{wid}\,\mathbf{x} < \varepsilon$) then
    push ($L_{pos}$, $\mathbf{x}$) {The box $\mathbf{x}$ is too small for bisection}
  end if
  if ($\mathbf{x}$ was discarded or $\mathbf{x}$ was stored) then
    if ($L == \emptyset$) then
      return $L_{ver}, L_{pos}$ {All boxes have been considered}
    end if
    $\mathbf{x}$ = pop ($L$)
  else
    bisect ($\mathbf{x}$), obtaining $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$
    $\mathbf{x} = \mathbf{x}^{(1)}$
    push ($L$, $\mathbf{x}^{(2)}$)
  end if
end loop
```

In the above formulae, by f we denoted the function, represented by the neural network, (x_i, y_i) were the pairs from the training set, and w – the vector of weights.

It is worth noting that virtually all non-interval methods, use the approach (9) for training neural networks. The objective function from (9) can be modified to contain some regularization terms, e.g., to obtain the sparsity or other features of the solution.

Only the interval calculus allows replacing optimization with the direct search of points satisfying certain constraints. In (8) they are equations, but inequality constraints are even more natural in the interval formulation:

$$\mathrm{f}(\mathbf{x}_i, \mathbf{w}) \subseteq \mathbf{y}_i, \text{ for } i = 1, \ldots, N \ . \tag{8'}$$

Also, instead of using the regularization terms, we can directly check (non)satisfiability of various constraints on subsequent boxes. This is the virtue of the interval calculus, that is a natural approach to seeking points satisfying certain logical conditions. Details can be found in [24].

B. The case of an autoencoder

As it has already been stated, interval methods have many natural advantages, when applied to training AEs.

Firstly, we can use CSP solving approach (8'), instead of the optimization problem (9), which is a more straightforward approach.

Secondly, handling all uncertainties, including the noise in a denoising AE is natural, by the virtues of the interval calculus.

Also, when using interval branch-and-bound type methods, we can restrict ourselves to using undercomplete AEs. The additional information, that would be processed for overcomplete AEs, will still be available for various parameterizations of the network, represented by various boxes in the B&BT procedure.

What is more, as stacked AEs can be trained subsequently, the problem under consideration is of lower dimensionality than for some other neural networks.

V. The solver and other software used

In a series of earlier papers, including [25], [26], [27], [28], the author has introduced his solver HIBA_USNE (Heuristical Interval Branch-and-prune Algorithm for Underdetermined and well-determined Systems of Nonlinear Equations) [29]. It was also used in [4] to analyze a Hopfield-like neural network.

HIBA_USNE in its original form could also be used to train a DAE, but it would not be the optimal tool for this purpose. Please note that now the CSP under consideration has a specific form. It contains both equations and inequalities, but:

- the equations refer to hidden layer(s), and they always have the form: $h_i = \sigma\left(\sum_{j=1}^{n} w^1_{ij} x_j\right)$,
- the inequalities refer to the output layer; they come from Formula (8'), and they have the form: $\sigma\left(\sum_{j=1}^{n} w^2_{ij} h_j\right) \subseteq [\underline{y}_i, \overline{y}_i]$.

Obviously, if we had more hidden layers, there would be analogous equations for each of them.

What is important is to note that the role of equations is mostly to establish the relations between the input x and the output y. The specific, 'explicit' form of these equations:

$$h_i = \sigma\left(\sum_{j=1}^{n} w^1_{ij} x_j\right) ,$$

makes the use of the Newton operator on them almost useless. The most important tool is the hull-consistency (HC) enforcing operator, which propagates the information forward and backward through the layers of the neural network.

There are also other tools that might improve the performance of computing the weights of the neural network, but they have not been implemented yet. They will be briefly mentioned in Section VII.

Nevertheless, due to the specific structure of the problem under consideration, the HIBA_USNE solver has been forked by the author. In this paper, the fork, called HIBA_TANN [30] is used. The name stands for HIBA_USNE applied to Training Artificial Neural Networks. It has been adapted for this specific applications. All irrelevant tools have been removed from it (precisely: all but the interval Newton operator and the HC enforcing procedure).

Also, the following additions and amendments have been made to it:

- Only variables representing the weights of the network, and not values propagated by the neurons, get bisected.
- A (primitive) procedure to construct a feasible box (i.e., satisfying all constraints) has been implemented.
- Irrelevant examples have been removed, and new ones have been added.

VI. Experiments

A. Environment

All experiments have been performed on the author's laptop computer, with AMD Ryzen 5-4600H CPU (6 cores, 12 hardware threads; 3GHz). The machine ran under control of a 64-bit Manjaro GNU/Linux operating system with glibc 2.37-2 and the Linux kernel 5.15.93-1-MANJARO (with SMP and PREEMPT options).

The software was written in C++ and compiled using the GCC compiler (GCC 12.2.1). The parallelization (8 threads) was done with TBB 2021.5.0-2 [31]. OpenBLAS 0.3.17 [32] was linked for BLAS operations.

As for the the author's libraries, the following versions have been used:

- ADHC 2.2.1,
- survive-CXSC 2.6.1,
- HIBA_TANN 0.9.1, the fork of HIBA_USNE.

B. The test data set

The experiments have been performed using one of the classical datasets used to test ML tools: *the Iris dataset* [33]. This popular dataset contains descriptions of 150 individuals of some Iris plants, belonging to three species: Iris-setosa, Iris-versicolor, and Iris-virginica. Each of the individuals is described by four numerical attributes: sepal length, sepal width, petal length, and petal width – all of them in centimeters.

Usually, this dataset is used for verifying classification tools; in our experiments, we shall attempt to train an AE to reproduce the attribute values. Adhesion to a specific class will be ignored.

C. Uncertainty

The Iris dataset in its original form contains no uncertainty. In the experiments, the author has induced it by adding to all attributes a random noise. It had a normal distribution with the mean-value zero, and a few values of the standard deviation σ.

Two versions of the uncertainty representation have been considered:

- *Probabilistic uncertainty*: the random values are generated, using the C++11 `std::normal_distribution` class. The size of the dataset is increased four times, to have four

instances of each individual, with various noise values added to its attributes.
- *Interval uncertainty*: the noise value is bounded by the interval $[-3\sigma, +3\sigma]$.

The first version results in significant increasing of the size of the problem under consideration. Instead of $150 \times 4 = 600$ (150 individuals times 4 attributes), we now have $150 \times 4 \times 4 = 2400$ inequality constraints plus the related equations.

The second version does not require increasing the problem dimension. Thanks to the virtues of the interval calculus, all possible values are bound by a single interval.

Remark 6.1: It is worth noting that formally, the interval $[-3\sigma, +3\sigma]$ does not bound the whole support set of the normal distribution $N(0, \sigma)$. Indeed, theoretically, this support is the whole set $\mathbb{R}$. Nevertheless, virtually all practical generators of the normally distributed points generate the points from this range, only.

D. The structure of the neural network

The AE we are training in the presented experiments has the structure precisely described by Fig. 1: there is a single hidden layer, and the numbers of neurons in the input, hidden, and output layers are: 4, 2, 4, respectively.

Each layer is dense, i.e., each neuron of the hidden layer is connected to all neurons of both the input and output layers.

E. Numerical results

We consider the following versions of the problem:

- The ReLU activation function for the neurons, and the noise with $\sigma = 1.0$.
- The ReLU activation function, and the noise with $\sigma = 0.1$.
- The sigmoid activation function, and the noise with $\sigma = 1.0$.
- The sigmoid activation function, and the noise with $\sigma = 0.1$.

All four problems are solved by two versions of the program – using the interval or probabilistic uncertainty description.

The following notation is used in all of the tables:

- eq.evals, grad.eq.evals – numbers of equations evaluations, and their functions' gradients (in the interval algorithmic differentiation arithmetic),
- ineq.evals, grad.ineq.evals – numbers of inequalities evaluations, and their functions' gradients (in the interval algorithmic differentiation arithmetic),
- bisecs – the number of boxes bisections,
- HC evals – numbers of times hull-consistency has been enforced on a box,
- pos.boxes, verif.boxes – number of elements in the computed lists of boxes containing possible and verified solutions,
- Leb.pos., Leb.verif. – total Lebesgue measures of both sets,
- time – computation time in seconds.

F. Analysis of the results

For the sigmoid function, it is not possible to obtain the result satisfying all constraints. Both versions of the program are able to determine it immediately (without any bisections, in a single HC enforcing step!). And it is worth noting that a non-interval algorithm would not be able to tell it for sure – even after a longer search.

For the ReLU function, the version bounding all uncertainty with a single interval, finds a solution quickly, yet it is not able to verify it. The version using a probabilistic representation of the uncertainty was not able to solve the problem in a reasonable time. This fact was surprising to the author – even provided the severely increased number of constraints.

Further studies should improve this version of the algorithm, as well.

VII. Conclusions and future work

This paper describes an attempt to use interval-based constraint-satisfaction algorithms for training denoising autoencoders. The results are interesting, yet only partially successful.

Even for the relatively simple and small problem, considered in the paper, only one version of the algorithm was able to deliver the solution in a reasonable time, and the solution has not been verified with certainty.

Still the results show several important prospects of the proposed approach, in particular the possibility of determining the non-existence of the AE with the given structure that would represent the given data with the assumed precision.

There are several tools that can be used to enhance the considered algorithms. In particular:

- Zonotopes [34] or the Taylor arithmetic [7] could be used to reduce the dependency problem in interval formulae.
- Also, they can be used to represent the covariance matrix of the noise; in the current implementation all attributes are assumed to have independent perturbances.
- A more sophisticated procedure for constructing feasible boxes should be delivered; in particular, it could use Kaucher arithmetic and related theorems, proven by Shary [23], [35].
- For the probabilistic representation, it might be worthwhile to consider only a random subset of the constraints. It is a technique analogous to using the stochastic gradient in optimization problems.
- Finally, processing various equations and inequalities can be parallelized. In the current version of the solver, processing different boxes is done in parallel, but the HC4 algorithm is serial. Cf. the discussion in Sect. 9 of [24].

Also, a similar study is planned for a CAE.

References

[1] S. P. Adam, D. A. Karras, G. D. Magoulas, and M. N. Vrahatis, "Solving the linear interval tolerance problem for weight initialization of neural networks," *Neural Networks*, vol. 54, pp. 17–37, 2014.

[2] S. Huang, Z. Ma, S. Yu, and Y. Han, "New discrete-time zeroing neural network for solving time-variant underdetermined nonlinear systems under bound constraint," *Neurocomputing*, vol. 487, pp. 214–227, 2022.

TABLE I
COMPUTATIONAL RESULTS FOR INTERVAL UNCERTAINTY)

	ReLU, $\sigma = 1$	ReLU, $\sigma = 0.1$	sigmoid, $\sigma = 1$	sigmoid, $\sigma = 0.1$
eq. evals	0	0	0	0
grad.eq.evals	300	300	300	300
ineq. evals	325,283	2,006,727	0	0
grad.ineq.evals	0	0	0	0
bisections	268	1,731	0	0
HC evals	340	3,194	0	0
pos. boxes	1	1	0	0
verif. boxes	0	0	0	0
Leb. pos.	3e-249	3e-249	0.0	0.0
Leb. verif.	0.0	0.0	0.0	0.0
time (sec.)	1	3	<1	<1

TABLE II
COMPUTATIONAL RESULTS FOR PROBABILISTIC UNCERTAINTY)

	ReLU, $\sigma = 1$	ReLU, $\sigma = 0.1$	sigmoid, $\sigma = 1$	sigmoid, $\sigma = 0.1$
eq. evals	n/a	n/a	0	0
grad.eq.evals	n/a	n/a	300	300
ineq. evals	n/a	n/a	0	0
grad.ineq.evals	n/a	n/a	0	0
bisections	n/a	n/a	0	0
HC evals	n/a	n/a	0	0
pos. boxes	n/a	n/a	0	0
verif. boxes	n/a	n/a	0	0
Leb. pos.	n/a	n/a	0.0	0.0
Leb. verif.	n/a	n/a	0.0	0.0
time (sec.)	>3600	>3600	<1	<1

[3] M. Beheshti, A. Berrached, A. de Korvin, C. Hu, and O. Sirisaengtaksin, "On interval weighted three-layer neural networks," in *Simulation Symposium, 1998. Proceedings. 31st Annual*. IEEE, 1998, pp. 188–194.

[4] B. J. Kubica, P. Hoser, and A. Wiliński, "Interval methods for seeking fixed points of recurrent neural networks," in *International Conference on Computational Science*. Springer, 2020. doi: 10.1007/978-3-030-50420-5_30 pp. 414–423.

[5] A. Rauh and E. Auer, "Interval extension of neural network models for the electrochemical behavior of high-temperature fuel cells," *Frontiers in Control Engineering*, vol. 3, 2022. doi: 10.3389/fcteg.2022.785123

[6] P. V. Saraev, "Numerical methods of interval analysis in learning neural network," *Automation and Remote Control*, vol. 73, no. 11, pp. 1865–1876, 2012.

[7] E. de Weerdt, Q. Chu, and J. Mulder, "Neural network output optimization using interval analysis," *IEEE Transactions on Neural Networks*, vol. 20, no. 4, pp. 638–653, 2009.

[8] L. V. Utkin, A. V. Podolskaja, and V. S. Zaborovsky, "A robust interval autoencoder," in *2017 International Conference on Control, Artificial Intelligence, Robotics & Optimization (ICCAIRO)*. IEEE, 2017, pp. 115–120.

[9] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[10] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, 2019.

[11] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," Institute for Cognitive Science, University of California, San Diego, Tech. Rep. 8506, 1985.

[12] ——, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[13] G. E. Hinton, "Connectionist learning procedures," in *Machine learning*, 1990, pp. 555–610.

[14] O. Kosheleva and V. Kreinovich, "Why deep learning methods use KL divergence instead of least squares: A possible pedagogical explanation," University of Texas at El Paso, Tech. Rep. UTEP-CS-17-95, 2017.

[15] X. Li, S. Feng, N. Hou, R. Wang, H. Li, M. ZGao, and S. Li, "Surface microseismic data denoising based on sparse autoencoder and Kalman filter," *Systems Science & Control Engineering*, vol. 10, no. 1, pp. 616–628, 2022.

[16] M. Chen, K. Weinberger, F. Sha, and YoshuaBengio, "Marginalized denoising auto-encoders for nonlinear representations," in *International conference on machine learning*. PMLR, 2014, pp. 1476–1484.

[17] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[18] E. Hansen and W. Walster, *Global Optimization Using Interval Analysis*. Marcel Dekker, New York, 2004.

[19] L. Jaulin, M. Kieffer, O. Didrit, and E. Walter, *Applied Interval Analysis*. Springer, London, 2001.

[20] R. B. Kearfott, *Rigorous Global Search: Continuous Problems*. Kluwer, Dordrecht, 1996.

[21] U. Kulisch, *Computer Arithmetic and Validity – Theory, Implementation and Applications*. De Gruyter, Berlin, New York, 2008.

[22] R. E. Moore, R. B. Kearfott, and M. J. Cloud, *Introduction to Interval Analysis*. SIAM, Philadelphia, 2009.

[23] S. P. Shary, *Finite-dimensional Interval Analysis*. Institute of Computational Technologies, Sibirian Branch of Russian Academy of Science, Novosibirsk, 2013.

[24] B. J. Kubica, *Interval methods for solving nonlinear constraint satisfaction, optimization and similar problems: From inequalities systems to game solutions*, ser. Studies in Computational Intelligence. Springer, 2019, vol. 805.

[25] ——, "Interval methods for solving underdetermined nonlinear equations systems," *Reliable Computing*, vol. 15, pp. 207–217, 2011, proceedings of SCAN 2008.

[26] ——, "Presentation of a highly tuned multithreaded interval solver for underdetermined and well-determined nonlinear systems," *Numerical Algorithms*, vol. 70, no. 4, pp. 929–963, 2015. doi: 10.1007/s11075-015-9980-y

[27] ——, "Parallelization of a bound-consistency enforcing procedure and its application in solving nonlinear systems," *Journal of Parallel and Distributed Computing*, vol. 107, pp. 57–66, 2017. doi: 10.1016/j.jpdc.2017.03.009

[28] ——, "Role of hull-consistency in the HIBA_USNE multithreaded solver for nonlinear systems," *Lecture Notes in Computer Science*, vol. 10778, pp. 381–390, 2018. doi: 10.1007/978-3-319-78054-2_36 Proceedings of PPAM 2017.

[29] "HIBA_USNE, C++ library," https://gitlab.com/bkubica/hiba_usne, 2023.

[30] “HIBA_TANN, C++ library,” https://gitlab.com/bkubica/hiba_tann, 2023.

[31] “Intel TBB,” https://github.com/oneapi-src/oneTBB, 2023.

[32] “OpenBLAS library,” https://www.openblas.net, 2023.

[33] “Iris Species dataset,” https://www.kaggle.com/datasets/uciml/iris?resource=download, 2023.

[34] B. J. Kubica, “Preliminary experiments with an interval Model-Predictive-Control solver,” *Lecture Notes in Computer Science*, vol. 9574, pp. 464–473, 2016, PPAM 2015 Proceedings.

[35] S. P. Shary, “Algebraic approach to the interval linear static identification, tolerance, and control problems, or one more application of Kaucher arithmetic,” *Reliable Computing*, vol. 2, no. 1, pp. 3–33, 1996.

Communication Papers of the 18th Conference on Computer Science and Intelligence Systems pp. 181–188

DOI: 10.15439/2023F9248
ISSN 2300-5963 ACSIS, Vol. 37

An Empirical Framework for Software Aging-Related Bug Prediction using Weighted Extreme Learning Machine

Lov Kumar[1], Vikram Singh[2]
Dept. Computer Engineering
National Institute of Technology, Kurukshetra
{lovkumar, viks}@nitkkr.ac.in

Lalita Bhanu Murthy[3]
Dept. CSIS
BITS Pilani Hyderabad Campus
bhanu@hyderabad.bits-pilani.ac.in

Sanjay Misra[4]
Department of Applied Data Science
Institue for Energy Technology, Halden, Norway
ssopam@gmail.com

Aneesh Krishna[5]
School of Elec Eng, Comp and Math Sci (EECMS)
Curtin University Perth, Australia
A.Krishna@curtin.edu.au

***Abstract*—Software ageing (SA) related bugs highlight the issue of software failure within continuously running systems, resulting in a decline in quality, system crashes, resource misuse, and more. To mitigate these bugs, software companies employ various techniques, including code reviews, bug-tracking systems deployment, and thorough testing. Nevertheless, the identification of aging-related bugs remains challenging through these conventional approaches. To address this predicament, early prediction of the affected software regions due to runtime failures can be immensely valuable for software quality assurance teams. By accurately identifying the vulnerable areas, these teams can strategically allocate their limited resources during the testing and maintenance processes. This proactive approach ensures a more efficient and effective bug detection and resolution, enhancing overall software reliability and performance. This study aims to develop aging-related bug prediction models using source code metrics as input. In particular, our objective is to investigate metrics selections, data balancing, and weighted ELM to detect software runtime failure. Experimental results show that ELM with data imbalance SMOTE technique performs the best compared to weighted ELM for addressing the class imbalance problem. The weighted ELM and ELM + SMOTE can predict SA bugs, and these models can be applied to the future releases of software projects for online failure prediction well in advance. The experimental finding shows that the models trained using normal ELM with SMOTE data sampling techniques have significant performance improvement.**

***Index Terms*—Functional Requirements, Non-Functional Requirements, Data Imbalance Methods, Feature Selection, Classification Techniques, Software Aging**

I. Introduction

In TODAY'S scenario, software (SW) companies are adopting Object-Oriented (OO) concepts to develop modern software systems. The primary reason for adopting these concepts is due to their efficient functionalities, like reusability (extending the code use again), inheritance, data abstraction, polymorphism, cohesion, and coupling. These functionalities help to design SW with high quality such as maintainability, reliability, portability, and reusability. Estimating the SW quality and finding its correlation with static code metrics can help testers, architects, and requirement analysts to analyze the source code concerning SW quality before deploying[1][2]. This point is our primary motivation for present work, with an aim to find the correlation between Software Aging (SA) related bugs and static code metrics as both cost and effort to fix run-time failures or Aging-related bugs increase exponentially if the reason for these failures is not identified prior to SW deployment [3] [4]. In such contexts, the utilization of software aging prediction models emerges as a valuable asset during the initial stages of the SW development life cycle (SDLC). Furthermore, their application holds the potential to enhance software quality, diminish testing expenses, and streamline maintenance efforts.

Software development companies often intend to consider different techniques, such as code reviews, deployment of bug-tracking systems, and various testing techniques for reducing SA bugs [3] [4]. However, it is quite an arduous task to discover the region of SW to be affected due to runtime failure [5]. This study aims to address the issue of predicting runtime failures related to SA by developing a model that incorporates source code metrics and aging-related data. By combining these elements, we strive to create a robust prediction model capable of identifying potential runtime failures in SW systems.

More specifically, this work aims to study the relationship between various source code metrics and SW aging-related bugs and develop a machine learning (ML) enabled prediction model to proactively predict these bugs. These ML models would steer the identification of patterns within the future versions of the SW system based on source code metrics for bug detection. However, we found two major issues in the development of aging-related bugs[6][7][8][9]:

- ***High-Dimensional Data:*** Before applying any technique for model development, it is essential to select relevant and right sets of features that are significant for the development of ML models. In this study, we have considered five different feature ranking techniques: Gain Ratio (GR), OneR, Relief-F(RF), Information Gain (IG), and Symmetric Uncertainty (SU) for ranking and selecting the significantly relevant features for the development of bug prediction models [6][7].

- ***Imbalanced Data:*** Developing an effective prediction model becomes very challenging, particularly with training over highly imbalanced data, it is another pertinent issue for designing SA-related bug prediction models. In these settings, ELM has emerged as a highly efficient and effective ML technique with interest across various domains in recent years. ELM utilizes least square learning methods within a single hidden layer neural network, forming the foundation of its concept for the creation of robust regression and classification models. In this study, the Weighted ELM (WELM) technique has been considered to handle the imbalanced data and develop an effective model for SW aging prediction [8][9]. Further, the performance of these models developed using WELM has been compared with the data imbalance technique and unweighted ELM [8][9].

This work focuses on conducting extensive experimentation to design intelligent models that utilize machine-learning techniques for the proactive prediction of ageing-related bugs in SW. To achieve this, various approaches were employed, including the selection of effective metrics, data balancing, and pattern identification using weighted ELM. Additionally, the data balancing techniques employed addressed class imbalance issues, ensuring that the models were trained on a well-represented dataset. By leveraging these techniques, this study aimed to enhance the proactive identification of SW ageing-related bugs, enabling developers to mitigate potential issues and improve overall SW reliability. Accordingly, the respective Research Questions (RQs) are justified in this study:

RQ1: *What benefits on the performance of aging prediction models after removing ineffective metrics?*

RQ2: *What benefits on the performance of aging prediction models after changing kernel functions?*

RQ3: *What is the benefit of using weighted ELM over ELM + SMOTE techniques for aging prediction models?*

The rest of the paper is organized as follows: Literature Review on methods used for SA bugs is presented in SectionII. The solution methodology, experimental datasets, as well as the various performance parameters used to compare the developed models, are described in SectionIII. The experimental finding of this work and comparative analysis of models developed using different methods are described in SectionIV. Finally, Section VIIwraps up the material presented and suggests research directions for future studies.

II. Related work

In today's scenario, the SW systems operate continuously to complete the assigned task. However, due to faults in design, development, testing, and inappropriate application environment, there is a chance of occurrence of bugs during run time that eventually causes software ageing (SA). Here, we present some key background concepts related to SA.

The idea of SA was first introduced by Huang et al. [10] and subsequently, other researchers have extended it in order to recognize this significant phenomenon [3][11]. Specifically, Parnas et al. [3] discussed the reason behind SA and characterized two types of SA: first, the malfunction of the manufactured goods prior to transforming it to gather transforming needs and second, the effect of the modifications that are prepared. Similarly, Alonso et al. [11] have performed a comparative study related to software rejuvenation and have demonstrated SA in 6 different ways from ground to granularity level. Further, Matias [4] focused on highlighting the potentially common problems that occur due to SA presence, such as data discrepancy, statistical errors, and exhaustion of operating system (OS) resources, which are sample demonstrations of SA. There are a good number of classes where SA effects are reported in the literature with associated impacts to running down of OS resources, and predominantly those connected to the functioning of main memory [12][13].

Zheng and his group developed different rejuvenation rules to find time-based constraints[14]. They have conducted a systematic study to measure these rules numerically and observed that they are better than Markovian arrival processes (MAPs). They observed that eliminating all bugs is not practically possible. Therefore, efforts are being made to estimate the run-time failure or SA of an SW system with the objective of avoiding future system failures. The process of identifying and predicting these bugs is a challenging task for software engineers. Padhy et al. [15] proposed an aging prediction system based on re-usability optimization. This prediction system estimates the re-usability level of the software components. They have applied the concept of re-usability risk management and found that aging-related systems are excessively reused systems. The other method to avoid aging failures is Rejuvenation. Sharma and Kumar have used different types of ensemble models to develop SA prediction models [16]. They have validated the effectiveness of ensemble learning for SA prediction using LINUX and MYSQL bug datasets. They have observed that ensemble methods can identify bugs at early stages and can help reduce the cost and damage caused due to SA.

Khanna and her team have used Artificial Immune Systems for developing SA bugs prediction classifiers [17]. They have used five different types of open-source SW systems to validate the proposed models and asserted that these models have the ability to predict SA bugs. Similarly, Fangyun Qin and his team [18] have examined the variation performance of the models for cross-project SA bugs prediction using different normalization methods, kernel functions, and ML-based classifiers. They have adopted the Scott-Knott test technique to validate their finding and observed that the performance of cross-project SA bugs prediction models depends on the classifiers and kernel functions, while normalization methods do not impact much on performance.

From the above studies and developments, we observed that many researchers have addressed the problem of SA bug prediction. However, the prediction models trained on imbalanced data have rarely been addressed in the literature. Thus, this research work will be a pioneer in the development of SA bugs on imbalanced datasets. In this work, we empirically investigate the performance of SA prediction models developed using weighted ELM and ELM with separate methods for data sampling i.e., SMOTE.

III. Methodology

This section presents the methodology adopted in this experiment in order to predict SA using various ML techniques. Figure 1 illustrates the proposed framework for the development of the SA bug prediction model considering publicly available datasets from seven large open-source software systems. In the discussion in the previous section, it is observed that there have been different types of static code metrics used for developing intelligent models to detect SA bugs[19][20] [21][22]. Therefore, we have applied different sets of static code metrics, such as McCabe's cyclomatic complexity, Halstead's set of metrics, metrics related to the size of software, and metrics associated with aging bugs for aging-related bug prediction. Since we are using these sets of metrics as input, so it is compulsory to remove ineffective metrics, which may help improve the models' performance.

The proposed solution's first phase is applying the feature selection concept to remove ineffective metrics and compute the best sets of effective metrics on pre-processed datasets. Here, we have used five techniques to remove ineffective features such as Gain Ratio, Symmetric Uncertainty, OneR, RELIEF, and Information Gain. The concepts of these techniques are based on performance parameters to rank the features and select top-ranked features for the analysis. In this work, we have used $\lceil log_2^n \rceil$ numbers of top features as the best sets of effective features i.e., $\lceil log_2^{82} \rceil$=7.

The next phase of the proposed framework involves balancing data using SMOTE techniques. We have used ELM with SMOTE and WELM to find patterns to predict SA-related bugs. Here, WELM is applied separately because it was observed that the WELM could handle class imbalance problems. Finally, the ability of the model prediction trained by using ELM with SMOTE and WELM is computed in terms of AUC and Accuracy. This research framework uses two projects, Linus and MySQL, which are downloaded from the PROMISE data repository. We have considered four variants of Linus and three variants of MySQL. Table I presents the description of Linux and MySQL project with the total number of classes(Total Classes), number of non-ageing classes(Non-Aging), number of Aging classes(Aging), % of non-ageing classes(% Non-aging), and % of ageing classes (%Aging).

TABLE I: Software Projects Description

	Linux				MySQL		
Name	**Driver Net**	**Ext3**	**Driver Scsi**	**Ipv4**	**Optimizer**	**Replication**	**Innodb**
ID	**Proj1**	**Proj2**	**Proj3**	**Proj4**	**Proj5**	**Proj6**	**Proj7**
Total Classes	2292	29	962	117	36	32	402
Non-Aging	2283	24	958	115	33	28	370
Aging	9	5	4	2	3	4	32
%Non-Aging	99.61	82.76	99.58	98.29	91.67	87.5	92.04
%Aging	0.39	17.24	0.42	1.71	8.33	12.5	7.96

Feature Ranking: In this work, we have considered 82 different source code metrics as input to the models used for SA prediction. In order to achieve better performance, five different feature ranking techniques are used for feature selection. In this method, a performance parameter is used to rank the features, and the top $log_2 n$ features out of n number of features are selected for aging prediction. The output of two feature ranking techniques i.e., Gain ratio and relief are shown in TableII. Table II presents the ranking of software metrics using two feature ranking techniques from rank 1 to 82 metrics for all 3 projects.

Similarly, the rank of software metrics is computed using other three feature ranking techniques i.e., oneR, infogain,

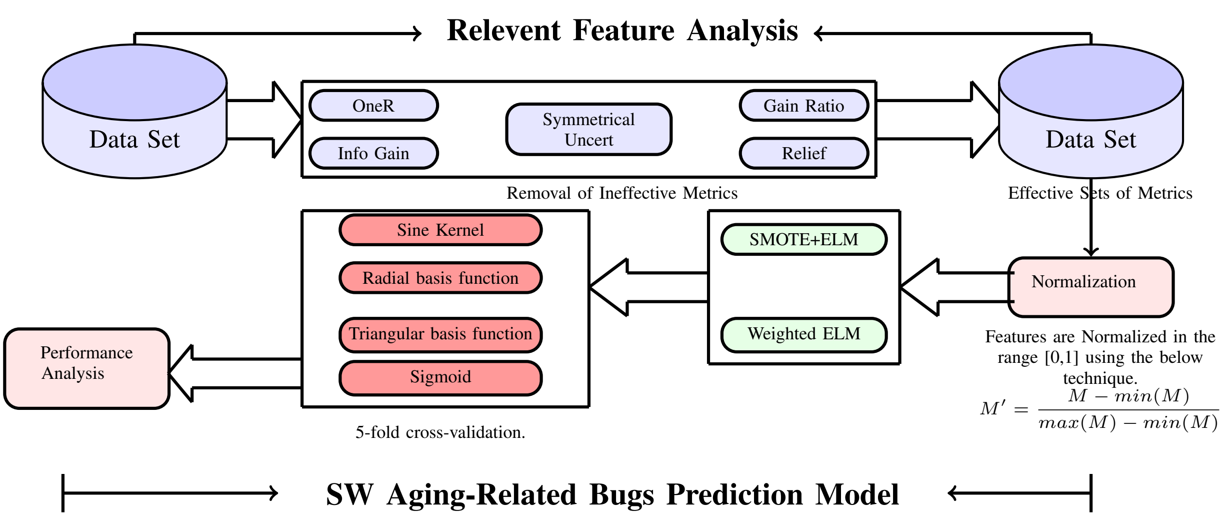

Relevent Feature Analysis
Data Set
OneR
Info Gain
Symmetrical Uncert
Gain Ratio
Relief
Data Set
Removal of Ineffective Metrics
Effective Sets of Metrics
Sine Kernel
Radial basis function
Triangular basis function
Sigmoid
SMOTE+ELM
Weighted ELM
Normalization
Performance Analysis
5-fold cross-validation.
Features are Normalized in the range [0,1] using the below technique.
$M' = \frac{M - min(M)}{max(M) - min(M)}$
SW Aging-Related Bugs Prediction Model

Fig. 1: Conceptual Scheme of Proposed Framework

Symmetric Uncertainty (SU) ranking techniques. It is observed from Table II that the rank of AltAvgLineBlank metrics is 38, 82, and 82 for projects Proj1, Proj2, and Proj3 respectively when gain-ratio feature ranking technique is applied.

IV. Experimental Results

RQ1: What benefits on the performance of aging prediction models after removing ineffective metrics? The AUC value is computed for every model over multiple cutoff points, these values of AUC curve along with Accuracy and F-Measure are listed in Tables III and IV. Here, the results are presented over different projects, different classification techniques, and one data imbalance technique. The feature ranking technique yielding the high AUC for a given project is depicted in green color. Further, inference from Tables III and IV, the AUC value of proj1 using oneR with the sigmoid kernel is better than the other feature ranking techniques and kernel methods. In most of the cases, the performance parameters values of the feature ranking techniques with ELM + SMOTE were found to be comparable or even better than feature ranking techniques with weighted ELM.

Comparison of Feature Ranking Techniques using Boxplots and Descriptive Statistics: Figures 2 and 3 depict boxplots for comparing the minimum, maximum, interquartile range, degree of dispersion, and outliers in the AUC, F-Measure, and Accuracy for all feature ranking techniques. Table refdst reports the descriptive statistics of the performance of different feature ranking techniques. From Figures 2 and 3, we infer that the accuracy and F-measure values of the case where feature ranking techniques are used with ELM + SMOTE is better than the one in which we use feature ranking techniques with weighted ELM. However, these performance parameters such as accuracy and F-measure are not those good parameters to validate the model developed using imbalanced data sets. So, in this experiment, Area under the ROC (Receiver Operating Characteristics) Curve (AUC) values have been considered to measure classifier performance. The line is represented using the red color of the Figures 2 and 3 displays the median points of the data and this line is used to divide the box into two segments. Similarly, Figures 2 and 3 depict the average AUC of RF in the case of ELM + SMOTE is higher than the corresponding values for other feature ranking techniques.

Comparison: Null Hypothesis: Feature Ranking Techniques: Statistical Significance Testing: After comparing different feature ranking techniques using boxplots and Descriptive Statistics, Wilcoxon signed-rank test with a Bonferroni correction has been applied for statistical hypothesis testing. The objective of this testing is to investigate the statistical difference between the pairs of different feature ranking techniques.

The results of the Wilcoxon signed-rank test with a Bonferroni correction of the feature ranking technique are shown in Figures 4 and 5. Figures 4 and 5 consist of a green and red dots. The rejected null hypothesis is represented using a red dot and the accepted null hypothesis is represented using

TABLE II: Ranking of Static Code Metrics for all Projects using Gain Ratio and Relief.

	Gain Ratio			Relief		
	Proj1	**Proj2**	**Proj3**	**Proj1**	**Proj2**	**Proj3**
CountDeclInstanceVariable	43	10	17	52	61	81
CountLineInactive	20	42	47	32	51	36
CountClassDerived	46	36	40	57	57	67
CountLineBlank	2	17	48	10	22	22
AvgCyclomaticStrict	53	31	33	50	13	47
CountDeclMethodPrivate	82	8	21	80	81	59
MaxCyclomaticModified	64	78	75	45	6	33
CountDeclClass	42	35	38	56	56	71
MaxCyclomatic	63	80	77	44	4	34
n2	23	44	10	3	37	4
N1	4	45	56	9	42	8
CountDeclMethodConst	74	5	23	78	79	70
CountLineCodeExe	7	15	71	26	38	2
MinEssentialKnots	76	64	53	71	71	52
CyclomaticStrict	62	73	80	74	60	64
DeallocOps	34	55	60	31	41	37
RatioCommentToCode	72	62	51	47	44	82
CountDeclInstanceVariablePublic	58	2	19	75	82	73
Dif	13	61	11	7	5	26
DerefUse	21	57	59	21	12	15
MaxInheritanceTree	67	76	64	62	64	57
CountStmtEmpty	70	67	67	51	46	50
CountDeclMethodProtected	77	12	25	81	76	51
CountStmtDecl	17	65	8	23	20	25
CountDeclFunction	26	13	26	6	8	28
CountDeclClassMethod	47	33	28	53	55	77
AllocOps	35	58	61	37	40	38
CountDeclMethod	71	6	16	82	77	69
MaxCyclomaticStrict	65	75	65	46	3	32
AvgLineCode	55	40	44	34	25	40
CountLineCode	31	14	66	20	35	10
CountStmtExe	24	66	74	16	9	5
SumCyclomatic	16	51	50	38	19	13
CountPath	79	71	68	66	65	61
CountDeclMethodPublic	78	18	20	76	73	56
MaxNesting	66	74	54	63	63	53
Essential	61	79	79	60	69	63
n1	10	47	4	2	2	24
DerefSet	11	23	1	19	28	20
CountClassCoupled	49	34	27	55	58	66
CountLine	27	19	41	4	32	16
CountClassBase	48	37	39	58	53	65
AvgCyclomaticModified	41	25	36	49	16	45
AvgEssential	54	28	37	41	26	49
AltCountLineComment	15	24	31	25	39	31
SumCyclomaticModified	12	50	55	39	15	19
PercentLackOfCohesion	57	63	52	72	59	79
AltAvgLineComment	40	1	34	35	23	39
AvgCyclomatic	52	26	30	48	7	48
UniqueDerefUse	32	56	2	14	18	23
CountLineComment	14	41	12	27	36	30
CountDeclInstanceMethod	44	21	18	59	52	80
AltAvgLineBlank	38	82	82	29	30	44
Cyclomatic	59	72	76	69	68	76
CountStmt	25	68	73	18	11	12
MaxEssentialKnots	68	77	57	65	66	55
AltAvgLineCode	51	30	35	33	27	41
AltCountLineBlank	1	29	32	8	33	21
SumCyclomaticStrict	19	49	49	42	10	9
Knots	60	81	78	70	70	58
Vol	9	59	63	13	48	1
CountDeclInstanceVariableProtected	50	7	13	67	75	74
CountOutput	80	46	69	64	62	60
AvgLineBlank	37	38	45	28	29	46
SumEssential	28	48	3	15	1	3
CountDeclMethodAll	75	3	22	77	78	68
AltCountLineCode	33	27	29	17	34	17
CountDeclInstanceVariablePrivate	56	9	14	61	67	75
CountInput	81	20	46	73	72	54
AvgLine	30	32	43	30	14	43
Eff	8	60	62	24	50	18
AvgLineComment	39	39	42	36	24	42
N2	5	52	9	12	45	6
Len	3	53	58	11	43	7
Voc	18	54	7	1	31	14
CountLinePreprocessor	36	43	70	43	47	35
CountLineCodeDecl	29	16	72	40	49	27
CountDeclClassVariable	45	22	24	54	54	78
CountDeclMethodFriend	73	11	15	79	80	72
CountSemicolon	22	70	6	22	21	11
CyclomaticModified	69	69	81	68	74	62
UniqueDerefSet	6	4	5	5	17	29

TABLE III: AUC: Accuracy: Weighted ELM

	AUC																			
	Sine					Radial basis function					Triangular basis function					Sigmoid				
	OneR	IG	GR	RF	SU	OneR	IG	GR	RF	SU	OneR	IG	GR	RF	SU	OneR	IG	GR	RF	SU
Proj1	0.63	0.60	0.60	0.58	0.43	0.58	0.50	0.50	0.50	0.50	0.48	0.50	0.50	0.50	0.50	0.8	0.70	0.66	0.72	0.67
Proj2	0.9	0.41	0.79	0.61	0.80	0.79	0.88	0.56	0.59	0.57	0.57	0.49	0.57	0.50	0.47	0.70	0.71	0.85	0.80	0.82
Proj3	0.67	0.61	0.55	0.53	0.51	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.67	0.71	0.58	0.69
Proj4	0.37	0.86	0.65	0.51	0.62	0.68	0.39	0.47	0.50	0.45	0.49	0.45	0.48	0.50	0.46	0.58	0.56	0.81	0.73	0.80
Proj5	0.59	0.44	0.47	0.57	0.48	0.50	0.50	0.49	0.50	0.49	0.52	0.50	0.51	0.50	0.49	0.76	0.67	0.74	0.66	0.71
Proj6	0.35	0.21	0.26	0.64	0.55	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.56	0.59	0.56	0.55	0.55
Proj7	0.54	0.55	0.70	0.46	0.54	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.77	0.86	0.70	0.61	0.77
	Accuracy																			
	Sine					Radial basis function					Triangular basis function					Sigmoid				
	OneR	IG	GR	RF	SU	OneR	IG	GR	RF	SU	OneR	IG	GR	RF	SU	OneR	IG	GR	RF	SU
Proj1	71.20	52.66	52.66	49.61	52.14	94.28	99.61	99.61	99.61	99.61	95.99	99.61	99.61	99.61	99.61	59.86	51.27	43.06	43.98	34.42
Proj2	80.56	81.81	83.26	71.52	84.82	83.06	75.99	87.32	92.41	88.15	89.71	98.13	89.19	98.96	93.04	65.38	67.15	70.06	60.19	63.93
Proj3	58.62	62.07	37.93	62.07	44.83	82.76	82.76	82.76	82.76	82.76	82.76	82.76	82.76	82.76	82.76	31.03	58.62	51.72	31.03	48.28
Proj4	72.65	72.65	78.63	52.14	73.50	85.47	76.07	92.31	98.29	88.89	95.73	88.03	94.87	98.29	90.60	64.96	61.54	63.25	47.86	61.54
Proj5	77.86	47.26	50.00	67.66	51.99	92.04	92.04	91.04	91.54	85.82	92.29	91.54	92.04	91.29	91.04	62.94	62.69	62.94	63.93	64.18
Proj6	36.11	38.89	47.22	61.11	72.22	91.67	91.67	91.67	91.67	91.67	91.67	91.67	91.67	91.67	91.67	19.44	25.00	19.44	16.67	16.67
Proj7	56.25	59.38	65.63	62.50	56.25	87.5	87.5	87.5	87.5	87.5	87.5	87.5	87.5	87.5	87.5	78.13	75.00	65.63	50.00	59.38

TABLE IV: AUC: Accuracy: ELM + SMOTE

	AUC																			
	Linear					Polynomial					Sigmoid					Radial basis function				
	OneR	IG	GR	RF	SU	OneR	IG	GR	RF	SU	OneR	IG	GR	RF	SU	OneR	IG	GR	RF	SU
Proj1	0.50	0.49	0.48	0.50	0.48	0.57	0.62	0.61	0.72	0.62	0.57	0.50	0.50	0.50	0.50	0.72	0.50	0.50	0.50	0.50
Proj2	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.59	0.68	0.69	0.55	0.65
Proj3	0.89	0.75	0.76	0.89	0.78	0.68	0.91	0.80	0.81	0.81	0.40	0.50	0.50	0.56	0.50	0.50	0.60	0.65	0.50	0.60
Proj4	0.50	0.50	0.50	0.48	0.50	0.49	0.50	0.49	0.83	0.49	0.50	0.50	0.50	0.50	0.50	0.48	0.48	0.49	0.50	0.48
Proj5	0.85	0.76	0.77	0.66	0.78	0.79	0.80	0.79	0.85	0.83	0.73	0.50	0.50	0.81	0.50	0.77	0.63	0.64	0.88	0.65
Proj6	0.66	0.96	0.96	0.88	0.84	0.63	0.65	0.60	0.66	0.51	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
Proj7	0.92	0.85	0.85	0.94	0.98	0.92	0.85	0.88	0.87	0.82	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
	Accuracy																			
	Linear					Polynomial					Sigmoid					Radial basis function				
	OneR	IG	GR	RF	SU	OneR	IG	GR	RF	SU	OneR	IG	GR	RF	SU	OneR	IG	GR	RF	SU
Proj1	98.44	97.30	95.05	98.30	95.13	98.57	98.39	76.40	95.60	78.63	93.31	98.43	98.39	98.43	98.43	98.22	98.39	98.44	98.43	98.43
Proj2	98.24	98.33	98.33	98.25	98.33	98.34	98.33	98.33	98.35	98.33	98.34	98.33	98.02	98.35	98.33	98.34	98.23	98.54	96.80	98.23
Proj3	89.19	77.78	77.78	89.19	80.56	67.57	91.67	80.56	81.08	80.56	40.54	55.56	55.56	56.76	55.56	48.65	55.56	61.11	48.65	55.56
Proj4	96.52	96.52	96.52	92.24	96.52	94.78	95.65	94.78	89.66	94.78	96.52	96.52	96.52	96.55	96.52	93.04	93.04	94.78	95.69	93.04
Proj5	87.58	80.70	79.33	64.73	78.68	85.33	86.62	86.44	89.51	88.13	72.46	71.93	71.56	75.45	71.87	85.33	79.39	79.33	92.19	79.78
Proj6	48.65	94.59	94.59	86.84	86.49	67.57	70.27	72.97	60.53	59.46	75.68	75.68	75.68	76.32	75.68	75.68	75.68	75.68	76.32	75.68
Proj7	90.48	83.33	80.95	92.86	97.44	90.48	80.95	85.71	83.33	76.92	61.90	61.90	61.90	61.90	64.10	61.90	61.90	61.90	61.90	64.10

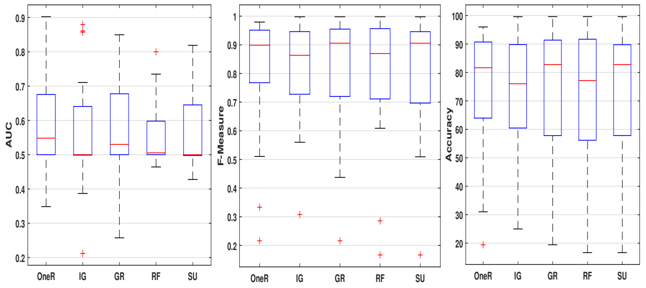


Fig. 2: Box-and-Whisker plot: AUC Value of Weighted ELM with Feature ranking techniques

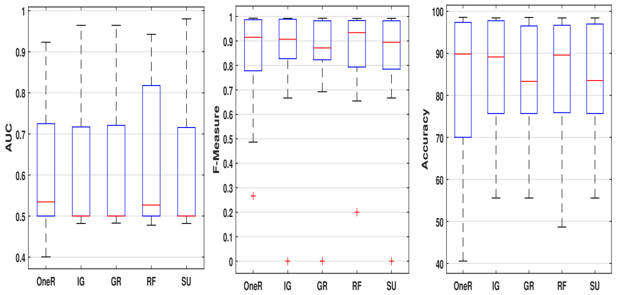


Fig. 3: Box-and-Whisker plot: AUC Value of ELM + SMOTE with Feature ranking techniques

a green dot. The null hypothesis in this experiment is that "there is no statistically significant difference between the two techniques". In this experiment, the standard cut-off value of 0.05/(*total number of unique pairs*)=0.05/10=0.005 is used to reject and accept this hypothesis. The results shown in the Figures 4 and 5 depict that all cells contain a green dot for feature ranking techniques in both weighted ELM and ELM + SMOTE cases. Based on these results, it is observed that the aging prediction model developed by considering different sets of metrics obtained using feature ranking techniques is not significantly different.

RQ2: What benefits on the performance of aging prediction models after changing kernel functions?

After finding relevant sets of features using five different feature ranking techniques, the aging prediction models are developed using weighted ELM with various kernels and ELM with SMOTE data imbalance technique. In this study, four different types of kernel functions are used to develop a model

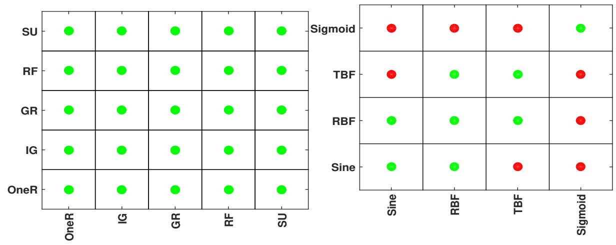


Fig. 4: Weighted ELM: Wilcoxon Signed-Rank Test + Bonferroni Correction: A Red Dot Means that H_0 is Rejected

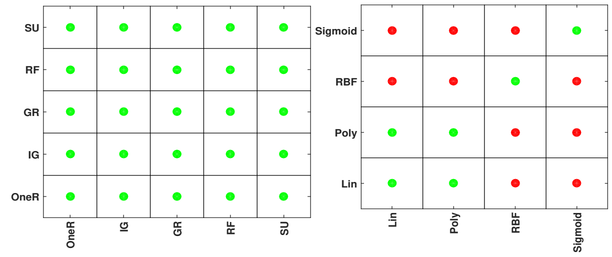


Fig. 5: ELM + SMOTE: Wilcoxon Signed-Rank Test + Bonferroni Correction: A Red Dot Means that H_0 is Rejected

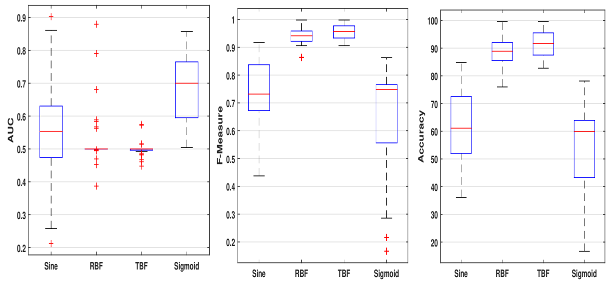


Fig. 6: box-and-whisker plot: AUC Value of Weighted ELM with different kernels

for predicting SA bugs. The developed models are validated using 5-fold cross-validation. The value of the AUC curve along with accuracy and F-Measure for different kernels are shown in Tables III and IV. The results are presented over different projects, different feature ranking techniques, andone data imbalance technique. From the Tables III and IV, it can be inferred that AUC values of models using sine and sigmoid kernels are better than the other radial basis function and triangular basis function kernels in the case of weighted ELM. Similarly, the AUC value of linear, polynomial, and radial basis function kernels are better than the sigmoid kernel in the case of ELM + SMOTE

Comparison of kernel functions using Boxplots and Descriptive Statistics: Figures 6 and 7 depict boxplots for comparing the minimum, maximum, interquartile range, degree of dispersion, and outliers in the AUC, F-Measure, and Accuracy for all considered kernel functions. From Figures 6 and 7, we infer that the accuracy and F-measure values of the case where kernel functions are used with ELM + SMOTE are better than the one in which we use kernel functions with weighted ELM. In this experiment, Area under the ROC Curve (AUC) values has been considered to measure classifier performance. Figures 6 and 7 depict the average AUC of sigmoid kernel function in case of weighted ELM is higher than the corresponding values for other kernel functions.

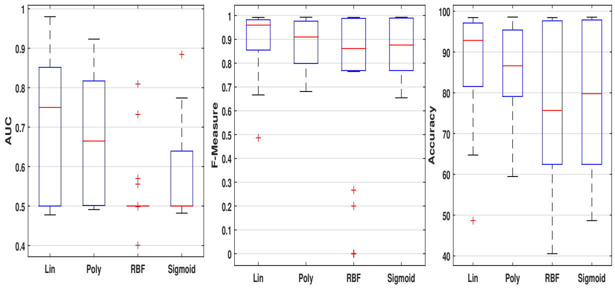


Fig. 7: box-and-whisker plot: AUC Value of ELM + SMOTE with different kernels

Comparison: Null Hypothesis: kernel functions: Hypothesis Statistical Significance Testing: The results of the Wilcoxon signed-rank test with a Bonferroni correction of the different pairs of kernels are shown in Figures 4 and 5. The rejected null hypothesis is represented using a red dot and the accepted null hypothesis is represented using the green dot. In this experiment, a standard cut-off value of $0.05/(\textit{total number of unique pairs})$=$0.05/6$ has been considered to reject and accept this hypothesis. The results shown in Figures 4 and 5 depict that the aging prediction models developed using weighted ELM with sine and RBF kernel function are not significantly different. Similarly, the SA prediction models developed using weighted ELM with sine, TBF, and sigmoid kernel functions are significantly different.

RQ3: What is the benefit of using weighted ELM over ELM + SMOTE techniques for aging prediction models? Zong et al.[8] mathematically proved that weighted ELM with various kernel functions is able to handle imbalanced data and also maintain better performance on balanced data as compared to unweighted ELM. In this work, we have considered weighted ELM and unweighted ELM with data imbalance techniques i.e., SMOTE (Synthetic Minority Oversampling Technique) to develop a SA bugs prediction model. SMOTE technique is based on the oversampling concept and aimed to increase the number of artificial instances that belong to the minority class. Specifically, artificial instances are created using oversampling.

The value of the AUC curve along with Accuracy and F-Measure for WELM and ELM + SMOTE listed in Tables III and IV. According to Tables III and IV, the performance parameters i.e., AUC, Accuracy, and F-Measure of the scenario where weighted ELM is used are lower or comparable with that of the one in which ELM with data imbalance SMOTE technique is used.

Comparison of weighted ELM and ELM + SMOTE using Boxplots and Descriptive Statistics: Figure 8 presents

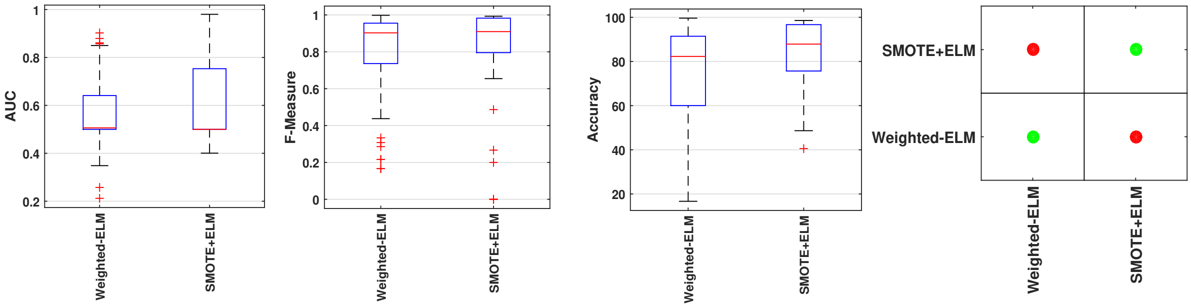


Fig. 8: Box-and-whisker plot: AUC Value of ELM + SMOTE and Weighted ELM

the pictorial representation of descriptive statistics containing Minimum, Maximum, Interquartile range, Degree of dispersion, and Outliers the AUC, F-Measure, and Accuracy for weighted ELM and ELM + SMOTE. An analysis of the Figure 8 indicates the performance parameter value of ELM + SMOTE is better than the one in which we use weighted ELM. So, the aging prediction model developed using ELM + SMOTE obtains better performance as compared to weighted ELM.

Comparison: Null Hypothesis: weighted ELM and ELM + SMOTE: Hypothesis Statistical Significance Testing: In this paper, we have also applied Wilcoxon signed-rank test with a Bonferroni correction for statistical hypothesis testing i.e., NULL Hypothesis: "There is no significant difference in the predictive ability of models trained using different machine learning techniques". The results of the Wilcoxon signed-rank test with a Bonferroni correction of the weighted ELM and ELM + SMOTE are shown in the last sub-figure Figure8, which consists of green and red dots. The rejected null hypothesis is represented using a red dot and the accepted null hypothesis is represented using the green dot.

The null hypothesis in this experiment is that "there is no statistically significant difference between the model developed using weight ELM and model developed using ELM + SMOTE". The results shown in Figure 8 depicts that the cell contains a red dot for weighted ELM and ELM + SMOTE. Based on this result, it may be observed that the aging prediction model developed using weighted ELM and ELM + SMOTE is significantly different.

V. Discussion of results

The proposed study conducted extensive experimentation with the application of the five feature ranking techniques to extract the relevant metrics against 7 projects, to analyse the impact on the accuracy and predictability of SA related Bugs Prediction models when SMOTE (data sampling) + ELM is used in comparison to WELM. Additionally, we examined the performance of sine, sigmoid, radial basis, and triangular basis function kernels in the case of WELM and linear, polynomial, radial basis, and sigmoid basis function kernels in the case of ELM + SMOTE to further investigate the finding with the use of different kernels. The outcome of the empirical experimentation reveals that in the majority of cases, the performance parameter values of feature ranking techniques with ELM + SMOTE were found to be equivalent to or outperformed than with WELM. The employment of OneR feature ranking techniques typically yields results that are competitive. The ELM + SMOTE with the application of Relief feature ranking has the highest F-measure, at 0.93, of all the combinations.

The experimental analysis of the effectiveness of various kernels manifests that the AUC values of models using sine and sigmoid kernels outperform other applied kernels in the case of weighted ELM, whereas for ELM + SMOTE, the sigmoid kernel underperforms other utilised kernels. The study established the improved performance post-implication of kernel techniques with ELM + SMOTE in comparison to weighted ELM. The ELM + SMOTE with linear kernel secures the highest performance among other developed models with 0.75 AUC value, 92.86% Accuracy, and 0.96 F-measure. This study noticed improved AUC, F-measure, and Accuracy values in the descriptive statics based on WELM and ELM + SMOTE results. The Accuracy value of ELM + SMOTE is 87.86, whereas 82.9 for WELM is 82.29, indicating an increase in accuracy of 5.57%.

VI. Threats to Validity

We have also expressed threats to the validity of the proposed work.

i. *Internal Validity:* In relation to internal validity the SA bugs datasets are utilized, to validate the proposed models, which were sourced from the tera-promise data repository. It is important to note that we cannot assert with absolute certainty that the provided data is 100 per cent accurate. However, we have confidence that it was collected consistently. Another factor affecting internal validity is that the sampled data obtained through the use of the SMOTE technique may not precisely reflect the characteristics of actual aging datasets. Both assertions may lead to a potential threat to internal validity because sampled data is used as an input of the trained models and not generalized for testing. However, in the proposed solutions we have been validated with different classification performance parameters such as Accuracy, AUC, and F-Measure, in order to reduce the validation bias.

ii. *Construct Validity:* Many researchers have already developed the SA bugs prediction methods using different sets of source code metrics, as highlighted within the related work section. These works successfully validated the SA bugs that we have also used in this experiment. So, the construct validity threat related to aging or run time failures does not exist.

iii. *External Validity:* The developed models are validated using 07 different datasets that have been designed using procedure language. The finding may vary for projects developed using other programming languages. Hence, a threat to external validity exists in this study. However, the argument setting of developed models helps to reduce the threats to generalizability.

VII. Conclusion

The development of SA prediction model using source code metrics steers the improved software quality and reduces runtime failure. In this paper, empirical experiments have been conducted on seven different applications and proposed to develop early SA bug prediction. The major contributions of this paper are (a) the development of aging prediction models using weighted ELM and ELM + SMOTE with various kernels, (b) the selection of significant right sets of features using different feature ranking techniques, (c) the handling of imbalanced data using SMOTE and weighted elm, and (d) analysis of the performance of the developed model to find the generalized and meaningful conclusion. The experimental assertions of the study are as follows:

- The effectiveness of feature ranking techniques employing ELM + SMOTE exceeds that of feature ranking techniques utilizing weighted ELM.
- The aging prediction model, developed by incorporating diverse metric sets obtained through feature ranking techniques, demonstrates no significant variation.
- The performance of the same kernel functions with ELM + SMOTE is better than kernel functions with weighted ELM.
- The aging prediction models developed using different kernel functions are significantly different.
- The aging prediction model developed using ELM + SMOTE outperforms the weighted ELM approach, and a prediction model based on weighted ELM and ELM + SMOTE exhibit significant dissimilarities.

Thus, we finally conclude that the better value of AUC for the models trained using weighted ELM and ELM + SMOTE confirms that the trained models have the ability to predict SA bugs. These models can be applied to future releases of the SW system's for proactive runtime failure prediction.

VIII. Acknowledgements

This research is funded by TestAIng Solutions Pvt. Ltd.

References

[1] S. R. Chidamber and C. F. Kemerer, "A metrics suite for object oriented design," *IEEE Transactions on software engineering*, vol. 20, no. 6, pp. 476–493, 1994.

[2] M. K. Thota, F. H. Shajin, P. Rajesh *et al.*, "Survey on software defect prediction techniques," *International Journal of Applied Science and Engineering*, vol. 17, no. 4, pp. 331–344, 2020.

[3] R. Pietrantuono and S. Russo, "A survey on software aging and rejuvenation in the cloud," *Software Quality Journal*, vol. 28, no. 1, pp. 7–38, 2020.

[4] R. Matias, B. E. Costa, and A. Macedo, "Monitoring memory-related software aging: An exploratory study," in *2012 IEEE 23rd International Symposium on Software Reliability Engineering Workshops*. IEEE, 2012, pp. 247–252.

[5] S. S. Chouhan, S. S. Rathore, and R. Choudhary, "A study of aging-related bugs prediction in software system," in *Proceedings of the International Conference on Paradigms of Computing, Communication and Data Sciences*. Springer, 2021, pp. 49–61.

[6] J. Dai, J. Chen, Y. Liu, and H. Hu, "Novel multi-label feature selection via label symmetric uncertainty correlation learning and feature redundancy evaluation," *Knowledge-Based Systems*, vol. 207, p. 106342, 2020.

[7] A. G. Karegowda, A. Manjunath, and M. Jayaram, "Comparative study of attribute selection using gain ratio and correlation based feature selection," *International Journal of Information Technology and Knowledge Management*, vol. 2, no. 2, pp. 271–277, 2010.

[8] W. Zong, G.-B. Huang, and Y. Chen, "Weighted extreme learning machine for imbalance learning," *Neurocomputing*, vol. 101, pp. 229–242, 2013.

[9] K. Li, X. Kong, Z. Lu, L. Wenyin, and J. Yin, "Boosting weighted elm for imbalanced learning," *Neurocomputing*, vol. 128, pp. 15–21, 2014.

[10] Y. Huang, C. Kintala, N. Kolettis, and N. D. Fulton, "Software rejuvenation: Analysis, module and applications," in *Twenty-fifth international symposium on fault-tolerant computing. Digest of papers*. IEEE, 1995, pp. 381–390.

[11] J. Alonso, R. Matias, E. Vicente, A. Maria, and K. S. Trivedi, "A comparative experimental study of software rejuvenation overhead," *Performance Evaluation*, vol. 70, no. 3, pp. 231–250, 2013.

[12] R. Matias and J. Paulo Filho, "An experimental study on software aging and rejuvenation in web servers," in *30th Annual International Computer Software and Applications Conference (COMPSAC'06)*, vol. 1. IEEE, 2006, pp. 189–196.

[13] J. Araujo, R. Matos, P. Maciel, R. Matias, and I. Beicker, "Experimental evaluation of software aging effects on the eucalyptus cloud computing infrastructure," in *Proceedings of the middleware 2011 industry track workshop*, 2011, pp. 1–7.

[14] J. Zheng, H. Okamura, L. Li, and T. Dohi, "A comprehensive evaluation of software rejuvenation policies for transaction systems with markovian arrivals," *IEEE Transactions on Reliability*, vol. 66, no. 4, pp. 1157–1177, 2017.

[15] N. Padhy, R. Singh, and S. C. Satapathy, "Enhanced evolutionary computing based artificial intelligence model for web-solutions software reusability estimation," *Cluster Computing*, vol. 22, no. 4, pp. 9787–9804, 2019.

[16] S. Sharma and S. Kumar, "Analysis of ensemble models for aging related bug prediction in software systems." in *ICSOFT*, 2018, pp. 290–297.

[17] M. Khanna, M. Aggarwal, and N. Singhal, "Empirical analysis of artificial immune system algorithms for aging related bug prediction," in *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, vol. 1. IEEE, 2021, pp. 692–697.

[18] F. Qin, X. Wan, and B. Yin, "An empirical study of factors affecting cross-project aging-related bug prediction with tlap," *Software Quality Journal*, vol. 28, no. 1, pp. 107–134, 2020.

[19] A. G. Koru and H. Liu, "An investigation of the effect of module size on defect prediction using static measures," in *Proceedings of the 2005 workshop on Predictor models in software engineering*, 2005, pp. 1–5.

[20] T. Mende, "Replication of defect prediction studies: problems, pitfalls and recommendations," in *Proceedings of the 6th International Conference on Predictive Models in Software Engineering*, 2010, pp. 1–10.

[21] T. Mende and R. Koschke, "Revisiting the evaluation of defect prediction models," in *Proceedings of the 5th International Conference on Predictor Models in Software Engineering*, 2009, pp. 1–10.

[22] A. Bovenzi, D. Cotroneo, R. Pietrantuono, and S. Russo, "Workload characterization for software aging analysis," in *2011 IEEE 22nd International Symposium on Software Reliability Engineering*. IEEE, 2011, pp. 240–249.

DOI: 10.15439/2023F5023
ISSN 2300-5963 ACSIS, Vol. 37

Mathematical Modeling of Water Regimes on Drained Lands

Lyudmyla Kuzmych
: 0000-0003-0727-0508
Institute of Water Problems and Land Reclamation,
37, Vasylkivska Str., Kyiv, 03022, Ukraine
Email: kuzmychlyudmyla@gmail.com

Halyna Voropai
0000-0002-5004-0727
Institute of Water Problems and Land Reclamation,
37, Vasylkivska Str., Kyiv, 03022, Ukraine
Email: voropaig@ukr.net

Stepan Kuzmych
0000-0001-8983-3882
National University of Water and Environmental Engineering,
11, Soborna Str., Rivne, 33000, Ukraine,
Email: kuzmych_vg17@nuwm.edu.ua

Abstract—**The model of the dynamics of the ground water level (CWL) in the area between the drains during pressure regulation in the drains in the conditions of a three-layer soil structure is proposed and implemented. Having the connection between ground water level and humidity in the aeration zone established on the basis of the conducted experiments, the issue of ensuring the necessary humidity in the aeration zone within the root system is resolved.**

As a result of the regulation of GWL in different modes (passive reduction and humidification) taking into account natural conditions, in particular, based on the received database on the amount of precipitation, the necessary parameters were obtained that characterize the water regime in the aeration zone. The analysis of the obtained results allows for establishing and proposing more effective resource-saving modes of moistening under the condition of a sufficient supply of moisture to the root layer.

In the conducted experiments, the accumulated precipitation in the active layer (0-0.6 m) of the soil in the mode of passive reduction of GWL, when an accumulative capacity for moisture retention is formed in the upper layers of the soil, was used as efficiently as possible.

Index Terms—**model, analysis, root layer, groundwater level, humidification mode, filtering, soil, aeration zone**.

I. Introduction

The complex conditions of the functioning of reclamation systems at the present stage necessitate the development and implementation of advanced, cost-effective, environmentally safe regimes and technologies that provide for the restoration of the ecological balance of natural landscapes and water ecosystems, the reproduction of soil fertility, as well as adaptation to new socio-economic conditions, advanced technologies of agricultural production, i.e. ensuring the conditions for the transition from extensive agriculture on the reclaimed lands to intensive on the basis of the use of the latest scientific achievements and best practices [1-4].

The purpose of the paper is the mathematical modeling of the groundwater level regime and the justification of resource-saving technological parameters of water regulation on drained lands, which take into account the peculiarities of the use of moisture by cultivated agricultural crops, which allows for ensuring their reliable moisture supply on drained lands.

II. Methods and Techniques

Well-known Ukrainian and foreign scientists studied various aspects of regulating the water regime on drained lands [3-6]. Regulation of the water regime on drained lands is based on two main approaches [7-11].

The first of them is water balance, the easiest to implement, but very close. In modern conditions, it does not satisfy the requirements of agricultural production on drained lands, since it does not allow to correct establish the amount of water exchange between soil layers in the aeration zone and practically does not provide information about the dynamic properties of the object (parameters and type of regulatory network, hydrogeological and geological conditions, hydrophysical soil parameters, etc.).

The second is based on the wide application of mathematical modeling methods, while the moisture transfer equation is used [12]. This approach makes it possible to take into account a complex of factors that determine the moisture regime in the root layer of the soil, including the intensity and distribution of moisture absorption by cultivated crops along the depth of the root zone.

The water-air regime of the soil, which is favorable for agricultural crops, is maintained thanks to the use of various regulation technologies, which must provide for such dynamics of the groundwater level (GWL), in which favorable soil moisture is maintained in the aeration zone, and at the same time, atmospheric precipitation can be accumulated in the root layer without harmful impact of flooding of the root system of cultivated crops.

Taking into account the modern requirements for the level of substantiation of technological schemes and melioration regimes, it becomes obvious that an important approach to solving the tasks is the use of mathematical modeling based on physically based models of hydrophysical processes, which will allow taking into account the complex of natural and technical factors that affect the quality functioning of reclamation systems.

Therefore, an extremely important component of research is the schematization of natural conditions and the construction of calculated filtration schemes that reflect the most important factors in the formation of water-physical processes in real conditions[10-14].

In the process of schematization, the form and structure of the groundwater movement area are established; the presence and intensity of pressurized power supply; the nature of water exchange between saturated-unsaturated zones and the atmosphere; significance, location, and nature of the action of the main elements of the drying-humidification system; the initial and final values of the flow characteristics, which must be calculated, and their relationship with other elements of the scheme. The method of schematizing natural conditions is described in detail in [6, 15-21].

III. Results and Discussion

Analysis of engineering-geological studies, that according to the geological structure, the aquifer layer of the experimental area is heterogeneous and can be reduced to a three-layer structure with a horizontal water table. The aquifer is pressureless with a free surface.

After carrying out the schematization of natural conditions, an estimated filtration scheme was built (Fig. 1). We will conduct an analysis of the Groundwater level (GWL) regime in a three-layer foundation when it is moistened as a result of raising the level to the required depth, in which case the formation of the necessary moisture regime is substantiated in the root layer.

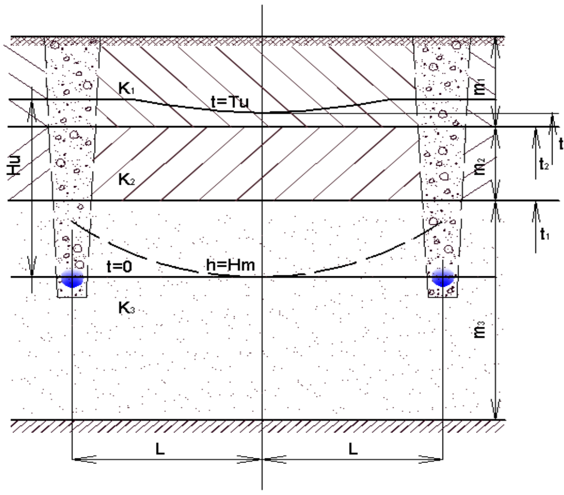


Fig. 1. Estimated filtering scheme

After the passive reduction of the GWL to the maximum depth h_m , the drainage system is switched to the humidification mode (Fig. 1). Water for humidification is supplied to the main and humidification network until a certain moment of time t=T$_u$, which must be calculated until the GWL in the middle between the drains reaches the mark $h = h_n$. In fact, the entire period of conditional release can be divided into three stages. During the first, preparatory stage, which is short in time (lasts, as a rule, several hours), the leading and regulating network is filled with water until the appropriate pressure H$_u$ is established above the drain.

In the second stage, the level above the drain rises almost to a mark equal to the pressure in the H$_u$ drain. At the same time, the pitted zone is saturated.

The third stage, the main one, is characterized by an intensive rise of GWL between the drains. As the results of long-term field observations in research and production conditions showed, its duration significantly exceeds the duration of the first two stages. In this regard, we assume that the level above the drain is established instantly. For the most general solution, we also assume that at the beginning of wetting, the GWL in the middle between the drains is in the third, lower layer, and the level of drainage backfill H$_u$=const is in the upper, first layer from the surface of the layer. The complexity of building a mathematical model of filtration in a three-layer is explained by the fact that when the GWL rises (decreases), the surface of the depression curve can be located in different layers, that is, the layer limits can be crossed at the interstices. Therefore, in the case of a three-layer base, for an approximate description of the change in GWL, we use the following level of the system, recorded with sufficient depth (Fig. 2):

$$\begin{cases} \mu_1 \frac{dh_1}{dt} = \frac{d^2V_1}{dx^2}; \\ \mu_2 \frac{dh_2}{dt} = \frac{d^2V_2}{dx^2}; \\ \mu_3 \frac{dh_3}{dt} = \frac{d^2V_3}{dx^2}, \end{cases} \tag{1}$$

where

$$V_1 = \frac{k_{\text{в}1}}{2} h_1^2;$$

$$V_2 = \frac{k_{\text{в}2}}{2} h_2^2 + \alpha_2 h_2;$$

$$V_3 = \frac{k_{\text{в}_1}}{2} h_3^2 + \alpha_1 h_3.$$

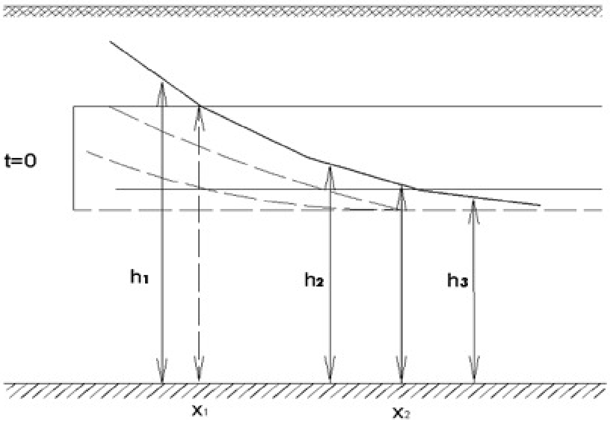


Fig. 2. Calculation scheme for determining the parameters of the humidification mode

The coefficients α_2 and α_1 (auxiliary parameters) are determined by the expressions:

$$\alpha_2 = (k_{\text{в}_3} - k_{\text{в}_2})m_3; \quad \alpha_1 = k_{\text{в}_2}m_1 + k_{\text{в}_3}m_3 - k_{\text{в}_1}(m_2 + m_3).$$

In addition, in the written system of equations (1) h_i - the level of groundwater within the i -th layer of the soil (Fig. 2), and μ_i - the coefficients of the averaged total lack of saturation of the i -th layer.

The initial and limit conditions of the problem are as follows:

$$t = 0; \; h_1^0 = H_m \tag{2}$$

$$x = L; \; \frac{dh_1}{dx} = 0 \tag{3}$$

$$x = x_1; \; k_{\text{в}_1}\frac{dh_1}{dx} = k_{\text{в}_2}\frac{dh_2}{dx}; \; h_1 = h_2 \tag{4}$$

$$x = x_2; \; k_{\text{в}_2}\frac{dh_2}{dx} = k_{\text{в}_3}\frac{dh_3}{dx}; \; h_2 = h_3 \tag{5}$$

$$x = 0;\ V_3 - 2\phi \frac{dV_3}{dx} = \frac{k_{\text{в}_1}}{2} H_u + \alpha_1 H_u \tag{6}$$

In the limit condition (6), the value ϕ is the filtration resistance of the drain. Other parameter values not given are clear from the corresponding figures.

Using the results of works [3], in the first approximation, instead of the system of equations (1), we have the following system:

$$\begin{cases} \mu_1 \dfrac{dh_1^{(0)}}{dt} = \dfrac{d^2V_1}{dx^2}; \\ \mu_2 \dfrac{dh_2^{(0)}}{dt} = \dfrac{d^2V_2}{dx^2}; \\ \mu_3 \dfrac{dh_3^{(0)}}{dt} = \dfrac{d^2V_3}{dx^2}, \end{cases} \tag{7}$$

where

$$h_i^{(0)} = h_i(L, t).$$

After the transformations of equations (1) – (6), the problem is reduced to finding the change in the groundwater level between the drains, which is the main thing in solving the given problem. The solution to this problem allows establishing the dynamics of the GWL in the middle between the drain lines within each s -th layer:

$$t = \mu_3(L^2 + 4L\phi)R_s(h_1^{(0)}) \tag{8}$$

The variable parameter R_s is generally determined by the dependence:

$$\begin{cases} \frac{1}{\sqrt{\delta_s}}\left(\varphi_s - arctg\frac{k_{\text{в}_s}h_1^{(0)}+\alpha_s}{\sqrt{\delta_s}}\right) \\ at\ \delta_s = -k_{\text{в}_3}V_3 - \alpha_s^2 > 0; \\ \frac{1}{k_{\text{в}_s}h_1^{(0)}+\alpha_s} - \frac{1}{k_{\text{в}_s}m+\alpha_s} \\ at\ \delta_s = 0; \\ \frac{1}{2\sqrt{-\delta_s}}\ln\left(\Psi_s\frac{\alpha_s+\sqrt{-\delta_s}+k_{\text{в}_s}h_1^{(0)}}{\alpha_s-\sqrt{-\delta_s}+k_{\text{в}_s}h_1^{(0)}}\right. at\ \delta_s < 0, \end{cases} \tag{9}$$

where

$V_3 = k_{\text{в}_1}H_u^2 + 2\alpha_1 H_u + \sum_{i=s}^{2}(k_{\text{в}_{i-1}} - k_{\text{в}_i})\mu_i^2$ - the layer with the number s will be completely saturated at the moment t^* of contact of the free surface between the drains on the limit s and $s-1$ the layers. The named point in time is defined by the expression:

$$t^* = \mu_s(L^2 + 4L\phi)R_s(\mu_s), \tag{10}$$

where $R_s(\mu_s)$ is determined by dependence (9) at :

$$h_1^{(0)} = \mu_2.$$

In the case of a three-layer layer, the time for the groundwater level to rise to the border of the second and third layers is determined by the dependence:

$$t_1 = \mu_3(L^2 + 4L\phi)R_1, \tag{11}$$

where $R_1 = \frac{1}{2\omega_1} ln \frac{k_{\text{в}_3}m-\omega_1}{k_{\text{в}_3}m+\omega_1} \cdot \frac{k_{\text{в}_3}m+\omega_1}{k_{\text{в}_3}m-\omega_1};$

$$\omega_1 = \sqrt{k_{\text{в}_3}\begin{bmatrix} k_{\text{в}_1}H_u^2 + 2\alpha_1 H_u + \left(k_{\text{в}_1} - k_{\text{в}_2}\right) * \\ * \ (m_2 + m_3)^2 + (k_{\text{в}_2} - k_{\text{в}_3})m_3^2 \end{bmatrix}};$$

$$\alpha_1 = k_{\text{в}_2}m_2 + k_{\text{в}_3}m_3 - k_{\text{в}_1}(m_2 + m_3);$$

$$\begin{aligned} &\mu_3 \\ &= \xi_1\frac{(\mu - m)^{\xi_1+1} - (m_2 + m_3 - m)^{\xi_1+1} - (m_1 + m_2)^{\xi_1+1} + m_2^{\xi_1+1}}{m_3 - m} \\ &+ \\ &\quad +\xi_2\frac{(m_2 + m_3 - m)^{\xi_2+1} - (m_3 - m)^{\xi_2+1} - m_2^{\xi_2+1}}{m_3 - m} \\ &\quad + \xi_3(m_3 - m)^{\xi_3}. \end{aligned}$$

The coefficient of the lack of saturation μ_3 when the GWL rises to the border of the second and third layers (and for further calculations) is determined by the linear expression for the current lack of saturation:

$$\mu_i = \xi_i h^{\xi}, , \tag{12}$$

where for peat soils: $\xi = 0{,}116k_{\text{в}}^{\ 0{,}375}$, $\zeta = 0{,}75$ [8]; for mineral soils: $\xi = 0{,}056k_{\text{в}}^{\ 0{,}5}$, $\zeta = 0{,}33$ [7]; for the conditions of drained and irrigated lands of Ukraine $\zeta = 0{,}5$, ξ is determined depending on the type of soil [6].

The duration of the GWL rise from the lower boundary of the layer interface (third and second) to the upper boundary of the layer interface (second and first) t_2 is determined by the expression:

$$t_2 = \mu_2(L^2 + 4L\phi)R_2, \tag{13}$$

where

$$R_2 = \frac{1}{2\omega_2}\ln\left(\frac{k_{\text{в}_2}m_3 - \omega_2 + \alpha_2}{k_{\text{в}_2}m_3 + \omega_2 + \alpha_2} \cdot \frac{k_{\text{в}_2}(m_2 + m_3) + \omega_2 + \alpha_2}{k_{\text{в}_2}(m_2 + m_3) - \omega_2 + \alpha_2}\right),$$

$$\omega_2\sqrt{k_{\text{в}_2}\beta + \alpha_2^2},$$

$$\beta = k_{\text{в}_1}H_u^2 + 2\alpha_1 H_u + (k_{\text{в}_1} - k_{\text{в}_2})(m_1 + m_2)^2,$$

$$\alpha_2 = \left(k_{\text{в}_3} - k_{\text{в}_2}\right)m_3,$$

$$\mu_2 = \xi_1\frac{(m_1 + m_2)^{\xi_1+1} - m_2^{\xi_1+1} - m_1^{\xi_1+1}}{m_2 - m_1} + \xi_2 m_2^{\xi_2}.$$

The duration of the rise of GWL in the upper layer of the soil to the mark is determined by the expression:

$$t_3 = \mu_1 (L^2 + 4L\phi) R_1, \tag{14}$$

where

$$R_1 = \frac{1}{2T_2^*} \ln\left(\frac{H_u - m_2 - m_3}{T + T_2^*} \cdot \frac{T_2 + T_2^* - k_{в_1} H_N}{H_u - m_1 - m_2 - m_3 + H_N} \right);$$

$$T = k_{в_1} m_3 + k_{в_2} m_3 + k_{в_3} m_3;$$

$$T_2^* = k_{в_1}(H_u + m_2 - m_3) + k_{в_2} m_2 + k_{в_3} m_3;$$

$$T_2 = k_{в_1} m_1 + k_{в_2} m_2 + k_{в_3} m_3;$$

$$\mu_1 = \xi_1 \frac{m_1^{\xi_1+1} - H_N^{\xi_1+1}}{m_1 - H_N}.$$

The total duration of the rise of GWL T_u is determined as the sum according to the expression:

$$T_u = t_1 + t_2 + t_3 \tag{15}$$

Thus, the obtained expressions (11) - (15) allow calculating the duration of the rise of the GWL T_u to a given mark during humidification, upon reaching which the supply of water to the conducting and regulating network is stopped. The drying-humidification network is switched to passive mode, which excludes, except for the urgent need during the period of torrential rains, the discharge of water outside the reclamation area.

According to functional characteristics, information support is divided into six blocks. In the first three blocks (meteorological, soil, biological) regulatory and reference information is formed. Filtration and hydrophysical characteristics, including total and minimum moisture capacity, which are formed in the soil block, are determined based on existing reference data and, if necessary, supplemented with field determinations. The necessary data for calculations related to the growth characteristics of specific crops and their water consumption are contained in the biological block. In particular, in this block, averaged data from the leaf index, the power of the root system of cultivated crops according to the phases of their development, and the function of the distribution of moisture absorption by the depth of the aeration zone is formed.

Operational information (the next three blocks) about the actual parameters of water regime regulation (actual GWL for the previous decade and soil moisture), plant growth parameters, and current meteorological parameters are formed based on the results of monitoring the production process on the reclaimed field.

Implementation of the obtained results of theoretical studies of the experimental site of the "Ikva" polder system

Analyzing the weather conditions of the growing season of 2022, it is possible to note a fairly uneven distribution of precipitation, the presence of long periods without rain with extreme values of temperature and air humidity deficit, and unfavorable conditions for growing crops, which also confirm the results of observations of biometric characteristics.

To determine the optimal modes of moistening, taking into account the patterns of moisture absorption by the root system, we used data obtained in the conditions of the experimental site with the help of a field tensiometry device.

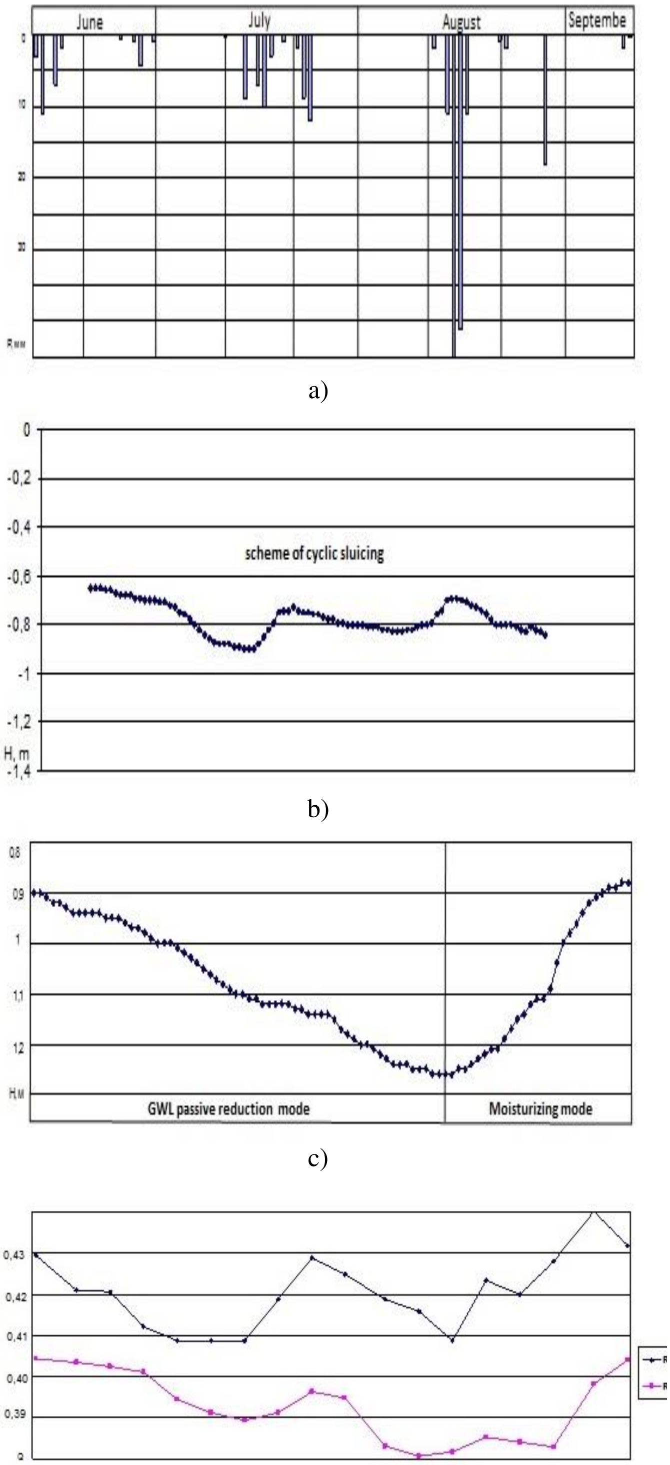
The technological parameters that characterize the process of subsoil moistening (draining) and all characteristics related to the regulation of the water regime are listed in Table 1.

TABLE I. INITIAL TECHNOLOGICAL PARAMETERS DURING THE IMPLEMENTATION OF RESOURCE-SAVING MODES OF HUMIDIFICATION IN THE EXPERIMENTAL AREA OF THE POLDER SYSTEM "IKVA"

<table>
<tr><td colspan="7">Remedial regime</td></tr>
<tr><td colspan="2">Start of passive reduction of GWL</td><td colspan="3">Transition from drying mode to humidification mode</td><td colspan="2">End of humidification mode</td></tr>
<tr><td colspan="7">Implementation date</td></tr>
<tr><td colspan="2">19.06.2022</td><td colspan="3">12.08.2022</td><td colspan="2">4.09.2022</td></tr>
<tr><td colspan="4">The duration of GWL reduction - 53 days</td><td colspan="3">The duration of hydration - 23 days</td></tr>
<tr><td colspan="7">Amount of precipitation, mm</td></tr>
<tr><td colspan="4">73.1</td><td colspan="3">117.9</td></tr>
<tr><td colspan="7">Groundwater level, m</td></tr>
<tr><td colspan="2">Initial</td><td colspan="3">Maximum</td><td colspan="2">Final</td></tr>
<tr><td colspan="2">0.94</td><td colspan="3">1.26</td><td colspan="2">0.91</td></tr>
<tr><td colspan="7">Soil moisture, in parts by volume</td></tr>
<tr><td>In the layer</td><td colspan="2">19.06.2022</td><td colspan="2">12.08.2022</td><td colspan="2">4.09.2022</td></tr>
<tr><td>0 – 0.1 м</td><td colspan="2">0.43</td><td colspan="2">0.41</td><td colspan="2">0.43</td></tr>
<tr><td>0.1 – 0.2</td><td colspan="2">0.44</td><td colspan="2">0.41</td><td colspan="2">0.42</td></tr>
<tr><td>0.2 – 0.3</td><td colspan="2">0.40</td><td colspan="2">0.38</td><td colspan="2">0.40</td></tr>
<tr><td>0.3 – 0.4</td><td colspan="2">0.40</td><td colspan="2">0.38</td><td colspan="2">0.40</td></tr>
<tr><td>0.4 – 0.5</td><td colspan="2">0.40</td><td colspan="2">0.38</td><td colspan="2">0.41</td></tr>
<tr><td>0.5 – 0.6</td><td colspan="2">0.41</td><td colspan="2">0.40</td><td colspan="2">0.41</td></tr>
<tr><td colspan="7">Soil moisture in the zone of maximum moisture absorption by the root system, in fractions of the volume</td></tr>
<tr><td colspan="2">0.42</td><td colspan="3">0.40</td><td colspan="2">0.42</td></tr>
<tr><td colspan="7">Leaf surface index / grass stand height, m</td></tr>
<tr><td colspan="2">0.82/0.15</td><td colspan="3">2.98/0.50</td><td colspan="2">5.4/0.65</td></tr>
<tr><td colspan="7">Average long-term experimental values of total evaporation (E, mm) for perennial grasses and atmospheric precipitation (P, mm)</td></tr>
<tr><td rowspan="3">Month</td><td colspan="6">Decade</td></tr>
<tr><td colspan="2">1</td><td colspan="2">2</td><td colspan="2">3</td></tr>
<tr><td>E</td><td>P</td><td>E</td><td>P</td><td>E</td><td>P</td></tr>
<tr><td>June</td><td>28.5</td><td>79.5</td><td>25.9</td><td>20.0</td><td>18.1</td><td>7.0</td></tr>
<tr><td>July</td><td>24.5</td><td>0.30</td><td>36.2</td><td>32.0</td><td>45.7</td><td>20.8</td></tr>
<tr><td>August</td><td>27.6</td><td>2.00</td><td>15.6</td><td>108.9</td><td>33.7</td><td>20.0</td></tr>
<tr><td>September</td><td>19.9</td><td>2.30</td><td>19.4</td><td>35.0</td><td>15.5</td><td>13.0</td></tr>
</table>

Perennial grasses were grown on the experimental site of the "Ikva" polder system. The first cut was made on June 18. The height of the grasses on the first slope was 0.66 m, and the leaf index was 3.66. As can be seen from the table, the growth dynamics of perennial grasses of the second slope are characterized by the following indicators of the leaf index:

0.82 – for the first decade of growth; 2.98 – at the end of the cycle of passive reduction of GWL and 5.4 – at the moment of the second slope. In 2012, the herb yield of the second cutting was higher compared to the first cutting (18.5 and 27.5 t/ha, respectively).



a)

b)

c)

d)

Fig. 3. Rainfall dynamics (a); groundwater level in the case of the cyclic sluicing scheme and the implementation of the developed regimes (b, c); average soil moisture by pentads, in fractions of the soil volume (Row 1 – in the layer of maximum moisture absorption (0–0.2 m), Row 2 – in the layer 0.3–0.6 m) in the growing season of 2012, "Ikva" polder system (d).

The initial GWL (for the period of approbation) was at a depth of 0.7 m, and the soil moisture in the zone of the most intense absorption of moisture by the root system (at a depth of 0.15–0.25 m) was 0.42 (in fractions of the soil volume).

The period of field research on the dynamics of GWL and soil moisture is divided into two stages. The first stage corresponds to the cycle of passive reduction of GWL, when due to a small amount of precipitation for 53 days, as well as evaporation and transpiration of cultivated perennial grasses, GWL decreases to the maximum possible - 1.26 m (Fig. 3). This period continued until August 12.

The end of the cycle of passive reduction of GWL under the action of evaporation and transpiration and the beginning of moistening was determined by soil moisture in the zone of its maximum absorption by the root system of perennial grasses at a depth of 0.15 - 0.25 m. When the moisture in a given soil layer decreases to the lower limit of its optimal value, which is established according to the recommendations [8], the first cycle (passive reduction of GWL ended and the drainage system was switched to the humidification mode. The value of groundwater pressure, respectively, was 3.23 kPa, which corresponds to the lowest soil moisture content of 0.38 (in parts of the soil volume). The position of the 1.26 m water well corresponds to the mentioned moment. According to traditional technology, the water wells were maintained at a depth of 0.7 - 0.9m.

The dynamics of the average humidity by pentads in the layer 0 - 0.2 m (Row 1), where the maximum absorption of moisture by the roots of cultivated perennial grasses is observed, and in the layer, 0.3 - 0.6 m (Row 2) during the growing season of 2022 is given in Fig. 3. Obtained plots of moisture distribution in the root zone during the experiment (Fig. 3).

The results of the analysis showed that during the implementation of the developed moistening regimes on the experimental site of the Ikva polder system throughout the cycle of passive reduction of GWL, the soil moisture in the root layer was within the optimal range. At the same time, the criterion of a sufficient supply of moisture was the condition of its maintenance within the necessary limits in the zone of maximum absorption of moisture by the roots of perennial grasses. The maximum efficiency of this technological scheme of water regulation was achieved precisely in the mode of passive reduction of GWL, when an accumulative capacity was formed in the upper layers of the soil to retain moisture from precipitation.

On August 12 (Fig. 4), the second cycle of implementation of the developed water regulation regimes began, when the drainage system was transferred to the humidification mode by supplying water to the collector and drainage network to ensure the necessary pressure. A significant amount of precipitation (117.9 mm) falls during this period, which is 30% of the total amount of precipitation during the growing season.

In the process of moistening, the soil moisture gradually increased. During the periods of rainfall, a sharp increase was observed, especially rapidly in the upper layer (0-0.20 cm). Atmospheric precipitation contributed to the acceleration of soil moisture equalization to its initial values in the depth of the aeration zone.

The analysis of field studies of the dynamics of soil moisture and GWL in the experimental area during the implementation of the developed water regulation regimes showed that even in the presence of a long dry period, which occurred during a passive decrease of GWL, soil moisture in the zone of its maximum absorption and in general in the active layer was maintained in the recommended range is long enough (53 days) even with a relatively small amount of precipitation (73.1 mm). Accumulated precipitation in the active layer (0-0.6 m) of the soil was used as efficiently as possible.

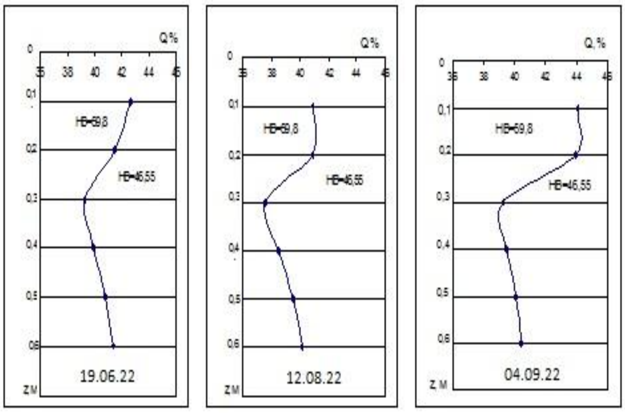


Fig.4. Plots of moisture distribution in the aeration zone during water regulation under experimental conditions at the experimental site of the “Ikva” polder system.

CONCLUSIONS

A mathematical model of the dynamics of GWL in the area between the drains during pressure regulation in the drains in the conditions of a three-layer soil structure is proposed and implemented. Having the connection between GWL and humidity in the aeration zone established based on the conducted experiments, the issue of ensuring the necessary humidity in the aeration zone within the root system is resolved.

As a result of the regulation of GWL in different modes (passive reduction and humidification) taking into account natural conditions, in particular, based on the received database on the amount of precipitation, the necessary parameters were obtained that characterize the water regime in the aeration zone. The analysis of the obtained results allows for establishing and proposing more effective resource-saving modes of moistening under the condition of a sufficient supply of moisture to the root layer. In the conducted experiments, the accumulated precipitation in the active layer (0-0.6 m) of the soil in the mode of passive reduction of GWL, when an accumulative capacity for moisture retention is formed in the upper layers of the soil, was used as efficiently as possible.

REFERENCES

[1] Alfonso, L., Lobbrecht, A., Price, R. (2010) Optimization of water level monitoring network in polder systems using information theory. *Water Resources Research*, 46 (12), p. 1–13.

[2] Basharin D, Polonsky A, Stankunavichus G. (2016) Projected precipitation and air temperature over Europe using a performance-based selection method of CMIP5 GCMs. *Journal of Water and Climate Change*. 2016;7(1), p.103–113.

[3] Dolid M.A. (1990) Optimal length conducting network of polder systems of Polissya of Ukraine. *Hydromelioration and hydro technical construction*: Lviv , p. 17-27.

[4] Kuzmych L., Voropay G., Moleshcha N., Babitska O. (2021): Improving water supply capacity of drainage systems at humid areas in the changing climate. *Archives of Hydro-Engineering and Environmental Mechanics*. Vol. 68. No. 1: 29–40.

[5] Kuzmych,L., Furmanets,O., Usatyi,S., Kozytskyi,O., Mozol,N., Kuzmych,A.,Polishchuk,V. & Voropai,H.(2022).Water Supply of the Ukrainian Polesie Ecoregion Drained Areas in Modern Anthropogenic Climate Changes. *Archives of Hydro-Engineering and Environmental Mechanics*,69(1) 79-96. https://doi.org/10.2478/heem-2022-0006

[6] Schultz Bart (2008) Water management and flood protection of the polders in the Netherlands under the impact of climate change and man-induced changes in land use. *Journal of water and land development*, No. 12, p.71–94.

[7] Shang, S.H. (2014) A general multi-objective programming model for minimum ecological flow or water level of inland water bodies. *Journal of Arid Land*, 7 (2), p. 166-176.

[8] Su, X., Chiang, P., Pan, S., Chen, G., Tao, Y., Wu, G., Wang, F., Cao, W. (2019) Systematic approach to evaluating environmental and ecological technologies for wastewater treatment. *Chemosphere*, 218, p. 778-792.

[9] Van Overloop, P.J. (2006) Drainage control in water management of polders in the Netherlands. *Irrigation and Drainage Systems*, 20 (1), p. 99-109.

[10] Andrić I., Koc M. and Al-Ghamdi S. G. 2019 A review of climate change implications for built environment: Impacts, mitigation measures and associated challenges in developed and developing countries *J. Clean. Prod.* 211 83-102

[11] Korobiichuk I., Kuzmych L., Kvasnikov V., 2019. The system of the assessment of a residual resource of complex technical structures, MECHATRONICS 2019: *Recent Advances Towards Industry* 4.0, 350–357. https://doi.org/10.1007/978-3-030-29993-4−43.

[12] Ahmed M. R., Hassan Q. K., Abdollahi M. and Gupta A. 2020 Processing of near real time land surface temperature and its application in forecasting forest fire danger conditions *Sensors* 20 984

[13] Kuzmych L. (2016). Currentr trends in creating water systems for measuring of mechanical quantities. *Collection of scientific works of the OSATRQ*. No 1(8). P. 95-99 DOI: https://doi.org/10.32684/2412-5288-2016-1-8-95-99

[14] Rózsás Á., Kovács N., Gergely Vigh L. and Sýkora M. (2016). Climate change effects on structural reliability in the Carpathian Region *Q. J. Hungarian Meteorol. Serv.* 120 103-25

[15] Chen K., Blong R. and Jacobson C. (2003). Towards an integrated approach to natural hazards risk assessment using GIS: With reference to bushfires *Environ. Manage.* 31 546-60

[16] Rokochinskiy, A., Kuzmych, L., Volk, P. (Eds.). (2023). *Handbook of Research on Improving the Natural and Ecological Conditions of the Polesie Zone*. IGI Global. https://doi.org/10.4018/978-1-6684-8248-3

[17] Rokochinskiy A., Volk P., Kuzmych L., Turcheniuk V., Volk L. and Dudnik A. [2019]. "Mathematical Model of Meteorological Software for Systematic Flood Control in the Carpathian Region*," 2019 IEEE International Conference on Advanced Trends in Information Theory (ATIT),* pp. 143-148, doi: 10.1109/ATIT49449.2019.9030455

[18] Yakymchuk A., Kuzmych L., Skrypchuk P., Kister A., Khumarova N., Yakymchuk Y.. (2022). Monitoring in Ensuring Natural Capital Risk Management: System of Indicators of Socio-Ecological and Economic Security. *16th International Conference Monitoring of Geological Processes and Ecological Condition of the Environment*, Nov 2022, Volume 2022, p.1 – 5. DOI: https://doi.org/10.3997/2214-4609.2022580047

[19] Kuzmych L., Voropai H. (2023). Enviornmentally Safe and resource-saving water regulation technologies on drained lands. *Handbook of Research on Improving the Natural and Ecological Conditions of the Polesie Zone*. IGI Global of Timely Knowledge. Hershey, Pennsylvania 17033-1240, USA. 2023. P. 75-96. DOI: 10.4018/978-1-6684-8248-3.ch005

[20] Kuzmych L, Yakymchuk A (2022). Environmental sustainability: economical and organizational aspects of WEF Nexus. *16th International Conference Monitoring of Geological Processes and Ecological Condition of the Environment*, Nov 2022, p.1 – 5. DOI: https://doi.org/10.3997/2214-4609.2022580009

[21] Prykhodko N., Koptyuk R., Kuzmych L., Kuzmych A. (2023). Formation and predictive assessment of drained lands water regime of Ukraine Polesie Zone. *Handbook of Research on Improving the Natural and Ecological Conditions*. IGI Global of Timely Knowledge. Hershey, Pennsylvania 17033-1240, USA. 2023.– p.51-74. DOI: 10.4018/978-1-6684-8248-3.ch004

DOI: 10.15439/2023F9346
ISSN 2300-5963 ACSIS, Vol. 37

Research Paper Blockchain (RPB) : A Blockchain for Checking Previously Published and Concurrently Submitted Research Papers

Nicholas Kniveton and Navid Shaghaghi
Ethical, Pragmatic, and Intelligent Computing (EPIC) Research Laboratory
Department of Computer Science and Engineering (CSEN)
Santa Clara University (SCU), Santa Clara, California, USA
{nkniveton, nshaghaghi}@scu.edu

***Abstract*—Thousands of technical conferences and peer reviewed journals around the world, each annually solicit hundreds of newly authored papers for indexing and publishing. Since no central indexing and publishing authority exists, each conference and journal maintains its own database of papers. It is thus relatively easy for authors to submit their papers to more than one conference or journal simultaneously despite being strictly prohibited by the various indexing authorities and publishers. A manual or even automated check of the hundreds of databases is unrealistic.**

Blockchain technology, however, provides a viable solution to this long-standing problem. This paper explores a potential implementation of a Research Paper Blockchain (RPB) which stores encrypted copies of all papers submitted for publication to participating indexing and publishing authorities. Conference and journal publication chairs would attempt to add the submitted paper as a new block to RPB (uploaded through a graphical web front-end or an automated back-end interface) and thus check for the uniqueness of the submission. Based on the uniqueness, the system would either add the paper to or reject it from RPB and report a uniqueness score for the paper back to the publisher. If a paper was added, it would be time stamped as well as stamped with a percentage of uniqueness for future reference. Publishers could set their own uniqueness threshold for acceptance and thus guarantee the originality and freshness of a new paper submitted before wasting reviewer time and accepting that paper.

***Index Terms*—Blockchain, Double Publication, Double Submission, Originality, Plagiarism Detection, Research Paper**

I. Introduction

THOUSANDS of technical conferences and peer reviewed journals around the world each annually solicit hundreds of newly authored papers for indexing and publishing. Since no central indexing and publishing authority exists, each conference and journal maintains their own database of papers. It is thus relatively easy for authors to submit their papers to more than one conference or journal simultaneously even though this is strictly prohibited by the various indexing authorities and publishers. A manual or even automated check of the hundreds of databases is unrealistic. And since the authors can simply pull their submissions from all the other conferences and journals after selecting the most highly ranked conference or journal that has accepted their paper, the authors are able to cover their tracks with no one ever knowing that they had double submitted their paper and thus wasted hours of reviewers' time from each of the conferences or journals.

Furthermore, cases of plagiarism where an author has simply copied another already published paper to some unacceptable percentage can go undetected until someone finds both publications and cares enough to alert the publishers. Therefore, plagiarizers can easily get away with publishing plagiarized versions of existing papers in a vastly different, lesser known/read, or different language conference proceeding or journal with little chance of anyone running into both the original and plagiarized versions of the same paper.

This paper presents a viable remedy to these long-standing problems using blockchain technology. Section II will delineate a typical scenario where this issue arises. Section III will go over some background information about blockchain technology and the different ways in which they are validated and section IV covers the need for research paper validation prior to publication. section V will cover the design of a Research Paper Blockchain (RPB) under research and development at Santa Clara University's Ethical, Pragmatic, and Intelligent Computing (EPIC) laboratory and section VI will delineate RPB's implementation details. Lastly in sections VII and VIII, the current ongoing research and development of RPB, and some results are reported on respectively.

II. A case study

Benjamin Carlisle describes an incident in which a blog post he authored was plagiarized and submitted to a research conference [1]. In particular, the article highlights the need for an automated system of checking with set protocols. When Benjamin identified the plagiarized content, the process to have the published work revoked was challenging. The plagiarized paper went through multiple human checks, but none of the reviewers recognized that much of the paper was not original.

Furthermore, there were multiple issues once the author requested the paper to be removed. Reviewers at the journal initially dismissed the claims of plagiarism as that would tarnish the reputation of the authors. Furthermore, the journal was worried that retracting an article would be embarrassing to the journal, showing that their review process was inadequate.

These reasons show the need for a pre-publication verification algorithm that follows strict rules in order to eliminate the human factor in the decision to publish or retract an article for reasons other than scientific merit. Furthermore, by using such pre-publication verification technology many of the instances of article retractions would be eliminated as articles would be reviewed prior to publication, and plagiarized articles would never be published in the first place.

III. The Need for Paper Verification

There are three important checks for paper verification: plagiarism, double submission, and duplicate publication.

A. Plagiarism

Plagiarism is defined as the representation or wrongful appropriation of another author's ideas, language, or work as one's own original work. With the plethora of online tools and software available today the checking of new work against already published work is often straightforward. However, due to the vast number of publications and databases which span almost all existing countries and languages, it is impossible to check each and every one of them before accepting a new contribution.

This can happen in both directions. It is possible that a researcher, having read a paper in a well-known journal or conference proceeding then replicates the paper and publishes it in an obscure journal in a different language and country which has loose or no copyright laws; and it is equally as likely that a researcher having read a paper published in an obscure journal or conference proceeding in some other country and language reproduces the paper and publishes it in a reputable journal and thus gains major recognition for work they did not originate and thus becomes a leader in that field. In the first scenario, the author(s) and/or publisher of the original work may never find the plagiarized version of their work, not have a method of recourse due to the lack of copyright laws in the country of the plagiarizer, or simply not care if the paper is plagiarized in such an obscure journal that results in no real loss in citations and readership. In the second scenario however, the original author(s) and/or publishers may not have the power to pursue any actions due to the wealth and strength of plagiarizers or the limitations faced by their country of origin due to international pressures such as sanctions just as an example.

Furthermore, how can publishers check work against papers which were never published? For instance, imagine a scenario in which an author submitted a paper for publication to a conference yet after the paper's acceptance, the author never registered for the conference and thus the paper was never actually published. What if then another individual, say a colleague or patent agent, having seen the work, reproduces the work and submits it to a different conference in an attempt to publish the work as his own original work? How would the second publisher be able to verify the originality of the work since no existing tool can search the space of never published work?

B. Double Submission

After plagiarism, double submission is the biggest sin an author can commit. Double submission occurs when an author submits their paper to multiple conferences or journals simultaneously. Double submission is distinct from duplicate publication: while double submission can lead to duplicate publication, it often does not, making it nearly impossible to detect using existing tools. Double submission is notoriously difficult to catch as authors are often able to pull their work from conferences. A typical case might go as follows: an author writes a research paper, then submits it to multiple conferences or journals with relatively close submission deadlines. When the author receives a notification of acceptance from one of the conferences or journals, the author pulls the paper from the remaining conferences. Or, the author waits till all the acceptance/rejection notifications come in and then picks the most prestigious conference/journal to send the camera-ready version of the paper to and pulls the submission from all the other conferences/journals. This is not a case of plagiarism as the author is not reproducing someone else's work as their own but rather trying to increase the odds of having their paper published even if it is rejected by one or more publishers. Furthermore, since journal reviews often take way too long, as documented and rejected in [2], [3], and [4] for instance, it has been time and time again proposed that authors be allowed to simultaneously submit their papers to multiple journals.

Even though this may seem to be a good strategy it is never the less problematic as it is unethical and can lead to legal disputes. It is unethical for two reasons: First, each paper undergoes reviews by several peers. Those Peers are faculty and researchers who are taking time out of their own work and research in order to serve as reviewers. Instead of spending time reviewing duplicate work, these peers could be reviewing unique works produced by others that may be worthy of publication. Second, the editors of the particular conference proceedings or journal issue will most likely be tailoring and balancing the publication with all the papers which have been accepted by the reviewers. If an author's paper is accepted in multiple conferences, the author will pull the paper from all but one conference. Pulling a paper from a conference in the last minute has the potential to throw off the carefully crafted balance created by reviewers and conference proceedings editors, harming other authors and the conference as a whole.

Double submission can also lead to potential legal disputes if the author is unable to pull their accepted work, turning the scenario into a case of duplicate publication explored below.

C. Duplicate Publication

Duplicate publication can occur with or without double submission. In the case of double submission, an author submits work that has already been published to a second publisher. When an author's paper is published, the author typically transfers an exclusive copyright of the paper to the publisher [5]. Since only one publisher can own the paper, the submission of a paper to multiple publishers could result in multiple exclusive owners of the work, creating a legal battle as to who is the true owner. Even though electronic publication has enhanced the chances of duplicate publication being detected, the tools and methods are no further reliable than plagiarism detection tools and methods. Mainly meaning that it is detectable after the fact or at best if one copy is already fully published and available before the second copy is submitted for publication. No detection mechanism exists for if both copies are under publication at the same time.

But, it should also be noted that duplicate publication is not always a problem and hence its detection alone is not enough. For example as noted by Janie Morse, editor of one of Sage publications' journals, some exceptions are the publication of a translation of an already published article into another language, a republishing of an article in an anniversary issue, or an invited republication of a particularly meritorious article in a book or special collection provided the author does not fail to provide, or the new publisher does not fail to obtain, copyright release from whoever holds the copyright - which usually is the publisher of the original article, and that the article is published with appropriate acknowledgement to the original source [6].

IV. Other Proposed Solutions

Even though various software solutions (such as Turnitin [7], Unicheck [8], Grammarly [9], etc.) for detecting plagiarism and double publication are utilized, unfortunately they are only applicable after the fact. No preventative technological solutions for those issues as well as the nefarious act of double submission have been proposed except for [10] in 2018 which the authors do also note as such, as part of the scarce results of their literature review. Their proposed solution relies on a central system with Application Programming Interface (API) access to all participating publishers' editorial systems and the sharing of the attributes - such as authors' names and email addresses, abstract, etc. - that are collected by the editorial system during the manuscript submission.

An obvious issue with that proposed system, which is not missed by the authors, is that it would require infrastructure not currently in existence. It would be a very difficult task to require publishers to create APIs for use with the system, let alone the fact that the data they would be sharing through their APIs would have Personally Identifiable Information (PII) that cannot be obfuscation for obvious reasons. Therefore, such a system would require an enormous security consideration and infrastructure. The labor, infrastructure, and maintenance costs of which would far outweigh the potential benefits of such a system.

Hence, the current best strategies in use so far have been to hold training sessions for graduate students and other researchers early in their career [2], for scientists to employ conscious efforts to ensure that plagiarism does not creep into any scientific work of theirs [11], adding measures to punish redundant publication and duplicate submission in author guidelines [12], and for conference submission systems to require authors to confirm that their submission conforms to the conference and society rule for double submission and plagiarism [2].

Any system with real potential, would have to work without exposing PII and be operable without the need for new infrastructure created by the publishers. The use of hashing and a shared blockchain with the ability for manual submission by journal editors and/or conference Technical Program Committee (TPC) members edges the development of such preventative technology closer to reality.

V. Technical Background

A blockchain is a decentralized, peer to peer method of data storage that is highly resistant to modification. Users upload their data into a system where participants of the peer to peer blockchain network compile the data into a block and then add it to a chain of blocks. [13] The main methodology for validating submitted blocks is the Proof of Work protocol.

In a Proof of Work system, each time a block is to be added to the chain, block validators, known as "miners", compete with each other to solve cryptography problems that are very difficult to solve but have solutions that are very easy to verify. [14]. Solving each problem requires a large amount of computational power, known as "work". Whoever solves the problem first is allowed to place the new block on the chain and is thus rewarded for their efforts with new cryptocurrency coins, or fractions thereof, based on the difficulty of the work. Since anyone can easily verify the miner's solution, users of the blockchain can consider the new block to be valid and thus the data within it to be trustworthy due to the work that was required to add the block.

This allows users of the blockchain network to know that their data is stored both securely and permanently without the need for a centralized arbiter of trust such as a bank, data warehouse company such as Google, or government agency [15].

RPB uses a Proof of Work protocol due to its known effectiveness and security.

VI. Design

RPB consists of two main parts: a blockchain that stores the data and a verification algorithm that analyzes the data.

A. Blockchain

RPB uses off the shelf, existing blockchain technology that has been proven secure and reliable. RPB's proof of concept uses the bitcoin blockchain, a well known system that utilizes a proof of work consensus mechanism [16], for its simplicity and security. However, for the working product, a more elastic, scalable, and efficient solution was needed. The security and reliability provided by proof of work based blockchains comes at a cost of high energy usage and a lack of elasticity. Proof of work blockchains have only one method of verification and it is very challenging to make changes.

Changes that seem minor often require what is known as a ”fork”, where a completely new blockchain path is created from the old chain. Due to the changing nature of conferences and publications, creating a fork every time a minor change is needed is not a viable option. In order to allow for elasticity and scalability, RPB needed a different consensus mechanism. Mechanisms designed for enterprises meet these requirements but they are typically complex and difficult to implement. One such mechanism is the Ethereum based Quorum Enterprise Blockchain Client [17]. Quorum allows for both elasticity and scalability that RPB needs, however, implementing a Quorum based blockchain solely for RPB is not feasible. To design the RPB, a commercial blockchain service that uses Quorum was chosen. Microsoft Azure Blockchain Service met the requirements for RPB, and a Quorum based blockchain using the RAFT consensus protocol was created. Using Azure allows RPB to make changes as needed and scale the product without sacrificing its efficiency. The Control of RPB's blockchain, should it be adopted by publishers, would be through a consortium of research conferences and journal publishers, such as the Association for Computing Machinery (ACM), European Alliance for Innovation (EAI), International Academy, Research, and Industry Association (IARIA), Institute of Electrical and Electronics Engineers (IEEE), Springer, etc. for computer science and engineering manuscripts for example.

B. *Verification Algorithm*

The second component of the RPB, the verification process, consists of the following protocol: 1. Verify that the ”transaction”, which includes the uploaded material, meets the general parameters of RPB (uploader, file type, etc). 2. Run an algorithm that compares the current paper to all others in the chain. The algorithm would iterate through all papers previously added to the chain, comparing the new paper to the old ones. Whenever a matching paper is discovered, the forger would note the transaction ID of the matching paper and the percentage of similarity. 3. Once all parameters are met and the matching algorithms are complete, the miner would compile the paper into a block and add it to the chain.

VII. Implementation

RPB is implemented in three phases: a proof of concept using an existing blockchain product for documents, a custom prototype built on a private blockchain, and a working product running on the Microsoft Azure Blockchain Service.

A. *Proof of Concept*

To show that RPB is viable, a proof of concept was needed that showed the two core parts of the system could be implemented: uploading papers to a blockchain and verifying the uniqueness of papers.

1) Blocksign: Creating a blockchain that accepts large research papers is not an easy task, so existing products were explored. A product produced by a company called Blocksign that verifies signatures on PDF documents by storing a hash of the document on a blockchain [18] was chosen. Blocksign piggy-backed their service off the pre-existing blockchain developed for Bitcoin. Using a function known as OP Return that is attached to the script for every bitcoin transaction, users can input up to 40 bytes of data that will be added to the bitcoin transaction, which will in turn be added to the blockchain. Blocksign utilizes OP Return to store the hash of a PDF document. Blocksign did not perform similarity comparisons between papers, which RPB must do, but Blocksign provides a means to upload papers to a blockchain. When a Blocksign user signs a paper, they upload it to the Blocksign website, which hashes the entire contents of the signed paper. The output hash is then added to a bitcoin transaction to be stored on the blockchain. Later on, a user can verify the document, its signatures, and the time it was signed. Since the hash was encoded on the blockchain, the user can trust the timestamp and signatures.

2) Verification Algorithm: To create RPB's proof of concept, a comparison algorithm that could run on top of Blocksign's services was created. The algorithm took the text of the documents that had been uploaded and hashed them using a SHA-1 hashing algorithm. While SHA-1 is not considered secure compared to SHA-256, for the purposes of a proof-of-concept SHA-1 is appropriate as it is simple and easy to implement. Next, the algorithm compared the hashes of the two in order to verify if the paper was unique or whether it was copied. This allows the algorithm to detect when two documents were identical, indicating plagiarism.

3) Proof of Concept Results: Using the verification algorithm comparing the hashes of two papers stored on blocksign, the proof of concept was able to successfully recognize when two papers were identical and when they were unique. However, due to the nature of hashing functions, the comparison was atomic. One slight change in the document would yield a completely different hash, meaning that a user could simply change one letter in the document to fool the system. A malicious actor for instance could submit a paper to a conference, and then change only one letter or word before submitting it to another. When the second conference uploads the paper to RPB, the verification algorithm would tell the user that there are no matches to the uploaded paper and hence accept the paper as a new block on the blockchain. For this reason, the use of the bitcoin blockchain is not ideal for the final version of our product and a more robust system needed to be developed.

B. *Prototype*

While the proof of concept showed the user whether or not two papers are identical, this information was not particularly useful because since a hashing function outputs a completely different string for each unique input, the hashes of two papers were completely different even if the papers differed by only a letter or word. Therefore, the verification algorithm would allow a malicious actor to write a paper, submit it to a conference, and then change only one letter or word before submitting it to another. When the second publisher uploads the paper, the prototype verification algorithm would tell the user that there are no matches to the uploaded paper. For this reason, the use of the bitcoin blockchain is not ideal for the final version of our product. RPB would need to use a system that goes beyond a simple hash of the complete document. To accomplish this. Thus a custom blockchain was implemented that uses blocks that have room for plain text so that the system can determine what degree of similarity the papers share rather than whether they are identical.

1) Blockchain Template: An IMB python blockchain template [19] was used to create a proof-of-work blockchain in python and then modified to store chunks of plain text inside the blocks. The blockchain did not include a cryptocurrency as that feature is not necessary for the operation of RPB. The goal of the prototype was to show that papers can be added to a blockchain and compared in a non-atomic manner. The details of creating the blockchain are unimportant to report in this paper as it was created from an existing template that uses the proof of work consensus mechanism. Proof of work blockchains are generally well understood and the security of blockchain technology has been proven through extensive research and real world use. Therefore instead, this subsection focuses on the modifications we made to the standard template to accommodate the features of RPB. As with any blockchain, the entire blocks in the RPB prototype are still hashed but RPB's custom chain allows for large amounts of data to be stored in each block. Each block has a size of 100kb, allowing for approximately 100,000 characters of text (minus the small amount used for the block's code and hash) to be added to the block. This size was chosen to allow for nearly all papers to fit on a single block, easing comparisons. This size can be manipulated based on the specific use case RPB is being used for. RPB stored the text data in JSON format, which is the format typically used for storing digital transactions. Each paper to be added to the blockchain was treated like a ”transaction” with a date, time, and uploader identification. In the ”new_transaction” section of the template, the plain text data was inserted from the uploaded paper into the blockchain block. Once inserted, each block was added to the chain as it normally would.

2) Comparison Algorithm: RPB's custom comparison algorithm reads the data inside each ”transaction”, which in this case is the text, and compares it to the data in another transaction. A python text comparison library [20] was used to compare the data in each transaction. After the comparison is run, a percentage of similarity is outputted to the user. This allows the user to decide whether or not the paper needs further investigation. RPB is intended to have a variable cut off point, as percentages of similarity may differ based on different use cases. For example, if a paper consists of large amounts of quoted and cited text, a high percentage of similarity would be expected. Another case in which two original reports are being compared might require a very low percentage of similarity. Hence, RPB allows the user to determine what is acceptable for them by adjusting the similarity level acceptable.

3) Prototype Results: Using the newly developed proof of work blockchain and verification algorithm, RPB was robustly tested. Multiple papers with varying degrees of similarity were uploaded to the prototype blockchain and then processed to reveal degrees of similarity. The prototype was tested at multiple different acceptable levels of similarity, and each time the prototype successfully accepted the papers that fell below the threshold and rejected the ones above the threshold. The results showed that RPB was a working concept and merited further development. The results also showed the challenges of using blockchain in an environment that varies consistently. To make adjustments to RPB, the source code had to be changed each time, requiring the blockchain to start over again. Restarting a local blockchain is not an issue, however, if changes needed to be made to RPB after it had been deployed to a large number of conferences and publishers, changing source code and restarting the chain would not be an option. For this reason, it became clear that RPB be best implemented using a commercial service that allows for active management of the blockchain.

C. Working Product

1) Commercial Blockchain Service Selection: After reviewing the results from the prototype and reviewing different consensus mechanisms, an enterprise block-chain service was pursued. Multiple services were reviewed, but it was decided to run RPB on Microsoft Azure's Blockchain Service. Azure allows RPB to be implemented in a manner that is scalable, secure, efficient, and easily adaptable to different organizations. Azure allows the uploading of RPB's comparison algorithm to the service and running it on top of Microsoft's Blockchain Workbench. Azure's Blockchain is an Ethereum based system that uses the RAFT consensus mechanism [21]. Through Ethereum's Raft mechanism, RPB will be able to grow and expand. By using Azure, RPB is much more scalable and will utilize a blockchain mechanism that is backed by a reputable organization.

2) Implementing RPB using Azure: After selecting Azure Blockchain Service, RPB's verification algorithm was converted into a JSON based application that could run on Azure. Azure was configured to run RPB, and then the original python code was translated as needed and uploaded into the Blockchain Workbench. The implementation process in Azure was simple and RPB was able to quickly run live.

3) Azure Results: After successfully implementing the blockchain in Azure, multiple tests were performed. Ten papers of varying degrees of similarity from 0 to 100, in increments of 10 percent, were uploaded to the chain and tested using RPB. RPB successfully identified the similarity percentages between the papers and either accepted or rejected them based on the configured threshold. The final results clearly displayed that RPB was a viable, functional product.

VIII. Work in Progress

1) A Better Verification Algorithm: While RPB's working product results showed that the Azure based app worked well, they also revealed room for improvement. The comparison algorithm used in RPB compares content in a quantitative way, disregarding the qualitative components of written works. Therefore, the algorithm has the potential to miss plagiarism that an author attempted to mask by replacing words with synonyms or rewriting the ideas of others without appropriate credit. To help solve these more complex cases, the RPB team is currently developing an algorithm that uses natural language processing to better detect plagiarism and reduce false negatives. For example, the new algorithm can recognize text in quotation marks that is cited properly, allowing this text to be removed from the tally of ”copied” or ”unoriginal” text. Furthermore, the algorithm can track groups of words and phrases, allowing for better detection of more complex cases of plagiarism.

2) Graphical User Interface: Currently, papers are uploaded to RPB through an online terminal. While this method is effective, it is not ideal for users who are unfamiliar with terminal commands. The RPB team is currently constructing a web based graphical user interface that allows conferences to more easily upload papers and access results.

By implementing these two changes, RPB’s ease of use and effectiveness will be increased.

IX. Conclusion

The implementation of the RPB proof of concept showed that RPB is able to add papers to a blockchain and verify whether the papers match. However, more importantly, the results revealed the need for the further development. While RPB’s prototype successfully revealed whether two papers had any differences, the use of a hash limited the scope of the prototype’s verification system. Since a hashing function outputs a completely different string for each unique input, the hashes of two papers were completely different even if the papers differed by only a letter or word. Therefore, the verification algorithm would allow a malicious actor to write a paper, submit it to a conference, and then change only one letter or word before submitting it to another. When the second publisher uploads the paper, the prototype verification algorithm would tell the user that there are no matches to the uploaded paper. For this reason, the use of the bitcoin blockchain is not ideal for the final version of our product.

RPB’s prototype solved the issues revealed in the proof of concept. The prototype results proved that RPB is functional and able to perform comparisons between papers that identify differences beyond a binary ”unique” or ”not unique” comparison. Files with 0, 33, 66, and 100 percent similarity were tested via the RPB prototype and the prototype performed as expected, revealing to the user the proper percentages of similarity. However, the percentage of similarity was both manipulable and had a high number of false positives. By performing superficial changes to a paper, such as replacing words with synonyms and changing the order of the text, RPB identified papers as unique, where deep down they were very similar. Furthermore, papers that had a high number of quotes were often identified as not unique, even when they had original content. These results encouraged the implementation of a more complex comparison algorithm using natural language processing for the final product.

Acknowledgment

Many thanks are due to Santa Clara university’s blockchain working group led by Professor Ahmed Amer of the Computer Science and Engineering (CSE) department for their belief in the project and help in ideation and brainstorming. Also, thanks to SCU’s Frugal Innovation Hub for their recognition of the project’s humanitarian goal and their continued support. And lastly to the department of Mathematics and Computer Science (MCS) of the Collage of Arts and Sciences as well as the department of Computer Science and Engineering (CSEN) of the school of Engineering at Santa Clara University for their support of the project.

References

[1] A. A. McCook, “Authors retract much-debated blockchain paper from f1000,” May 2017. [Online]. Available: https://retractionwatch.com/2017/05/24/authors-retract-much-debated-blockchain-paper-f1000/

[2] H. Schulzrinne, “Double submissions: Publishing misconduct or just effective dissemination?” *SIGCOMM Comput. Commun. Rev.*, vol. 39, no. 3, p. 40–42, Jun. 2009. [Online]. Available: https://doi.org/10.1145/1568613.1568622

[3] U. Cem, “Multiple submission, duplicate submission and duplicate publication,” *Balkan medical journal*, vol. 30, no. 1, p. 1, 2013.

[4] S. Pressman, “Simultaneous multiple journal submissions: The case against,” *American Journal of Economics and Sociology*, vol. 53, no. 3, pp. 316–333, 1994.

[5] P. Samuelson, “Self-plagiarism or fair use,” *Communications of the ACM*, vol. 37, no. 8, pp. 21–25, 1994.

[6] J. M. Morse, “Duplicate publication,” 2007.

[7] Turnitin, LLC, “Empower students to do their best, original work,” 2020. [Online]. Available: https://www.turnitin.com

[8] Unicheck, “Plagiarism checker that prefers results over numbers,” 2017. [Online]. Available: https://unicheck.com

[9] Grammarly Inc., “Plagiarism checker by grammarly,” 2021. [Online]. Available: https://www.grammarly.com/plagiarism-checker

[10] M. Kolhar, A. Alameen, and S. B. B. AlMudara, “A proposal to detect the double submission of a manuscript sent for review,” *Science and Engineering Ethics*, vol. 24, pp. 1315–1329, 2018.

[11] S. S. Khadilkar, “The plague of plagiarism: Prevention and cure!!!” 2018.

[12] J. Wei, “The countermeasures and thinking on redundant publication and duplicate submission,” *Journal of Liaoning Normal University (Natural Science Edition)*, no. 3, p. 26, 2012.

[13] S. Nakamoto, “Bitcoin: A peer-to-peer electronic cash system,” *Decentralized business review*, 2008.

[14] R. Wottenhofer, *Blockchain Science*, 3rd ed. Inverted Forest Publishing, 2019.

[15] T. D. Washington, *Blockchain Technology*. Content Arcade Publishing, 2019.

[16] “How does bitcoin work?” [Online]. Available: https://bitcoin.org/en/how-it-works

[17] Quorum, “Enterprise ethereum client.” [Online]. Available: http://docs.goquorum.com/en/latest/

[18] K. Cruz, “Blocksign: Signing documents on the blockchain,” 2014. [Online]. Available: https://bitcoinmagazine.com/articles/blocksign-signing-documents-on-the-blockchain-1416508388

[19] S. Kansal, “Develop a blockchain application from scratch in python,” Jan 2020. [Online]. Available: https://developer.ibm.com/technologies/blockchain/tutorials/develop-a-blockchain-application-from-scratch-in-python/

[20] “7.4. difflib - helpers for computing deltas.” [Online]. Available: https://docs.python.org/2/library/difflib.html

[21] M. Iansiti and K. R. Lakhani, “The truth about blockchain,” Aug 2019. [Online]. Available: https://hbr.org/2017/01/the-truth-about-blockchain

Communication Papers of the 18th Conference on Computer Science and Intelligence Systems pp. 201–207
DOI: 10.15439/2023F3416
ISSN 2300-5963 ACSIS, Vol. 37

Using Features of PLRs to Chromatic Light Pulse Irradiations of Either Eye to Detect Dementia in Elderly Persons

Minoru Nakayama*
0000-0001-5563-6901
School of Engineering,
Tokyo Institute of Technology,
Tokyo, Japan 152-8552
Email: nakayama@ict.e.titech.ac.jp

Wioletta Nowak and Anna Zarowska†
0000-0002-4135-2526 0000-0003-4544-9082
Biomedical Eng. and Instrumentation
Wrocław University of Science and Technology,
Wrocław, Poland 50–370
Email: wioletta.nowak, anna.zarowska@pwr.edu.pl

***Abstract*—A procedure for detecting symptoms of dementia was developed using waveform features of pupil light reflexes (PLR) of both eyes, in response to blue or red light pulses directed toward either eye. The experiment was conducted using elderly people with Alzheimer's disease, mild cognitive impairment, and a normal control group who were not patients. This paper focuses on the differences between the features of irradiated and non-irradiated eyes, and two combined metrics were produced in addition to the three factor scores in our previous work. The level of dementia was estimated using two regression functions with the extracted features. The performance of the procedure developed was evaluated using two sets of data, and its validity was confirmed.**

***Index Terms*—Pupil, Pupil Light Reflex, Alzheimer's disease, feature extraction, both eyes**

I. Introduction

One major clinical assessment for dementia is a medical diagnosis known as the Mini-Mental State Examination (MMSE). The rating classifies participants into groups with Alzheimer's Disease (AD) or mild cognitive impairment (MCI). However, as this clinical test is based on face-to-face surveillance, sufficient verbal communication is required. The authors have been trying to develop a metric using reactions to pupil light reflex (PLR) activity in order to improve the early clinical diagnostic procedure [1], [2], [3]. Approaches using PLR have been studied during previous research [4], [5], [6], [7]. In particular, PLR responses based on Melanopsin ganglion cells can be an index for the detection of dementia symptoms [8], [9], [10], [11].

Some of the previous studies have suggested that dementia may be influenced by the optic nerve which transfers light detection signals from the retina to a unit of the Edinger-Westphal nuclei in the pretectum [12], [13], as illustrated in Figure 1. In order to evaluate the effect of these functions on PLRs, the responses of both eyes were employed in the evaluation, instead of a single eye [14], [15]. The specific metrics for some of the differences of each eye should be introduced in order to evaluate the activity of these functions. Of course, a participant's age is the major factor in the diagnosis of dimentia, and along with features of PLRs is a key piece of information.

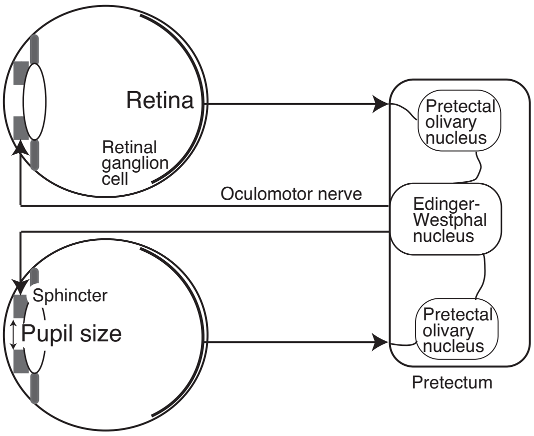


Fig. 1. PLR reaction system based on reference [16]

This paper will focus on the differences of PLR waveform features of both eyes, and their contributions to the level of dementia is evaluated, in addition to experimental factors such as the colour of the light pulses, whether the left or right side eye is irradiated, and the sequences of the light pulses.

The following topics are addressed in this paper.

1) Features of PLRs to blue and red light pulses of the irradiated and non-irradiated eyes are compared, and the differences are evaluated.
2) In order to classify the participants as AD/MCI or normal control group (NC), two types of functions predicting the probabilities of patients are developed using the extracted features of PLRs.

II. Method

PLRs of dementia (AD or MCI) patients and elderly persons not diagnosed with the disease (NC: Normal Control) were

This work was partially supported by the Japan Science and Technology Agency (JST) [JPMJTM20CQ, 2020-2022].

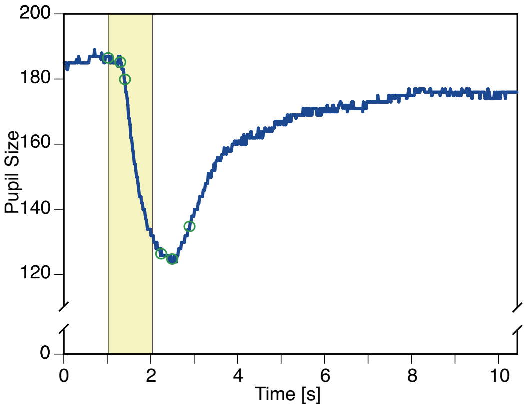


Fig. 2. Example of PLR [14], [15]

measured.

A. Measuring pupil reactions

Pupil diameters of both eyes were measured for 10 sec. in a temporal darkened space under the following conditions [14].

1) Condition 1: Observe static pupil oscillation without light pulses
2) Condition 2: Blue light pulses to the right eye
3) Condition 3: Blue light pulses to the left eye
4) Condition 4: Red light pulses to the right eye
5) Condition 5: Red light pulses to the left eye

The 1 second light pulses of blue (469nm, 14.3cd/m^2, 6.5lx) or red (625nm, 12.3cd/m^2, 10.5lx) light were irradiated on either eye using Condition 2 ∼ 5, as producing PLR waveforms as shown in the example in Figure 2. The waveform diameters were measured in pixels using a piece of specialized measuring equipment (URATANI, HITOMIRU).

B. Participants

Selected participants were clinically examined using MMSE tests, and all participants were classified into three levels with regard to score, such as AD (Alzheimer's disease, with MMSE<=23), MCI (Mild cognitive impairment, with 23<MMSE<=27) and NC (Normal control, including no MMSE scores). Two sets of measured data were prepared for the 2 different periods.

1) Data set 1: Blue light sessions such as Conditions 1 ∼ 5 were assigned first [14], [15].
   - AD: 31 (Mean age:83.0, SD:6.3).
   - MCI: 9 (Mean age:82.1, SD:6.3).
   - NC: 61 (Mean age:75.6, SD:9.2).
2) Data set 2: All participants were measured twice in two sequential sessions which were conducted beginning with Blue light sessions (Conditions 1 ∼ 5) or Red light sessions (Conditions: 1,4,5,2,3) in random combinations.
   - AD: 12 (Mean age:80.7, SD:5.5).
   - MCI: 2 (Mean age:83.5, SD:7.8).

The measurement observations were conducted by a clinical physician at two medical institutions, and the procedure was approved by an ethics committee at Osaka Kawasaki Rehabilitation University.

TABLE I
FEATURES OF PLR

Variables	Definitions
RA	Relative Amplitude of miosis
t_min	Time at minimum size
diff_min	Minimum differential of size
t_diff_min	Time at minimum differential
diff_max	Maximum differential of size
t_diff_max	Time at maximum differential
diff2_min	Minimum acceleration
t_diff2_min	Time at minimum acceleration
diff2_max	Maximum acceleration
t_diff2_max	Time at maximum acceleration

TABLE II
FACTOR LOADING MATRIX FOR PLR FEATURES [14], [15]

Variables	Factor1	Factor2	Factor3
diff_min	**0.87**	-0.13	0.09
diff2_min	**0.76**	0.06	0.16
diff2_max	**-0.83**	-0.17	0.22
diff_max	-0.36	0.08	0.15
RA	-0.24	**0.78**	-0.09
t_min	0.22	**0.73**	0.14
t_diff2_min	-0.13	-0.00	**0.49**
t_diff_min	-0.05	-0.03	0.36
t_diff_max	-0.11	0.23	0.36
t_diff2_max	0.06	0.07	0.30

Factor1: differential & acceleration
Factor2: miosis size and minimum time
Factor3: time components

C. PLR waveforms and feature extraction

Condition 1 was designed to observe the degree of pupillary oscillation while at rest. Subsequently, the frequency powers of both eyes were calculated separately [17]. The frequency powers are summarized for each eye in three frequency bands as PSD1:∼1.6Hz, PSD2:1.6∼4.0Hz, and PSD3: 0.35∼0.7Hz.

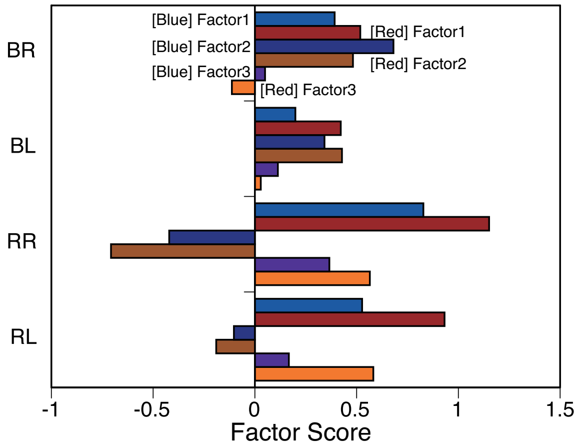


Fig. 3. Comparison of factor scores in Data set 2 in order to extract the order effect of the light pulses (Blue or Red light)

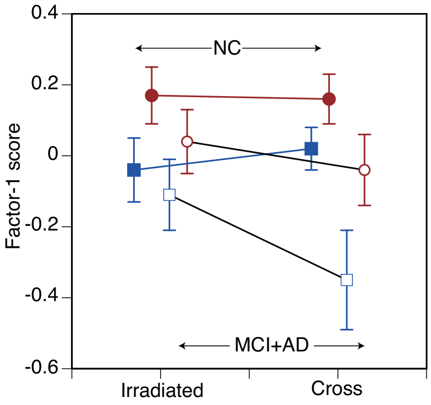


Fig. 4. Data set 1:Factor1(differential & acceleration)

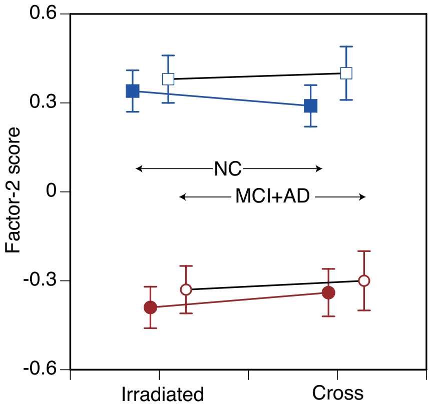


Fig. 5. Data set 1:Factor2(miosis size and the time)

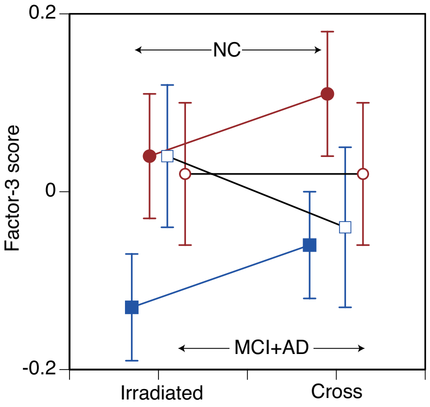


Fig. 6. Data set 1:Factor3(time components)

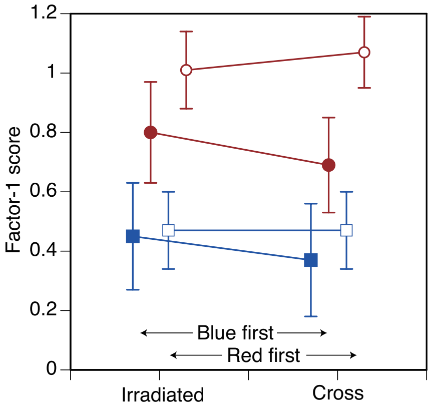


Fig. 7. Data set 2:Factor1(differential & acceleration)

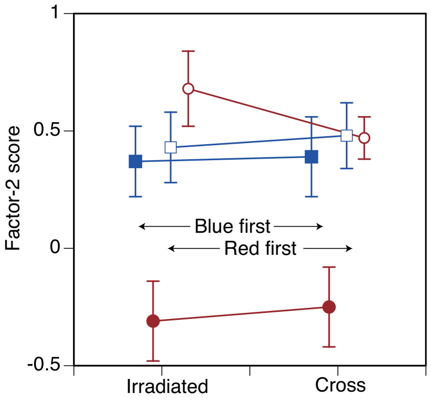


Fig. 8. Data set 2:Factor2(miosis size and the time)

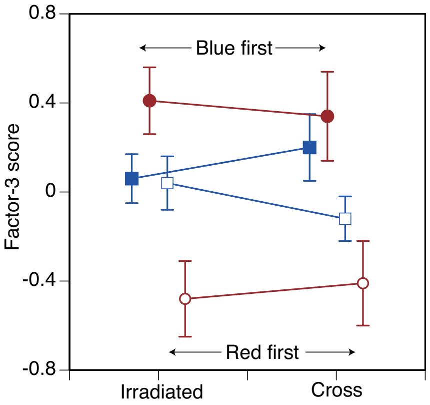


Fig. 9. Data set 2:Factor3(time components)

A typical PLR reaction is illustrated in Figure 2, and several features of waveforms were extracted, as shown in Table I, which shows variable names and their short descriptions [3], [14]. The latent factors were extracted using factor analysis, and the contribution ratio is 45.5%. The factor structure and loading matrix are shown in Table II. Their factor scores are calculated as meta features [3], [14].

III. Results

A. Order effect of coloured light pulses

During an analysis of Data set 1, an order effect of light pulse presentation was suspected since some differences were observed between the two coloured light pulse conditions. In order to examine the effect, means of factor scores in two sequences were compared using Data set 2. The means of blue first and red first sessions are summarised in Figure 3. The means of factor scores in the two sequences were compared. The differences between the two sequences were tested statistically, and the differences were not significant ($p > 0.05$). Also, there are no significant difference in factor scores of patients in the two sets of data. Patient's response data in the two data sets is comparable ($p > 0.05$). Therefore, the order of the light pulses (blue or red light first) of the two data sets is not significant. However, as there are some differences in factor scores for each eye, these contributions were analyses as follows.

B. Differences of factor scores for irradiated and non-irradiated eyes

Factor scores of features of PLRs are compared to reactions of eyes which have been irradiated (Irradiated) or have not been irradiated, such as Cross reaction with the EW nucleus. Factor scores of Data Set 1, which uses blue and red coloured light pulses (blue vs. red) as two conditions with the groups of participants (NC vs. MCI+AD), are summarized and the responses of each eye when irradiated directly or in cross condition are compared. The results are illustrated in Figures 4~6. Responses to light pulse irradiation of both eyes are summarized using the hypothesis that there are a few differences between the two eyes. For each factor score, the light pulse colour affects the differences. The factor of the light pulse colour is significant for each subject group according to the results of Two-way ANOVA, except for Factor3 in the MCI+AD group. The factor for observed eyes (Irradiated vs. Cross) is significant for Factor1 scores of the MCI+AD

TABLE III
LOGISTIC REGRESSION MODELS FOR MCI AND AD WITH NC GROUPS FOR DATA SET-1

Model	Variables of extracted features	N of variables	SEL	AUC
2n*	(R)(L)[3factor x 4 cond. + PSD1-3], age	31	31	0.95
2ns*	age,brr1-2,brl3,blr1-3,bll1-3,sPSD3,dPSD3	31	14	0.92
2a*	the same condition for 2n [AD vs. MCI+NC]	31	31	0.95
2as*	brr1-2,brl3,blr1-3,bll1,bll3,rrr2,rrl2,PSD3	31	13	0.89
3n	age,(Irradiated & Cross)*3Factors*2Colour,PSD1-3	16	16	0.83
3ns	age,IbF3,IrF1,XbF1-3,XrF3,PSD1,PSD3	16	9	0.81
3a	age,(Irradiated & Cross)*3Factors*2Colour,PSD1-3	16	16	0.80
3as	IbF3,IrF1,XbF2,XrF1,PSD2	16	5	0.77
4n	age,(Products + Rates)*3Factors*2Colour,PSD1-3	16	16	0.80
4ns	age,PbF1-2,PSD1,PSD3	16	5	0.77
4a	age,(Products + Rates)*3Factors*2Colour,PSD1-3	16	16	0.86
4as	PrF2,DbF3,DrF1,PSD2	16	4	0.77
5n	age,(Irradiated & Cross + Products + Rates)*3Factors*2Colour,PSD1-3	28	28	0.89
5ns	age,IbF2-3,IrF1,XbF1-3,XrF1,PrF1,DrF1,ff1-ff3	28	13	0.84
5a	age,(Irradiated & Cross + Products + Rates)*3Factors*2Colour,PSD1-3	28	28	0.89
5as	IbF2-3,IrF1,XrF1,PbF1,PrF2,DrF1,DrF3	28	8	0.84
6nL	age,(Irradiated & Cross + Products + Rates)*3Factors*2Colour,PSD1-3	28	28	0.89
6nLs	age,IbF1-3,IrF2,XbF2-3,PbF2,PrF2,DbF2,DrF3,PSD1-3	28	13	0.87
6nR	age,(Irradiated & Cross + Products + Rates)*3Factors*2Colour,PSD1-3	28	28	0.93
6nRs	age,IbF1-3,XbF3,XrF2,PbF2,DbF2,DrF3,PSD1,PSD3	28	11	0.88
6aL	age,(Irradiated & Cross + Products + Rates)*3Factors*2Colour,PSD1-3	28	28	0.92
6aLs	age,IbF2-3,XrF3,PrF2,PSD1	28	6	0.81
6aR	age,(Irradiated & Cross + Products + Rates)*3Factors*2Colour,PSD1-3	28	28	0.97
6aRs	age,IbF3,IrF3,XbF2-3,XrF1-2,PbF1,PrF1-3,DbF1,DrF3,PSD2	28	14	0.93

SEL: the number of selected variables; *: reported in our previous work [15]

group ($F(1, 40) = 5.64$). This means that there are significant differences in Factor1 scores for velocity and acceleration of PLRs between the two eyes. In addition, there are some differences in Cross Factor1 and Irradiated Factor3 scores for blue light pulses, while there are few differences in Factor2 scores. There are few differences in factor scores for red light pulse except for Cross-Factor3 as well.

These behavioural characteristics are confirmed in considering two types of light pulse sequences using another data Set, Data Set 2. The results are summarized in Figures 7∼9 using an identical format as in Figure 4∼6. In Data Set 2, two kinds of plots represent two light sequences since all participants are patients. In comparing the means of the two sequences, those for blue light pulses are almost similar, but there are some differences in means for red light pulses. The responses to red light pulses seem unstable. Also, the factor of the light pulse colour was significant for factor scores of the first sequence of red light pulses. The effect of irradiated eyes (Irradiated or Cross) on means of factor scores is significant for Factor1 scores of the first blue sequence, and Factor2 scores of the first red sequence. The results of Factor1 coincide with the results of Data Set 1 in Figure 4.

C. Relationship between responses of irradiated and non-irradiated eyes

Previous results suggest that the deviations in the factor scores for two eyes, when directly or cross irradiated, are relatively smaller than the deviations in factor scores with other conditions. In order to evaluate the relationships between the responses of the two eyes, some new metrics are introduced in order to classify patients and NC participants. In regards to previous studies, the relationship may represent a disorder related to PLR reactions. Here, one pair of original metrics is introduced as a result of trial and error evaluations. The pair may represent asynchronous response characteristics as a dissimilarity between the responses of the two eyes.

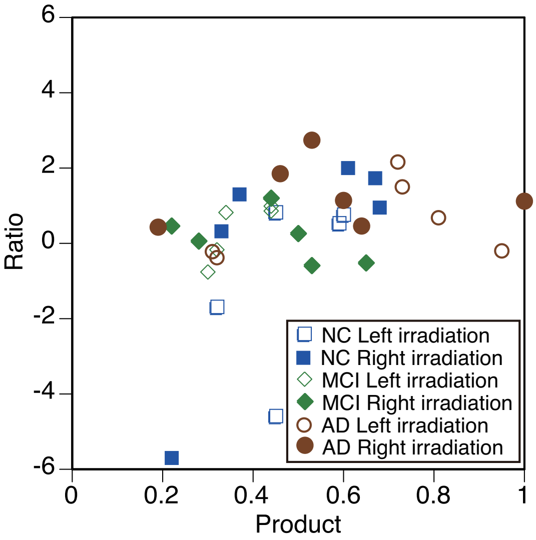


Fig. 10. Product and ratio of factor scores between irradiated and non-irradiated eyes.

$$Product_{\{b|r,F_j\}} = iFactor_{b|r}j \times xFactor_{b|r}j, (j = 1, ., 3)$$

$$Rate_{\{b|r,F_j\}} = \frac{iFactor_{b|r}j}{xFactor_{b|r}j}, (j = 1, ., 3)$$

In the equations, $iFactor_{b|r}j$ means jth ($j = 1, 2, 3$), for factor scores of eyes irradiated using blue or red light pulses. $xFactor_{b|r}j$ means jth factor scores on blue or red non-irradiated (cross reaction) eye.

These metrics are summarized in Figure 10. The horizontal axis represents the product value, and the vertical axis represents the rate value. Pairs of products and rates are calculated from 3 sets of factor scores for two colours of light pulses on either eye (2×2). They are summarized for the three levels of participants, where plots for left irradiation are illustrated as symbols with no fill and plots for right irradiation are illustrated as solid symbols. All plots deviate along both product and rate axes. Most of the AD patients show higher values for the products, and some of MCI patients show lower values for the products. Some of the NC participants deviate along the rate axis. In comparing the three scatter grams, the distributions of the two-dimensional values for the the three levels of participants are slightly different. In particular, the mean of the products for AD patients (0.61) is higher than the mean for MCI patients (0.40), and the mean for NC participants (0.48) is the middle of the two groups of patients. Though the differences are small, these positive contributions to the classification of patients will be confirmed in the next section of the paper.

D. Classification of dementia levels of patients

In order to classify participants into three levels of dementia, two logistic regression functions were introduced using the extracted factor scores of both eyes. Also, power spectrum densities of pupil oscillation (PSD1 ∼ PSD3) are introduced. The two regression functions consist of classifying NC or MCI+AD participants and NC+MCI or AD patients in Data Set 1 [15]. The logistic regression function provides the probability of each classification, such as whether NC or patient. Since both regression functions are used for binary classification, prediction performance is evaluated using AUC (Area Under the Curve) of a ROC (Receiver Operation Characteristics) curve. The AUC for extracting AD patients is sometimes unstable, as the number of AD patients is limited. The performance of the previous regression models is summarized at the top of Table III. Performance is based on Data Set 1. The model "2n" is used for classifying NC or MCI+AD patients, and model "2a" is used for extracting AD patients from NC+MCI participants. The factor scores of the four conditions 2 to 5 are employed in these functions. In assessing the contribution of factor scores, selection of variables using a step-wise procedure is employed to calculate the function values for models "2ns" and "2as". Most AUCs for validation of the performance of Data Set 1 are over 0.9. Prediction performance of patient detection rates for Data Set 2 are around 50% when the threshold for probability is set to 0.5 on models "2n" and "2a". Performance is different for the two sequences of colour light pulses.

In the previous section, factor scores for the left and right eyes are summarized into the scores for irradiated or non-irradiated (cross) eyes. The number of variables for both eyes can be reduced by half, because the factor scores for irradiations of the left and right eyes using each colour of light pulse have been averaged. When these factor scores are applied to two types of regression functions for models "3n" and "3a", the performance AUCs show lower performance values than those for models "2n" and "2a". Another set of metrics to measure product ($Product\_b|r, F_j$) and rate ($Rate\_b|r, F_j$) of factor scores of irradiated and non-irradiated (cross) eyes are also employed for the two regression functions. However, detection performance is lower than the performance in the previous study.

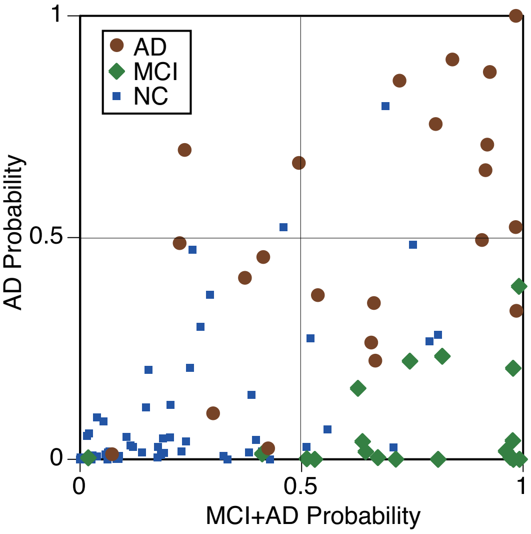


Fig. 11. Probabilities of trained data for irradiated left eyes using models "6nL" and "6aL".

In the next step, two types of metrics are introduced simultaneously using combinations of models "3n" and "4n" or "3a" and "4a". The performance of model "5" improved in comparison with models "3" and "4". Here, three sets of models employ mean factor scores of left and right eye light pulses. Most variables are based on the responses to light irradiated or non-irradiated eyes. Therefore, these variables can be transformed into two sets for irradiated left or right eyes. The extracted metrics may represent features of the reaction mechanism for both eyes, since the metrics are generated from reactions of both eyes. In the section for model "6", AUCs of the sets of data for both the left and right eyes are summarized independently. The performance of model "6" is comparable with values for model "5". As the data is divided into sets for right and left irradiated eyes, and both sets provide similar performance, estimation may be possible using two colour light pulses instead of four conditions.

Classification results for Data Set 1 using models "6nL" (NC detect) and "6aL" (AD detect) are summarized in Figure 11, where the horizontal axis represents the probability of MCI+AD and the vertical axis represents the probability of AD. The two-dimensional region is divided into four sub-regions by two thresholds of probability of 0.5 each. Most NC participants are classified into the low probability sub-regions,

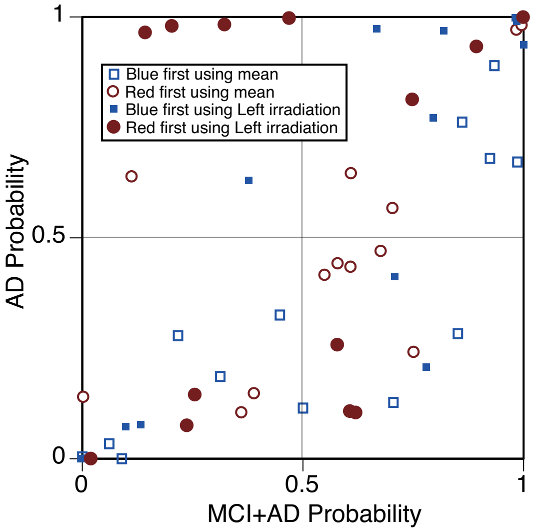


Fig. 12. Two sets of prediction result using two models: Mean features and features of irradiated left eyes using two sets of models, with symbols with no fill for models "5n" and "5a", and solid symbols for models "6nL" and "6aL".

as shown in Figure 11. Some AD patients are classified into the bottom-right sub-region, such as those with MCI, but most MCI patients are classified into the appropriate sub-region.

In order to validate the performance of the models, extracted features of two sequences in Data Set 2 are applied to the models, and the predictions are compared. As mentioned above, 54% of patients are detected using a probability threshold of 0.5 for the first blue sequence, and 38% of ADs are detected using the first red sequence when model "2n" is introduced.

The updated results for Data Set 2 are summarized in Figure 12 using model sets "5" and "6". In Figure 12, the symbols with no fill are the predicted probabilities using models "5n" and "5a", and the solid symbols are the predicted probabilities using models "6nL" and "6aL". When models "5n" and "5a" are employed, 54% of patients are detected using the first blue light sequence, and 77% of patients are detected using the first red sequence. In addition, 77% of patients are detected during both sequences when models "6nL" and "6aL" are introduced. These results suggest the possibility of reducing measurement trials to one light pulse to either eye using two colours, and improving prediction performance of other data sets.

The robustness of this procedure should be examined using other data sets which consist of patients and NC participants of various ages. In this study, the classification of participants was based on only MMSE scores as mentioned in the section about participants. Since there are many indices of dementia, additional information should be considered in order to diagnose the symptoms of the disease, including participants' personal histories. The confirmation of the contribution of these additional procedures to the diagnosis of dementia will be a subject of our further study.

IV. Summary

In order to detect persons who may have symptoms of dementia, additional features of PLRs are defined and applied to classify the level of dementia using two types of logistic functions. Prediction performance was evaluated using two sets of surveyed data obtained from clinical institutes. In particular, the following points were discussed.

1) PLR observation procedures were evaluated using two sequences of combinations of light pulses. In the results, the order of the colour of the light irradiation was not significant for the extracted factor scores.
2) Two metrics were introduced to represent the characteristics of irradiated and non-irradiated eye reactions of PLRs. Since these metrics can be generated when either eye is irradiated by chromatic light, another type of classification procedure was performed using the features of PLRs and the two functions.
3) Prediction performance of patients with dementia is evaluated using several conditions which are based on two types of functions for identifying two levels of patients such as those with MCI or AD. These models are trained using Data Set 1, and their performance was validated using Data Set 2.

A more accurate prediction procedure and method of analysis of the response mechanisms will be subjects of our further study.

Acknowledgement

The authors would like to thank Prof. Masatoshi Takeda and Prof. Takenori Komatsu of Osaka Kawasaki Rehabilitation University, Toshinobu Takeda, MD at the Jinmeikai Clinic, Yasuhiro Ohta and Takato Uratani of the Uratani Lab Company Ltd. for their kind contributions.

References

[1] A. J. Oh, G. Amore, W. Sultan, S. Asanad, J. C. Park, M. Romagnoli, C. L. Morgia, R. Karanjia, M. G. Harrington, and A. A. Sadun, "Pupillary evaluation of melanopsin retinal ganglion cell function and sleep-wake activity in pre-symptomatic Alzheimer's disease," *PloS ONE*, vol. 14, no. 12, pp. 1–17, December 2019.

[2] W. Nowak, M. Nakayama, T. Kręcicki, E. Trypka, A. Andrzejak, and A. Hachoł, "Analysis for extracted features of pupil light reflex to chromatic stimuli in Alzheimer's patients," *EAI Endorsed Transactions on Pervasive Health and Technology*, vol. 5, pp. 1–10, November 2019, e4.

[3] W. Nowak, M. Nakayama, T. Kręcicki, and A. Hachoł, "Detection procedures for patients of Alzheimer's disease using waveform features of pupil light reflex in response to chromatic stimuli," *EAI Endorsed Transactions on Pervasive Health and Technology*, vol. 6, pp. 1–11, December 2020, e6.

[4] D. F. Fotiou, V. Setergiou, D. Tsiptsios, C. Lithari, M. Nakou, and A. Karlovasitou, "Cholinergic deficiency in Alzheimer's and Parkinson's disease: Evaluation with pupillometry," *International Journal of Psychophysiology*, vol. 73, pp. 143–149, 2009.

[5] D. M. Bittner, I. Wieseler, H. Wilhelm, M. W. Riepe, and N. G. Müller, "Repetitive pupil light reflex: Potential marker in Alzheimer's disease?" *Journal of Alzheimer's Disease*, vol. 42, pp. 1469–1477, 2014.

[6] J. K. H. Lim, Q.-X. Li, Z. He, A. J. Vingrys, V. H. Wong, N. Currier, J. Mullen, B. V. Bul, and C. T. O. Nguyen, "The eye as a biomarker for Alzheimer's disease," *Frontiers in Neurology*, vol. 10, no. 536, pp. 1–14, 2016.

[7] S. Asanad, F. N. Ross-Cisneros, E. Barron, M. Nassisi, W. Sultan, R. Karanjia, and A. A. Sadun, "The retinal choroid as an oculavascular biomarker for Alzheimer's dementia: A histopathological study in severe disease," *Alzheimer's & Dimentia: Diagnosis, Assessment & Diesease Monitoring*, vol. 11, pp. 775–783, 2019.

[8] P. D. Gamlin, D. H. McDougal, and J. Pokorny, "Human and macaque pupil responses driven by melanopisn-containing retinal ganglion cells," *Vision Research*, vol. 47, pp. 946–954, 2007.

[9] A. Kawasaki and R. H. Kardon, "Intrinsically photosensitive retinal ganglion cells," *Journal of Neuro-Ophthalmology*, vol. 27, pp. 195–204, 2007.

[10] A. J. Zele, P. Adhikari, D. Cao, and B. Feigl, "Melanopsin and cone photoreceptor inputs to the afferent pupil light response," *Frontiers in Neurology*, vol. 10, no. 529, pp. 1–9, 2019.

[11] P. S. Chougule, R. P. Najjar, M. T. Finkelstein, N. Kandiah, and D. Milea, "Light-induced pupillary responses in Alzheimer's disease," *Frontiers in Neurology*, vol. 10, no. 360, pp. 1–12, 2019.

[12] L. Scinto, M. Frosch, C. Wu, K. Daffner, N. Gedi, and C. Geula, "Selective cell loss in Edinger-Westphal in asymptomatic elders and Alzheimer's patients," *Neurobiology of Aging*, vol. 22, no. 5, pp. 729–736, 2001.

[13] C. L. Morgia, F. N. Ross-Cisneros, J. Hannibal, P. Montagna, and A. A. Sadun, "Melanopsin-expressing retinal ganglion cells: implications for human diseases," *Vision Research*, vol. 51, pp. 296–302, 2011.

[14] M. Nakayama, W. Nowak, and A. Zarowska, "Detecting symptoms of dementia in elderly persons using features of pupil light reflex," in *Proceedings of the Federated Conference on Computer Science and Information Systems (FedCSIS)*, 2022, pp. 745–749.

[15] ——, "Prediction procedure for dementia levels based on waveform features of binocular pupil light reflex," in *Proceedings of ACM Eye-Tracking Research & Applications (ETRA)*, 2023, pp. 1–6.

[16] D. H. McDougal and P. D. Gamlin, "Autonomic control of the eye," *Comprehensive Physiology*, vol. 5, no. 1, pp. 439–473, 2015.

[17] W. Nowak, M. Nakayama, E. Trypka, and A. Zarowska, "Classification of Alzheimer's disease patients using metric of oculo-motors," in *Proceedings of the Federated Conference on Computer Science and Information Systems (FedCSIS)*, 2021, pp. 403–407.

DOI: 10.15439/2023F4426
ISSN 2300-5963 ACSIS, Vol. 37

A machine learning approach for automatic testing

Felix Petcusin
Computer Science and
Information Technology,
University of Craiova
Craiova, Romania
felix.petcusin@edu.ucv.ro

Cosmin Stoica Spahiu
Computer Science and
Information Technology,
University of Craiova
Craiova, Romania
cosmin.spahiu@edu.ucv.ro

Liana Stanescu
Computer Science and
Information Technology,
University of Craiova
Craiova, Romania
liana.stanescu@edu.ucv.ro.

***Abstract*—Systems' complexity has exponentially increased in recent years. Security and safety have become crucial in critical systems, and end-users now demand clear traceability to ensure protection against errors and external attacks. Meeting this requirement necessitates significant effort in testing. Although automated test sequences can handle a large portion of testing, it is crucial to identify as many errors as possible within the initial hours or days of the testing period. This paper introduces a machine learning-based solution that utilizes learned patterns to determine the test order. It analyzes which functionalities are more susceptible to errors and recursively generates the test sequence to be executed at each step.**

***Index Terms*—automated tests, machine learning, tests prioritization**

I. Introduction

TODAY, people's safety, entertainment, business decisions, and lives rely heavily on computers and various software tools. Therefore, it is crucial to ensure their proper functioning. The most effective approach to achieve this is by testing the products before they are released on the market. Software testing has become an essential component of any software project being the only way to guarantee high-quality applications that meets customer requirements, almost defects-free.

The complexity of software systems is increasing rapidly, with approximately one third of development costs currently being spent on electric/electronic development, a figure that continues to rise. Multiple variants of components are developed and tested through a series of prototyping phases, often with different schedules. Consequently, the level of complexity in specification activities has surpassed what can be effectively handled by traditional testing methods systems that are reliant on human input [1].

Most system development projects include a separate stage devoted to requirements specification, another stage for development and another for testing. All functional requirements must be validated and verified against the implementation. Testing is crucial especially for safety related industry, and software requirements are usually categorized as either "Integration Test" or "Software Test" requirements based on the content of requirements information. Specific testing techniques and methods are employed depending on this classification.

It is crucial to conduct testing before deploying software to end users to identify and fix errors in a timely manner and ensure the software is functional as intended. While testing aims to minimize errors, it is almost impossible to achieve a software product that is 100% error-free. Product quality is dependent on various parameters, such as performance, reliability, correctness, testability, and reusability, which can only be ensured through testing. Although testing can be time-consuming and expensive, it is better to invest in it early rather than after customer issues have arisen [2]. Balancing project costs and benefits/quality is important to make a development business case feasible. The key elements that indicate how much testing is enough are provided test coverage, time, and cost.

Manual testing was the traditional process, but it was time-consuming and limited by the available timeframe [3]. Hence, most testing activities have shifted to automated software testing execution with different tools that are more efficient, reduce time and cost, and increase test coverage. There is a wide range of testing tools available on the market that can be customized based on the complexity of the projects [4].

Recently, the authors have proposed a new paradigm in testing by extending classic tools to incorporate artificial intelligence (AI). This approach offers several benefits, such as faster and easier test creation, simpler test execution and analysis, and reduced test maintenance [5]. AI has transformed the testing approach by simplifying test documentation steps, decreasing maintenance effort, and providing new ways to interpret the results.

Thematic track: Software Engineering for Cyber-Physical Systems

The aim of the current work paper is to present the results obtained by a prototype system that uses Machine Learning algorithms for test-cases prioritization. The input is represented by an existing database of testcases together with the results obtained from past releases, in an Automotive project. The system will analyze existing testcases (executed in past releases), together with the newly created ones in order to classify the ones with higher risk of failure. They will be executed first in the list.

The remainder of this article is organized as following: chapter 2 presents the motivation leading to the present work and summarizes the prior research in this domain, chapter 3 presents an original proof-of-concept application which was developed and the results obtained during experiments and in the final chapter there are presented the conclusion and future directions.

II. Related Work

In recent years, several machine learning (ML) algorithms have been used to solve a few particularly difficult problems in the field of automate classifications for systems. Two aspects are considered: requirements and tests classification.

An example of such a problem is the identification and classification of non-functional requirements in requirements documents. ML-based solutions have shown promising results that go beyond those of traditional Natural Language Processing (NLP) approaches.

In [6], the authors work on automatic classification of requirements by performing a systematic review of 24 ML-based solutions for identifying and classifying NFRs. The authors selected 24 research papers that use 16 different ML algorithms. These algorithms can be divided into three categories: supervised learning (7 algorithms), unsupervised learning (4 algorithms), and semi-supervised learning (5 algorithms). The supervised learning algorithms were used in 17 papers (71%), with SVM being the most popular algorithm in 11 studies (45.8%). The authors come to the following conclusions: ML-based solutions have potential in classifying and identifying NFRs; collaboration between RE and ML researchers is needed to address open challenges in the development of ML systems for real-world use; The use of ML in RE opens up exciting possibilities for the development of novel expert and intelligent systems to support RE tasks and processes.

Another article [7] presents an approach to automatically classify content elements of a requirements specification into requirements or information. The presented approach could be used in the following ways: classification of content elements in previously unclassified documents; Perform an analysis on a previously classified document and assist the user in identifying elements that are not correctly classified.

The authors propose Convolutional Neural Networks, a machine learning algorithm that is receiving more and more attention in the field of natural language processing. To train the neural network, the authors used a collection of over 10,000 content elements extracted from 89 requirements specifications from their industry partner. By using 90% of the content elements as training data and the remaining 10% as test data, the authors' approach was able to achieve a stable classification accuracy of about 81%.

In [13], the authors introduce a methodology for automatically assessing the quality of requirements based on input from field experts who utilize the methodology. The main objective of this methodology is to predict the quality of new requirements. To accomplish this, the experts provide an initial set of requirements that have been previously classified according to their quality and deemed appropriate. For each requirement in the set, the authors extract metrics that quantify the quality value of the requirement. The methodology suggests employing a Machine Learning technique called rule inference to learn the value ranges for these metrics and determine how they should be combined to interpret the quality of requirements, as perceived by domain experts

Strictly related to tests classification, the state-of-the-art approach involves using machine learning algorithms for test creation and maintenance. This has led to improvements in reducing maintenance efforts and enhancing product quality. Machine learning can be used throughout the software testing life cycle, from test creation to issue management. So, how can machine learning be applied in testing? There are several ways to enhance the process:

A. *Handling issues:*

ML algorithms can be used to classify issues based on severity levels, predict assignees for issues based on past experience, cluster issues based on common features, and prioritize cases based on their relation to issues. Various tools have been developed in this area [8]:

- classify issues according to severity levels
- predict assignee of an issue according to previous experience
- cluster issues to see whether they heap together on specific features
- prioritization of cases by relating to issues

B. *Software maintenance*

Machine learning techniques can reduce maintenance efforts by automating the review process. Self-healing methods can save time by automatically detecting and suggesting resolutions for broken test cases caused by code changes [9].

C. *New tests generation*

As software systems become more complex, it becomes challenging to define tests that can check the entire product spectrum. Investing in an ML system that can automatically generate pattern-based tests is a cost-effective solution. Once the ML system is trained, it can learn patterns and generate tests automatically in the long run. Innovative solutions for automated test generation have been proposed in [10] and [11] for embedded systems, while [12] proposes a hybrid ML algorithm to manage test scenarios for printed-circuit boards.

Experimental results have shown an increase in fault identification from 57.3% to 78.9%

III. The Method

The testing phases for successive software versions in complex projects are in most cases particularly challenging, both in terms of cost and time. Several thousand test cases can be executed during a full test cycle, and the execution time can be extended to weeks. Given this long duration of testing, a serious bug significantly increases the cost of the project if it is discovered in the final phase of the timeframe. This includes implementation costs and retesting costs. A big advantage would be to find a way where topics with a high chance of failing to be executed first. This would make it possible to evaluate first those functions where there is a high risk of failure, then those with lower risk, and so on. At the end of the test phase, only the functions with the lowest risk of failure remained.

This article presents an original solution that performs an automated classification of tests based on their scope and determines step by step which tests need to be run next. The decision is based on previous experience and the results of the current cycle, based on tests already carried out. The solution uses machine learning algorithms for automated classification, result analysis, and determination of the test sequence for each next step. The application has been designed as a tool that supports the testing process in the following areas:

- Automated classification of test cases
- Risk assignment - learning phase
- Selection of the test sequence.

A. *Testcases definition*

The application was designed to be compatible with systems where tests are designed based on a template that provides detailed information about the scope and steps of the tests. The following elements need to be defined for each test specification and will be considered as input:

Test scope description: This specifies what will be verified in the current test and is expressed as free text. For example, "The test is verifying the system's reaction in case of a damaged LED."

Test preconditions: This describes the initial state of the system, also in free text. For example, "The system is running with no active errors."

Test steps description: This provides a step-by-step specification of the tasks that need to be performed to verify the test scope. For example:

Step 1: Start the system diagnosis and disconnect the load.

Step 2: Turn on the light.

Step 3: Verify if the system detects any errors.

Pass/fail criteria: This defines how to interpret the results of the previous steps. For example, "If the system detects the fault, the test is considered passed."

Automation: This refers to a script that can be executed automatically to perform the specified steps and evaluate the results.

During a regular test cycle, the tests are grouped into sequences based on functionalities. Each sequence is executed sequentially, and the results are analyzed and documented. This means that each test is assigned to only one main functionality and is executed when that functionality is the focus.

The proposed solution use the TensorFlow algorithm to parse each test individually. Based on the analysis of the defined scope and description, it generates a list of functionalities that are directly or indirectly verified. This classification creates a graph where all the tests are linked to each other based on predefined keywords, which represent the functionalities in focus.

B. *Dataset*

We have generated a dataset using an existing set of real Testcases used for the validation of an Electronic Control Unit (ECU) in the automotive industry. Each individual Testcase in the dataset possesses the following attributes:

- Test scope description
- Test preconditions
- Test steps description
- Pass/fail criteria

Starting from these attributes we have generated two new attributes that are used for Testcases prioritization.

The first attribute, called Test Priority is obtained by concatenating the text information from the Test scope description, Test preconditions and Test steps descriptions. The second attribute is called Test Added Value and is obtained based on the Pass/fail criteria and the results of the interpretation of the Testcases executed in each of the five releases. The Test Added Value is a label having values with 0 and 1.

0 - Testcase low value added

1 - Testcase high value added

These labels from the Test Added Value were automatically assigned via a script to each of the Testcases. Preprocessing was performed on the Test Priority attribute to eliminate extraneous information, such as logical expressions and statements that could not be converted into lexical tokens.

The resulting dataset comprises 5000 Tescases , which serve as the raw input for our Test prioritization process.

C. The Application description

We utilized two computational methods for our dataset. The first method is checking the traditional "bag of words" (BoW) representation, where word occurrences are counted to create a vector for each sentence. The second method utilizes "word embeddings," a newer approach that assigns each word a vector preserving semantic meaning. Our experiments aimed to evaluate the potential of word embeddings (VE) compared to the classic BoW representation when combined with deep neural networks for automated Testcases priorization.

IV. Experiments and Results

The experiments performed followed a two-step processing pipeline:

Text vectorization: Each Testcase document was transformed into a numerical vector using either the BoW or VE method.

Deep learning classification: A suitable deep neural network (NN) was defined, trained, and validated to classify the vectorized representations obtained in the previous step.

For both models, we applied the standard approach of cross-validation. Our dataset was split into training and testing sets, with 75% of the examples used for training and the remaining 25% for testing to measure model accuracy. The split was performed using the "train_test_split" method from the scikit-learn package.

A. First Model:

Initially, we made the BoW representation of text. To accomplish this, we constructed a vocabulary from our dataset consisting of a unique word list. Each word was assigned an index, and every Testcases (example) was then associated with a vector of dimensions equivalent to the vocabulary size, which was 2135 in our specific case. Within the vector, each element indicates the count of occurrences for the corresponding word in our dataset.

For the Testcases classification using deep neural networks (NN), we utilized Keras. The NN architecture consisted of an input layer, one hidden layer with 10 nodes, and an output layer. The hidden layer employed a densely-connected NN layer of type layers. Dense with the ReLU activation function. Since we were dealing with a binary classification problem, we used the sigmoid activation function with a dimensionality of 1 for the output layer. The optimization of the NN was performed using the Adam algorithm, and binary cross-entropy served as the loss function.

Using the constructed model, we trained it using our training data. The training process involved 10 samples per gradient update, and we performed 20 iterations. The first layer had 21,360 parameters, while the second layer had 11 parameters. The total number of parameters was calculated as follows: each feature vector had 2135 dimensions, which required weights for each feature dimension and each node, resulting in 2135 * 10 (adding 10 times bias for each node). The layer had 10 weights and one bias. During the training process, all 21,371 parameters were determined. The results are presented in Figure 1.

To evaluate the performance of the trained network, we measured accuracy on both the training and test sets, as well as the training and validation loss. The first model achieved an accuracy of 85%.

B. Second Model

In our second model, we employed word embeddings to represent the requirements, departing from the previous model that used the Bag-of-Words (BoW) approach to map each requirement to a single feature vector. Instead, we represented each word as a numeric vector. There are two options to acquire word embeddings: training them separately on the new corpus or using pre-trained versions. In our experiment, we opted to utilize pre-trained GloVe word embeddings, specifically the glove6B.50d.txt file, which encompasses 400,000 unique words and has a total size of 822MB. However, we filtered the embeddings to include only the words present in our dataset.

To prepare the data for our word embeddings model, we utilized the pre-processed test and training data and performed tokenization. The Keras Tokenizer utility class was employed to convert the dataset into a list of integers. Each integer in the list corresponds to a word in the dictionary that

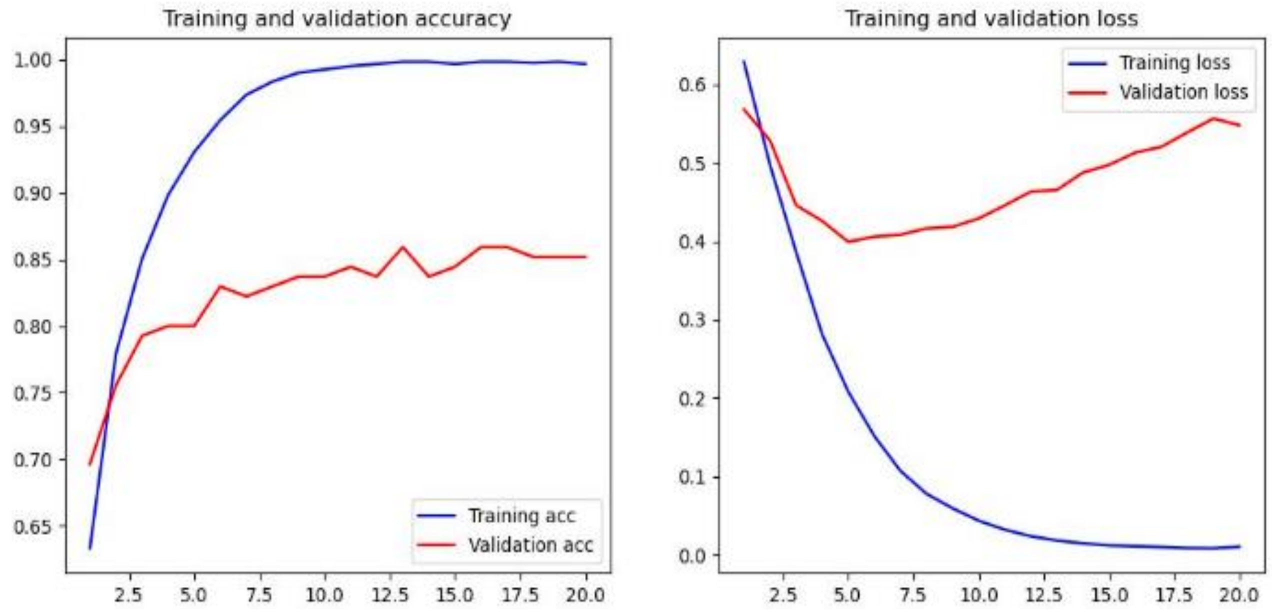


Fig. 1 Results obtained using first model

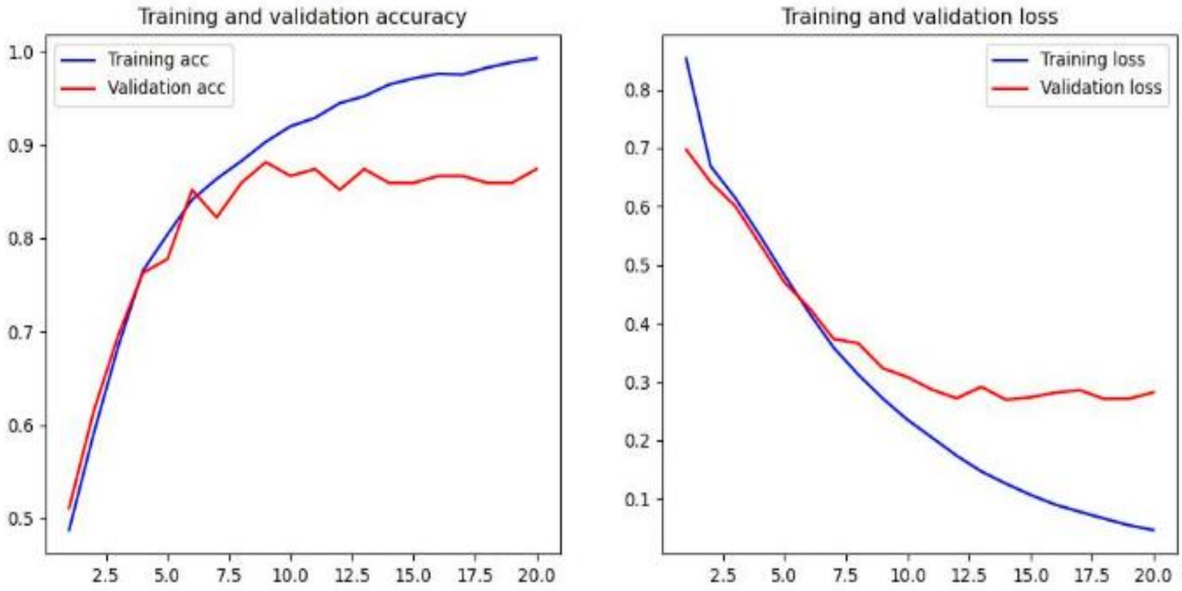


Fig. 2 Results obtained using the second model

represents the entire corpus. As the length of each requirement (example) may differ, we padded the word sequences with zeros to ensure a consistent length of 100 for our experiment.

For the architecture of our second model, we employed a deep neural network (NN) comprising an input layer, three hidden layers, and an output layer. The first layer utilized the layers.Embedding data type, enabling us to map the examples represented as lists of integers to a suitable representation for processing by the subsequent GlobalMaxPool1D layer. The embedded layer was configured with the following parameters:

Input dimension: 2135, representing the vocabulary size.

Output dimension: 50, indicating the size of the dense vector.

Input length: 100, denoting the length of the word sequence.

The second hidden layer was of type GlobalMaxPool1D, employing default parameters to downsample the incoming feature vectors by selecting the maximum value across each feature dimension.

The third hidden layer was of the Dense Layer type, similar to the first model.

As we were dealing with a binary classification problem, akin to the first experiment, we employed the sigmoid activation function with an output dimension of 1 for the NN model's output layer. Binary cross-entropy was used as the loss function to measure the discrepancy between the actual output and the predicted output.

We optimized our NN using the Adam optimizer. With the model created, we proceeded to train it using our training data. We used 10 samples per gradient update and conducted 50 iterations. The results obtained are depicted in Figure 2.

To evaluate the performance of the trained network, we measured the accuracy on the training and test sets, as well as the training and validation loss. This model showcased improvement over the first model, achieving an accuracy of 89%.

V.CONCLUSION

In the last years, artificial intelligence started to transform the testing paradigm in ways that could not have been considered possible some years ago. In the current paper it is presented an innovative solution applied in the testing domain based on machine learning algorithms for tests execution.

It is considered the first results and experiments towards automating classification of new written Testcases in software engineering for automotive industry towards test execution. We have investigated the potential of combining deep NN models with word representations for improving the performance of this task.

Our results are preliminary, nevertheless, we were able to obtain an improvement in performance for the word embeddings approach, as compared with the baseline bag of words approach. In the near future we plan to strengthen our results by expanding our experiments to include different and larger data sets, as well as to use different and suitably trained deep NN architectures.

REFERENCES

[1] M.Weber, J.Weisbrod (2003) Requirements engineering in automotive development: experiences and challenges. IEEE Software,2003, 20(1):16–24

[2] F. Azaïs, S. Bernard, M. Comte, B. Deveautour, S. Dupuis, H. El Badawi, M.-L. Flottes, P. Girard, V. Kerzerho, L. Latorre, F. Lefèvre, B. Rouzeyre, E. Valea, T. Vayssadel, A. Virazel, "Development and

Application of Embedded Test Instruments to Digital Analog/RFs and Secure ICs", IEEE 26th International Symposium on On-Line Testing and Robust System Design (IOLTS), pp. 1-4, 2020

[3] A. Adekanmi, "Research on software testing and effectiveness of automation testing", 2019

[4] H. Gamido, M. Gamido "Comparative Review of the Features of Automated Software Testing Tools", International Journal of Electrical and Computer Engineering, vol 9, pp. 4473-4478, 2019

[5] https://www.functionize.com/machine-learning-in-software-testing

[6] Binkhonain M., Zhao L, "A review of machine learning algorithms for identification and classification of non-functional requirements", 2019, Expert Systems with Applications: X, 1, doi: 10.1016/j.eswax.2019.100001

[7] Winkler J., Vogelsang A. (2016) Automatic Classification of Requirements Based on Convolutional Neural Networks, In: IEEE 24th International Requirements Engineering Conference Workshops (REW), 39–45 doi: 10.1109/REW.2016.021.

[8] K. Sneha, G. M. Malle, "Research on software testing techniques and software automation testing tools," International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), pp. 77-81, 2017

[9] https://www.infoq.com/news/2021/03/machine-learning-testing/

[10] S.Roy, S. K. Millican, V.D. Agrawal, "Training Neural Network for Machine Intelligence in Automatic Test Pattern Generator", 20th International Conference on Embedded Systems (VLSID) 2021, pp. 316-321, 2021

[11] S. Roy, S.K. Millican, V.D. Agrawal, "Principal Component Analysis in Machine Intelligence-Based Test Generation", Microelectronics Design & Test Symposium (MDTS), pp. 1-6, 2021

[12] M. Liu, F. Ye, X. Li, K. Chakrabarty, X. Gu, "Board-Level Functional Fault Identification Using Streaming Data", Computer-Aided Design of Integrated Circuits and Systems, vol 40, no. 9, pp. 1920-1933, 2021

[13] Parra E., Dimou C., Llorens J., Moreno V., Fraga, A. (2015) A methodology for the classification of quality of requirements using machine learning techniques, Information and Software Technology, 67:180–195, doi: 10.1016/j.infsof.2015.07.006.

[14] M.N. Velev, C. Zhang, P. Gao, and A.D. Groce, "Exploiting Abstraction, Learning from Random Simulation, and SVM Classification for Efficient Dynamic Prediction of Software Health Problems," 16th International Symposium on Quality Electronic Design (ISQED '15), March 2015, pp. 412–418

DOI: 10.15439/2023F3165
ISSN 2300-5963 ACSIS, Vol. 37

Generation of Benchmark of Software Testing Methods for Java with Realistic Introduced Errors

Tomas Potuzak
0000-0002-8140-5178
Department of Computer Science and Engineering/
NTIS – New Technologies for the Information Society,
European Center of Excellence, Faculty of Applied
Sciences, University of West Bohemia
Univerzitni 8, 306 14 Plzen, Czech Republic
Email: tpotuzak@kiv.zcu.cz

Richard Lipka
0000-0002-9918-1299
NTIS – New Technologies for the Information
Society, European Center of Excellence/Department
of Computer Science and Engineering, Faculty of
Applied Sciences, University of West Bohemia
Univerzitni 8, 306 14 Plzen, Czech Republic
Email: lipka@kiv.zcu.cz

***Abstract*—This paper deals with a benchmark of automated test generation methods for software testing. The existing methods are usually demonstrated using quite different examples. This makes their mutual comparison difficult. Additionally, the quality of the methods is often evaluated using code coverage or other metrics, such as generated tests count, test generation time, or memory usage. The most important feature – the ability of the method to find realistic errors in realistic applications – is only rarely used. To enable mutual comparison of various methods and to investigate their ability to find realistic errors, we propose a benchmark consisting of several applications with wittingly introduced errors. These errors should be found by the investigated test generation methods during the benchmark. To enable an easy introduction of various errors of various types into the benchmark applications, we created the Testing Applications Generator (TAG) tool. The description of the TAG along with two applications, which we developed as a part of the intended benchmark, is the main contribution of this paper.**

***Index terms*—Benchmark, software testing methods, automated test generation, application generation, Java code parsing, error introduction.**

I. Introduction

THE testing is a very important part of software development. It improves the probability of the correct functioning of an application, as it helps to uncover and fix errors unwittingly introduced into it during its development. As the manual creation of the tests is a lengthy and error-prone process, there is an intensive research on automated test generation methods for more than two decades (see, for example, [1] or [2]). In existing scientific papers, automated test generation is proposed or used on various testing levels. The lowest level is unit testing, which focuses on individual basic functional elements of the tested application, such as methods or functions. The middle level is the regression and integration testing focused on the correct cooperation of larger parts of the application. The highest level is the testing of the functionality of the entire application, its cooperation with its environment, and its adherence to its specification. At all levels, the expected advantages of the automated testing is the reduced time spent by programmers on the tests preparation and/or execution, decreased number of errors within the tests themselves and increased code coverage of the tested application. Nevertheless, there are also disadvantages. For example, it is inherently difficult to automatically verify whether the tested parts of the application give correct results, as it requires knowledge or generation of correct results. Another problem (discussed for example in [3]) is the combinatorial explosion. This means that the number of generated tests can be very high in order to cover all combinations of representative input values (e.g., values of tested method parameters). This can lead to long running times.

A different problem is the testing of automated test generation methods themselves. In many scientific papers, the method functioning is often demonstrated only on small examples (e.g., in [4] or [5]), from which the usability in a real project cannot be concluded. Although there are also methods tested on more realistic examples (e.g., in [6] or [7]), these examples are not mutually similar and do not enable direct comparison of the features of the methods.

It should also be noted that, from the pragmatic point of view, the most important feature of the method in a real project is the ability to find realistic errors of various types [8]. Nevertheless, in scientific papers, this feature is virtually never used as the metric for method assessing (with some exceptions, e.g., [9]). A quite common approach used for automated test generation methods evaluation is mutation testing [10]. Using this approach, several versions of the tested program (so-called mutants) with small changes imitating real errors introduced by mutation operators are generated. The quality of the automated test generation method can be then assessed by the number of mutants the method is able to identify [10]. However, a large portion of the papers uses only code coverage for the methods evaluation (e.g., [11] or

The work was supported by the Ministry of Education, Youth and Sports of the Czech Republic university specific research project SGS-2022-016 Advanced methods of data processing and analysis.

[12]) or other metrics, such as generated tests count, test generation time, or memory usage (e.g., in [13]).

In order to enable mutual comparison of various automated test generation methods and to investigate their ability to find realistic errors, we propose to create a benchmark consisting of several various applications with wittingly introduced errors. The numbers of the errors discovered and not discovered by each method can be then used for a direct assessment of the quality of each method. In order to enable an easy introduction of various errors of various types into the benchmark applications, we created a tool called Testing Applications Generator or TAG. The TAG is designed in and for Java language, similarly to many automated test generation methods (e.g., [4] or [14]). It enables to introduce errors of various types into the source codes of methods bodies of an application. The description of the TAG and its functioning and the first two applications, which we plan to utilize as a part of the benchmark of the automated test generation methods, are the main contributions of this paper.

The remainder of the paper is structured as follows. Section II briefly discusses the existing automated test generation methods. Section III is focused on related work. In Section IV, the TAG is described in detail. The two benchmark applications are described in Section V. The tests of the TAG and their results are described in Section VI. The conclusions and the future work are discussed in Section VII.

II. Automated Test Generation Methods

There is a large number of existing test generation methods, which are based on various technologies and use various inputs for their functioning (see [15] for details). Some examples are summarized in following subsections.

A. Commonly Used Technologies in Testing Methods

The test generation methods can be based on a single technology, but often employ multiple technologies. Control-flow-based methods utilize control flow diagrams created by static analysis, for example in [1]. The diagrams are used to generate tests covering all branches of the program, often in conjunction with random input data generation, as in [16].

Specification-based methods are somewhat similar to control-flow-based methods as the tests can be generated from diagrams utilized for the description of the application, such as UML diagrams, as in [17], [18]. Different forms of specification can be used as well, for example use case descriptions employed in [17], [19] or contracts employed in [20].

The search-based methods typically employ a search meta-heuristic to generate tests including genetic algorithms (e.g., in [4], [11]), particle swarm optimization (e.g., in [5]), or ant colony optimization (e.g., in [21]). The meta-heuristic is typically combined with a technology enabling to evaluate the found solutions, for example with control-flow diagrams (e.g., in [22]) or program instrumentation (e.g., in [23]).

Program-execution-based methods employ real or symbolic executions of the tested program for test generation. If the real program is executed, there is usually some form of code instrumentation, such as in [24] or [25]. Examples of symbolic-execution-based methods can be found in [26] or [27].

B. Commonly Used Inputs for Testing Methods

The aforementioned and other existing test generation methods use various primary inputs. In many cases, it is the source code of the tested program, as for example in [1], [4], [11], [26], or [27]. The source code does not have to be used directly. For example, in [1], the static analysis of source code is used for the generation of control-flow diagrams, which are in turn used for the generation of the tests. In [4], the source code is used for the determination of the method parameters, which are then used by a genetic algorithm.

The primary input can also be an instrumented execution of the program (e.g., in [24], [25]), or Java bytecode (e.g., in [12]). Yet another primary input can be a description of the tested program in some form, for example the UML diagram (e.g., in [17], [18]) or the contracts description (e.g., in [20]).

It should be noted however that, regardless of utilized primary input, the resulting generated tests are used for the testing of a real tested program (i.e., not its model nor description). That means that, although the source code and/or executable version of the program may not be necessary for the generation of the tests, it is required for the execution of the generated tests. The executed tests should then discover errors present in the program.

III. Related Work

As we are working on the creation of automated test generation methods benchmark, which would solve the difficulties of test generation method comparison (see Section I), we investigated the existing research in this area.

A. Benchmarks of Testing Methods

The benchmarks of testing methods are quite rare, but there are a few examples. A benchmark was used for a competition of Java unit tests generating tools (Java Unit Testing Tool Contest 2018) [28]. The benchmark consisted of 59 real-life Java classes from 7 open-source projects. The projects were selected randomly from a pool of GitHub repositories, which met predefined criteria, such as having enough stars, being able to be built by Maven, and containing JUnit 4 tests [28]. From the total number of 2 566 classes of the 7 projects, only classes with at least 1 method with at least 2 condition points were considered further. From them, 59 randomly selected classes were used as the benchmark [28]. For the selection of the projects, an unspecified script was used. For the class filtering, an extended CKJM library was used [28]. No intentional introduction of errors was reported in [28]. For the evaluation of the contesting tools, code coverage computed by the JaCoCo tool and mutation testing analysis performed by PIT tool were used [28].

A similar benchmark was used for the Java Unit Testing Tool Contest 2020 [29]. The selection of the projects was a bit different, the predefined criteria were being able to be built by Gradle or Maven and containing JUnit 4 tests [29]. In the end, 4 projects were selected. Only 1 094 classes with at least 1 method with at least 4 condition points were considered further. These classes were further filtered by trying to generate tests for them using Randoop tool with 10 seconds time budget for each class. Only the classes, for which at least one test was generated, were considered further. Using this filter, 382 classes remained. From them, 60 classes were randomly selected for the benchmark, while another 10 were selected based on the past experience [29]. For the class filtering, JavaNCSS tool was used [29]. Again, no intentional introduction of errors was reported in [29]. Similarly to [28], code coverage computed by the JaCoCo tool and mutation testing analysis performed by the PIT tool were used for the evaluation of contesting tools [29].

In [10], the creation of a repository of artifacts, usable for standardized evaluation of mutation-based testing methods, is described. The authors used a relational database as the storage of the artifacts and created import scripts for them. The basis of the repository is a set of Java classes taken from 4 open source projects from GitHub and from a set of simple Java programs. From the ca. 2 000 classes, ca. 50 000 test cases and ca. 195 000 mutants were generated using the existing EvoSuite and PIT tools, respectively. These test cases and mutants are also stored in the repository [10].

In [8], a benchmark testbed application with artificial error injection for the evaluation of testing methods is described. The application is the University Information System Testbed (TbUIS), a fictional, but functional university study information system, which includes students, teachers, management of exams, and related processes. It is a layered J2EE-JSP-Spring web application with relational database storage and object-relational mapping (ORM) using Hibernate. The application consists of 87 `.java` and 18 `.jsp` files with more than 10 000 lines of code in total. The TbUIS source code is highly covered by automated unit and frontend functional tests in order to reduce the number of errors introduced during the development of the application [8].

To introduce errors into the TbUIS application, the Error seeder application is used. It operates on the bean (i.e., class) level. Each bean of the TbUIS application can be replaced by a version with introduced error or errors. The errors, which shall be introduced, are selected from a predefined set. The resulting version of the TbUIS application with the beans with introduced errors can be then compiled and used as a part of the benchmark of testing methods [8].

B. Assessing and Comparability of Testing Methods

The diversity of examples, on which the functionality of the automated test generation methods is demonstrated in scientific literature (see Section I), is mentioned in several review papers. A thorough review paper [30] deals with search-based test generation methods. One of the conclusions is that there is a lack of standardized rigorous way to assess and compare various methods. Moreover, it is pointed out that, while many of the test-generating methods can achieve high code coverage, it is not clear whether the tests are actually able to find errors in the source code [30].

Similarly, the review paper [31], which is focused on mutation testing, concludes that the experimental material used in the papers describing various test generation methods is typically non-standardized, lacks reusability, and is rarely available to be shared to support further experiments [31]. One of the conclusions of the review paper [32] focused on search-based and mutation testing methods is that the comparability of the automated methods is difficult [32].

IV. Testing Applications Generator

In order to address the difficult comparability of the automated test generation methods, we decided to create a benchmark, which would consist of several various applications with wittingly introduced errors. The number of the errors discovered and not discovered by each automated test generation method can be then used for a direct assessment of the quality of each method. Nevertheless, since various methods can be focused on specific types of applications and/or errors, the creation of a single benchmark application with hardwired errors would be of limited usefulness. Hence, we created a prototype implementation of the Testing Applications Generation (TAG) tool. The TAG enables to introduce errors of various types into the source codes of imported applications. The TAG is inspired by the TbUIS [8] (see Section III.A), but is different in many ways (see below).

A. Usage of TAG

The TAG is a Java desktop application with a graphical user interface (GUI) enabling to import multiple Java applications. The entire project can be imported (see Section IV.C), but the source codes are required. The source code files of each imported application are parsed and the entire structure of packages, classes, interfaces, and other code artifacts are stored down to the level of individual methods.

Each method has a single imported body, but additional copies of the body can be created on user request. The user can then introduce one or multiple errors into each copy (see Section IV.D for details). All the created copies are stored.

In order to export an application with selected introduced errors, the user then only selects the method bodies containing the required errors and the application is created in a selected folder. The exported application can be used as a part of the benchmark of automated test generation methods, as it contains known introduced errors and, inevitably, other errors already present in the application prior its import.

So, the TAG is distantly similar to the Error seeder of the TbUIS (see Section III.A). However, unlike the Error seeder, the TAG is not designed for a single application. Multiple applications can be stored and virtually any Java application with source codes can be imported. There are no requirements for a specific technology, such as Spring, the applica-

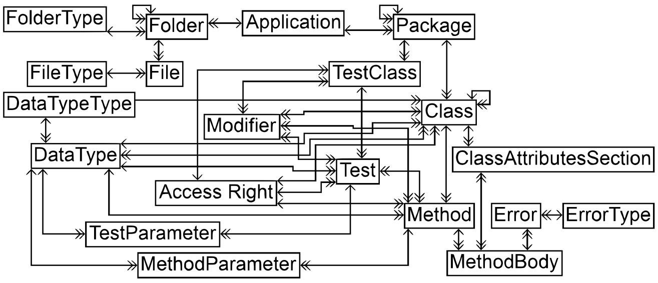


Fig. 1 Scheme of the TAG data model

tion must be only compilable by a standard Java compiler (currently version 11). The introduced errors are also not limited to a predefined set. Lastly, the Error seeder operates at the Java beans level (i.e., an entire bean is replaced with a faulty one), while the TAG operates at the method level (i.e., a method body is replaced with a faulty one).

B. Data Model

All the data utilized by the TAG are stored in a relational database using ORM via Hibernate. The scheme of the data model is depicted in Fig. 1. The database enables to save the entire structure of an application project including folder structure with various types of files (e.g., libraries, resources, documentation, or build scripts) and the package structure with classes, interfaces, and other source code artifacts. The files other than source code files are stored only as type, name, content, and parent folder. On the other hand, the folders representing package structure are also stored as packages and the contained source code files are parsed and stored as classes, interfaces, methods, and their bodies. The content of the class outside any method (i.e., typically attributes), so-called *class attributes section* are stored as well. For each method, multiple bodies can be stored and, for each class, multiple class attributes sections can be stored. The test classes (or test cases in JUnit terminology) containing individual tests are parsed as well. However, each test (corresponding to a method of the test class) has only one body and each test class has only one class attributes section. The reason is that the introduction of errors into tests is not expected.

Besides the structures of the imported applications, the database also contains all necessary code lists, such as access rights, modifiers, file and folder types, or error types. Some of them, for example the access rights and modifiers are expected to hold constant sets of values. To the others, such as file and folder types or error types, new values can be added as needed. The database also contains all the errors, which are introduced into the applications.

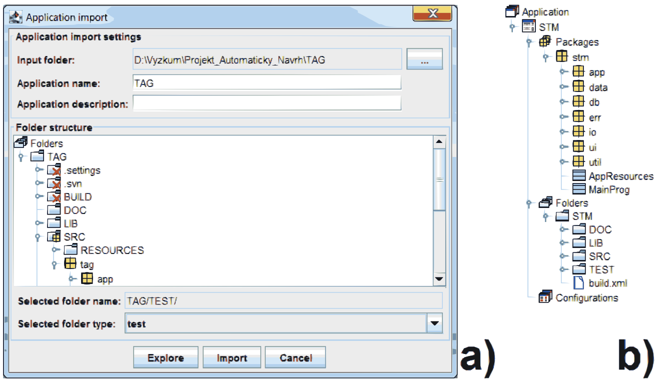
C. Application Import

For each application, its entire project can be imported including the folder structure, source codes, resources files, libraries, build scripts, and other files. Nevertheless, only the source codes are required. The contents of other files are only stored into the database, while the contents of source code files (i.e., `.java` files) are parsed down to the method body level, but not further. That means that the content of the body of a method is not parsed and is stored as a text segment. Similarly, a class attributes section (containing mainly attributes) is stored as a text segment. On the other hand, the headers of classes and methods are parsed including method parameters, return values, type parameters, and so on. Test classes are parsed and stored similarly to normal classes.

During the import, the content of the selected folder with the imported application is explored and displayed as a tree to the user, who must mark the source code and tests subfolders. He or she can also choose the type of other folders or mark some not required folders as ignored (see Fig. 2a). Then, the import including the parsing of the source code files is performed automatically. There are no specific requirements for the folder structure, it is possible to import Eclipse or Maven/Gradle styles or customized structures.

If there is a problem during parsing a source code file, the import is not stopped. Instead, the source code file is stored into the database as a general file with its entire content "as is" (similarly to, for example, a resource file). The import then resumes with the next source code file. This way, one (or multiple) file with a parsing error does not hinder the entire import. It is not possible to introduce errors into the files, whose parsing failed (unless the application is edited later directly using the GUI of the TAG), but the other correctly parsed files are not negatively affected. The parsing error can be caused by syntax errors or by using an unexpected construction, such as constructions added to newer versions of Java. Currently, the parser is set to Java 11.

Once the application is imported, its folder and package structures are displayed as a tree (see Fig. 2b). It is possible to display the details of its individual items and add/edit/delete them. Theoretically, it is possible to create the entire application by adding its individual items one by one (i.e., without the import), but this approach would be lengthy and error-prone and it is not recommended. The TAG is no Integrated Development Environment (IDE), its editing capabilities are intended only for little changes, which might be necessary during the introduction of the errors (see Section IV.D) or during other minor adjustments of the application (e.g., correction of the failed parsing – see above).


Application import

Application import settings

Input folder: D:\Vyzkum\Projekt_Automaticky_Navrh\TAG

Application name: TAG

Application description:

Folder structure

Selected folder name: TAG/TEST/

Selected folder type: test

Fig. 2 Application import (a) and imported application structure (b)

D. Error Introduction

The errors introduction is manual in the sense that the user must manually edit a method body or a class attributes section, to which he or she intends to insert the error. The added error should be also added to the list of errors. Each specific error has its type, name, description and corresponding method body and/or class attributes section. This way, each method body and class attributes section can be described by the errors it contains. This enables easy selection of intended method bodies and class attributes sections during the export of the application (see Section IV.E).

Because, within a class, errors can be introduced into various bodies of the same or of various methods and also into various class attributes sections, it is possible that some of the method bodies would not be compatible with some of the class attributes sections. Hence, the information, which bodies are compatible with which class attributes sections, is stored and then utilized during export (see Section IV.E)

The types of the errors can be selected from a list and the user can readily add new types. Currently, the list is largely empty, since the common and realistic error types determination is the part of our future work (see Section VII).

E. Application Export

Each application can be exported in multiple versions, with different errors or entirely free of artificially introduced errors. The errors present in the application during the import (i.e., unwittingly introduced during application development) will be of course present. Each version is represented by so-called *configuration*. The configuration is created by specifying, which method body shall be used for each method with multiple method bodies and which class attributes section shall be used for each class with multiple class attributes sections. The stored information about the compatible method bodies and class attributes sections is used for checking whether only compatible method bodies and class attributes sections are used together. There can be multiple configurations per application. All available configurations are displayed as part of the application tree (see Fig. 2b).

The version of the application corresponding to the configuration can be exported to a selected folder. The export is automatic. The generation of the folder and package structure is straightforward. The `.java` files are generated from the classes and their contents in folders corresponding to their packages. For this purpose, the user can specify several features of the code style, such as the usage of spaces or tabulators for the indentation. The other files are only created in corresponding folders and filled with their content stored in the database.

The exported application can be directly used as a part of a benchmark. If its compilation is required, for example for the generation or execution of the tests by the benchmarked test generation methods, it can be performed using a standard build script, which is usually present. The contemporary prototype version of the TAG cannot perform the compilation automatically, but it is part of our future work. It is currently possible to import end export `.class` files with bytecode. However, without automatic compilation, the exported `.class` files do not correspond to exported `.java` files if some errors were wittingly introduced using the TAG. Hence, the manual compilation after the export is recommended.

V. Benchmark Applications

Concurrently with our work on the implementation of the TAG, we are working on our own benchmark for test generation methods. This work consists of two main branches – creation or selection of the applications used for the benchmark and selection of the artificially introduced errors into these applications.

The selection of the errors is part of our future work (see Section VII). Regarding the benchmark applications, we decided to create new applications rather than using existing projects, similarly to the TbUIS (see Section III.A). The main reason is the consequent full control over these applications enabling us, among other things, to prepare and perform their very thorough testing.

Besides the thorough testing, there were several other requirements for the applications:

- Usage of relational database
- Usage of ORM
- Usage of web services
- Usage of file input and output
- Usage of command line interface (CLI)
- Optionally usage of third party libraries
- Optionally usage of simple GUI for debugging purposes and simple data input/output.

Based on these requirements, the resulting applications should use various common technologies and there should be an opportunity to introduce errors of very different types (such as a database error versus a web service error). The GUI was not considered essential, since the test generating methods are usually not focused on GUI testing. It is also not considered necessary for all the resulting applications to meet all the requirements.

Currently, there are two applications, which were developed by two of our bachelor students (see Acknowledgment section). The applications are two parts (frontend and backend) of a single system – a school agenda of an elementary or a high school. Both parts are described in Sections V.A and V.B. During the development of both applications, approximately half of the development time was devoted to the unit, integration, and functional testing to limit the number of errors unwittingly introduced during the development. Even then, the applications are expected to contain some errors. However, these errors are likely to be discovered during the usage of the applications as a part of the benchmark sooner or later. Once an unwittingly introduced error is discovered, it will be only documented and its discovery in the further usages of the benchmark will be observed.

A. Benchmark Application 1 – School Agenda Backend

Benchmark application 1 (BA1 – School Agenda Backend) manages data of elementary or high school agenda including students, teachers, classrooms, absences, and so on. It is a realistic application in sense that it would be utilizable for a real school, but some aspects may be missing in its data model. Additionally, there is only a basic HTTP authentication for the access of the web service.

The BA1 is a standard Java application with a layered architecture utilizing Spring Boot. The data are stored in a relational database using ORM via Hibernate. There is no GUI, the only interface of the application is the REST (Representational State Transfer) web service. The data transferred using the web service is in JSON format. Besides the Java Core API classes, the application utilizes several third-party libraries, such as Jackson, Hibernate, Spring boot, or JUnit among others.

The source code of the application consists of 77 classes in 10 packages with a total length of 328 kB. The source code of the unit and integration tests consists of 57 test classes with 655 tests with a total length of 448 kB. Functional testing consisting of 111 scenarios was performed manually.

B. Benchmark Application 2 – School Agenda Frontend

Benchmark application 2 (BA2 – School Agenda Frontend) provides the user interface for the school agenda. It communicates with the BA1 using the REST web service. It provides the JavaFX GUI for manual management of the data and CLI for bulk data import and export.

The BA2 is a standard Java application with a layered architecture. All the data are acquired from the BA1 REST web service, there is no direct access to the database. The data can be imported and exported from and into `.json` and `.xml` files. The application utilizes Jackson and JUnit among other libraries.

The source code of the application consists of 141 classes in 44 packages with a total length of 489 kB. The source code of the unit and integration tests consists of 38 test classes with 524 tests with a total length of 239 kB. Functional testing consisting of 437 scenarios was performed manually.

VI. Tests and Results

The prototype implementation of the TAG was tested in order to verify its ability to import and parse and export the applications, which are intended for the benchmark.

A. Testing Environment and Applications

The tests were performed on a standard notebook. Its hardware consists of dual-core Intel i5-6200U at 2.30 GHz, 8 GB of RAM, 250 GB SSD, and 500 GB HDD with 7 200 RPM. All the imports and exports were performed using the 500 GB HDD (not the 250 GB SSD). The installed software was Windows 7 SP1 64bit, and Java 11 (64 bit).

Five applications were used for testing. Each application was represented by a single folder with the entire project. The applications were developed using various IDEs and build tools leading to various folder structures. Also, the applications were developed by four different authors leading to different Java code styles and different utilized Java versions. All these features increase the variability of the applications and hence improve the quality of testing.

First two applications were the BA1 (see Section V.A) and the BA2 (see Section V.B). Second two applications were taken from a project focused on traffic assignment problem – the Dynamic Traffic Assignment (DTA) and Static Traffic Modeler (STM). The last application was the TAG itself. The features of the applications are summarized in Table I. The "Folder structure" describes the folder structure of the project. Three applications utilize a Maven/ Gradle-based structure while the two remaining utilize an Eclipse-based style. That does not mean that the same-based structures are identical, there are slight variations. The "Size to import" shows the total size of the folders and files, which are not ignored during the import. The "`.java` to import count" is the number of `.java` files contained in the imported folders and the "Others to import count" is the number of all other files contained in the imported folders.

B. Tests Description

The tests were performed the same way for each application. In the TAG, the application import was started and the root folder of the application project was selected as the input folder. Then, the structure of the project was explored and displayed as a tree. The tester marked the source and test folders and also the ignored folders. The ignored folders were the project settings folders, build and output folders, and version control folders. The application was then automatically imported. No errors were introduced into the imported application. Rather, the imported application was exported to a different folder without any functional changes.

Several parameters were observed – the number of folders and files, the number of packages and classes, the number of unsuccessfully parsed files (see Section IV.C), the import time, and the export time. Because of the time measurement, import and export of each application were performed four times. First time measurement was discarded, as it was significantly higher than the others, because the data from the disk was not present in the cache. Three other time measurements were averaged. Although only three measurements do not offer significant precision, it is enough to get a good idea of how long the export and import approximately last, which is the main purpose of the time measurement. The other parameters did not change between attempts, since the import and export are both deterministic.

Table I Features of the Applications Used for the Testing

Feature	BA1	BA2	DTA	STM	TAG
Folder structure	Maven/Gradle			Eclipse	
Size to import [kB]	835	2 856	899	9 945	11 415
`.java` to import count	134	179	78	305	62
Others to import count	25	408	97	34	39

Since the introduction of errors, which is the purpose of the TAG, was not part of these tests, the exported application should be identical to its imported counterpart. To determine this, the exported application was compiled and executed and manual functional testing of randomly chosen functionalities was performed. Moreover, all unit and aggregation tests present in the application were executed. Direct comparison of the imported and exported (i.e., generated) source code was not performed since there are non-functional differences, such as different indentation, empty lines, methods order, and so on.

C. Tests Results

The results of the testing are summarized in Table II. It can be observed that the import and export times are quite similar for a single application, with the export time being slightly higher in all instances. The times are also quite low, under a second in four of five applications, and under 2.5 seconds in the case of the STM. As such, the import and export times do not pose any problem for the TAG usage. The times seem to be influenced mainly by the number of parsed (and generated) `.java` files.

The parsing of `.java` files during the import works very well. There were no parsing errors in two applications, namely the BA1 and DTA. There were 5 files (2.9%), which were not parsed correctly, for the BA2, 13 files (3.6%) for the STM, and 10 (7.6 %) for the TAG. The parsing errors were caused by Java 14 `record` construction in the case of the BA2. In all other cases, the parsing error was caused by the usage of methods with type parameters (e.g., `<T> void foo(T t)`), which our parser currently does not support. This setback will be corrected as part of our future work (see Section VII). The unsuccessfully parsed files were stored as general files with their entire contents (see Section IV.C) and were correctly recreated similar to resource or library files during the export. The counts of these files were added to “Files count” row in Table II. After this adjustment, the numbers of actually imported files precisely correspond to the expected numbers of imported files (compare Table II “Files count” row and Table I “Others to import count” row).

The testing of the exported applications as described in Section VI.B was performed for all applications successfully, no errors unwittingly introduced by the TAG were found. This indicates that even the unsuccessfully parsed `.java` files during the import do not pose a problem as long as their number is low enough. A high number of unsuccessfully parsed files (e.g., 50 %) would significantly reduce the amount of source code, to which an error can be intentionally introduced. This, in turn, would reduce the usefulness of the application as a part of the benchmark.

TABLE II RESULTS OF THE APPLICATIONS EXPORT AND IMPORT

Feature	BA1	BA2	DTA	STM	TAG
Packages count	10	44	5	24	8
Classes count	134	174	84	363	132
Folders count	27	183	41	50	21
Files count	25	413	97	47	49
Unsuccessfully parsed files count	0	5	0	13	10
Import time [ms]	688	725	717	2 231	582
Export time [ms]	757	833	841	2 477	664

VII. Conclusion and Future Work

In this paper, we described the proposal for a benchmark of automated test generation methods consisting of realistic applications with artificially introduced errors. We focused primarily on the TAG tool, which enables the error introduction into imported applications and the export of multiple versions of multiple applications with various sets of introduced errors. The tests of the prototype implementation of the TAG were also described and its ability to import and export application was demonstrated using five applications. We also described first two applications, which are planned to be part of the benchmark.

In our future work, we will continue to work on the implementation of the TAG. These works include updating the parser to include newer Java constructions and the methods with type parameters. We also plan to improve user experience by automatically analyzing the structure of the imported folder and presetting all the types of files and folders to the correct types. The automatic compilation of the exported applications will be added as well.

We will also add common and realistic types of errors, which will be then introduced into the applications for their usage as the part of the benchmark. We plan to semi-automatically process publicly available contents of bug tracking tools to determine the common types of errors in developed and maintained applications and their frequency of occurrence. Then, we will utilize the obtained information to introduce realistic errors into our benchmark applications and finish the benchmark of automated test generation methods. Both the resulting benchmark and the TAG applications are planned to be made public, once they are finished.

Acknowledgment

The authors wish to thank to Martin Dub and Martin Jakubašek, who created the BA1 and BA2 applications, respectively, as part of their bachelor thesis.

References

[1] N. Gupta, A. P. Mathur, and M. L. Soffa, “Generating test data for branch coverage,” in Proceedings ASE 2000 - Fifteenth IEEE International Conference on Automated Software Engineering, Grenoble, September 2000, https://doi.org/10.1109/ASE.2000.873666

[2] P. Fröhlich and J. Link, “Automated Test Case Generation from Dynamic Models,” in ECOOP '00: Proceedings of the 14th European Conference on Object-Oriented Programming, Cannes, June 2000, pp. 472–491, https://doi.org/10.1007/3-540-45102-1_23

[3] B. S. Ahmed, K. Z. Zamli, W. Afzal, and M. Bures, “Constrained Interaction Testing: A Systematic Literature Study,” in IEEE Access, vol. 5, November 2017, pp. 25706–25730, https://doi.org/10.1109/ACCESS.2017.2771562

[4] Z. J. Rashid and M. F. Adak, “Test Data Generation for Dynamic Unit Test in Java Language using Genetic Algorithm,” in 2021 6th International Conference on Computer Science and Engineering

(UBMK), Ankara, September 2021, pp. 113–117, https://doi.org/10.1109/UBMK52708.2021.9558953

[5] R. J. Cajica, R. E. G. Torres, and P. M. Álvarez, "Automatic Generation of Test Cases from Formal Specifications using Mutation Testing," in 2021 18th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), Mexico City, November 2021, https://doi.org/10.1109/CCE53527 .2021.9633118

[6] H. Homayouni, S. Ghosh, I. Ray, and M. G. Kahn, "An Interactive Data Quality Test Approach for Constraint Discovery and Fault Detection," in 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, December 2019, pp. 200–205, https://doi.org/10.1109/BigData47090.2019.9006446

[7] A. Alsharif, G. M. Kapfhammer, and P. McMinn, "DOMINO: Fast and Effective Test Data Generation for Relational Database Schemas," in 2018 IEEE 11th International Conference on Software Testing, Verification and Validation (ICST), Västerås, April 2018, pp. 12–22, https://doi.org/10.1109/ICST.2018.00012

[8] M. Bures, P. Herout, and S. A. Bestoun, "Open-source Defect Injection Benchmark Testbed for the Evaluation of Testing," in Proceedings of the 13th IEEE International Conference on Software Testing, Validation and Verification (ICST), Porto, October 2020, pp. 442–447, https://doi.org/10.1109/ICST46399.2020.00059

[9] M. Kelly, C. Treude, and A. Murray, "A Case Study on Automated Fuzz Target Generation for Large Codebases," in International Symposium on Empirical Software Engineering and Measurement (ESEM), Porto de Galinhas, Septemmber 2019, https://doi.org/10.1109/ESEM.2019.8870150

[10] A. V. Pizzoleto, F. C. Ferrari, and G. F. Guarnieri, "Definition of a Knowledge Base Towards a Benchmark for Experiments with Mutation Testing," in SBES '21: Proceedings of the XXXV Brazilian Symposium on Software Engineering, Joinville, September 2021, pp. 215–220, https://doi.org/10.1145/3474624.3477060

[11] S. Varshney and M. Mehrotra, "A differential evolution based approach to generate test data for data-flow coverage," in 2016 International Conference on Computing, Communication and Automation (ICCCA), Greater Noida, April 2016, pp. 796–801, https://doi.org/10.1109/CCAA.2016.7813848

[12] J. Zhang, S. K. Gupta, and W. G. Halfond, "A New Method for Software Test Data Generation Inspired by D-algorithm," in 2019 IEEE 37th VLSI Test Symposium (VTS), Monterey, April 2019, https://doi.org/10.1109/VTS.2019.8758641

[13] H. V. Tran, L. N. Tung, and P. N. Hung, "A Pairwise Based Method for Automated Test Data Generation for C/C++ Projects," in 2022 RIVF International Conference on Computing and Communication Technologies (RIVF), Ho Chi Minh City, December 2022, https://doi.org/10.1109/RIVF55975.2022.10013824

[14] M. Motan and S. Zein, "Android App Testing: A Model for Generating Automated Lifecycle Tests," in 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), Istanbul, October 2020, https://doi.org/10.1109/ISMSIT50672.2020.9254285

[15] T. Potuzak and R. Lipka, "Current Trends in Automated Test Case Generation," in FedCSIS 2023, September 2023, to be published

[16] S. Poulding and J. A. Clark, "Efficient software verification: Statistical testing using automated search," in IEEE Transactions on Software Engineering, vol. 36, no. 6, February 2010, pp. 763–777, https://doi.org/10.1109/TSE.2010.24

[17] L. Bao-Lin, L. Zhi-shu, L. Qing, and C. Y. Hong, "Test Case Automate Generation from UML Sequence diagram and OCL expression," in 2007 International Conference on Computational Intelligence and Security (CIS 2007), Harbin, December 2007, pp. 1048–1052, https://doi.org/10.1109/CIS.2007.150

[18] Meiliana, I. Septian, R. S. Alianto, Daniel, and F. L. Gaol, "Automated Test Case Generation from UML Activity Diagram and Sequence Diagram using Depth First Search Algorithm," in Procedia Computer Science, vol. 116, October 2017, pp. 629–637, https://doi.org/10.1016/j.procs.2017.10.029

[19] M. Zhang, T. Yue, S. Ali, H. Zhang, and J. Wu, "A Systematic Approach to Automatically Derive Test Cases from Use Cases Specified in Restricted Natural Languages," in LNCS, vol. 8769, 2014, pp. 142–157, https://doi.org/10.1007/978-3-319-11743-0_10

[20] D. Xu, W. Xu, M. Tu, N. Shen, W. Chu, and C. H. Chang, "Automated Integration Testing Using Logical Contracts," in IEEE Transactions on Reliability, vol. 65, no. 3, November 2016, pp. 1205–1222, https://doi.org/10.1109/TR.2015.2494685

[21] H. Sharifipour, M. Shakeri, and H. Haghighi, "Structural test data generation using a memetic ant colony optimization based on evolution strategies," in Swarm and Evolutionary Computation, vol. 40, June 2018, pp. 76–91, https://doi.org/10.1016/j.swevo.2017.12.009

[22] T. Shu, Z. Ding, M. Chen, and J. Xia, "A heuristic transition executability analysis method for generating EFSM-specified protocol test sequences," in Information Sciences, vol. 370–371, November 2016, pp. 63–78, https://doi.org/10.1016/j.ins.2016.07.059

[23] S. Khor and P. Grogono, "Using a genetic algorithm and formal concept analysis to generate branch coverage test data automatically," in Proceedings. 19th International Conference on Automated Software Engineering, 2004, Linz, September 2004, pp. 346–349, https://doi .org/10.1109/ASE.2004.1342761

[24] C. Fetzer and Z. Xiao, "An automated approach to increasing the robustness of C libraries," in Proceedings International Conference on Dependable Systems and Networks, Washington D.C., June 2002, pp. 155–164, https://doi.org/10.1109/DSN.2002.1028896

[25] H. Tanno, X. Zhang, T. Hoshino, and K. Sen, "TesMa and CATG: Automated Test Generation Tools for Models of Enterprise Applications," in 2015 IEEE/ACM 37th IEEE International Conference on Software Engineering, Florence, May 2015, pp. 717–720, https://doi.org/10.1109/ICSE.2015.231

[26] T. Su, G. Pu, B. Fang, J. He, J. Yan, S. Jiang, and J. Zhao, "Automated Coverage-Driven Test Data Generation Using Dynamic Symbolic Execution," in 2014 Eighth International Conference on Software Security and Reliability (SERE), San Francisco, June 2014, pp. 98–107, https://doi.org/10.1109/SERE.2014.23

[27] L. Hao, J. Shi, T. Su, and Y. Huang, "Automated Test Generation for IEC 61131-3 ST Programs via Dynamic Symbolic Execution," in 2019 International Symposium on Theoretical Aspects of Software Engineering (TASE), Guilin, July 2019, https://doi.org/10.1109/TASE.2019.00004

[28] U. R. Molina, F. Kifetew, and A. Panichella, "Java Unit Testing Tool Competition: Sixth round," in SBST '18: Proceedings of the 11th International Workshop on Search-Based Software Testing, Gothenburg, May 2018, pp. 22–29, https://doi.org/10.1145/3194718.3194728

[29] X. Devroey, S. Panichella, and A. Gambi, "Java Unit Testing Tool Competition: Eighth Round," in ICSEW'20: Proceedings of the IEEE/ACM 42nd International Conference on Software Engineering Workshops, Seoul, June 2020, pp. 545–548, https://doi.org/10.1145/3387940.3392265

[30] S. Ali, L. C. Briand, H. Hemmati, and R. K. Panesar-Walawege, "A systematic review of the application and empirical investigation of search-based test case generation," in IEEE Transactions on Software Engineering, vol. 36, no. 6, August 2009, pp. 742–762, https://doi.org /10.1109/TSE.2009.52

[31] A. V. Pizzoleto, F. C. Ferrari, A. J. Offutt, L. Fernandes, and M. Ribeiro, "A Systematic Literature Review of Techniques and Metrics to Reduce the Cost of Mutation Testing," in Journal of Systems and Software, vol. 157, November 2019, https://doi.org/10.1016/j.jss.2019.07.100

[32] R. Jeevarathinam and A. S. Thanamani, "A survey on mutation testing methods, fault classifications and automatic test cases generation," in Journal of Scientific and Industrial Research (JSIR), vol. 70, no. 2, February 2011, pp. 113–117.

DOI: 10.15439/2023F1425
ISSN 2300-5963 ACSIS, Vol. 37

Efficient Feature Selection On Adversarial Botnet Detection

Farsha Bindu[1], Sheikh Sanjida[2], Nafiza Tabassoum[3], Tamima Binte Wahab [4], Raqeebir Rab[5], and Anonnya Ghosh[6]
[1,2,3,4,5]Department of Computer Science and Engineering
[1,2,3,4,5]Ahsanullah University Of Science and Technology (AUST), Dhaka, Bangladesh
[6]Software Developer, SOFTEKO Bangladesh
Email- farshabindo,sanjidasheikh7,nafisatabassum2016,wahabtamima@gmail.com
raqeebir.cse@aust.edu, 114anonnya@gmail.com

***Abstract*—Botnet attacks now pose a significant hazard to the security and integrity of computer networks and information systems. However, due to technological advancements and the proliferation of malware, machine learning-based Intrusion Detection Systems (IDS) are incapable of protecting against such cyberattacks. IDS cannot detect novel bots because the vast majority of them are programmed systems. Keeping IDS up-to-date with new malware varieties is, therefore, a crucial task. In this paper, we employ Generative Adversarial Networks (GANs) in which two neural networks compete and endeavor to outperform each other, which will serve as self-training for IDS. Our paper's primary objective is to develop an IDS capable of detecting novel malware with fewer attributes in real-time. To accomplish this, we present a method for feature selection that trains GAN models with a minimal subset of features so that the Generator can generate similar false bots with fewer features and the discriminator's ability to identify fake data improves. We used Pearson Correlation, the Wrapper method, and Mutual Information to select the best training model characteristics. The experimental evaluation suggests the GAN model in conjunction with Mutual Information is superior at detecting novel malware.**

***Index Terms*—Generative Adversarial Network, feature selection, Mutual Information, Wrapper Method, CNN**

I. Introduction

CURRENTLY, there is a growing interest in cyber security globally. As technology advances, hackers face new threats and opportunities for criminal activities. As more people, devices, and programs are added to modern business, as well as more data, the majority of which are sensitive or confidential, the significance of cyber security will only increase. This issue is exacerbated by the increase in the quantity and sophistication of hackers' attack methods. In its 2023 Cyber Security Report, the Check Point reflects on the challenging year 2022, when cyberattacks peaked as a result of the Russo-Ukrainian War [1]. A group of Ukrainian hackers has been interfering with Russian web services as a form of retaliation for Russia's invasion of the country. Compared to 2021, cyberattacks worldwide increased by 38 percent by 2022 [3]. We must be aware of the most significant intrusions of the previous year and what we learned from them as we approach 2023. [2]

Every business requires a secure digital infrastructure for conducting transactions. To achieve this goal, network architects and researchers are continuously attempting to create systems that provide impenetrable security for commercial websites. To promote economic growth, prosperity, efficiency, and security, governments must secure global digital infrastructure. With the rise of cyberattacks, machine learning and data mining have become crucial tools for addressing these problems. An anomalous network flow consists of outliers that deviate from typical user traffic patterns. Machine learning and data extraction enable more precise and rapid network traffic detection. They captured the data, analyzed the network flow, and classified the flows for detection purposes. However, data can be abundant, leading to low levels of precision, high computations, overfitting, and other issues. Only the correct selection of features can capture the correct network trace patterns. In other words, the essential characteristics of the network packets must be chosen. Additionally, redundant or irrelevant features must be eliminated.

We use generative adversarial networks (GANs) in this paper to build an adversarial machine learning attack on machine learning or deep learning-based intrusion detection systems (IDSs) when the adversary is uninformed of the ML technique used by the IDS. GANs are a type of generative model that is built on generator networks with recognizable outputs. A generator network and a network of discriminators compete in an interactive environment similar to that of game theory. The discriminator network's goal is to distinguish between samples from the original data and created data, whereas the generator network's goal is to build the best approximation of the training data.

Our contribution is an inclusion-exclusion-based feature selection integrated with Mutual Information (MI) for detecting bots. The objective of the proposed generative adversarial networks (GANs) model is to eliminate insignificant and redundant features as well as improve accuracy in detecting bots. We evaluated a number of feature selection strategies, including Pearson correlation, the wrapper method, and Mutual Information, to determine the optimal solution. Mutual Information combined with the exclusion-inclusion method demonstrates the optimal solution. In addition, GAN was used to generate false data and evaluate certain features with Convolutional Neural Networks (CNN). The experimental results based on the dataset CSE-CIC-IDS2018 [18] and dataset KDD-99 [16], demonstrate that the GAN model combined with the mutual information selection performs exceptionally well for IDS in detecting novel bots, with an accuracy of 85% and 83%, respectively.

II. Related Works

An intrusion detection system (IDS) is a system that monitors and analyzes data to detect any intrusion in the system or network. As they detect network attacks, intrusion detection systems (IDS) are currently one of the most crucial security solutions. Numerous machine learning and deep learning-based intrusion detection strategies have been proposed over the years [5] [6]. However, the majority of these methods have demonstrated significant false-positive rates and class imbalance problems. Muhammad Usama et al. [8] proposed a generative adversarial network (GAN)-based adversarial machine learning (ML) attack capable of evading an ML-based IDS. They extracted four crucial features necessary for a successful intrusion attack. A GAN framework includes three elements: a generator network, a discriminator network, and a classifier. The IDS model is trained using a generative model against known and unknown adversarial attacks. As evaluation criteria for the evasion assault, they used accuracy, precision, recall, and the F1 score. Among them, the Logistic Regression algorithm had the highest accuracy at 86.64%.

Chuanlong Yin et al. [7] presented the concept of modified GAN for creating false adversarial samples in order to improve the network system's performance in detecting bots. The name of the modified model is BOT-GAN. It is a framework for augmenting botnet detection models with generative adversarial networks, thereby enhancing detection performance and decreasing false positives. It retains the essential features of the original model. However, this paradigm is inefficient because it does not optimize network flow characteristics. The primary objective of this work is to detect novel botnets that are indifferent to network payloads and reduce the false-positive rate. Similarly, Rizwan Hamid et al. [9] proposed a technique called "Botshot" that generates plausible botnet traffic data using GANs to enhance detection. Two GANs (vanilla and conditional) are utilized to generate realistic botnet traffic. Using the classifier two-sample test (C2ST) with a 10-fold cross validation, the effectiveness of the generator is determined. In terms of average accuracy, precision, recall, and F1 score across six distinct ML classifiers, they evaluated the achieved results with benchmark oversampling techniques that included additional botnet traffic data. The showed the using the recall method, and the result was 98.65%. This system will detect a greater number of novel bots, and performance uncertainty will decrease. Francisco Villegas Alejandre et al. [10] proposed a genetic algorithm (GA) and a machine learning algorithm (C4.5), a novel technique for selecting features to detect botnets in their command and control (C&C) phase is presented. Their results demonstrated a reduction in the number of features and an increase in the detection rate. Giovanni Apruzzese et al. [11] proposed research in which they re-trained and re-tested each classifier with feature sets that do not contain the flow duration, sent bytes, received bytes, or exchanged packets, as well as all derived features. Multiple botnet detectors based on distinct machine learning classifiers were utilized. The accuracy increased from 72% to 75% by utilizing multilayer perceptrons and k-nearest neighbors.

So, based on previous research, we can conclude that there have been numerous studies on the performance improvement of IDS with GAN or feature selection separately. However, no work has proposed combining these two approaches for an effective solution for feature selection to detect botnets.

III. Dataset

To evaluate our model, we used two datasets one is known as KDD-99 and the CSE-CIC-IDS2018. KDD-99 has 42 features with binary class levels. The data are divided into two classes: Anomaly (53.1%) and Normal (46.1%). The data distribution of the KDD-99 is shown in Fig.1.

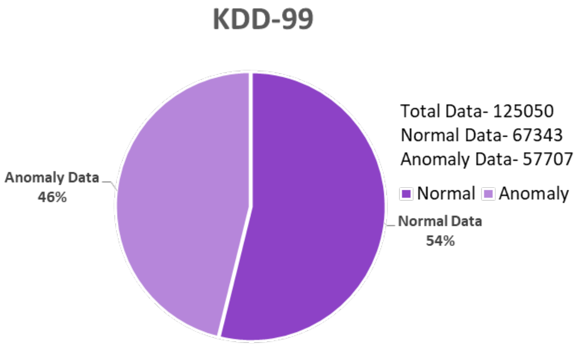


Fig. 1. Data Distribution of KDD-99 dataset

CSE-CIC-IDS2018 dataset contains 80 features for two binary classes. One class is named BOT, whereas another is named as benign. It is an imbalanced dataset. Benign data is 72.7% and Bot data is 27.3%. The data distribution of the CSE-CIC-2018 is shown in Fig.2.

Typically, a network connection consists of two flows, one for the uplink and the other for the downlink. Both the dataset contained the combination of a pair of up-and-down link flows. A short overview of some of the important features of both datasets is given in Table.I and Table.II

IV. Generative Adversarial Networks(GAN)

A generative adversarial network (GAN) is a well-known machine learning model for approaching generative artificial intelligence. In June 2014, Ian Goodfellow and his associates first conceived of the idea [4]. When two neural networks compete against one another in a GAN, a zero-sum game in

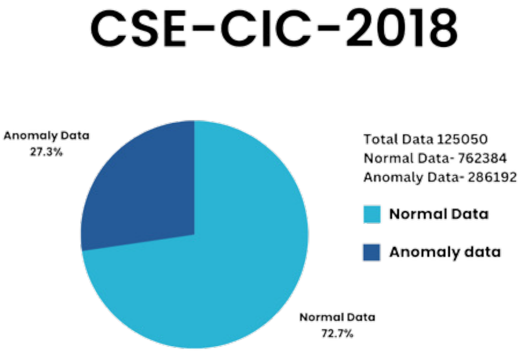


Fig. 2. Data Distribution of CSE-CIC-2018 dataset

TABLE I
DESCRIPTION ABOUT SOME FEATURES OF CIC-IDS2018

Features	Description
Dst Port	Destination port address
Flow Duration	The time elapsed between receiving the first and last packets in the flow
TotLen Fwd Pkts	Total number of forward packets
Fwd Pkt Len	forward packet length
Bwd Pkt Len Max	Maximum backward packet length
Flow Byts/s	Flow-Byte stands for the number of bytes transmit in one flow per second
Flow IAT	IAT is the arrival time difference between two packets. Flow IAT Mean is the average time gap between all sent packets in the flow(18).
Fwd IAT Mean	The average time between two packets sent in the forward packets flow.
Fwd IAT Std	The standard deviation of the time between two packets sent in the forward flow.

TABLE II
DESCRIPTION ABOUT SOME FEATURES OF KDD-99

Features	Description
duration	The length (number of seconds) of the connection
service	The network service on the destination, e.g., http, telnet, etc.
flag	The connection status (normal or problem)
src bytes	Quantity of data bytes transferred from source to destination
dst bytes	Quantity of data bytes from destination to source
wrong fragment	The amount of “wrong” fragments
logged in	1 if successfully logged in; 0 otherwise
is guest login	1 if the If the login is a "guest" login, 1; otherwise, 0.
count	Number of previous connections to the same host as the current connection in the last two seconds
srv count	Number of connections in the last two seconds to the same service as the current connection

which one agent’s gain is another agent’s loss occurs. The GAN training procedure is iterative, with the generator and discriminator networks trained in succession. The overview of the GAN model is shown in Fig. 3. The generator G learns to deceive the discriminator by transforming noise variables z into samples G(z), whereas the discriminator D is trained to maximize the probability of distinguishing between training examples and G(z). Both D and G use the following expression to maximize and minimize V (D, G) in an effort to enhance their learning process.

$$\min\max \mathrm{V(D,G)} = E\_x \sim p\_data(x)[logD(x)] + E\_z \sim p\_z[1 - logD(G(z))]$$

where V (D, G) = binary cross entropy function for binary classification problems, Pdata(x) = real data and Pz(z) = noise variable. [19]

V. DATA PREPROCESSING

The data must be processed prior to being fed into the machine learning models. Regarding the CSE-CIC-IDS2018 dataset for data processing, we initially converted all string-type data into numeric values. To accomplish this, we parsed the object data type into date and time data types of the timestamp feature. Subsequently, we converted the data and

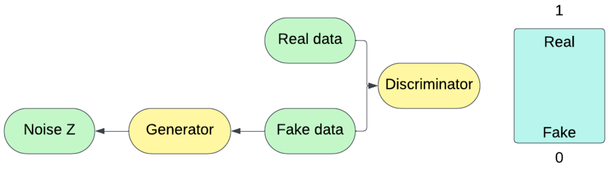


Fig. 3. Architecture of Generative Adversarial Networks

time data types to floats. Then, we normalized the binary label classes to zero and one. Here, Benign is regarded as zero, and Bot as one. (All samples of all features were also converted between 0 and 1). First, we identified constant characteristics, yielding a total of 12 constant characteristics. By eliminating 12 constant characteristics, we were left with 68 of 80 features. Additionally, duplicate attributes were identified. Five pairs of duplicate samples were identified. From each combination, the second characteristic was eliminated. Thus, we selected 63 out of the 68 features. In the final step of data preprocessing, we searched for null values and obtained no results. Then, we cell-transformed our dataset, which resulted in column names being converted to numbers. By completing all of these procedures, we have completed the preprocessing of our dataset. Our dataset was cleansed, normalized, and preprocessed so that our proposed models could be utilized effectively. Fig.4 depicts the data preprocessing processes.

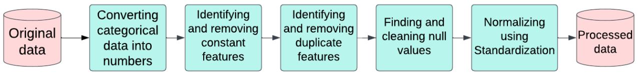

Original data
Converting categorical data into numbers
Identifying and removing constant features
Identifying and removing duplicate features
Finding and cleaning null values
Normalizing using Standardization
Processed data

Fig. 4. Data Preprocessing

Second, the KDD-99 dataset contained a single constant feature with no duplicates or missing values. Therefore, one feature was eliminated from the original 42 features, leaving us with 41 features. There were three categorical features in this dataset that were converted to numeric values. Additionally, we changed our class identifiers to 0 and 1. Here, 0 is an anomaly and 1 is normal. Then, we normalized all the values in the range 0 to 1 as a concluding step.

VI. Methodology

The objective of feature selection approaches in machine learning is to identify the optimal set of characteristics that permits the development of efficient models of studied phenomena. To get efficient features we employed three different feature selection methods–Correlation method, Wrapper method, and Mutual information.

A. Correlation method

We used the Pearson feature selection correlation method on our dataset CSE-CIC-IDS2018, to find the best-correlated feature pairs among the 63 features [13]. We took the pairs that had correlated values above 90%. The Pearson feature selection correlation method was used to select the best-correlated feature pairs, resulting in 38 features with 75% discriminator accuracy. To calculate the Pearson correlation coefficient, we take the covariance of the input feature X and output feature Y and divide it by the product of the standard deviation of the two features.

$$\rho X, Y = \frac{\sigma XY}{\sigma X \sigma Y}$$

B. Wrapper method

The feature selection wrapper method was tested on the dataset CSE-CIC-IDS2018 where firstly we did the forward selection method, which works with a p_value. It started with a null model and fitted it with each individual feature one at a time, selecting the feature with the minimum p_value. This process was repeated until the set of selected features had a p_value of individual features less than the significance level. However, 55 features were obtained from 63 using the forward selection, which did not produce the desired feature selection outcome [12]. Backward elimination was used to remove insignificant features from the discriminator model, resulting in 48 features out of 63 with the highest accuracy of 85%, almost the same as the initial accuracy. The same dataset CSE-CIC-IDS2018 was used for this method[12].

C. Exclusion/Inclusion with Mutual_Information

In this research, we present a novel method for selecting features that combine mutual information with feature exclusion and inclusion depending on the accuracy generated by the GAN model. The comprehensive overview of the suggested model is shown in figure5. Algorithm1 of our model are discussed below:

- In the CSE-CIC-IDS2018 dataset, we first applied mutual information (MI). Mutual information basically estimates the information about the amount of data one variable relates to another [14]. This allowed us to select the top 30 features with the highest information dependencies.
- From 30 features we started working top 5 features in the discriminator model which gave us 68% accuracy.
- We then included one-by-one features and checked if the accuracy of the discriminator was increased and continued the inclusion-exclusion process until the accuracy reached the initial accuracy with all 63 features, which was 85%. As a result, we got 20 features with 85% accuracy.

To verify the validity of our model, we used another binary dataset, KDD-99. The final methodology was used for this dataset. We estimated the mutual information of 41 features of this dataset after data preprocessing. The initial accuracy was 83% with 41 features. Then, using mutual information, we selected the top 30 features and started working with the top five features. Then, we included one-by-one features and checked if the accuracy was increased and continued the inclusion-exclusion process until the accuracy reached the initial accuracy with all 41 features, which was 83%. Consequently, we obtained 24 features with 83% accuracy.

Algorithm 1 Feature Selection Algorithm for proposed methodology

```
Require: Features
Ensure: Feature Set
  Initial accuracy is calculated by all features;
  FinalFeatureSet ← 5 features;
  Accuracy $a_m$ ← Top 5 feature Set;
  Set N ← Top features for which accuracy $a_m \approx$ Initial feature;
  i=6;
  while i=N features of Set do
    Calculate accuracy $a_i$;
    if $a_i$ ← > $a_m$ then
      FinalFeatureSet ← add i feature;
    else {}
      excludeSet ← i feature;
    end if
  end while
  FinalFeatureSet;
```

In our proposed algorithm Algorithm. 1 initial accuracy with all the features was calculated in Initial accuracy . Then FinalFeatureSet with the top 5 features, selected from mutual information was declared. Accuracy a_m with these features was calculated. In set N , the Top features with whom accuracy reached equal to the initial accuracy were stored . One by one feature was taken from this set and checked accuracy including or excluding it in the FinalFeatureSet and then it was added to FinalFeatureSet or excluded according to its performance.

D. Evaluating features using GAN

The Generator sub-model of GAN generated false novel data that resembled the original data, which was then optimized using feature selection and sent to the Discriminator sub-model to evaluate the accuracy of each feature. Discriminator made use of the Convolutional Neural Network (CNN). A CNN contains multiple layers, each of which learns to recognize the various characteristics of input data. A filter or kernel is applied to each data layer to generate an output that is progressively more accurate and detailed [15] [19].

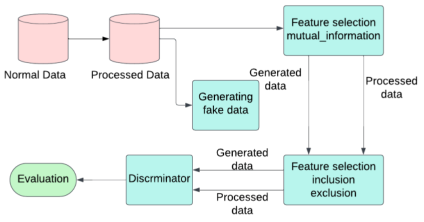


Fig. 5. Proposed Feature Selection Model

VII. Result Analysis

Both datasets were evaluated with the proposed feature selection model. As demonstrated in Table. III Mutual Information Feature Selection in conjunction with exclusion inclusion produces an optimal feature set with the utmost accuracy, comparable to the initial accuracy of all feature sets. The top five features selected by Mutual Information for Dataset CSE-CIC-IDS2018 began with an accuracy of 68%. Using inclusion and exclusion, we subsequently obtained 15 additional features with an initial accuracy of 85%. The same results were obtained for the KDD-99 dataset. The best five features of this dataset had a 56% accuracy rate. Using inclusion and exclusion, we subsequently obtained 19 additional features and attained an accuracy of 83%.

TABLE III
Performance Analysis with GAN

Dataset	**Number of selected features**	**Accuracy before feature selection**	**Accuracy after feature selection using MI with exclusion & inclusion**
CSE-CIC-IDS2018	20 from total 63	85%	85%
KDD-99	24 from total 41	83%	83%

Table IV displays the names of the features with the highest accuracy after exclusion and inclusion with Mutual Information.

A number of researches have been conducted to detect botnets implementing GANs. TableV outlined the comparative analysis of our model with other GAN-based models.

VIII. Conclusion and Future Works

As Internet usage increases, a growing number of threats are posing increasingly severe information security problems. There have been works of feature selection in network anomaly detection [22]. Despite their great potential, few IDS employ a class of algorithms known as generative adversarial networks. Therefore, we propose a GAN-based model capable of detecting novel malware with fewer features and greater accuracy. The maximum accuracy of the few previous studies that used GAN and feature selection was 74% [5]. Using the GAN and the Mutual Information Method, we achieved an accuracy of 85%. In the subsequent phase of our work, we intend to employ multiclass datasets in addition to binary-labeled datasets.

References

[1] Check point, title=Cyber security report 2023, url=https://pages.checkpoint.com/cyber-security-report-2023.html, note=(Date last accessed 27-July-2023)

[2] Michali, title = Biggest Cybersecurity Challenges in 2022, Check Point Software url=https://www.checkpoint.com/cyber-hub/cyber-security/what-is-cybersecurity/biggest-cybersecurity-challenges-in-2022, note=(Date last accessed 27-July-2023)

[3] Jnguyen, title =What is cyber security? the different types of cybersecurity, Check Point Software, url = https://pages.checkpoint.com/cyber-security-report-2023.html, note=(Date last accessed 27-July-2023)

TABLE IV
SELECTED FEATURES OF BOTH DATASETS USING EXCLUSION-INCLUSION WITH MUTUAL INFORMATION

NO	CSE-CIC-IDS2018	KDD-99
1	Dst Port	duration
2	Flow Duration	service
3	TotLen Fwd Pkts	flag
4	Fwd Pkt Len Max	src bytes
5	Fwd Pkt Len Mean	dst bytes
6	Bwd Pkt Len Max	wrong fragment
7	Flow Byts/s	logged in
8	Flow IAT Mean	is guest login
9	Flow IAT Max	count
10	Flow IAT Min	srv count
11	Flow IAT Mean	srv serror rate
12	Fwd IAT Std	rerror rate
13	Bwd PSH Flags	srv rerror rate
14	Fwd Header Len	same srv rate
15	Bwd Header Len	diff srv rate
16	Fwd Pkts/s	dst host count
17	Bwd Pkts/s	dst host srv count
18	RST Flag Cnt	dst host same srv rate
19	Bwd Pkt Len Std	dst host diff srv rate
20	Init Fwd Win Byts	dst host same src port rate
21		dst host srv diff host rate
22		dst host serror rate
23		dst host terror rate
24		dst host srv terror rate

TABLE V
COMPARISON ANALYSIS WITH OTHER WORK

Title	Methodology	Performance
[7]	GAN & Original Features	74%
[11]	GAN & Features Exclusion Based on ML models	75%
This work[18]	GAN, Exclusion inclusion in feature Mutual Information	85%
This Work[16]	GAN, Exclusion inclusion in feature Mutual Information	83%

[4] Goodfellow, Ian, et al. "Generative adversarial nets in advances in neural information processing systems (NIPS)." Curran Associates, Inc. Red Hook, NY, USA (2014): 2672-2680, doi = 10.1145/3422622, https://doi.org/10.1145\%2F3422622,year = 2020

[5] Chih-Fong Tsai and Hsu, Yu-Feng and Lin, Chia-Ying and Lin, Wei-Yang. "Intrusion detection by machine learning: A review." expert systems with applications 36.10 (2009): 11994-12000, https://www.sciencedirect.com/science/article/abs/pii/S0957417409004801, doi = 10.1016/j.eswa.2009.05.029, https://doi.org/10.1016/j.eswa.2009.05.029

[6] Modi, Chirag, et al. "A survey of intrusion detection techniques in cloud." Journal of network and computer applications 36.1 (2013): 42-57, doi = 10.1016/j.jnca.2012.05.003, https://doi.org/10.1016\%2Fj.jnca.2012.05.003,year = 2013.

[7] Yin, Chuanlong, et al. "An enhancing framework for botnet detection using generative adversarial networks." 2018 International Conference on Artificial Intelligence and Big Data (ICAIBD). IEEE, 2018, doi = 10.1109/icaibd.2018.8396200, https://doi.org/10.1109\%2Ficaibd.2018.8396200

[8] Usama, Muhammad, et al. "Generative adversarial networks for launching and thwarting adversarial attacks on network intrusion detection systems." 2019 15th international wireless communications & mobile computing conference (IWCMC). IEEE, 2019, doi = 10.1109/iwcmc.2019.8766353, https://doi.org/10.1109\%2Fiwcmc.2019.8766353

[9] Randhawa, Rizwan Hamid, et al. "Security hardening of botnet detectors using generative adversarial networks." IEEE Access 9 (2021): 78276-78292,doi = 10.1109/access.2021.3083421,https://doi.org/10.1109\%2Faccess.2021.3083421

[10] Alejandre, Francisco Villegas, Nareli Cruz Cortés, and Eleazar Aguirre Anaya. "Feature selection to detect botnets using machine learning algorithms." 2017 International Conference on Electronics, Communications and Computers (CONIELECOMP). IEEE, 2017,doi = 10.1109/conielecomp.2017.7891834, https://doi.org/10.1109\%2Fconielecomp.2017.7891834

[11] Apruzzese, Giovanni, Michele Colajanni, and Mirco Marchetti. "Evaluating the effectiveness of adversarial attacks against botnet detectors." 2019 IEEE 18th International Symposium on Network Computing and Applications (NCA). IEEE, 2019,doi = 10.1109/nca.2019.8935039, https://doi.org/10.1109\%2Fnca.2019.8935039

[12] Vikas Verma, title = A comprehensive guide to Feature Selection using Wrapper methods in Python, url=https://shorturl.at/gnqET, note=(Date last accessed 27-July-2023)

[13] Mehreen Saeed, title= Calculating Pearson Correlation Coefficient in Python with Numpy, url=https://shorturl.at/abfim, note=(Date last accessed 27-July-2023)

[14] Guhanesvar, title= Feature Selection Based on Mutual Information Gain for Classification and Regression, url = https://bit.ly/3ofYzmt note=(Date last accessed 27-July-2023)

[15] IBM — ibm.com, title= What are Convolutional Neural Networks?, url = https://www.ibm.com/topics/convolutional-neural-networks. note=(Date last accessed 27-July-2023)

[16] kDD-99 dataset url = http://kdd.ics.uci.edu/databases/kddcup99/task.html, note=(Date last accessed 27-July-2023)

[17] Choudhary, Sarika, and Nishtha Kesswani. "Analysis of KDD-Cup’99, NSL-KDD and UNSW-NB15 datasets using deep learning in IoT." Procedia Computer Science 167 (2020): 1561-1573., doi = 10.1016/j.procs.2020.03.367, https://doi.org/10.1016\%2Fj.procs.2020.03.367 ,year = 2020.

[18] CSE-CIC-IDS2018 dataset, urlhttps://www.unb.ca/cic/datasets/ids-2018.html?fbclid=IwAR2dCUq0TM0kzlKG6eTX23TkEueKSUwmUh5coYQJidfMLn7rcs-4ICt4Fy8, note=(Date last accessed 27-July-2023),

[19] Zhang, Xiran. "Network intrusion detection using generative adversarial networks." (2020). https://ir.canterbury.ac.nz/bitstream/handle/10092/100016/Zhang,\%20Xiran_Master’ s\%20Thesis.pdf?isAllowed=y&sequence=1

[20] Paolo Caressa, title=How to build a GAN in Python , url=hhttps://www.codemotion.com/magazine/ai-ml/deep-learning/how-to-build-a-gan-in-python/, note=(Date last accessed 27-July-2023),

[21] K. Cabaj and J. Wytrębowicz, S and Kukliński ,P. Radziszewski and K. Truong Dinh. "SDN Architecture Impact on Network Security" Position papers of the 2014 Federated Conference on Computer Science and Information Systems pp. 143–148,year=2014, doi = 10.15439/2014F473, http://dx.doi.org/10.15439/2014F473

[22] Ghosh, Anonnya and Ibrahim, Hussain Mohammed and Mohammad, Wasif and Nova, Farhana Chowdhury and Hasan, Amit and Rab, Raqeebir. "CoWrap: An Approach of Feature Selection for Network Anomaly Detection" In: Barolli, L., Hussain, F., Enokido, T. (eds) Advanced Information Networking and Applications. AINA 2022. Lecture Notes in Networks and Systems, vol 450. Springer, Cham., year=2022, doi = 10.1007/9783030995874_47

DOI: 10.15439/2023F7757
ISSN 2300-5963 ACSIS, Vol. 37

To propose an attack detection model for enhancing the security of 5G-enabled Vehicle-to-Everything (V2X) communication for smart vehicle

Prince Rajak,
Dept. of Information Technology
National Instiute of Technology Raipur
Raipur, India
rajakprince123@gmail.com

Pavan Kumar Mishra
Dept. of Information Technology
National Instiute of Technology Raipur
Raipur, India
pavan_km.it@nitrr.ac.in

***Abstract*—The evolution of 5G technology has revolutionized the communication landscape, enabling faster speeds, low latency, and increased capacity. The integration of 5G technology and the emergence of the 5G-enabled V2X communication network are driving the transformation of the automotive industry. Connected cars and the software-defined vehicle concept enable new business models and enhanced safety measures. However, ensuring the security of the 5G-enabled V2X communication network is crucial to mitigate potential attacks and protect the integrity of the ecosystem. To identify this potential attack, we have proposed a novel deep learning-based attack detection model (ADM) for detecting attack in 5G-enabled V2X communication network. In this we have used correlation coefficient as the feature selection method and used the deep learning-based stacked LSTM model for attack detection. The performance metrices are detection rate, accuracy, precision and F1-score.**

***Index Terms*—5G, V2X Communication Network, Stacked LSTM, ADM.**

I. Introduction

THE AUTOMOTIVE industry is undergoing a profound transformation, driven by the integration of 5G technology and connected cars. 5G has become a pivotal component in revolutionizing how automotive OEMs (original equipment manufacturers) design, build, and operate their vehicles, as well as how customers interact with them [1]. This evolution is not limited to hardware advancements but extends to the software-defined vehicle concept, where a significant portion of a vehicle's functionalities are implemented through software that can be updated or even upgraded over time. This paradigm shift allows automotive OEMs to embrace new business models, such as subscription-based or on-demand "as-a-service" offerings, and opens up opportunities for monetization.

A key enabler of this transformative journey is the 5G-enabled Vehicle-to-everything (V2X) communication network. V2X leverages the power of 5G to create a highly connected ecosystem where vehicles can communicate not only with other vehicles but also with various entities, including infrastructure, pedestrians, and smart city systems. This connectivity enables the real-time exchange of critical information, leading to enhanced road safety, improved traffic management, and a more efficient use of resources. There are several types of 5G-enabled V2X communication networks [2] that relate to different aspects of V2X connectivity and some of them are as follow:

- Vehicle-to-Vehicle (V2V): It enables direct communication between vehicles. This allows vehicles to share information such as speed, position, acceleration, and braking, fostering cooperative behaviour, enhancing safety on the road, and also providing the necessary updates on collisions, etc.
- Vehicle-to-Infrastructure (V2I): V2I communication networks involve the exchange of information between vehicles and infrastructure components such as traffic lights, road signs, and toll booths and receive real-time traffic updates, traffic signal timing information, and road condition alerts, optimizing traffic flow and improving overall efficiency.
- Vehicle-to-Pedestrian (V2P): V2P communication networks focus on the interaction between vehicles and pedestrians. These networks allow vehicles to detect the presence of pedestrians and provide warnings to both the driver and the pedestrian, reducing the risk of accidents and enhancing pedestrian safety.
- Vehicle-to-Network (V2N): V2N communication networks involve the interaction between vehicles and the cellular network infrastructure. Vehicles can access cloud-based services, download software updates, and leverage network resources to enhance their functionalities and performance.
- Vehicle-to-Device (V2D): V2D communication networks encompass the connectivity between vehicles and external devices, such as smartphones, wearables, or smart home systems.

However, as with any connected system, the 5G-enabled V2X communication network faces potential security threats and attacks. V2X communication networks are susceptible to various types of attacks due to their interconnected and wireless nature. Adversaries may exploit vulnerabilities in the network infrastructure, software, or hardware components to launch attacks with malicious intent. Some common attack on V2X communication networks include: Man-in-the-Middle (MitM) attack, DoS, Distributed Denial of Service (DDoS), Spoofing attack, etc [3]. These attacks can have severe consequences, including compromised vehicle control, privacy breaches, and even physical harm to road users. Therefore, it is imperative to address the security challenges and implement effective prevention techniques to ensure the integrity and reliability of the 5G-enabled V2X communication network [4]. To handle these issues in this paper we have presented the novel ADM based on deep learning techniques which uses correlation coefficient-based feature selection method as its core and stacked Long Short-Term Memory (LSTM) based attack detection model. The model is tested with newly release dataset.

The remaining portions of this paper are organised as follows: In Section II, we provide a brief overview of related work. In Section III, we discuss occurrences of attack on 5G enabled V2X communication network. In Section IV, we illustrated the architecture of the proposed attack detection model is described. In Section V, the effectiveness of the proposed detection model is evaluated using AIoT-SoL dataset. In section VI of the paper served as its conclusion.

II. Literature Survey

Many researchers have been interested in V2X communication network and 5G network security in the past, and many researchers are working on different parts of 5G technology right now. Many studies on 5G network attack detection have been conducted. There are various feature selection and reduction methodologies discussed in the literature for attack detection. The author [5] proposed an IDS for connected vehicles for smart cities environment. In this Deep Belief Network is used as the dimension reduction method and uses Decision Tree as the classifier for attack detection model. The dataset used are NSL-KDD and NS-3 Network simulator for model evaluation.

The author [6] proposed the anomalous event detection for intelligent transportation systems. They proposed an LSTM-based Autoencoder. Here they have extracted the feature in two phase such as Feature extraction phase (FEP) and Statistical FEP and then train using the LSTM autoencoder. The dataset used are car hacking dataset and UNSW-NB15. The author [7] proposed an IDS for IoV. The hybrid deep learning method consist of LSTM and GRU (Gated Recurrent Unit) layers are developed. They combined the DDoS dataset which is the combination of CIC_DoS 2016, CIC_IDS 2017, and CSE-CIC_IDS 2018 and make the binary label dataset consist of normal and attack labels. The car attack dataset is also used which shows higher performance.

The author [8] proposed an intrusion detection approach to early detect the cyber–physical attacks targeting Fast Charging Station (FCS) considering Vehicle-to-Grid (V2G) operation. In this discreate power samples data is used and use the Gini index for calculation of power mid-point and then use the DT for the detection of DoS attack. They create their own test environment with 3 DoS attack profile and generate data for 4 Scenario and the major feature is power which help in identification of attack on FCS. Author [9] work on in-vehicle network to proposed the IDS based on DCNN (Deep Convolution Neural Network) to protect CAN bus of the vehicle. The DCNN is built by removing the unwanted layer from the ResNet architecture and reducing the model complexity. The dataset is self-generated which consist of five types of DoS attack label.

The authors of [10] proposed tree-based ensemble intrusion detection using the stacking ensemble methodology. They used the selectkbest feature selection method and selected the top 20 significant features from NSLKDD and UNSW-NB15. The dataset consists of binary classes. The author [11] suggested a software-defined network with an IDS that is smart for the 5G network. Firstly, we performed feature importance to select the significance feature using random forest. Then we used k-means clustering to divide the traffic into five clusters, i.e., normal and attacking clusters, and used Adaboost as a classifier. For the evaluation, they used the KDD Cup 1999 dataset. There are many cases discussed by different researchers [12] that the KDDCup1999 is biased towards eliminating redundancies, which helps all of them achieve higher accuracy.

III. Occurance of attack over the 5G-enabled V2X communication network scenrio

The 5G-enabled V2X communication network is a promising technology that has the potential to revolutionize automobile industry and many other services. The transition to 5G technology brings both opportunities and challenges. With increased bandwidth, lower latency, and improved connectivity, 5G enables faster and more efficient communication between vehicles and the surrounding infrastructure. To protect the integrity, privacy, and security of the v2X communication ecosystems, new security issues are also raised that must be resolved.

5G-enabled V2X communication networks are vulnerable to a variety of attacks. These attacks can be carried out by malicious individuals or organizations, and they can have a significant impact on the safety and security of vehicles and infrastructure. Some of the most common attacks that can be carried out on 5G-enabled V2X communication networks include [13]:

- *Denial of Service (DoS) and Distributed Denial of Service (DDoS) Attack:* These attacks aim to disrupt the availability and performance of the V2X network by overwhelming it with a flood of illegitimate requests or malicious traffic. By overloading the network resources or specific V2X components, attackers can prevent legitimate communication and potentially cause safety risks on the road.

- *Man-in-the-Middle (MitM) Attack: In* this type of attack, an unauthorized entity intercepts and alters the communication between two legitimate parties. In the context of a 5G-enabled V2X network, an attacker could position themselves between a vehicle and another vehicle, infrastructure, or device, and manipulate the information being exchanged. This could lead to unauthorized access, data tampering, or the injection of malicious commands.

- *Spoofing Attacks:* In a spoofing attack, an attacker impersonates a legitimate entity or device to deceive other participants in the V2X network. This could involve forging the identity of a vehicle, an infrastructure unit, or even a traffic management system, leading to unauthorized access or the manipulation of information. For example, an attacker could send false traffic information or alter the location of a vehicle, causing confusion or accidents.

- *Eavesdropping Attacks:* As V2X communication transmits sensitive information, such as location data or personal details, eavesdropping attacks pose a significant threat. Attackers may attempt to intercept and capture this information to gain insights into driver behaviour, track vehicle movements, or conduct targeted attacks based on the obtained data.
- *Malware and Code Injection:* Attackers could develop and deploy malware specifically designed to exploit vulnerabilities in the software or firmware of V2X components. Once compromised, the attacker can take control of these devices, potentially enabling unauthorised access, data theft, or further network exploitation.
- *Sybil Attacks:* It occurs when an attacker generates multiple fake identities or vehicles in the V2X network, effectively multiplying their influence and capabilities. This can lead to malicious activities such as flooding the network with false information, manipulating traffic patterns, or disrupting the overall operation of the system.
- *Physical Attacks:* In addition to cyber threats, physical attacks on the infrastructure supporting the 5G-enabled V2X network could also disrupt its functionality. For example, an attacker could physically damage communication equipment, power sources, or sensor arrays, leading to communication failures, misdirection, or safety hazards.

IV. Proposed Methodology

In this section, Fig. 2 shows a detailed description of the proposed framework. It is also showing a detailed description of an attack detection model (ADM) for V2X communication network environment with 5G-enabled network. This ADM used the correlation-based feature selection for removal of highly correlated features which will help in improving the performance and reducing the computational cost. The phases and the functional component of this framework include data pre-processing of

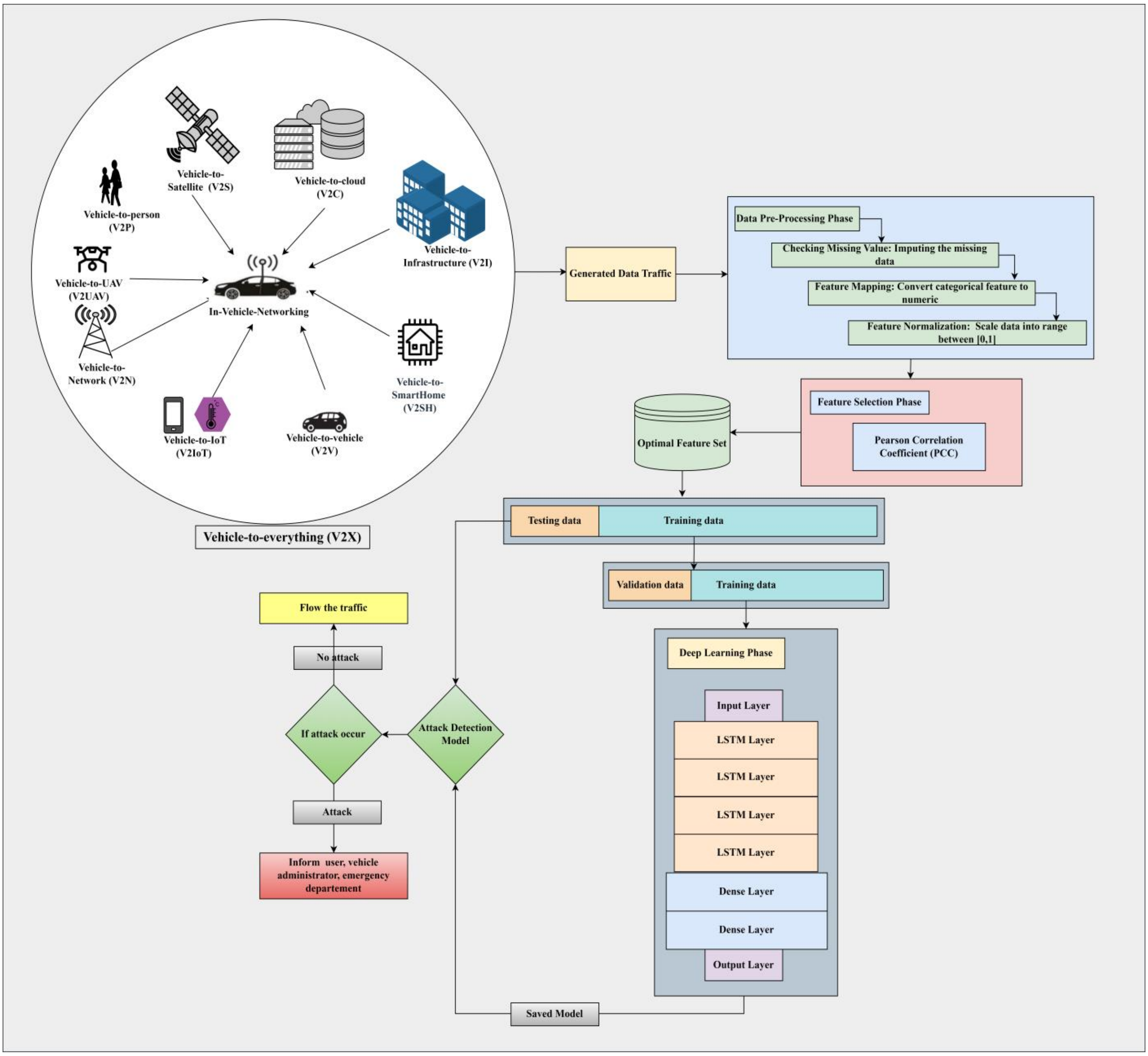


Fig. 1. Working architecture of proposed framework for an ADM for 5G-enabled V2X communication network using the Stacked LSTM

incoming V2X communication network traffic and used correlation coefficient-based feature selection method. Then the selected features are used optimal feature set to train the deep learning based Stacked LSTM-based attack classification model to identify the attack in a 5G-enabled V2X communication network. Each phase of the proposed model contributes significantly to the improvement of the overall performance. Following is a concise description of each phase of the proposed work:

A. Data Preprocessing Phase

This section describes in detail the data preparation steps for the proposed ADM. Due to the fact that V2X communication network traffic is generated by various smart sensors and smart vehicles in the 5G-enabled V2X communication environment, it includes both categorical and numeric values. It is necessary to pre-process this traffic in order to design an effective and robust ADM. Following is a brief summary of the discussion:

1) Checking Missing Values: The incoming network traffic data is missing a few values. These missing values should be handled correctly, as they will affect the model's overall performance. There are numerous methods for dealing with missing values. In the proposed detection system, missing values are substituted with the mean of the specific characteristics present in the dataset. Imputing these values will improve the quality of the data and increase the classifier's predictive ability.

2) Feature Mapping: The 5G-enabled V2X communication network traffic may frequently include various categorical variables. Some components of the ML and DL model are incompatible with categorical variables. As a solution, one must convert these categorical values to numeric values, we have used one-hot encoder techniques. This method will make the dataset sparser and increase the number of features to match the number of distinct categorical values present in each categorical feature.

3) Feature Normalization: It is observed in the network traffic data that attributes have drastically different scales, which will result in slow data-driven learning. By comparing the characteristics of each data point, a number of machine learning algorithms attempt to discover patterns hidden in the data. Different scales can result in increased processing and resource consumption. MinMaxScaler normalization techniques are employed in the proposed attack-detection model in order to discover hidden patterns, reduce training time, and enhance convergence. The MinMaxScaler techniques transform all dataset values by scaling each feature to a given range, in this case, 0 to 1. The following formula, as shown in equation 1, is utilized.

$$f_n = \frac{f_i - f_{\min}}{f_{\max} - f_{\min}} \tag{1}$$

Where f_i is the feature to scaled down, f_{min} is the minimum value and f_{max} is the maximun value for a particular feature present in the dataset and f_n is the new value.

B. Feature Selection Phase

Feature selection is the process of identifying and selecting the best subset of features from a dataset. For a machine learning and deep learning model to function well and have a quicker predictive power, these features are essential. Since the V2X communication network traffic contains a variety and a very large number of features, processing this data would require a significant amount of power and time if all features were used. The issue is resolved by using feature selection techniques to decrease the power and time without significantly affecting the classifier's predictive ability. Processing time and testing are also reliable, and feature selection aids in reducing the size of the training dataset. Numerous studies were compared and showed that data with repetitive and irrelevant features had an adverse effect on the accuracy of the learning model.

Correlation Coefficient: In next step correlation coefficient-based feature selection techniques is used. Most of the incoming V2X communication network traffic contains different features which are irrelevant and have same pattern of values. The independent features which are uncorrelated with the target attribute are the best candidate to be removed. In addition, if two independent features are highly correlated with each other than using both of them will not increase the performance but will make detection model more complex and will increase the computation time. Therefore, the proposed model removes all the features from the raw dataset and provides a simpler model for easy and efficient detection of attacks. In this model, PCC (Pearson Correlation Coefficient) is used. It works on the principle that tries to calculate the linear dependencies of two features. If the features are independent, then PCC value will be 0 and if features are dependent then the PCC value will be ± 1. So, it tries to calculate the covariance of the two features divided by the product of their standard deviation, the representation is given in equation 2.

$$p(f,t) = \frac{\mathrm{cov}(f,t)}{\sigma_f * \sigma_t} \tag{2}$$

where p is PCC, $f \epsilon \{F_1, F_2, F_3, F_{4, \dots}, F_n\}$, t target variable, cov is the covariance between the features f and t and σ_f , σ_f is the standard deviation of features and target variable.

The covariance between the features and the target is computed using equation 3, standard deviation σ_f , σ_t can be calculated using equation 4. and the mean μ_f , μ_t of feature and target is calculated in equation 5.

$$\mathrm{cov}(f,t) = \frac{1}{N}\sum_{i=1}^{N}(f_i - \mu_f)(t_i - \mu_t) \tag{3}$$

$$\sigma_k = \sqrt{\mathrm{var}\left(\sigma_k^2\right)} = \frac{1}{N}\sum_{i=1}^{N}\left(f_i - \mu_k\right)^2 \tag{4}$$

$$\mu_k = \frac{\sum_{i=1}^{N} val_i}{N} = \frac{(val_1 + val_2 + val_3 + \cdots + val_N)}{N} \tag{5}$$

Where $k \epsilon (f,t)$ for feature and target, N is the total number of values present in each feature, var(σ_k^2) is the variance and val_i is the values for the corresponding features in the dataset.

Above all the equation are used as a function to rank and to obtain the optimal feature set.

C. Attack Detection using Deep Learning Techniques

The proposed V2X communication network AMD uses the deep learning based Stacked LSTM techniques. It is the type of recurrent neural network (RNN) that has multiple LSTM layers which allows the model to learn increasingly complex pattern and representation. The advantages of using a Stacked LSTM model are that improved the performance and can able to handle complex data and pattern. Second it learns multiple level of abstraction from input data which lead to better feature representation and improved detection power and has more efficient than training than CNN and other deep learning method. The Stacked LSTM are less prone to overfitting than other deep learning architectures because of their ability to learn multiple levels of abstraction from the input data.

In the proposed deep learning architecture, we have used the combination of LSTM and Dropout layers. Four LSTM layers are configured to return sequences, meaning they pass their outputs to then next LSTM layers. The input shape of the first LSTM layer is specified based on the shape of the training data. Each LSTM layer consists of 64 units, which represent the number of hidden units or memory cells in the layer. After each LSTM layer, a dropout layer is added. Dropout is a regularization technique that randomly sets a fraction of the input units to zero during training, preventing overfitting by reducing interdependencies between neurons. Following the LSTM layers, two dense layers are added. Dense layers are fully connected layers where each neuron is connected to every neuron in the previous layer. The first dense layer has 32 units and uses the ReLU activation function, which introduces non-linearity into the model. A dropout layer is added after the first dense layer to further prevent overfitting. The second dense layer has 16 units with the ReLU activation function. Finally, a dense layer with five units and the softmax activation function are added. The softmax function converts the model's outputs into probabilities, indicating the likelihood of each class. The model is trained using the categorical cross-entropy loss function, which is suitable for multi-class classification problems. The model is compiled with the Adam optimizer, which is an efficient optimizer for gradient-based optimization algorithms. During training, the model's performance is evaluated using accuracy as a metric.

V. Experiment Result

This section presents a brief discussion about the AIoT-SoL dataset used with the proposed work. The proposed experiment is conducted using the Python programming language, and the system configuration is described in Table I.

TABLE I. SYSTEM CONFIGRATION DESCRIPTION

System	Configuration
Processor	Intel(R) Xeon(R) CPU E3-1240v6 @ 3.70GHZ
RAM	16 GB
GPU	NVIDIA Quadro P1000 4 GB
Operating System	Window 10
Programming Language	Python 3.9.12

A. Dataset Description

AIoT-SoL Dataset: For constructing the proposed attack detection model, we have used the AIoT-SoL dataset. The authors [14] created this new dataset because they noticed that the existing data does not more closely relate to web- and IoT-specific attacks. However, they have released the AIoT-SoL dataset, which contains a greater variety of attack types than others but does not include botnet attacks because they are available in the existing dataset. Some of the attack types may overlap with the existing ones, but they crafted a more realistic and comprehensive attack scenario for data generation. The dataset is generated from the testbed, which consists of various IoT devices, vulnerable applications as victim machines, MQTT components, and different software to connect the testbed. They generated both benign and attack traffic. The attack traffic consists of 16 types of attacks, which are discussed in Table II. These 16 attacks are classified into 4 categories: web, denial of service, network, and web IoT Message Protocol attacks. They have used CIC FlowMeter [14] for extracting features from the pcap file and generating them into the csv file

TABLE II. THE AIOT-SOL DASET ATTACK TYPE CATEGORIZATION

Binary Categories	Categorical Categories	Sub-Categorical Categories	Instance
Benign	Benign	Benign	2403450
Anomaly	Denial of Service Attack	SSL Regression Attack	10454
		SYN Flood Attack	1000000
		SSDP Flood Attack	970418
	Network Attack	ARP MITM Attack	2307
		Network Logan Bruteforce	44915
		Network Scanning	105417
	Web attack	Cross- site Request Forgery	5117
		Cross-site Scripting	2676
		XML External Entity	47459
		Open Redirect	1237
		Directory Traversal	4998

		Server-side Request Forgery	1756
		Command Injection	12894
		SQL Injection	16046
		Web Directory Bruteforce	23031
	Web IoT Message Protocol Attack	MQTT Bruteforce	482558

The AIoT-SoL dataset is publicly available at GitHub [16]. The size of the dataset is 1.82 GB in a csv file, and it consists of 85 attributes and 5134728 instances, with 2403450 benign instances and 2731278 attack instances. The label attributes consist of 5 different types of categories, such as Benign, DoS Attack, Network Attack, Web Attack and Web IoT Message Protocol Attack.

In the proposed work, we have only considered categorical category data because for sub-categorical dataset is imbalance in nature. So, we try to evaluate our proposed attack detection model on a five categorical data respectively. We have not chosen binary categories as more work has been done on binary attack detection.

B. *Description of Evaluation Metrics*

The most commonly used evaluation metrics for attack detection are accuracy, precision, detection rate, and F1-Score For computing these metrics, various parameters are required and used, which are given as follows:

- True Positive (T_P): It calculates the number of attack instances in the MassiveIoT network traffic that are correctly classified as attacks by the detection model.
- True Negative (T_N): It calculates the number of benign instances in the MassiveIoT network traffic that are correctly classified as benign by the detection model.
- False Positive (F_P): It calculates the number of benign instances in the MassiveIoT network traffic that are incorrectly classified as attacks by the detection model.
- False Negative (F_N): It calculates the number of attack instances in the MassiveIoT network traffic that are incorrectly classified as benign by the detection model.

From the above discussed parameter, we can evaluate the detection model through various metrics. These metrics are discussed below:

Accuracy: It calculates the ratio of correctly classified instances by the detection model to the total number of instances present in the test set. It takes both into account when calculating the accuracy of the model.

$$\text{Accuracy} = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{7}$$

Precision: It calculates the number of attack activities detected by the detection model divided by the total number of instances detected as attacks by the detection model.

$$\text{Precision} = \frac{T_P}{T_P + F_P} \tag{8}$$

Detection Rate: It calculates the number of attack activities detected by the detection model divided by the total number of activities present in the test set and is also known as recall.

$$\text{Detection Rate} = \frac{T_P}{T_P + F_N} \tag{9}$$

F1-Score: It calculates the weighted average of precision and recall. It is primarily used when the class distribution is imbalanced and is more useful than accuracy as it takes decision-making into account.

$$\text{F1 Score} = 2 * \frac{(\text{ Precision } * \text{ Recall })}{(\text{ Presion } + \text{ Recall })} \tag{10}$$

C. *Result Analysis*

For performance evaluation of the Stacked LSTM classifier, we split the AIoT-SoL into two parts in the ratio of 70:30 which is known as 70 for training and 30 for testing. Again, we use the 70 percent training data and spit it into the ratio of 80: 20 for generating the validation set for deep learning model training. The size of training data contains 2846748 rows, validation data contain711688 rows and testing data contain 1525044 rows. In Table III we have shown the performance evaluation of proposed work with validation and testing set. In Table IV we have shown the class-wise detection rate, precision, and F1-Score of validation and testing set. In Table V we have shown the comparison of proposed work with other existing work. Figures 2 and 3 represent the training and validation loss and accuracy for Stacked LSTM model. In Figs. 4 and 5 we have shown the confusion matrix of testing set and validation set. Our proposed model work well and have a good detection rate to detect the attack.

TABLE III. PERFORMANCE OF PROPSED WORK WITH AIOT-SOL VALIDATION SET AND TESTING SET DATA

Evaluation Set	**Accuracy**	**Precision**	**Detection Rate**	**F1-Score**
Validation	99.8	99.1	99.1	99.1
Testing	99.8	99.2	99.2	99.1

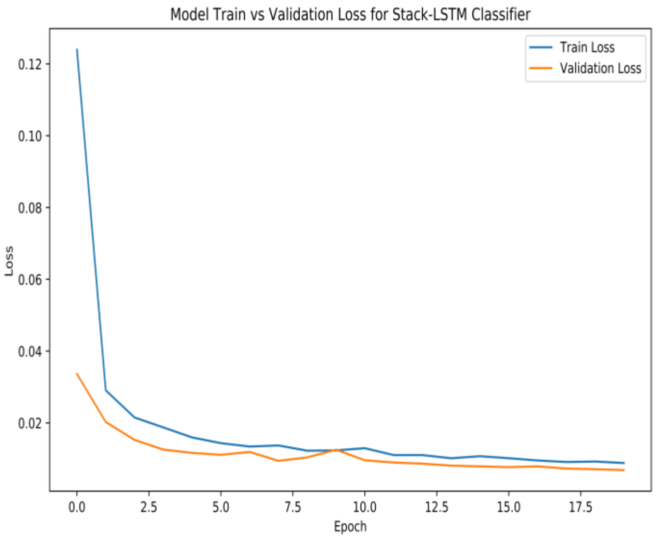


Fig. 2. Training and validation loss of the Stacked LSTM model

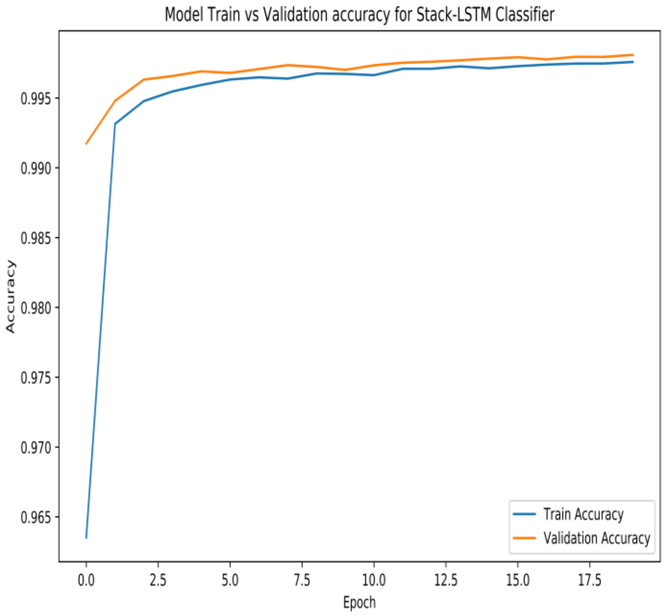
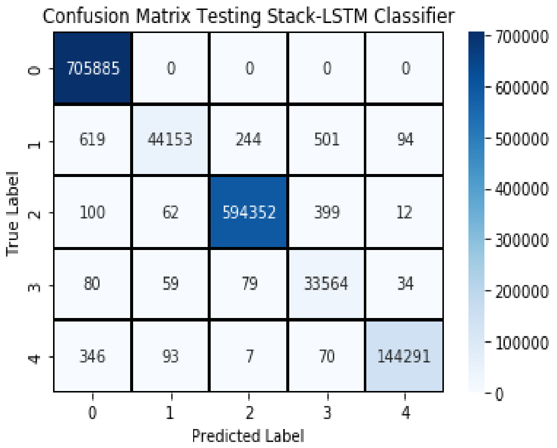


Fig. 3. Training and validation accuracy of the Stacked LSTM model

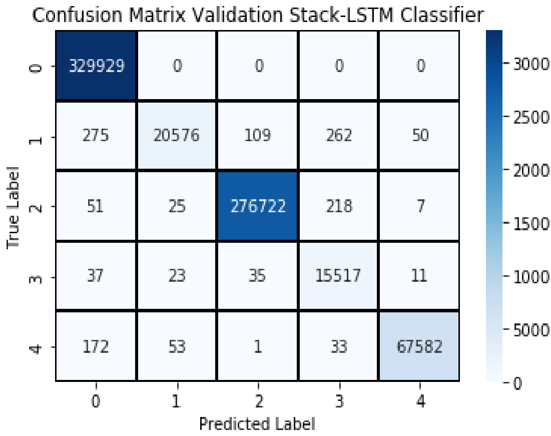
TABLE IV. PERFORMANCE OF PROPSED WORK WITH TESTING SET CLASS-WISE PRECISON, DETECTION RATE AND F1-SCORE

Evaluation set	**Attack**	**Precision**	**Detection rate**	**F1-Score**
Testing set	0	99.83	1.00	99.91
	1	99.51	96.80	98.14
	2	99.94	99.90	99.92
	3	97.19	99.25	98.20
	4	99.90	99.64	99.76
Validation set	0	99.83	1.00	99.91
	1	99.51	96.70	98.08
	2	99.94	99.89	99.91
	3	96.79	99.32	98.03
	4	99.80	99.61	99.70



	0	1	2	3	4
0	705885	0	0	0	0
1	619	44153	244	501	94
2	100	62	594352	399	12
3	80	59	79	33564	34
4	346	93	7	70	144291

Fig. 4. Confusion matrix of testing set



	0	1	2	3	4
0	329929	0	0	0	0
1	275	20576	109	262	50
2	51	25	276722	218	7
3	37	23	35	15517	11
4	172	53	1	33	67582

Fig. 5. Confusion matrix of validation set

TABLE V. PERFORMANCE COMPARISION OF PROPSED WORK WITH OTHER EXISTING WORK

Author	Model	Dataset	Feature selection	Target label	Detection Rate
proposed	Stacked LSTM	AIoT-SoL	Yes	5	99.2
[5]	DBN and DT, RF,	NSL-KDD	Yes	5	96
[6]	Autoencoder LSTM	UNSW-NB15	N.A.	9	97
[17]	Stacking Ensemble	CIC-IDS2017	Yes	7	99.05

VI. CONCLUSION

This paper proposes a novel ADM for 5G-enabled V2X communication network using DL techniques for smart vehicles. A feature selection method is used to find the optimal feature, which is the core of the proposed ADM. The proposed framework works at different stages. In the first stage, various pre-processing methods are used to convert the categorical values to numerical form and to scale the network traffic within a specific range. In the second stage, feature selection is applied, which consists of correlation coefficient-based feature selection methods to get the optimal feature set. Now we get the optimal feature set, which is used as the reduced feature set from the original data. In the last stage, Stacked LSTM-Based DL techniques are deployed as a detection tool that enables quick and effective decision-making in massively connected V2X communication networks in order to analyze large amounts of data and ensure a secure and reliable V2X environment. The performance of the proposed novel ADS framework is compared and evaluated with some of the recent state-of-the-art frameworks using the recently published AIoT-SoL dataset. The proposed framework outperforms the current state-of-the-art detection framework in terms of detection rate, accuracy, precision, and F1 scores, according to extensive result analysis. We intend to expand the suggested framework in the future by utilizing large real-time 5G-V2X Communication network data.

ABBREVIATIONS

ADM	Attack Detection Model
5G	Fifth Generation
V2X	Vehicle-to-everything
LSTM	Long Short-Term Memory
IDS	Intrusion Detection System
IoV	Internet of Vehicles
CNN	Convolutional Neural Network
DT	Decision Tree
V2S	Vehicle-to-Satellite
V2C	Vehicle-to-Cloud
V2I	Vehicle-to-Infrastructure
V2SH	Vehicle-to-Smart Home
V2V	Vehicle-to-Vehicle
V2IoT	Vehicle-to-IoT
V2N	Vehicle-to-Network
V2P	Vehicle-to-Person
V2UAV	Vehicle-to-UAV
UAV	Unmanned Aerial Vehicle
IoT	Internet of Thing
CAN	Controller Area Network

ACKNOWLEDGMENT

This work (Project No: 13(23/2020-CC & BT)) has been sponsored for R&D in Convergence Communications & Broadband Technologies (CC & BT) and Strategic Electronics., Ministry of Electronics & Information Technology (MeitY), Government of India.

REFERENCES

[1] 5G Connected Cars Changing Automotive Experiences - Ericsson, www.ericsson.com/en/blog/2021/10/powering-connected-cars-with-5g. Accessed 22 May 2023.

[2] Chen, Huimin, et al. "Towards Secure Intra-Vehicle Communications in 5G Advanced and Beyond: Vulnerabilities, Attacks and Countermeasures." *Vehicular Communications* (2022): 100548.

[3] Dibaei, Mahdi, et al. "Attacks and defences on intelligent connected vehicles: A survey." *Digital Communications and Networks* 6.4 (2020): 399-421.

[4] Noor-A-Rahim, Md, et al. "6G for vehicle-to-everything (V2X) communications: Enabling technologies, challenges, and opportunities." *Proceedings of the IEEE* 110.6 (2022): 712-734.

[5] Aloqaily, Moayad, et al. "An intrusion detection system for connected vehicles in smart cities." *Ad Hoc Networks* 90 (2019): 101842.

[6] Ashraf, Javed, et al. "Novel deep learning-enabled LSTM autoencoder architecture for discovering anomalous events from intelligent transportation systems." *IEEE Transactions on Intelligent Transportation Systems* 22.7 (2020): 4507-4518.

[7] Ullah, Safi, et al. "HDL-IDS: A hybrid deep learning architecture for intrusion detection in the internet of vehicles." *Sensors* 22.4 (2022): 1340.

[8] Warraich, Z. S., and W. G. Morsi. "Early detection of cyber–physical attacks on fast charging stations using machine learning considering vehicle-to-grid operation in microgrids." *Sustainable Energy, Grids and Networks* 34 (2023): 101027.

[9] Song, Hyun Min, Jiyoung Woo, and Huy Kang Kim. "In-vehicle network intrusion detection using deep convolutional neural network." *Vehicular Communications* 21 (2020): 100198.

[10] Rashid, Mamunur, et al. "A tree-based stacking ensemble technique with feature selection for network intrusion detection." *Applied Intelligence* 52.9 (2022): 9768-9781.

[11] Li, Jiaqi, Zhifeng Zhao, and Rongpeng Li. "Machine learning-based IDS for software-defined 5G network." *Iet Networks* 7.2 (2018): 53-60.

[12] Sapre, Suchet, Pouyan Ahmadi, and Khondkar Islam. "A robust comparison of the KDDCup99 and NSL-KDD IoT network intrusion detection datasets through various machine learning algorithms." *arXiv preprint arXiv:1912.13204* (2019).

[13] Hakak, Saqib, et al. "Autonomous Vehicles in 5G and beyond: A Survey." *Vehicular Communications* (2022): 100551.

[14] Min, Nay Myat, et al. "OWASP IoT Top 10 based Attack Dataset for Machine Learning." *2022 24th International Conference on Advanced Communication Technology (ICACT)*. IEEE, 2022.

[15] “Search UNB.” University of New Brunswick Est.1785, www.unb.ca/cic/research/applications.html. Accessed 22 May 2023.

[16] NayMyatMin. “Naymyatmin/Aiot-Sol.” GitHub, github.com/NayMyatMin/AIoT-Sol. Accessed 22 May 2023.

[17] Rani, Preeti, and Rohit Sharma. "Intelligent transportation system for internet of vehicles based vehicular networks for smart cities." *Computers and Electrical Engineering* 105 (2023): 108543.

[18] Banafshehvaragh, Samira Tahajomi, and Amir Masoud Rahmani. "Intrusion, anomaly, and attack detection in smart vehicles." *Microprocessors and Microsystems* 96 (2023): 104726.

DOI: 10.15439/2023F2826
ISSN 2300-5963 ACSIS, Vol. 37

Parallelized Population-Based Multi-Heuristic System with Reinforcement Learning for Solving Multi-Skill Resource-Constrained Project Scheduling Problem with Hierarchical Skills

Piotr Jędrzejowicz
0000-0001-6104-1381
Gdynia Maritime University
ul. Morska 83, 81-225 Gdynia, Poland
Email: p.jedrzejowicz@umg.edu.pl

Ewa Ratajczak-Ropel
0000-0002-3697-6668
Gdynia Maritime University
ul. Morska 83, 81-225 Gdynia, Poland
Email: e.ratajczak-ropel@wznj.umg.edu.pl

***Abstract*—In this paper the Parallelized Population-based Multi-Heuristic System controlled by the Reinforcement Learning based strategy is proposed to solve the Multi-Skill Resource Constrained Scheduling Problem with Hierarchical Skills, denoted as MS-RCPSP. It is an extension of the classical RCPSP where some given pool of skills has been assigned to the resources. The MS-RCPSP as well as the RCPSP belong to the class of strongly NP-hard optimization problems. To solve the MS-RCPSP the approach consisting of evolving a population of solutions and using a set of several heuristic algorithms controlled by the reinforcement learning strategy, and executed in parallel, has been proposed. To implement the system and take advantage of the speed-up offered by the parallel computations the Apache Spark platform has been used. The system has been tested experimentally using benchmark problem instances from the iMOPSE dataset with the makespan as the optimization criterion. The proposed approach produces good quality solutions often outperforming the existing approaches.**

I. Introduction

RESOURCE MANAGEMENT plays an important role in different domains. It involves planning, scheduling, and allocating various resources such as machines, technology, money, people, or teams to a project. In a majority of organizations, the task of determining some schedules occurs regularly, often daily. Deciding on schedules requires the allocation of resources which, usually, are limited and not freely available. In project management, there are three basic types of constraints imposed on the availability of resources. These are time, cost, and scope constraints. Therefore, effectively utilizing scarce resources is important for the success of any project. Extensive research in project management has led to the proposal of different models and methods aimed at optimizing resource utilization to achieve project goals. Among several possible project management problem formulations the best known and intensively researched is the Resource Constrained Project Scheduling Problem (RCPSP) and its numerous extensions. In recent years a high amount of papers have reported on various methods of solving the RCPSP and its variants. Extensive reviews of this research effort can be found in [5], [6]. Among possible RCPSP extensions focusing on the use of human resources is the idea of considering problems where to complete a project various skills on the part of human resources are needed. The idea of considering the multi-skill resource-constrained problems has been motivated by the practical needs of projects where staff with different skills is required and needs to be scheduled and assigned. In the MS-RCPSP human resources are considered each possessing a particular set of skills, which can be applied to these activities in the project that require such skills. The primary Multi-Skill Resource Constrained Project Scheduling Problem (MSRCPSP) with skillsets has been introduced in [24] and next considered, for example in [3], [17], [1]. The most recent classification for the MSRCPSP and its extensions can be found in [27]. One of such extension is MSRCPSP with hierarchical skills proposed in [2] and commonly denoted in the literature as MS-RCPSP. It is based on both the classical RCPSP and the Multi-Purpose Machine Model Problem to find a schedule that optimizes a performance criterion like, for example the project duration i.e. makespan.

Both, the MSRCPSP and MS-RCPSP as the generalizations of RCPSP belong to the class of strongly NP-hard optimization problems [4]. Hence, most of the approaches in the literature consider applying metaheuristic algorithms. Example successful approaches include Ant Colony Optimization [20], Greedy Randomized Adaptive Search Procedure [21], [22] Teaching–learning–based optimization algorithm [28], Differential Evolution and Greedy Algorithm [22], Genetic Programming [16], Genetic Programming Hyper-Heuristic [16]. In [19] the bicriteria MS-RCPSP optimization variant was proposed including project duration and cost. In [23] a new benchmark dataset was made available for public use. The approaches proposed and made available in [19], [23] involved a Greedy Algorithm that optimizes schedule duration and a Greedy guided search controlled by a Genetic Algorithm, for minimizing schedule duration and cost [23]. The Decomposition-Based Multi-Objective Genetic Programming Hyper-Heuristic has been proposed in [29].

Heuristic and meta-heuristic approaches have been important and intensively expanding area of research and development for many years. With the emergence of advanced technologies, the multi-heuristics and hyper-heuristics are commonly used in various fields, including optimization problems, search algorithms, scheduling, routing, and more generally various artificial intelligence applications. They often involve selecting, combining, or switching between different heuristics based on certain conditions, problem characteristics, or performance metrics. They are also used in solving project scheduling problems [15], [18], [25], [29]. The idea behind a multi-heuristic approach is that different heuristics may be more effective or efficient in different parts of a problem solution space. By employing a combination of heuristics, the approach can exploit their complementary strengths and mitigate their individual weaknesses. This can lead to improved problem-solving performance, such as faster convergence to a solution, better quality solutions, or increased robustness to different problem instances.

Multi-heuristic approaches can be more efficient using parallel computations which are commonly used in optimisation. They provide significant advantages in terms of speed, scalability, solution space exploration, and handling complex problem structures. By leveraging multiple processing units or distributed computing resources, parallel algorithms can efficiently solve optimization problems, leading to an improved performance, faster convergence, and better quality solutions. One of the tools for parallel computing is Apache Spark. Apache Spark is an open-source framework for processing big data in parallel across clusters or cloud architectures. Spark's core data structure is called Resilient Distributed Datasets (RDDs), which improves the performance of iterative algorithms and data mining tools. The platform automatically handles program distribution and data splitting. Spark's scheduler optimizes operations using data locality and lazy evaluation.

In this paper a Parallelized Population based Multi-Heuristic (PPMHRL) for solving MS-RCPSP is proposed, implemented and validated. The approach belongs to the population-based metaheuristics class. It is based on using four types of optimization heuristic algorithms controlled by a strategy based on Reinforcement Learning technique. The heuristic algorithms include three types of local search algorithms, the path relinking algorithm and exact solution based heuristic. To implement this approach the Scala language, Apache Spark framework and RDD collections have been used. The proposed approach has been tested experimentally using benchmark instances from the iMOPSE [30] library. The makespan minimization has been used as the optimization criterion.

The paper is constructed as follows: Section II contains the formulation of the MS-RCPSP problem. Section III provides a description of the proposed Multi-Heuristic Population Based Approach with Reinforcement Learning for solving instances of the MS-RCPSP. The section contains also a description of the optimization heuristic algorithms used: local search and path relinking. In section IV the computational experiment carried out has been described, including parameter settings experiment plan, experiment results, and comparisons of results with some other approaches. Finally, Section V contains conclusions and suggestions for future research.

II. Problem Formulation

In the paper, we consider the project management problem where activities to be executed require skills, and the available multi-skilled resources possess these skills.

The considered Multi-Skill Resource-Constrained Project Scheduling Problem with hierarchical skills can be described using classification scheme proposed in [7] for scheduling problems. An extension of this classification scheme that allows the representation of multi-skilled resource-constrained project scheduling problems and their extensions was proposed in [27] recently. The considered problem class is denoted as $ms, 1, H, TR, Flex|cpm, 1|C_{max}, C$.

In the MS-RCPSP problem the set of n activities (tasks) and m renewable resource types are considered. Each activity has to be processed without interruption to complete the project. The duration of activity a_j, $j = 1, \dots, n$ is denoted by d_j. The types of resources represent human staff with different skills. Every resource r_k, $k = 1. \dots, m$ possesses a subset of skills Q^k from the skill pool Q defined in a project and the salary paid for performed work as hourly rate (cost) c_k. In a given period of time, only one resource can be assigned to a given activity.

Each activity requires a set of skills to be executed denoted as Q_j, but not every resource can be applied to its realization. Each resource skill is labelled with familiarity level, that is the resource r_k is capable of performing the activity a_j only if r_k disposes skill required by a_j at the same or higher level.

There are precedence relations of the finish-start type with a zero parameter value (i.e. $FS = 0$) defined between the activities in the project. In other words activity a_j precedes activity a_i if a_i cannot start until a_j has been completed. S_j (P_j) is the set of successors (predecessors) of activity a_i, $j = 1, \dots, n$.

The objective is to find a schedule S of the project activities finishing times $[f_i, \dots, f_n]$, where the resource and precedence constraints are satisfied, such that the schedule duration (makespan) $MS(S) = s_n$ is minimized.

Since the MS-RCPSP is a generalization of the RCPSP, it belongs to the class of the strongly NP-hard problems [4], [19].

More detailed description and formal definition of the MS-RCPSP can be found in [2], [19], [23].

III. Parallelized Population-Based Multi-Heuristic System with RL for MS-RCPSP

A. Apache Spark based Implementation

To implement the proposed system Scala language and Apache Spark environment have been used. Apache Spark is an open-source framework designed for processing big data in parallel across clusters or cloud architectures. It prioritizes ease of use and leverages data locality to optimize computations while maintaining the required fault tolerance. Apache Spark is currently one of the most popular and fastest distributed computing frameworks, and it stands-out as the largest open-source project in data processing.

The architecture of Spark involves a master node and multiple worker nodes. The master node handles task scheduling, resource allocation, and error management, while the worker nodes perform parallel processing of Map and Reduce tasks. The platform automatically handles program distribution and data splitting for the users.

The core data structure in Spark is called Resilient Distributed Datasets (RDDs). RDD collections enhance distributed, parallel computation of iterative algorithms and interactive data mining tools. RDDs enable using parallel data structures and parallel computing.

Spark's scheduler efficiently executes operations specified by RDDs, exploiting data locality to avoid producing unnecessary data copies between nodes. RDDs are so called lazy structures evaluated, meaning the operations are performed only when the result is requested. This allows to increase the efficiency of parallel computing. Spark's built-in constraint solver can optimize the transformation graph by eliminating certain operations. RDDs also enable efficient fault tolerance by tracking the history of transformations rather than duplicating data between nodes.

The proposed Parallelized Population-based Multi-Heuristic system controlled by Reinforcemant Learning (RL) strategy is denoted as PPMHRL. The PPMHRL uses the parallel computing capabilities of Spark in order to solve MS-RCPSP problem instances stored in a population. In this approach to use the Spark capabilities efficiently its build-in parallelization mechanism has been used. To solve the MS-RCPSP the population of solutions, optimization heuristic algorithms and control strategy have been proposed and implemented. Individuals from the population of solutions are improved by optimization heuristic algorithms controlled by the RL strategy. The proposed optimisation heuristic algorithms are described in the following subsection.

B. Optimization Heuristic Algorithms Solving MS-RCPSP

To solve the MS-RCPSP with makespan minimalization the heuristic algorithms coded in Scala language have been used. The algorithms proposed in [12] have been improved and adjusted to the new system. Hence, five kinds of optimization heuristic algorithms are used:

- `LSAm` - Local Search Algorithm based on activities moving,
- `LSAe` - Local Search Algorithm based on activities exchanging,
- `LSAc` - Local Search Algorithm based on one point crossover operation,
- `PRA` - Path Relinking Algorithm based on activities moving,
- `EPTA` - Exact Precedence Tree Algorithm.

All proposed algorithms search for feasible solutions only and feasible solutions only are stored in the population.

The above mentioned `LSA` is a simple local search algorithm which finds the local optimum by moving (`LSAm`) activities or exchanging (`LSAe`) pairs of activities in the solution schedule. Simultaneously, the necessary change of assigned resources is checked and performed. In one iteration all possible moves or exchanges are checked and the best one is carried out. The best solution found is remembered. The only parameter of these algorithms is:

- $maxIt_{LSAm}$ - the maximum number of iterations without improvement for activities moving,
- $maxIt_{LSAe}$ - the maximum number of iterations without improvement for activities exchanging.

The `LSAc` is LSA based on one-point crossover operator applied to the pair of solutions. The crossover operation can be applied in each crossing point. Hence for project with n activities maximum $n-2$ crossing points can be checked. Because for some projects it may be too time consuming the algorithm stops after fixed number of iteration without improvement. The best solution found is remembered. The only parameter of this algorithm is:

- $maxIt_{LSAc}$ - the maximum number of iterations without improvement.

The `PRA` is a path-relinking algorithm where for a pair of solutions from the population a path between them is constructed. Next, the best of the feasible solutions from the path is selected. To construct the path of solutions the activities are moved to other possible places in the schedule. All possible moves are checked. Only feasible solutions are accepted. The best solution found is remembered. The algorithm has no parameters.

The `EPTA` is an exact precedence tree algorithm based on the concept of finding an optimum solution by enumeration for a partition of the schedule consisting of some activities. The implementation proposed for RCPSP [9] has been adopted for solving MS-RCSP by adjusting constraints for multi hierarchical skill levels. An exact solution for a part of the schedule beginning from activity on chosen position is found. The activity position is chosen randomly without repetition. The best solution found is remembered. The algorithm has two parameters:

- $maxIt_{EPTA}$ - the maximum number of iterations without improvement,
- $nPart_{EPTA}$ - the size of schedule partition for which the exact solution is found.

C. Architecture of the PPMHRL System

The Parallelized Population-based Multi-Heuristic system controlled by Reinforcement Learning strategy (`PPMHRL`) searches for solutions of MS-RCPSP using a set of improvement heuristic algorithms. The initial population of solutions (individuals) is generated using random priority rule and serial forward SGS (Schedule Generation Scheme). An individual is represented by the sequence of activities with resources assigned. To generate a solution from the sequence, the serial forward SGS is used. Individuals from the population are, at the following computation stages, improved by optimization heuristic algorithms described in section III-B. The behaviour of the system is controlled by the strategy. The control strategy defines parameters and methods for the whole system and is based on Reinforcement Learning.

The set of used priority rules includes ones known for RCPSP and proposed for MS-RCPSP [13], [14], [15], [12]:

- SPT - Shortest Processing Time first,
- LPT - Longest Processing Time first,
- EST - Earliest Start Time first,
- EFT - Earliest Finish Time first,
- LST - Latest Start Time last,
- LFT - Latest Finish Time last,
- HLSR - Highest Level of Skill Required first - activities are sorted by the level of skill required and SPT, which means that activities with the same level are sorted according to SPT,
- LLSR - Lowest Level of Skill required last,
- MRS - Most Required Skills first - for each skill in the project the sum of durations of activities which need this skill is calculated, next the activities are sorted by the duration of required skill and LPT,
- LRS - Least Required Skills last - for each skill in the project the sum of durations of activities which need this skill is calculated, next the activities are sorted by the duration of required skill and LPT.

To implement the approach in Spark two main RDD collections are used, one to store individuals in the population and the second one to store tuples. Each tuple contains a solutions and the algorithm that has been assigned to improve them. For selecting solutions and assigning algorithms to them the control strategy is responsible. The system state is stored and used by the control strategy to manage effectively the process of searching for the best solution. The general schema of the proposed approach can be seen in Fig. 1.

The Reinforcement Learning (RL) based cooperation strategy to control agents was proposed in [10], [11] for RCPSP and next partly adopted in the approach of solving MS-RCPSP by Asynchronous Team (A-Team) of agents in Multi-Agent System (ATMAS) described in [12]. In the approach proposed in this paper the concept of the RL based strategy has been used to control the execution of optimization heuristic algorithms using RDD collections in Spark environment. To describe the approach in a more detailed manner the following notation is used:

- P - population of $|P|$ solutions (individuals),
- S - solution,
- $nSGS$ - number of SGS procedure calls,
- $maxSGS$ - maximum number of SGS procedure calls,
- $angDiv$ - average diversity in the population P,
- $minAvgDiv$ - minimum average diversity in P,
- nS_{new} - number of newly generated solutions,
- pRA - percent of solution to be removed from population P, it is equal to the percent of the new ones to generate and add
- $selMet$ - selection method used to select the individuals to improve from P,
- $genMet$ - method generating new individuals,
- $merMet$ - method merging two populations,
- $f(S)$ - fitness function.

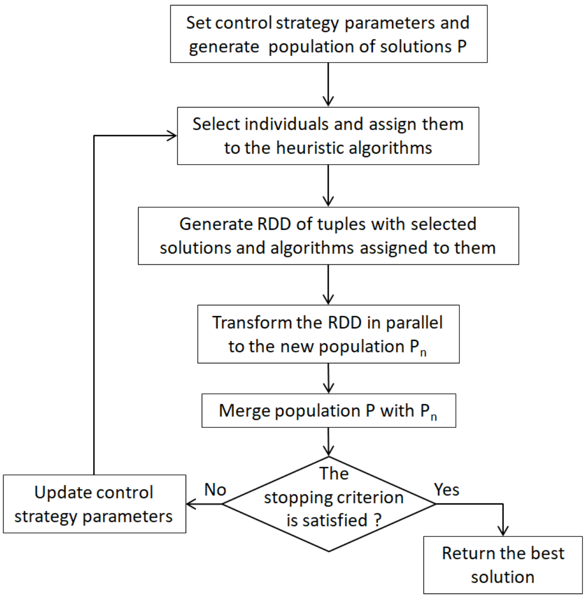

Set control strategy parameters and generate population of solutions P
Select individuals and assign them to the heuristic algorithms
Generate RDD of tuples with selected solutions and algorithms assigned to them
Transform the RDD in parallel to the new population P_n
Merge population P with P_n
The stopping criterion is satisfied ?
No
Yes
Update control strategy parameters
Return the best solution

Fig. 1. Proposed PPMHRL system architecture schema

Additionally to control the system, two probability measures have been used:

- p_{mg} - probability of selecting the method $mg \in Mg$ to generate a new individual,
- p_{ma} - probability of selecting the optimisation algorithm $ma \in Ma$ used to improve individuals in P.

To generate new individuals we have proposed the following four possible methods:

- mgr - randomly,
- $mgrp$ - randomly using random priority rule,
- mgb - random changes of the best individual in P,
- mgw - random changes of the worst individual in P.

For each method the weight w_{mg} is calculated, where $mg \in Mg$, $Mg = \{mgr, mgrp, mgb, mgw\}$. The w_{mgr} and w_{mgrp} are increased when the population average diversity decreases and they are decreased in the opposite case. The w_{mgb} and w_{mgw} are decreased where the population average diversity increases and they are increased in the opposite case.

There are five optimization heuristic algorithms described above. For each of them the weight w_{ma} is calculated, where $ma \in Ma$, $Ma = \{maLSAm, maLSAe, maLSAc, maPRA, maEPTA\}$. The w_{ma} is increased if the optimization agent received the improved solution and is decreased if this not the case. Additionally, the weights for $maLSAc$ and $maPRA$ are increased where the average diversity of the population decreases and they are decreased in the opposite case. The weights for $maLSAm$, $maLSAe$ and $maEPTA$ are increased where the average diversity of the population increases and they are decreased in the opposite case.

The probabilities of selecting the method are calculated as following:

$$p_{mg} = \frac{w_{mg}}{\sum_{mg \in M} w_{mg}}, \quad p_{ma} = \frac{w_{ma}}{\sum_{ma \in M} w_{ma}}.$$

The parameters settings and the resulting probabilities allow to control the system behaviour and balance the exploration and exploitation processes. The p_{ma} is used in $selMet$ for selecting the optimization algorithm for individuals from the population that are subject to an intended improvement. The method is showed as Algorithm 1.

Algorithm 1 $selMet(P)$

```
generate RDD collection
arrange the solutions in $P$ in random order
for all solutions in $P$ do
    $as = \emptyset$
    copy to $as$ one or two consecutive solutions from $P$
    if $|as| == 1$ then
        select $ma$ from $\{maLSAm, maLSAe, maEPTA\}$
        with the probability $p_{maLSAm}$, $p_{maLSAm}$ and
        $p_{maEPTA}$, respectively
    else
        select $ma$ from $\{maLSAc, maPRA\}$ with the prob-
        ability $p_{maLSAc}$ and $p_{maPRA}$, respectively
    end if
    add the tuple ($as$, $ma$) to RDD
end for
return RDD
```

The p_{mg} is used in $merMet$ method to merge the old population with the new one created in improvement stage by RDD transformation. The pseudocode of the $merMet$ method is presented as Algorithm 2.

Algorithm 2 $merMet(P, P_n, pRA)$

```
for each solution $S_n$ in $P_n$ do
    if $S_n$ is obtained from $S_k$ in $P$ then
        if $f(S_n) < f(S_k))$ then
            add $S_n$ to $P$
        end if
    end if
    if $S_n$ is obtained from $S_{k1}$ and $S_{k2}$ in $P$ then
        if $f(S_n) < S_{k1}$ or $f(S_n) < S_{k2}$ then
            add $S_n$ to $P$
        end if
    end if
end for
remove from $P$ the worst $pRA \cdot |P|$ solutions
add to $P$ the $pRA \cdot |P|$ of new solutions generated by
$genMet$
return $P$
```

It can be noticed that, as a result, the better solutions received from optimization algorithms replace the worse ones in the population. Moreover, the new solutions are generated in each stage with the calculated probability according to $genMet$ presented as Algorithm 3.

Algorithm 3 $genMet(P, pRA, Mg)$

```
generate empty RDD collection
for each $mg$ in $Mg$ do
    generate $pRA \cdot |P| \cdot p_{mg}$ solutions using $mg$ method
    and add them to RDD
end for
return RDD
```

All decreasing–increasing operations are performed following the proposed control strategy. As the stopping criterion the average diversity in the population $avgDiv(P)$ and the maximal number of SGS procedure calls are used.

IV. Computational Experiment

A. Problem instances

To evaluate the effectiveness of the proposed approach the computational experiment has been carried out using the benchmark instances of MS-RCPSP accessible as a part of Intelligent Multi Objective Project Scheduling Environment (iMOPSE) [30]. The test set includes 36 instances representing projects consisting from 78 to 200 activities. The file names of the instances are in the form $n\_m\_pr\_st$.def, where n means the number of activities, m the number of resources, pr the number of precedence relations and st the number of skill types. The detailed descriptions and benchmark data analyses can be found in [19], [23].

B. Settings

The computational experiment has been carried out using Intel Core i7 Quad Core CPU 2.6 GHz, 16 GB RAM. The PPMHRL is coded in Scala using Apache Spark environment. In the experiment the following values of parameters have been used:

- Population P of 30 and 50 solutions,
- 5 optimization heuristic algorithms: `LSAm`, `LSAe`, `LSAc`, `PRA`, `EPTA` using the following parameters:
  - $maxiIt_{LSAm} = 20$,
  - $maxIt_{LSAe} = 20$,
  - $maxIt_{LSAc} = 10$,
  - $maxIt_{EPTA} = 10$,
  - $nPart_{EPTA} = 3$,
- $maxSGS = 100000$ - maximal number of SGS procedure calls,
- $minAvgDiv = 0.01$ - minimal average diversity in the population,
- $pRA = 10\%$ - given initial value is decreased when the $avgDiv$ has increased and increased in the opposite case.

TABLE I
PERFORMANCE OF THE PROPOSED PPMHRL SYSTEM IN TERMS OF SCHEDULE DURATION (MAKESPAN).

	ATMAS($\|P\| = 50$)			PPMHRL($\|P\| = 30$)			PPMHRL($\|P\| = 50$)		
Instance	$Best$	AVG	STD	$Best$	AVG	STD	$Best$	AVG	STD
100_5_20_9_D3	392	394	1.67	393	394.8	**1.47**	**388**	**392.4**	2.94
100_5_22_15	**484**	**484.2**	**0.4**	485	485.4	0.49	**484**	484.6	0.49
100_5_46_15	529	530	1.55	529	530.6	1.02	**528**	**529.2**	**0.75**
100_5_48_9	491	491.4	0.49	492	492.2	**0.4**	**490**	**491**	0.63
100_5_64_15	482	483	0.89	482	**482.2**	**0.75**	**481**	482.8	0.98
100_5_64_9	475	475.2	**0.4**	475	475.6	0.8	**474**	**474.8**	0.75
100_10_26_15	237	238.2	**1.17**	**234**	**237.6**	2.73	**234**	238.4	2.58
100_10_27_9_D2	216	225.6	6.47	216	225	**6.42**	**207**	**220.4**	9.97
100_10_47_9	257	257.2	**0.4**	253	255.2	1.72	**252**	**254**	2.28
100_10_48_15	245	246.6	0.8	**244**	**244.2**	**0.4**	**244**	245.4	0.8
100_10_64_9	245	250	3.9	**243**	**246.2**	4.07	244	247.2	**2.32**
100_10_65_15	247	247.4	**0.8**	**244**	246.2	2.04	**244**	**245.6**	1.62
100_20_22_15	136	136	**0**	**130**	**133.2**	2.32	131	**133.2**	1.83
100_20_23_9_D1	174	174.6	**0.8**	**172**	**174**	1.41	174	174.6	**0.8**
100_20_46_15	164	164	**0**	164	164	**0**	**161**	**162.8**	0.98
100_20_47_9	132	133.4	**1.36**	127	132.6	3.01	**124**	**128**	3.41
100_20_65_15	240	240	**0**	240	240	**0**	**220**	**224.8**	3.19
100_20_65_9	140	140.8	**0.75**	140	134	3.35	**124**	**129.2**	5.31
200_10_128_15	464	464	**0**	461	**462.6**	1.02	**460**	**462.6**	1.74
200_10_135_9_D6	710	733.6	**13.4**	642	687.6	38.61	**550**	**603.2**	45.27
200_10_50_15	487	487.8	**0.75**	485	**486.4**	1.74	**484**	487	1.9
200_10_50_9	**488**	**489.6**	1.96	490	491	**1.1**	**488**	490.8	1.72
200_10_84_9	514	514.8	**0.75**	**507**	512.2	3.19	509	**511.2**	2.04
200_10_85_15	476	477.8	1.33	**475**	**477.6**	2.06	477	479	**1.1**
200_20_145_15	245	**246**	**1.1**	245	246.6	1.5	**244**	**246**	1.79
200_20_150_9_D5	**900**	**900**	**0**	910	948.2	20.72	**900**	**900**	**0**
200_20_54_15	268	269.6	**1.02**	263	268.4	2.8	**258**	**262.6**	3.77
200_20_55_9	252	257.6	**3.07**	251	258	4.29	**247**	**256.6**	5.82
200_20_97_15	**336**	**336**	**0**	**336**	**336**	**0**	**336**	**336**	**0**
200_20_97_9	251	251.8	**0.75**	**242**	246.6	4.03	**242**	**245.6**	3.72
200_40_130_9_D4	**513**	**513**	**0**	**513**	**513**	**0**	**513**	**513**	**0**
200_40_133_15	151	156.2	**3.71**	149	154.8	3.87	**135**	**143.8**	8.7
200_40_45_15	164	164	**0**	164	164	**0**	**160**	**162.4**	1.36
200_40_45_9	150	161	9.65	154	160.6	**4.84**	**144**	**152.8**	5.15
200_40_90_9	150	158.6	**6.18**	148	158	7.69	**135**	**148.2**	9.54
200_40_91_15	138	147.4	5.75	138	144.8	**4.66**	**132**	**143.2**	9.37
Average	331.8	334.5	2	328.8	333.6	**3.7**	**322.7**	**327.8**	4

Computations are stopped when the average diversity in the population is less then $minAvgDev$ or the number of SGS procedure calls is grater than $maxSGS$.

C. Results

During the experiment the following characteristics of the computational results have been calculated and recorded: best schedule duration (makespan) ($Best$), average schedule duration (AVG) and standard deviation (STD). Each problem instance has been solved 10 times and the results have been averaged over these solutions.

The computational experiment results for proposed Parallelized Population-based Multi-Heuristic system controlled by Reinforcement Learning strategy (PPMHRL) have been obtained for population size including 30 and 50 individuals are presented in Table I.

The results obtained by PPMHRL are good and promising. The results for the population with 50 individuals are better than for the population with 30 ones. The average $Best$ result is better by an average of 1.9%, the AVG by 1.7%, and simultaneously the STD is slightly lower. Results for both considered population sizes outperform the earlier approaches based on A-Team Multi-Agent Algorithm [12] but in this approach the optimization heuristic algorithms have been modified and one additional optimization algorithm has been used.

Comparison of the obtained results with the results known

TABLE II
COMPARISON WITH THE RESULTS KNOWN FROM THE LITERATURE IN TERMS OF SCHEDULE DURATION (MAKESPAN).

	GRASP			DEGR			GP-HH			PPMHRL($\|P\| = 50$)		
Instance	*Best*	*AVG*	*STD*	*Best*	*AVG*	*STD*	*Best*	*AVG*	*STD*	*Best*	*AVG*	*STD*
100_5_20_9_D3	401	402	**0.6**	392	393.2	0.92	387	387	0	**388**	**392.4**	2.94
100_5_22_15	503	504	1.2	**484**	**484.5**	0.53	**484**	484	0	**484**	484.6	**0.49**
100_5_46_15	552	556	3.9	529	529	**0**	**528**	**528**	**0**	**528**	529.2	0.75
100_5_48_9	510	510	0.3	**490**	490.1	0.32	**490**	**490**	**0**	**490**	491	0.63
100_5_64_15	501	502	1.2	481	483	0.82	481	481	**0**	481	482.8	0.98
100_5_64_9	494	494	0.2	474	474.9	0.32	474	**474**	**0**	474	474.8	0.75
100_10_26_15	251	258	7.1	234	235	1.05	**233**	**233**	**0**	234	238.4	2.58
100_10_27_9_D2	221	223	1.6	215	220.3	2.5	**207**	**207.9**	**0.5**	**207**	220.4	9.97
100_10_47_9	263	264	0.9	255	256.4	0.7	**252**	**252.2**	**0.4**	**252**	254	2.28
100_10_48_15	256	257	1.2	244	245	0.67	**243**	**243.7**	**0.5**	244	245.4	0.8
100_10_64_9	255	257	2.1	243	245.8	1.32	241	**242.2**	**0.6**	244	247.2	2.32
100_10_65_15	256	260	3.8	244	245.3	1.16	**243**	**243.9**	**0.3**	244	245.6	1.62
100_20_22_15	134	137	3.7	130	130.7	0.67	**126**	**126.5**	**0.5**	131	133.2	1.83
100_20_23_9_D1	**172**	172	**0**	**172**	**172**	**0**	**172**	**172**	**0**	174	174.6	0.8
100_20_46_15	170	174	3.9	164	164	**0**	**161**	**161**	**0**	**161**	162.8	0.98
100_20_47_9	133	140	6.8	125	127.5	3.31	**123**	**123**	**0**	124	128	3.41
100_20_65_15	213	213	0.1	240	240	**0**	**205**	**205**	**0**	220	224.8	3.19
100_20_65_9	135	135	0.4	126	129.1	2.73	**123**	**123.8**	0.4	124	129.2	5.31
200_10_128_15	491	496	5.2	461	463.1	0.88	**460**	**460.9**	**0.5**	**460**	462.6	1.74
200_10_135_9_D6	584	617	20.3	608	694.8	67.9	**534**	**534**	**0**	550	603.2	45.27
200_10_50_15	524	528	3.8	487	487.9	0.74	**484**	**484**	**0**	**484**	487	1.9
200_10_50_9	506	508	1.9	485	487.8	1.62	**484**	**484**	**0**	488	490.8	1.72
200_10_84_9	526	527	0.8	507	509.3	2.11	**505**	**505.8**	**0.4**	509	511.2	2.04
200_10_85_15	496	498	1.7	475	478	1.56	**473**	**473.7**	**0.5**	477	479	1.1
200_20_145_15	262	271	8.5	237	238.5	0.71	**236**	237.1	**0.5**	244	246	1.79
200_20_150_9_D5	**900**	913	13.6	**900**	906.9	11.82	**900**	**900**	**0**	**900**	**900**	**0**
200_20_54_15	304	308	3.7	258	261	1.89	258	258.3	0.5	258	262.6	3.77
200_20_55_9	257	258	0.6	249	257.8	10.37	**246**	**246.8**	**0.4**	247	256.6	5.82
200_20_97_15	347	347	**0**	336	**336**	**0**	336	**336**	**0**	336	**336**	**0**
200_20_97_9	253	256	3.8	**241**	247.6	8.93	**241**	**241.4**	**0.5**	242	245.6	3.72
200_40_130_9_D4	**513**	513	**0**	**513**	**513**	**0**	**513**	**513**	**0**	**513**	**513**	**0**
200_40_133_15	163	170	6.5	141	151.4	8.26	**135**	**136.8**	**1**	**135**	143.8	8.7
200_40_45_15	164	164	0.3	164	164	**0**	**159**	**159**	**0**	160	162.4	1.36
200_40_45_9	144	147	3.2	153	182.5	17.83	**137**	**138**	**0.4**	144	152.8	5.15
200_40_90_9	148	153	4.9	148	181.3	22.07	**134**	**135.1**	**0.5**	135	148.2	9.54
200_40_91_15	153	159	5.7	136	144.8	9.44	**130**	**131.6**	**1.1**	132	143.2	9.37
Average	337.6	341.4	3.4	326.1	332.5	5.1	**320.5**	**320.9**	**0.3**	322.7	327.8	4

from the literature are presented in Table II. It can be noticed that the results produced by the proposed approach are comparable with the results from several recently published papers. Among several algorithms proposed for solving MS-RCPSP instances, one seems outstanding and outperforms all others including the proposed one. It is also a population-based algorithm with the search for the best solution enhanced by a hyper-heuristic proposed in [24]. The best makespan value for the GP-HH algorithm is better on average by 0.7% as compared with our approach. It should be noted that the difference in performance between the proposed approach and the GP-HH one gets smaller or even nonexisting as the number of activities increases.

V. Conclusion

Results of the computational experiment show that the proposed Parallelized Population-based Multi-Heuristic System control by Reinforcement Learning strategy (PPMHRL) is an efficient and competitive tool for solving MS-RCPSP instances. The obtained results are comparable with solutions presented in the literature.

We believe that there is still room for further improvement of the proposed approach. Future research will focus on finding more effective methods for tuning optimization algorithms parameters. The use of reinforcement learning techniques could be further refined by finding better rules for controlling the number of iterations, population merging, and generating new individuals. Another performance improvement can be expected from running the solution procedure on a powerful computer cluster that can easily handle a bigger population of individuals and thus profit from the scale and synergy of interactions between optimization agents. It would also be worthwhile to investigate using different types and numbers of optimization algorithms.

REFERENCES

[1] B. F. Almeida, I. Correia and F. Saldanha-da Gama, "Modeling frameworks for the multi-skill resource-constrained project scheduling problem: A theoretical and empirical comparison.", *International Transactions in Operational Research*, vol. 26 (3), 2019, pp. 946–967, https://doi.org/10.1111/itor.12568

[2] O. Bellenguez and E. Néron, "Lower bounds for the multi-skill project scheduling problem with hierarchical levels of skills", *in Proceedings of the international conference on the practice and theory of automated timetabling*, *Lecture Notes in Computer Science*, vol. 3616, Springer, 2004, pp. 229—243, https://doi.org/10.1007/11593577_14

[3] F. Bellifemine, G. Caire and D. Greenwood, "Developing multi-agent systems with JADE." John Wiley & Sons, Chichester, 2007, DOI:10.1002/9780470058411

[4] J. Błażewicz, J. Lenstra and A. Rinnooy, "Scheduling subject to resource constraints: Classification and complexity", *Discrete Applied Mathematics*, vol. 5, 1983, pp. 11–24, https://doi.org/10.1016/0166-218X(83)90012-4

[5] S. Hartmann and D. Briskorn, "A survey of variants and extensions of the resource-constrained project scheduling problem", *European Journal of Operational Research*, vol. 207 (1), 2010, pp. 1–14, https://doi.org/10.1016/j.ejor.2009.11.005

[6] S. Hartmann and D. Briskorn, "An updated survey of variants and extensions of the resource-constrained project scheduling problem", In *European journal of operational research*, vol. 297 (1), 2021, pp. 1–14, https://doi.org/10.1016/j.ejor.2021.05.004

[7] W. Herroelen, E. Demeulemeester, and B. De Reyck, "A classification scheme for project scheduling," *Project scheduling*, 1999, Springer, pp. 1–26, https://doi.org/10.1007/978-1-4615-5533-9_1

[8] A. H. Hosseinian and V. Baradaran, "An evolutionary algorithm based on a hybrid multi-attribute decision making method for the multi-mode multi-skilled resource-constrained project scheduling problem", *Journal of Optimization in Industrial Engineering*, 12 (2), 2019, pp. 155—178, DOI:10.22094/JOIE.2018.556347.1531

[9] P. Jędrzejowicz and E. Ratajczak, "Population Learning Algorithm for Resource-Constrained Project Scheduling," *in Pearson D.W., Steele N.C., Albrecht R.F. (eds) Artificial Neural Nets and Genetic Algorithms*, Springer, Viena, 2003, pp. 223-228, https://doi.org/10.1007/978-3-7091-0646-4_40

[10] P. Jędrzejowicz and P. and E. Ratajczak-Ropel, "Reinforcement Learning Strategies for A-Team Solving the Resource-Constrained Project Scheduling Problem." *Neurocomputing*, vol. 146, 2014, pp. 301–307, doi:10.1016/j.neucom.2014.05.070

[11] P. Jędrzejowicz and E. Ratajczak-Ropel, "Dynamic cooperative interaction strategy for solving RCPSP by a team of agents." *in Nguyen, N.T., Manolopoulos, Y., Iliadis, L., Trawiński, B. (eds) Computational Collective Intelligence*. *Lecture Notes in Artificial Intelligence*, vol. 9875, 2016, pp. 454–463, https://doi.org/10.1007/978-3-319-45243-2_42

[12] P. Jędrzejowicz and E. Ratajczak-Ropel, "A-Team solving multi-skill resource-constrained project scheduling problem", *Procedia Computer Science*, vol. 207, 2022, pp. 3294-–3303 [26th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2022)], https://doi.org/10.1016/j.procs.2022.09.388

[13] R. Kolisch, "Efficient Priority Rules for the Resource-Constrained Project Scheduling Problem" *Journal of Operations Management*, vol. 14, 1996, 179–192, https://doi.org/10.1016/0272-6963(95)00032-1

[14] R. Kolisch, "Serial and Parallel Resource-Constrained Project Scheduling Methods Revisited: Theory and Computation" *European Journal of Operational Research*, vol. 90, 1996, pp. 320–333, https://doi.org/10.1016/0377-2217(95)00357-6

[15] R. Kolisch and S. Hartmann, "Experimental Investigation of Heuristics for Resource-Constrained Project Scheduling: An Update", *European Journal of Operational Research*, vol. 174 (1), 2006, pp. 23–37, https://doi.org/10.1016/j.ejor.2005.01.065

[16] J. Lin, L. Zhu, K. Gao, "A genetic programming hyper-heuristic approach for the multi-skill resource constrained project scheduling problem", *Expert Syst. Appl.*, vol. 140, 2020, art. 112915, https://doi.org/10.1016/j.eswa.2019.112915

[17] C. Montoya, O. Bellenguez-Morineau, E. Pinson and D. Rivreau, "Branch-and-price approach for the multi-skill project scheduling problem." *Optimization Letters*, vol. 8 (5), 2014, pp. 1721–1734, https://doi.org/10.1007/s11590-013-0692-8

[18] P. B. Myszkowski, M. E. Skowroński and Ł. Podlodowski, "Novel heuristic solutions for Multi–Skill Resource–Constrained Project Scheduling Problem", *in M. Ganzha, L. Maciaszek, M. Paprzycki (eds) Proceedings of the 2013 Federated Conference on Computer Science and Information Systems FedCSIS*, IEEE, 2013, pp. 159–166, ISBN 978-1-4673-4471-5

[19] P. B. Myszkowski, M. E. Skowroński and K. Sikora, "A new benchmark dataset for Multi-Skill Resource-Constrained Project Scheduling Problem" *in M. Ganzha, L. Maciaszek, M. Paprzycki (eds) Proceedings of the 2015 federated conference on computer science and information systems, ACSIS*, vol. 5, 2015, pp. 129-–138, DOI: 10.15439/2015F273

[20] P. B. Myszkowski, "Hybrid ant colony optimization in solving multi–skill resource–constrained project scheduling problem" *Soft Computing*, vol. 19 (12), 2015, pp. 3599—3619, https://doi.org/10.1007/s00500-014-1455-x

[21] P. B. Myszkowski and J.J. Siemieński, "GRASP applied to Multi-Skill Resource-Constrained Project Scheduling Problem", *in Nguyen, NT., Iliadis, L., Manolopoulos, Y., Trawiński, B. (eds) Computational Collective Intelligence (ICCCI 2016)*, *Lecture Notes in Computer Science*, vol. 9875, 2016, pp. 402—411, https://doi.org/10.1007/978-3-319-45243-2_37

[22] P. B. Myszkowski, Ł.P. Olech, M. Laszczyk and M. E. Skowroński, "Hybrid Differential Evolution and Greedy Algorithm (DEGR) for Solving Multi-Skill Resource-Constrained Project Scheduling Problem", *Applied Soft Computing*, vol. 62, 2018, pp. 1–14, https://doi.org/10.1016/j.asoc.2017.10.014

[23] P. B. Myszkowski, M. Laszczyk, I. Nikulin and M. Skowroński, "iMOPSE: a library for bicriteria optimization in Multi-Skill Resource-Constrained Project Scheduling Problem", *Soft Computing*, vol. 23 (10), pp. 3397–3410, https://doi.org/10.1007/s00500-017-2997-5

[24] E. Néron and D. Baptista, "Heuristics for multi-skill project scheduling problem", *International Symposium on Combinatorial Optimization (CO'2002)*, 2002.

[25] R. Pellerin, N. Perrier, F. Berthaut, "A survey of hybrid metaheuristics for the resource-constrained project scheduling problem", *European Journal of Operational Research*, vol. 280, 2020), pp. 395–416, DOI: 10.1016/j.ejor.2019.01.063

[26] E. Ratajczak-Ropel, "Agent-Based Approach to the Single and Multimode Resource-Constrained Project Scheduling. Population-Based Approaches to the Resource-Constrained and Discrete-Continuous Scheduling," *in Janusz Kacprzyk (ed.) Studies in Systems Decision and Control*, vol. 108, Springer International Publishing, 2018, pp. 1–100. doi:10.1007/978-3-319-62893-6.

[27] J. Snauwaert, and M. Vanhoucke, "A classifiction and new benchmark instances for the multi-skilled resource-constrained project scheduling problem," *European Journal of Operational Research*, vol. 307, 2023, pp. 1–19, https://doi.org/10.1016/j.ejor.2022.05.049

[28] Hy. Zheng, L. Wang and Xl. Zheng, "Teaching–learning based optimization algorithm for multi-skill resource constrained project scheduling problem," *Soft Computing*, vol. 21, 2017, pp. 1537–1548. doi.org/10.1007/s00500-015-1866-3

[29] L. Zhu, J. Lin, Y.-Y. Li, and Z. J. Wang, "A decomposition-based multi-objective genetic programming hyper-heuristic approach for the multi-skill resource constrained project scheduling problem," *Knowledge-Based Systems*, vol. 225, 2021, art. 107099, https://doi.org/10.1016/j.knosys.2021.107099

[30] iMOPSE (intelligent Multi Objective Project Scheduling Environment) project homepage, containing description of investigated methods, dataset definition, solution validators, references and best found solutions. http://imopse.ii.pwr.wroc.pl/

Communication Papers of the 18th Conference on Computer Science and Intelligence Systems pp. 245–250 DOI: 10.15439/2023F8389 ISSN 2300-5963 ACSIS, Vol. 37

Ensemble-based versus expert-assisted approach to carbon price features selection

Bogdan Ruszczak, Katarzyna Rudnik
0000-0003-1089-1778, 0000-0002-0653-6610
Opole University of Technology ul. Prószkowska 76, 45-758 Opole, Poland
Email: {b.ruszczak, k.rudnik}@po.edu.pl

***Abstract*—The paper comments on two main issues. First, on a model for estimating the carbon price using multi-year market data. And second, on the consideration of two approaches to feature set exploitation. On the one hand, two ensemble machine-learning models with randomly selected feature sets are employed. On the other hand, a hybrid feature selection strategy follows domain expertise on which features should be explored. This minimizes the number of feature set combinations to be tested. The additional information for the predictions was the data from other commodity contracts, which could be easily introduced into the collection, as too many of them do not necessarily improve the estimates. The results of the experiments are promising: for the model based on SVR, the MAPE obtained was 2.09% and 5.6% for the following day and week price forecasts, respectively.**

I. Introduction

THE European Union Emissions Trading Scheme (EU ETS) was established in 2005 to promote the cost-effective and economically efficient reduction of greenhouse gas emissions. According to the International Energy Agency, global CO_2 emissions reached record highs in 2021 (over EUR 60 per ton) and the price is still volatile. The costs of European Union Allowance (EUA) is increasing, not only for the environment but also for the European economy. For this reason, understanding the problem of carbon price volatility and being able to predict it has become essential for profitable business decisions in companies that emit CO_2 and are obliged to buy carbon credits, as well as in companies that are considering switching to renewable energy sources. The literature analysis shows the breadth of the scope of the topic and the potential correlation of EUA price volatility with many factors [1], [11]. Therefore, there is a need to identify a carbon price prediction model using determinants that have a particular impact on the EUA price forecasting in a dynamically changing environment.

The article is a continuation of the recent research presented by the authors in [16]. This paper describes the day-ahead carbon price prediction based on a wide range of fuel and energy indicators traded on the Intercontinental Exchange market. In the proposed approach, by combining the Principal Component Analysis (PCA) method and various methods of supervised machine learning, the possibilities of prediction in the period of rapid price increases are shown. The PCA method reduced the number of variables from 37 to 4, which were inputs for predictive models, so it reduced the complexity of models but did not improve the prediction errors [16].

Following these considerations, in this paper we propose a hybrid approach for feature selection and identification of the carbon price prediction model. In this approach, we combine a wrapper and an embedded method using different supervised machine learning methods and different time horizons of EUA price forecasting. We attempted to employ the wrapper methods, which would potentially increase the predictive power of the model, and alternatively tested the embedded approach, which allows for automatic reduction of the feature space. However, after some initial testing for the described forecasting case, we combined both approaches and supplemented them with domain expertise on which features are valuable, thus helping to reduce the number of feature set combinations.

II. Literature Review

For the purpose of testing the considered feature selection approaches, we decided to run some real data experiments. This has been performed for the field of European Carbon Emission Allowance Futures (EUA) price forecasting, and machine learning methodology has been exploited for the required estimation generation.

Different approaches to EUA price forecasting can be found in the literature. According to [27], carbon price forecasting models can be divided into the following types of models: econometric prediction model, artificial intelligence algorithms and combined prediction model.

An approach proposed in [3] employed a non-parametric method to estimate carbon prices and found that the method could reduce the prediction error by about 15% compared to linear autoregression models. In [10] a hybrid model combining the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model and a long-term memory network was presented, while in [1] the authors proposed the GARCH models and the k-nearest neighbor models. There are many approaches to EUA price prediction using machine learning methods. In [28], the authors proposed a novel paradigm of multiscale nonlinear ensemble learning, involving empirical mode decomposition and a least squares support vector machine with a kernel function prototype. An extreme learning machine optimized by the Kidney algorithm with

This work was supported by National Science Centre, Poland funds awarded to Katarzyna Rudnik under grant no. 2021/05/X/ST6/01693.

a coefficient of proportionality and cooperation is proposed in [9].

Due to the non-linearity and non-stationarity of EUA prices, the authors in [22] developed a system consisting of an analytical module and a forecasting module. There are relatively few publications that use determinant analysis in EUA price prediction models. In [11], a theoretical model was developed and presented that combines the energy sector (crude oil, natural gas, coal, electricity prices, and the share of fossil fuels in electricity generation), economic activity, and the market for CO_2 emission allowances. In [1], the authors suggested that Brent oil, coal, and electricity can be used to forecast the volatility of coal futures. In this paper, our research fills the gap in the related literature, where we take into account a wider range of data from the fuel and energy sector in order to perform feature selection and identify the EUA price prediction model in the short term. The short term prediction is especially important for traders and other market participants who, when buying EUA prices on a regular basis, follow price trends to buy EUAs at the cheapest price [15]. Prediction over a longer period of time can be useful in making strategic decisions for market participants and carbon-based companies. However, due to the nonstationary, nonlinear, and irregular EUA price, it is a particularly difficult issue that requires a more sophisticated approach to analysis. It seems that the solution in this respect could be models using the enoising procedure [5] and deep learning [4]. This will be the subject of further research.

III. Methodology

Fig. 1 presents a general overview of the applied research methodology that has been followed in this paper. We initially collected and prepared the data, which is discussed in section IV. As a result, we obtained time series for the EUA price and 16 factors related to the contracts for fuel and energy products (current values and values of the last change), which gives a total of 33 market-driven features. Since the data acquisition for the analyzed case could be costly, we tested several ways of feature selection that could lead to finding those that are really necessary and contribute to the final model performance. On the one hand, we used ensemble machine learning models with randomly selected feature sets. Providing the model with many features at the beginning, hoping that for some models this would be beneficial for the final result [14], [24]. Such a process was automated using so-called ensemble learning models, such as random forests or extra trees, which are reported by many to work well for datasets with many features [18]. These models could return even better estimates, outperforming Support Vector Machine based prediction [8]. And on the other hand, we used selected machine learning models and iteratively added successive model features, which were selected by an expert. For this part of the experiments aimed at investigating the importance of the features, as it has been advised [7], we applied linear models, support vector machines with linear kernel and simple linear regression models with additional regularization.

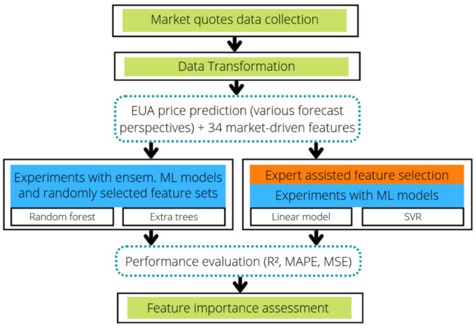

Market quotes data collection
Data Transformation
EUA price prediction (various forecast perspectives) + 34 market-driven features
Experiments with ensem. ML models and randomly selected feature sets
Random forest
Extra trees
Expert assisted feature selection
Experiments with ML models
Linear model
SVR
Performance evaluation (R², MAPE, MSE)
Feature importance assessment

Fig. 1. The experiment evaluation procedure.

All the planned experiments were conducted under a recommended cross-validation regime [26], which allows for credible performance analysis. To evaluate the tested models, we employed several widely adopted metrics in the following experiments: the coefficient of determination (denoted as R^2), which is commonly used to compare the performance of different models [19], [21], the mean square error (MSE) or its' square root [20], the mean absolute percentage error (MAPE) [21] which should provide a good intuition on the relative scale average model's prediction error. For each tested model and each generated price prediction $p_{i,pred}$ and the actual reference prices $p_{i,ref}$ we checked the statistics according to the formulas:

$$MAPE = \frac{1}{n}\sum_i \left|\frac{p_{i,ref} - p_{i,pred}}{p_{i,ref}}\right| \quad (1)$$

$$R^2 = 1 - \frac{\sum_i (p_{i,ref} - p_{i,pred})^2}{\sum_i (p_{i,ref} - \overline{p_i})^2} \quad (2)$$

Finally, it is worth mentioning the software packages we used. We used the `Scikit-learn` framework [13] to develop the necessary machine learning models and to perform all the planned experiments. For the visualization of the results and the representation of the features, we used the `Matplotlib` library [2] and the `Seaborn` package [25], all in the `Python 3.10` environment.

IV. Market data collection and transformation

The research was carried out upon data collection that gathers daily carbon futures of the EU ETS from the Intercontinental Exchange market over a long period of time. The analyzed delivery date is December of the same year (for trade dates from January to October) and December of the next year (for trade dates from November to December). The data set comes from the Fixed Income Trading Analytics web portal [6] (accessed on 8 August 2022). To be specific it spans from 2013-10-22 to 2020-12-16. For all working days on which transactions were quoted during this period, we put a total number of 1842 rows into our data set in a standardized format. Missing data were replaced with the average factor prices for the last three days.

For the purpose of the EUA price modeling, we supplemented the collection with more than a dozen of additional factors. We acquired data reflecting other fuel and energy factors from the Intercontinental Exchange, for the same time period as the target that could potentially provide useful information. Finally, the main series - f_1: "EUA Future" that shows the previous value of the modeled EUA prices, has been concatenated with the following series [6]:

- f_2: "AFR-Richards Bay Coal Future",
- f_3: "ATW-Rotterdam Coal Future",
- f_4: "M-UK Natural Gas NBP Future",
- f_5: "N-New York Harbor Unleaded Gasoline Future",
- f_6: "NCF-Newcastle Coal Future",
- f_7: "O-New York Harbor Heating Oil Future",
- f_8: "T-West Texas Intermediate Light Sweet Crude F.",
- f_9: "UBL-UK Power Baseload Future (Gregorian)",
- f_{10}: "G-Gasoil Future (Low Sulphur Gasoil Futures from February 2015 contract month)",
- f_{11}: "DPB-Dutch Power Base Load Futures",
- f_{12}: "TFM-Dutch TTF Natural Gas Base Load Futures",
- f_{13}: "B-Brent Crude Future",
- f_{14}: "CRF-CFR South China Coal Futures",
- f_{15}: "BPB-Belgian Power Base Load Futures",
- f_{16}: "GER-German GASPOOL Futures",
- f_{17}: "GNM-German NCG Futures".

The above factors are related to the contracts for fuel and energy products such as natural gas (f_4, f_{12}, f_{16}, f_{17}), coal (f_2, f_3, f_6, f_{14}), power (f_9, f_{11}, f_{15}), crude (f_8, f_{13}), heating oil (f_7), unleaded gasoline (f_5) and gasoil (f_{10}).

Besides the above-mentioned basic values of all commodities (f_1 ... f_{17}), the collection has been supplemented with the values of the last change of all these indices. These derivative values denote as follows: f_{18} for the last change in f_1, f_{19} for f_2, ..., and f_{34} reflects the last change in f_{17}.

V. Experiments

In order to provide reliable and low error forecasts, we conducted several experiments, starting with ensemble machine learning models and ending with linear models, operating on narrowed sets of covariates.

When modeling the prices of such commodities, in addition to the required highest possible performance, the narrowed data set in the sense of a smaller number of necessary collection features would be an advantage, and thus these experiments were focused on reducing the model inputs.

A. Experiments with ensemble models

Since the ensemble machine learning models are often reported to handle complex datasets successively [23], we tested their performance against investigated price collection. To estimate EUA prices using 34 market-driven features, we first utilized the random forest and extra trees algorithms.

We have trained models for various forecast horizons, for the next day's price, over the next several days' data, and up to ten days in advance each time. We expected lower performance for a longer forecast perspective but wanted to check the exact increase in the estimation error to get at least an approximation of how the model would perform for such several days-long predictions.

TABLE I
The statistics for random forest (RF) and extra trees (ET) algorithms for 10 forecast perspectives. We indicate if the metric should be minimized(↓) or maximized(↑).

Model	**MAPE (↓) for various forecast perspectives**									
	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}
ET	10.64	10.94	11.54	11.49	12.83	13.24	13.93	14.65	14.36	13.78
RF	3.920	5.595	6.889	9.852	10.83	11.77	12.33	13.09	13.31	13.91
	R^2 (↑) for various forecast perspectives									
	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}
ET	-0.049	-0.033	-0.160	-0.090	-0.393	-0.533	-0.456	-0.800	-0.567	-0.453
RF	0.928	0.854	0.782	0.567	0.475	0.372	0.324	0.250	0.232	0.164

It is worth noting, that during the cross-validation training runs, we allowed both models to select the best-performing configuration for the number of estimators used. The models of lowest errors in this experiment were most often configured for the maximum number of 25 estimators (although they may have used as many as 500).

The results for this investigation are provided in Table I. We have denoted the resulting metrics (MAPE and R^2) for the forecasts for the following 10 days (d_1, d_2, ..., d_{10}).

We tested two various ensemble algorithms because they approach feature selection differently, hoping that one of them would manage to find a profitable subset of columns and return good forecasts. It is clear from Table I that random forest performed better in this round of experiments. However, the measured errors for these forecasts were not impressive. The mean absolute percentage error for the next day's forecast was 3.9%, and it was lower than 10% only up to the fourth day in advance. The coefficient of determination was less than 0.5 for the random forest models from the fifth day on. And for the extra trees based algorithms, all models have negative R^2 indicating very poor performance. The better results of the random forest were probably related to the applied sample bootstrapping applied, but still, both of those architectures left room for improvement for other models.

B. Expert assisted feature selection experiment

Since the experiment with the automated approach to feature selection resulted in a rather disappointing forecasting performance we decided to test another one. For this experiment instead of testing models against all feature combinations, we trained models on the feature subsets that were recommended by another research. We tested 5 different setups.

The first one was based on prices only, and the second one utilized mainly price derivatives. And other subsequent subsets (S_3, S_4, and S_5) that used specifically indicated features. The sets S_1 and S_2 could be rephrased as simply working on all prices or working on all price derivatives. The latter three sets reflect the top-ranked features reported in the study [16], but to provide a deeper understanding of which features are really beneficial for final estimation, we tested the top five, top ten, and best fifteen features from that ranking to build S_3, S_4,

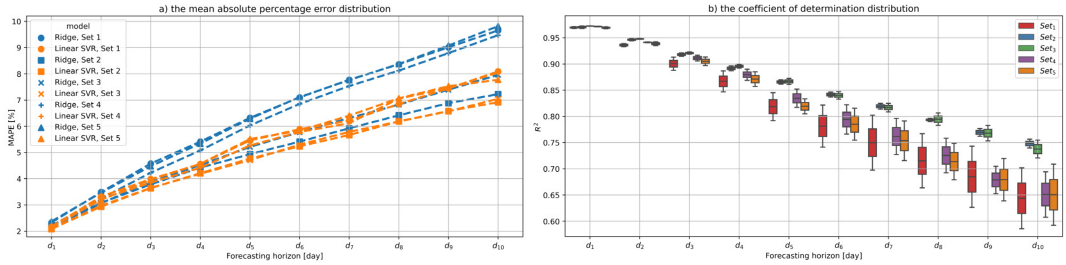


Fig. 2. A comparison of the results: (a) MAPE for two model architectures and (b) R^2 for various feature subsets.

and S_5 respectively. The detailed configuration of these subsets was as follows:

- Set 1, utilized all current prices, so the set denotes: $S_1 \in \{f_1, f_2, ..., f_{17}\}$,
- Set 2, consisted of the current EUA price value (f_1) and all features reflecting changes in the investigated prices, thus: $S_2 \in \{f_1, f_{18}, f_{19}, ..., f_{34}\}$,
- Set 3 was built on the top 5 recommended features, thus: $S_3 \in \{f_1, f_{11}, f_{15}, f_{14}, f_9, f_6\}$,
- Set 4 was built on the top 10 recommended features, thus: $S_4 \in \{f_1, f_{11}, f_{15}, f_{14}, f_9, f_6, f_3, f_2, f_4, f_5, f_{16}\}$,
- and lastly Set 5, which gathers the top 10 recommended features, with another 5, totaling the following 15 features: $S_5 \in \{f_1, f_{11}, f_{15}, f_{14}, f_9, f_6, f_3, f_2, f_4, f_5, f_{16}, f_{17}, f_{12}, f_{10}, f_8, f_7\}$,

In Fig. 2 we have plotted the results of the models trained on these 5 feature subsets. We used two various model architectures. The linear model was supported with L_2 regularization (known also as ridge regression), and the Support Vector Machine for regression with a linear kernel-based model (SVR) was also has been regularized. The comparison of the results is shown in the upper part of Fig. 2(a), and both approaches are also compared in detail in Table II. As can be seen, the models using the SVR architecture produced slightly lower errors for all subsets tested and for almost all forecast perspectives.

But the main reason for this experiment was related to the feature subsets. As we found out, in most cases the second and third subsets returned the lowest errors. This is clearly visible in the right part of Fig. 2, where the blue and green boxes represent the better coefficient of determination for all of the tested forecasting perspectives. The difference in performance for models based on S_2 and S_3 is so small that it requires another look at the results Table II. When using the R^2 as the decision criterion, the better choice would be the third subset, which has the best R^2 for 9 out of 10 forecast perspectives.

TABLE II
THE RESULTS FOR VARIOUS FEATURE SUBSETS AND FOR THE FOLLOWING DAYS' FORECASTS. THE BEST RESULTS ARE **BOLDFACED** AND WE INDICATE IF THE METRIC SHOULD BE MINIMIZED (↓) OR MAXIMIZED (↑).

Model	Set	MAPE (↓) for various forecast horizons									
		d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}
Linear	S_1	2.329	3.489	4.577	5.415	6.323	7.101	7.774	8.365	9.010	9.656
	S_2	2.227	3.080	3.785	4.424	4.935	5.414	5.924	6.417	6.883	7.224
	S_3	2.142	3.067	3.874	4.522	5.213	5.777	6.313	6.828	7.394	7.960
	S_4	2.203	3.246	4.229	5.080	6.032	6.853	7.536	8.123	8.790	9.475
	S_5	2.373	3.466	4.476	5.343	6.288	7.111	7.763	8.375	9.079	9.817
SVR	S_1	2.166	3.320	3.988	4.560	5.459	5.890	6.216	6.849	7.442	8.092
	S_2	**2.085**	**2.937**	**3.642**	4.211	4.768	**5.236**	**5.668**	6.189	6.580	**6.919**
	S_3	2.086	2.974	3.650	**4.187**	**4.709**	5.300	5.787	**6.182**	**6.573**	7.047
	S_4	2.132	3.226	3.853	4.419	5.269	5.783	6.116	7.054	7.457	8.047
	S_5	2.192	3.197	3.962	4.567	5.513	5.834	6.407	7.068	7.529	7.774
		R^2 (↑) for various forecast perspectives									
		d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}
Linear	S_1	0.968	0.933	0.888	0.847	0.792	0.741	0.697	0.664	0.627	0.586
	S_2	0.968	0.944	0.915	0.887	0.862	0.837	0.814	0.790	0.762	0.740
	S_3	0.972	0.947	0.919	0.892	0.861	0.833	0.809	0.783	0.753	0.721
	S_4	0.971	0.942	0.906	0.869	0.817	0.766	0.727	0.692	0.653	0.608
	S_5	0.967	0.935	0.897	0.857	0.805	0.755	0.716	0.679	0.639	0.592
SVR	S_1	0.971	0.939	0.913	0.887	0.845	0.822	0.802	0.767	0.743	0.702
	S_2	**0.973**	**0.949**	0.921	0.896	0.870	**0.847**	**0.825**	0.796	0.777	**0.757**
	S_3	**0.973**	**0.949**	**0.923**	**0.899**	**0.872**	**0.847**	**0.825**	**0.807**	**0.783**	0.755
	S_4	0.972	0.941	0.917	0.890	0.852	0.822	0.796	0.758	0.705	0.694
	S_5	0.971	0.943	0.913	0.885	0.834	0.816	0.791	0.748	0.720	0.709

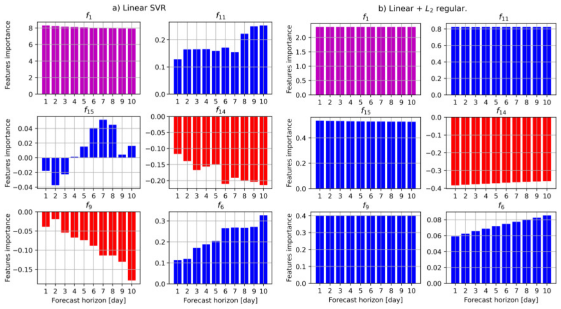


Fig. 3. The feature importance comparison for different forecast perspectives: (a) for SVR, and (b) for the linear model with L_2 regularization.

C. Features close-up

The interesting fact regarding the best-performing variant of the tested models is that it is based on the smallest subset of features. The subset S_3 consists of only 6 features. We took a closer look at the contribution of these features to the final prediction. We compared two tested variants of the linear models we trained and depicted the feature importance for both approaches (see Fig. 3). For the slightly better SVR models, the features appear to be dynamic over time dimension. Besides the most important first feature (the last denoted EUA price), other SVR features (Fig. 3a) radically change their values for the next day's forecasts. The linear model utilizes the features in a more stable manner, and only

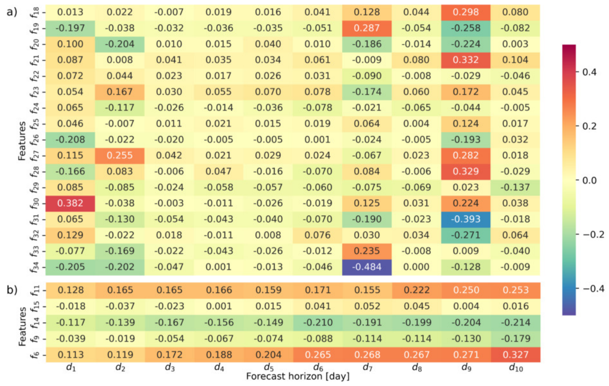


Fig. 4. The feature importance for linear SVR models and: (a) subset S_2, and (b) subset S_3. The first feature was omitted to improve chart readability.

for the last one (f_6) does the feature importance change by more than a few percent.

A closer look at the feature importance of the other feature set reveals an even more dynamic character for the set S_2 (see Fig. 4). The first feature for both sets was omitted because its importance is very dominant and would reduce the readability of the graph. The comparison of the two best-performing models using set S_2 and set S_3 which is presented there shows that the SVR-based model differs from variant to variant (when comparing variants for d_1, d_2, and the following).

The above analysis indicate that the most important are factors from two groups of individual products: power and coal. In particular, when using the SVR-based model, the current prices of the following futures have the greatest influence on the short-term forecast: “NCF-Newcastle Coal Future”, "DPB-Dutch Power Base Load Futures", "CRF-CFR South China Coal Future". It can also be seen that as the forecast horizon increases (from d_1 to d_{10}), the importance of these factors also increases. This proves that EUA price is the most sensitive to changes caused by energy markets. EUA price will largely reflect the demand for a given type of fuel in this sector. If all factors are used in the SVR-based model in the form of their last change, it is difficult to clearly rank the importance of individual factors. Their importance depends on the horizon of the forecast. In the case of the day-ahead forecast d_1, the importance of the "B-Brent Crude Future" price change is emphasized, for d_7 - "GNM-German NCG Futures", and for d_9 - "CRF-CFR South China Coal Futures". However, it is noticeable that in the case of a forecast 9 days in advance, the significance of most of the analyzed factors increases.

VI. Conclusion

As reported for the presented experiments, expertise-based feature selection could lead to better model results in some situations. For the analyzed case of EUA prices, it resulted in a lower prediction error than a more automated approach that was based on ensemble machine learning models. Such a piece of expert advice on which features to focus on could save a lot of time that would otherwise be spent experimenting with a potentially large number of different feature sets.

The apparent limitation of such an approach might be a lack of information about which covariates to focus on during modeling. This should not usually be a real concern when training a price prediction model for a relatively popular asset.

Further work should hit the time series based forecasting techniques, which would possibly lead to the ultimate performance improvement. These could be, reported to be effective for price modeling NBEATSx [12] or Temporal Convolutional Networks [17]. However, as these methods can be more time consuming, the precisely selected feature set as commented by this paper should be considered a strong asset.

References

[1] Byun, S.J., Cho, H.: Forecasting carbon futures volatility using garch models with energy volatilities. Energy Economics **40**, 207–221 (2013). https://doi.org/10.1016/j.eneco.2013.06.017

[2] Caswell, T.A., et al.: matplotlib/matplotlib: Rel: v3.7.0 (2 2023). https://doi.org/10.5281/zenodo.7637593

[3] Chevallier, J.: Nonparametric modeling of carbon prices. Ener. Econom. **33**(6), 1267–1282 (2011). https://doi.org/10.1016/j.eneco.2011.03.003

[4] Duong, V.T., et al.: Comparative study of deep learning models for predicting stock prices. In: Annals of Computer Science and Information Systems. vol. 33, pp. 103–108 (2022). https://doi.org/10.15439/2022R02

[5] E, J., Ye, J., He, L., Jin, H.: A denoising carbon price forecasting method based on the integration of kernel independent component analysis and least squares support vector regression. Neurocomputing **434**, 67–79 (2021). https://doi.org/https://doi.org/10.1016/j.neucom.2020.12.086

[6] Fixed Income Trading Analytics: https://www.theice.com/products (8 2022), www.theice.com, visited on 2022-08-02

[7] Ghosh, S., et al.: A study on support vector machine based linear and non-linear pattern classification. In: 2019 ICISS. pp. 24–28 (2019). https://doi.org/10.1109/ISS1.2019.8908018

[8] Ho, W.K., Tang, B.S., Wong, S.W.: Predicting property prices with machine learning algorithms. Journal of Property Research **38**(1), 48–70 (2021). https://doi.org/10.1080/09599916.2020.1832558

[9] Huang, Y., et al.: Carbon price forecasting with optimization prediction method based on unstructured combination. Science of The Total Env. **725**, 138350 (2020). https://doi.org/10.1016/j.scitotenv.2020.138350

[10] Huang, Y., et al.: A hybrid model for carbon price forecastingusing garch and long short-term memory network. Applied Energy **285**, 116485 (2021). https://doi.org/10.1016/j.apenergy.2021.116485

[11] Lovcha, Y., et al.: The determinants of co2 prices in the eu emission trading system. Applied Energy **305**, 117903 (2022). https://doi.org/10.1016/j.apenergy.2021.117903

[12] Olivares, K., et al.: Neural basis expansion analysis with exogenous variables: Forecasting electricity prices with nbeatsx. Int. J. of Forec. **39** (2023). https://doi.org/10.1016/j.ijforecast.2022.03.001

[13] Pedregosa, F., et al.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)

[14] Raschka, S., et al.: Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2 (2019)

[15] Rudnik, K., Hnydiuk-Stefan, A., Li, Z., Ma, Z.: Short-term modeling of carbon price based on fuel and energy determinants in eu ets. Journal of Cleaner Production **417**, 137970 (2023). https://doi.org/10.1016/j.jclepro.2023.137970

[16] Rudnik, K., et al.: Forecasting day-ahead carbon price by modelling its determinants using the pca-based approach. Energies **15**(21) (2022). https://doi.org/10.3390/en15218057

[17] Ruszczak, B.: Evolutionary configuration of the temporal convolutional network for the electricity price prediction. GECCO ’23, ACM (2023). https://doi.org/10.1145/3583133.3596429

[18] Ruszczak, B., et al.: The detection of Alternaria solani infection on tomatoes using ensemble learning. Journal of Ambient Intell. and Smart Env. **12**(5), 407–418 (2020). https://doi.org/10.3233/AIS-200573

[19] Sun, W., et al.: A carbon price prediction model based on secondary decomposition algorithm and optimized back propagation neural network. Journal of Cleaner Production **243**, 118671 (2020). https://doi.org/10.1016/j.jclepro.2019.118671

[20] Sun, W., et al.: An ensemble-driven long short-term memory model based on mode decomposition for carbon price forecasting of all eight carbon trading pilots in china. Energy Science & Engineering **8**(11), 4094–4115 (2020). https://doi.org/10.1002/ese3.799

[21] Sun, W., et al.: Factor analysis and carbon price prediction based on empirical mode decomposition and least squares SVM optimized by improved particle swarm optimization. Carbon Management **11**(3), 315–329 (2020). https://doi.org/10.1080/17583004.2020.1755597

[22] Tian, C., et al.: Point and interval forecasting for carbon price based on an improved analysis-forecast system. Applied Mathematical Modelling **79**, 126–144 (2020). https://doi.org/10.1016/j.apm.2019.10.022

[23] Trajanoska, M., Gjorgovski, P., Zdravevski, E.: Application of diversified ensemble learning in real-life business problems: The case of predicting costs of forwarding contracts. In: Ganzha, M., Maciaszek, L.A., Paprzycki, M., Slezak, D. (eds.) Proceedings of the 17th Conference on Computer Science and Intelligence Systems, FedCSIS 2022, Sofia, Bulgaria, September 4-7, 2022. Annals of Computer Science and Information Systems, vol. 30, pp. 437–446 (2022). https://doi.org/10.15439/2022F297, https://doi.org/10.15439/2022F297

[24] Trajanoska, M., et al.: Application of diversified ensemble learning in real-life business problems: The case of predicting costs of forwarding contracts. In: FedCSIS 2022. vol. 30, pp. 437–446 (2022). https://doi.org/10.15439/2022F297

[25] Waskom, M.L.: seaborn: statistical data visualization. Journal of Open Source Soft. **6**(60), 3021 (2021). https://doi.org/10.21105/joss.03021

[26] Xie, M., et al.: Efficient cross validation for svr based on center distance in kernel space. In: IRCE 2019. pp. 128–132 (2019). https://doi.org/10.1109/IRCE.2019.00033

[27] Xu, H., et al.: Carbon price forecasting with complex network and extreme learning machine. Physica A: Statistical Mechanics and its Applications **545**, 122830 (2020). https://doi.org/10.1016/j.physa.2019.122830

[28] Zhu, B., et al.: A novel multiscale nonlinear ensemble leaning paradigm for carbon price forecasting. Energy Economics **70**, 143–157 (2018). https://doi.org/10.1016/j.eneco.2017.12.030

Communication Papers of the 18th Conference on Computer Science and Intelligence Systems pp. 251–257

DOI: 10.15439/2023F8155
ISSN 2300-5963 ACSIS, Vol. 37

Architecture Analysis Cloud-Based Using MQTT Protocol for Braille Literacy

Josimar dos Santos
Program for Graduate Studies in
Computer Science (PROCC)
Federal University of Sergipe (UFS)
Aracaju-SE, Brazil 49100-000
Email: josimarsts@academico.ufs.br

Gilton José Ferreira da Silva
Program for Graduate Studies in
Computer Science (PROCC)
Federal University of Sergipe (UFS)
Aracaju-SE, Brazil 49100-000
Email: gilton@dcomp.ufs.br

Admilson de Ribamar Lima
Program for Graduate Studies in
Computer Science (PROCC)
Federal University of Sergipe (UFS)
Aracaju-SE, Brazil 49100-000
Email: admilson@ufs.br

***Abstract*—The use of assistive technologies and social inclusion is becoming increasingly important in both traditional education and active learning methodologies. In innovative education, technologies can play a crucial role in teaching people with disabilities (PWD), offering new perspectives in learning. In this context, this work is specifically aimed at individuals with visual impairments, specifically those with total loss of vision, whether congenital or acquired, or even for individuals interested in learning Braille.**

The learning needs in this case are based on sensory perception, such as touch, taste, smell, and hearing. Therefore, the objective of this study is to present a cloud-based architecture that integrates a tactile reading device and Braille display, which is simple, low-cost, and uses a lightweight messaging protocol like the *Message Queuing Telemetry Transport* (MQTT), widely used in Internet of Things (IoT) architectures.

In this architecture, the entire process of converting text into Braille occurs in the cloud. The used device has voice commands issued, received, and processed through an Application Programming Interface (API), commonly known as API, which performs the conversion of text to Braille. Then, the device prepares to display the points in high or low relief in a Braille cell, representing each character of the text. In this way, the visually impaired person can read the character through touch sensitivity.

This architecture could provide a practical and accessible solution for learning Braille and for reading by visually impaired individuals, enabling greater inclusion and active participation in the educational process.

I. Introduction

THE Information and Communication Technologies (ICT) have become consolidated through the processes of globalization, bringing significant changes and characteristics to the perspective of consumerism that permeate digital and social media. ICTs are present in our daily lives, with many of these technologies already integrated into humanity. We use them often without even realizing the level of evolution we have reached, constantly seeking to create, improve, or correct processes and connections, sending or receiving data, a subject for the next topic. Kenski (2007) states that “technologies are as old as the human species.” Hence, it is essential to reflect on the relationship between education and technology. About the formation of human society with technology [...] [1]. This excessive connectivity allows devices to also communicate, interact, and work together to achieve a specific objective or make decisions, as in Internet of Things (IoT) sensor networks. Consequently, with the development process, many devices, tools, and computational systems become increasingly capable of being ubiquitous, with dynamic and virtual connections that define the characteristics of Cloud Computing [2].

In this case, how to obtain a basic and low-cost architecture that allows conversion to the Braille system and obtain an interpretable output?

According to the research by [3], one of the highest rates among disabilities – visual, auditory, motor, and mental or intellectual – is visual impairment, with 20.1% in the age range between 15 and 64 years. This rate represents all individuals who declared having some level of difficulty in seeing. However, this study will be inclined to favor people with total loss of vision, but educators and people with sight interested in learning Braille can also benefit.

This study is organized as follows: firstly, aspects related to people with disabilities (PWDs) are discussed, with a specific focus on visual impairment. In addition, topics such as the Internet of Things, MQTT protocol, cloud computing, 3D modeling, maker culture, as well as their implications in the project using Arduino to compose the device, are addressed.

Subsequently, the processes of the research methodology adopted in this study are presented. Then, the proposed architecture is exposed, highlighting its main elements and operation. The results obtained from the application of this architecture are then discussed.

The final considerations of the study are presented, addressing the main insights, conclusions, and possible directions for future work. Finally, the bibliographic references that underpinned and substantiated the study are listed. This structure ensures a comprehensive and systematic approach to the topic, providing a detailed and well-founded analysis of the project at hand.

II. Methodology

According to the statements by [4], the methodology based on systematic mapping involves planning research with the intention of mapping all literature of empirical and non-empirical studies in a specific thematic area that can be

Thematic track: Internet of Things – Enablers, Challenges and Applications

submitted to accounting, selection, qualification, and finally, data extraction to not only answer the main question but also derived questions. In the following sequence, the authors relate systematic mapping to five process steps in the following order: defining research questions; Conducting Primary Studies Research; Screening articles based on inclusion/exclusion criteria; article classification; data extraction and aggregation.

In this chapter, the processes used to return the results, designated through a systematic mapping conducted between September 2021 and December 2021, are gathered.

A. Internet of Things

For [5], IoT is defined as: "[...] a system of cooperation between connected smart devices." These devices can be anything, as long as they can connect to the internet, with the purpose of sending and receiving data using the cloud and protocols to compose an "ecosystem." Defining actions, displaying data, performing tasks. It is with this intention that IoT fits into the architecture of this project. Having a terminal that can receive data, process it, and display it in a way that the user can interpret is the idea we need to "display" the conversion of points into Braille, that is, high and low-relief points so that the user can learn or "read" what is being "displayed."

B. Cloud Computing

The proposal of cloud computing enables exploring services and obtaining availability, mobility, flexibility, security, sharing, and low infrastructure cost, as reported by [6].

[7] addresses how the infrastructure facilitates the scalability of these resources, and maintenance can be outsourced. "[...] Resources such as processing and storage can be contracted/reserved according to demand." [7].

C. Modeling, 3D Printing, and Maker Culture

3D printing has proven to be a revolutionary technology that has positively impacted the lives of its users. By offering the possibility of building prototypes and customized objects to meet specific needs, either individually or in collaboration with others, it has opened new creative and practical perspectives [8].

The journey of 3D printing began in 1980 with Chuck Hull, a pioneer in creating stereolithography technology. This process uses laser heating of liquid elements that solidify to form the desired object. The construction with a 3D printer occurs by adding fragments, layer by layer, until the object is completed [8].

Among the various technologies available for building 3D objects, the use of plastic filament stands out, where the filament is melted and injected at specific positions to construct the object in the printer. The general process of creating 3D parts can be summarized in two steps: software modeling and printing of the developed piece.

Currently, there are various accessible software options for 3D modeling, with *AutoDesk® Tinkercad* being an example recommended for beginners due to its usability and free availability. In this type of modeling, the approach is based on adding and modifying positive and negative three-dimensional shapes, generating or "destroying" content in the physical world [8].

In the context of this project, 3D modeling and printing are fundamental to creating a final product, even if it is a prototype, with a more appealing presentation. The combination of modeling and 3D printing technology allows the development of a *case* structure to protect wires, microcontrollers, and used parts, as well as to build the parts that make up the mechanics of the Braille relief points, for example.

The modernization of the market and accessibility to new technological products, such as the 3D printer, have driven the maker culture. This movement encompasses people who seek to learn concepts from various areas, such as design, carpentry, and technology, through practical experiences. The maker culture encourages individuals to build new products, disassemble objects to understand their manufacturing processes, and thus satisfy their creative and functional needs [8].

An important aspect of the maker movement is the promotion of creative learning through experiences in maker spaces, where people create prototypes and develop new products that meet their specific needs or those of a certain audience [8].

In the context of product prototyping in the maker culture, we can understand that Arduino is a technology widely used to build prototypes for electronic and computational devices. It is important to note that the Arduino technology was developed for rapid prototyping of products for computational physics. In the development of the Arduino technology, we must be clear that it is characterized by three aspects: 1. A software: which is a platform for program development. 2. A legal aspect: which consists of using the technology without worrying about copyright issues. 3. A hardware: boards with electronic components. There are different types of boards when it comes to hardware, just as there are different software options for product development [8].

In summary, 3D printing has played a significant role in promoting creativity and people's ability to materialize their ideas and needs. The emerging maker culture strengthens this movement, driving creative learning and encouraging the development of personalized and innovative solutions. With technology constantly evolving, we can expect 3D printing to continue to have a positive impact on society, empowering people to turn their ideas into reality.

D. Braille System

In order to understand how the conversion process will be performed and displayed on the Braille display, we need to comprehend the Braille system itself. [9] explain that the formation of a three-by-two (3 rows and 2 columns) matrix of raised dots constitutes a Braille cell, as shown in Figure 1. Different combinations of raised or non-raised dots form representations of letters, numbers, and symbols for tactile mapping. Furthermore, when Braille cells are arranged side by side, they form words [9]. The Braille alphabet can be seen in Figure 2.

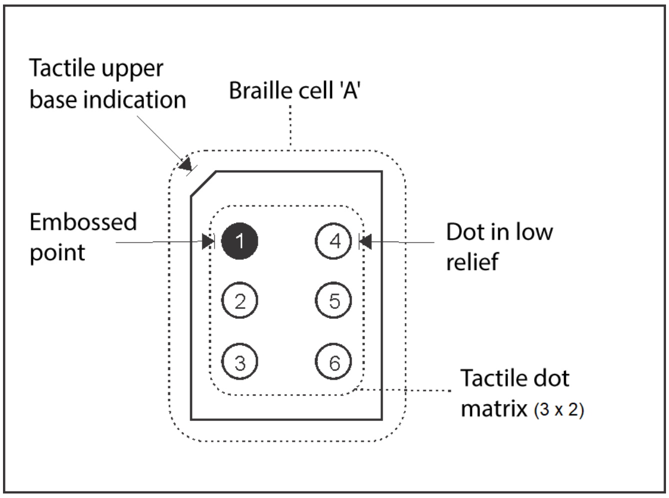


Figure 1. Braille Cell – Representation of the letter A – Authors (2021)

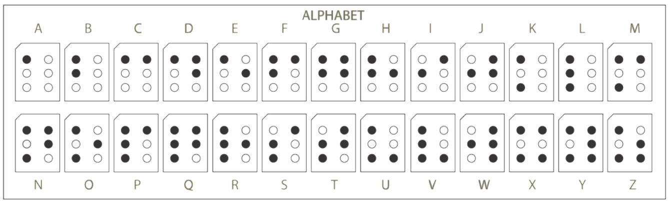


Figure 2. Braille Alphabet – Authors (2021)

E. Structural Concept

The work by [10] aims to transform the inclusion of visually impaired individuals in social discussion forums where transmission is done through speech. With the availability of three main components – Voice to Text, Text to Braille, and Text to Voice - and the ability to save, listen, and delete, according to the presented architecture, the stored file can be converted to Braille and printed.

The study conducted by [9] discusses available technologies that assist visually impaired individuals, ranging from smartphones to tactile Braille input devices and applications. However, the authors note that despite existing innovations, some tutors still use boards with raised or indented points to provide access to the Braille code, requiring more time to teach each student individually and specialized tutor training. [9] propose mass teaching of the Braille code, making the teaching process flexible, fast, and easy to use. The Microcontroller-based Actuator and Cortex-M4 with an Embedded System, the Braille Cell, is a project that allows the tutor to teach multiple students at once.

F. Discussion about the Works

For a better understanding, we have classified subsections that direct the data extracted from each selected study and their respective authors.

1) Architectures and Applicability: According to [10], the application is aimed at tutors/presenters who assist visually impaired individuals in their forums. This system contains the following main functionalities: it recognizes speech and converts it to text during execution, and displays the text on the screen. Once the text is shown on the screen, the user may be able to save the text, listen to the text, review the text, or they can convert this text to Braille. If the text is converted to Braille, the user can print it, as shown in Figures 3 and 4.

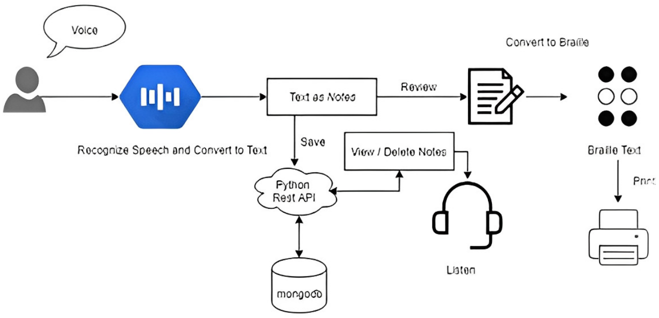


Figure 3. Architecture – Voice o Braille - [10]

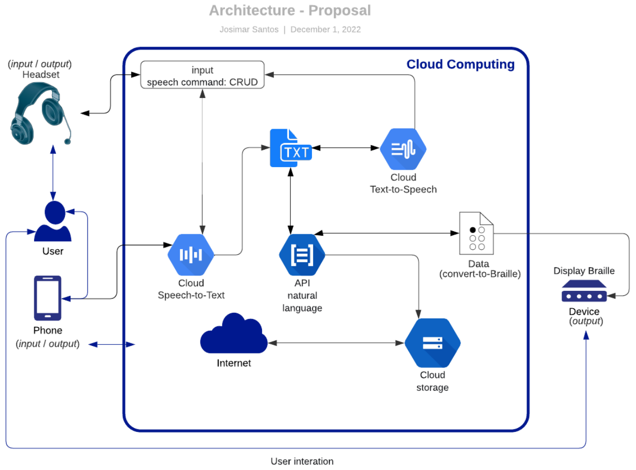


Figure 4. Architecture – Proposal – Authors - 2021

[9] mentions the usage of their project: The teacher will hold the student's hand and guide their finger on the matrix of raised dots. With the help of instructions and tactile interface, the student gradually learns the Braille code for the alphabet. Therefore, a single tutor can only teach one student at a time, and the tutor needs specialized training to teach Braille codes. This teaching method consumes a considerable amount of time when teaching a group of students, as each student needs to be taught individually. Hence, a concept of mass teaching system is proposed in this article to minimize the time required for teaching. This system has wireless features to avoid the use of cables so that students are not restricted to sitting in any specific position and at a particular distance. The tutor can type the character on the touchscreen keyboard, and this character is wirelessly transmitted to the entire student panel.

In the work of [10], beforehand, their proposal introduces the possibility of storage but not sharing of these Braille notations. Another aspect is that although it allows saving and listening to the saved text, we did not identify the possibility of editing or correcting the stored notations in MongoDB; if there is an error, the only available option is deletion. Another storage possibility, to adapt a modeling based on the studies of [10], could be tested in relation to the studies of [6], which report on low-latency and flexible cloud storage due to serverless architecture.

III. Architecture Proposal

In the work of [11], the development of an affordable Braille cell for the local context is discussed. In the initial iterations, the control system kept the solenoids active indefinitely to represent the Braille letters, but this caused excessive energy dissipation in the form of heat, resulting in high energy consumption and risk of damage. To solve this problem, the solenoids were programmed to be activated for only four seconds, allowing the interpretation of symbols and keeping them inactive for the remaining time.

The project faced difficulties in choosing the solenoids, as options with desirable characteristics were scarce in Mexico. In future iterations, hardware components that allow the miniaturization of the Braille cell and reduce energy consumption are analyzed.

In conclusion, the study demonstrated that it is possible to create an affordable Braille cell for the local context. The next steps involve reducing the size of the prototype and conducting tests with users to validate its usability, usefulness, and effectiveness. The work was partially funded by the Human-Computer Interaction Laboratory (IHCLab) of the School of Telematics at the University of Colima, Mexico. There was no need for ethical review for this research.

The study by [12] developed an assistive technology called PINDOTS, with a low-cost and easy-to-use Braille device for students with visual impairment and special education teachers. The technology consists of a six-dot Braille cell and six buttons for basic notation writing. A mobile application was created for teachers, allowing them to send exercises to the Braille device through a wireless connection. The device is capable of spelling three-letter words and has the option to emboss and record Braille dots. To improve the technology, they suggest increasing the number of spelled letters, adding contractions, including numbers, and testing with different types of visual impairments.

The studies by [10] and their architecture are very promising. The architecture of this work provides a vision that guides the structure of the Voice o Braille project, applying some differentials, such as:

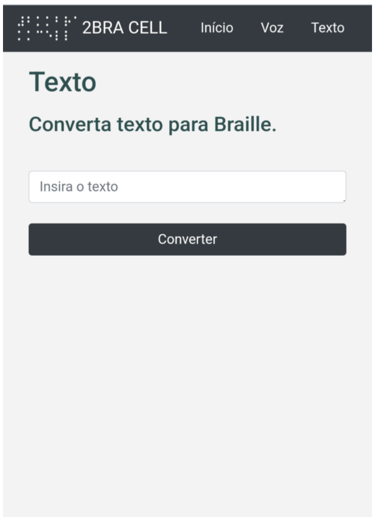
Output: Developing a Braille display with the idea of [9], to read characters on Arduino and use it as an output terminal, provides more cost-effectiveness compared to the Braille printer in the author's architecture [10]. As protection for the device, a box, as well as the tactile points, can be 3D printed.

Cloud: Brokers for MQTT publish/subscribe, as well as storage and execution tools for the conversion API, are deployed in the cloud.

The first two works were included after the research and during the experiment, and it is important to cite them as a way to complement the content and add knowledge.

The next steps contextualize and specify how the third-party libraries found and tested in the architecture can assist in each process of the architecture.

The architecture will provide instructions through voice commands, where an "assistant" will speak the process and request commands for decision-making. The user will respond with the desired command to continue the flow. The conversion from voice to text will be done using the SpeechRecognition module [13] in version 3.8.1. The gTTS library [14] in version 2.3.0 will perform the text-to-speech conversion, with the help of the PyAudio library [15] in version 0.2.12. A Python API in version 3.9 will manage a text file with the extension ".txt", enabling resuming the file later to continue, asking the assistant to "read" the text in the file, and also allowing the user to edit parts of the text or delete it. Among these functions, it will also be possible to use the command "convert" to convert the text into a format that will be interpreted by the Braille display device. The API was tested and executed locally using Flask [16], where the results were verified as shown in Figures 5 and 6.



Texto

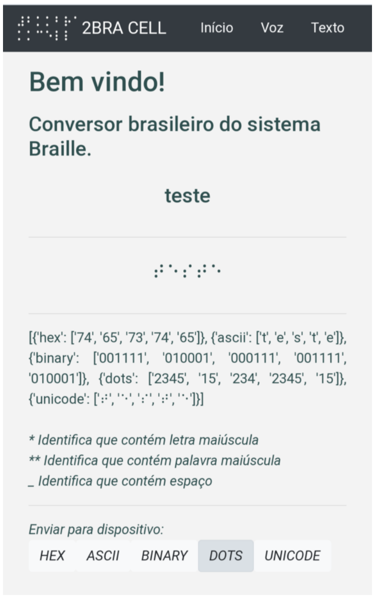
Converta texto para Braille.

Insira o texto

Figure 5. Architecture – API: text conversion screen – authors - 2022

In turn, this Braille display device will be structured with Arduino, specifically an ESP32, servomotors, pushbuttons, among other components for simulation. The display will "show" only one character at a time, and the control to move to the next character can be managed by buttons on the Braille display. For this purpose, we tested platforms like Tinkercad[1], where we did not obtain availability for the ESP32

[1] https://www.tinkercad.com/dashboard



Bem vindo!

Conversor brasileiro do sistema Braille.

teste

⠞⠑⠎⠞⠑

[{'hex': ['74', '65', '73', '74', '65']}, {'ascii': ['t', 'e', 's', 't', 'e']}, {'binary': ['001111', '010001', '000111', '001111', '010001']}, {'dots': ['2345', '15', '234', '2345', '15']}, {'unicode': ['⠞', '⠑', '⠎', '⠞', '⠑']}]

** Identifica que contém letra maiúscula*
*** Identifica que contém palavra maiúscula*
_ Identifica que contém espaço

Enviar para dispositivo:

Figure 6. Architecture – API: converted text – authors - 2022

or simulation functionalities with internet communication due to security reasons.

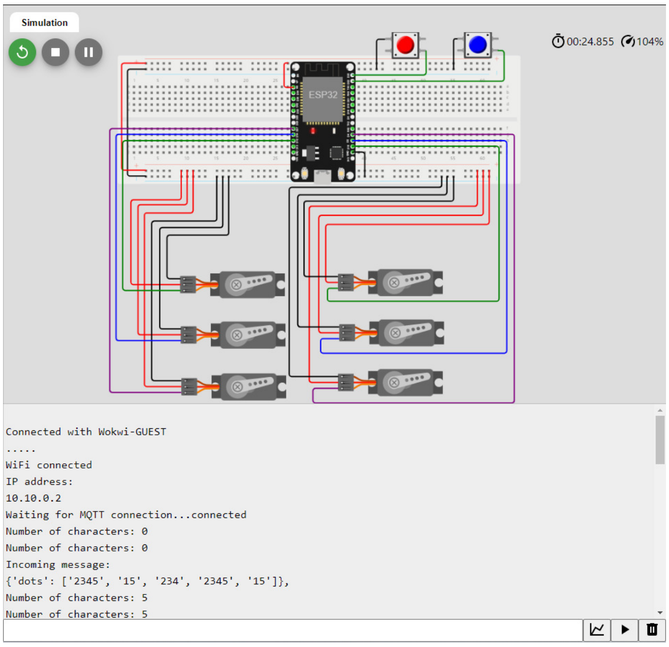
IV. Preliminary Results

We obtained satisfactory results for initial tests with the Wokwi platform[2], where we managed to structure the device with the necessary components. However, some delays were noticed while using the libraries for both speech-to-text and text-to-speech conversions.

During the execution of pre-defined voice texts from the "assistant," if the audio was being generated for the first time, there was a longer delay for the "assistant to speak." This delay occurs because the strings or phrases mentioned to the user are directly executed in the functions of the libraries used. To overcome this problem, it is possible to store these audios after the first execution, as they are reusable, and play them each time they are called. However, this requires sufficient storage space for these files.

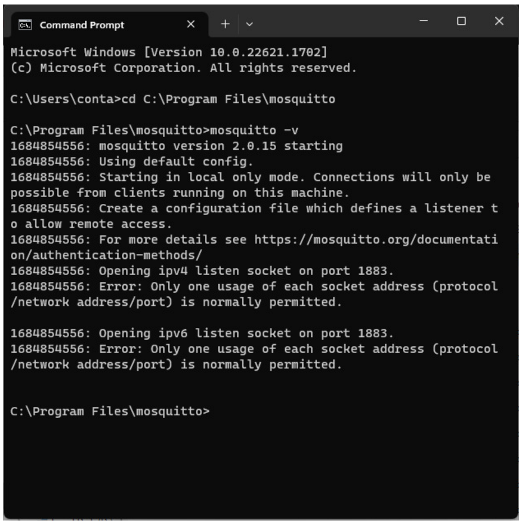
For voice capture from the user, even with the noise reduction applied in the library, some words could not be identified, or the capture was not successful. The words or phrases stored in text files on the server can be converted later and sent in JSON format over the internet to a broker, which can be published by Mosquitto[3]. The communication was tested through local simulation and the functionalities provided by the Wokwi platform.

One problem encountered regarding the return message in MQTT is that, since the API prepares a JSON with the text of the file requested by the user, we do not know the size of the text that will be converted. As a result, MQTT does not support sending messages above 268,435,455 bytes, approximately 260 MB. Therefore, a solution to use MQTT would be to limit the number of characters stored in each text file. The JSON is sent and received by the Arduino simulator, which handles the message received via MQTT and manages each character, separating them into sets of pins that will move the servomotors representing each Braille dot, as shown in Figure 7.



```
Connected with Wokwi-GUEST
.....
WiFi connected
IP address:
10.10.0.2
Waiting for MQTT connection...connected
Number of characters: 0
Number of characters: 0
Incoming message:
{'dots': ['2345', '15', '234', '2345', '15']},
Number of characters: 5
Number of characters: 5
```

Figure 7. Architecture – Arduino ESP32 Simulator – authors - 2022

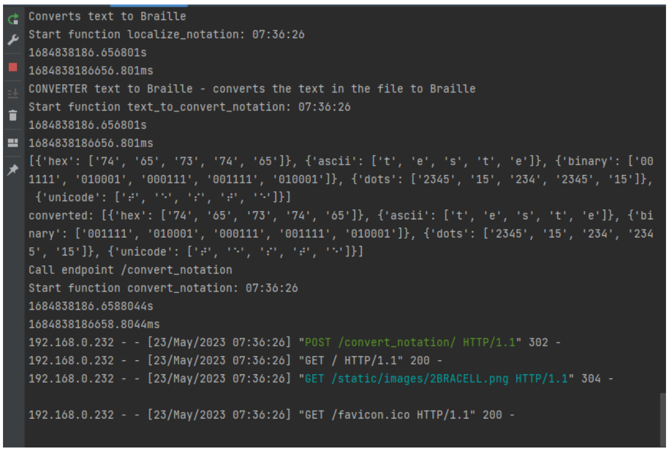


```
Microsoft Windows [Version 10.0.22621.1702]
(c) Microsoft Corporation. All rights reserved.

C:\Users\conta>cd C:\Program Files\mosquitto

C:\Program Files\mosquitto>mosquitto -v
1684854556: mosquitto version 2.0.15 starting
1684854556: Using default config.
1684854556: Starting in local only mode. Connections will only be
possible from clients running on this machine.
1684854556: Create a configuration file which defines a listener t
o allow remote access.
1684854556: For more details see https://mosquitto.org/documentati
on/authentication-methods/
1684854556: Opening ipv4 listen socket on port 1883.
1684854556: Error: Only one usage of each socket address (protocol
/network address/port) is normally permitted.

1684854556: Opening ipv6 listen socket on port 1883.
1684854556: Error: Only one usage of each socket address (protocol
/network address/port) is normally permitted.


C:\Program Files\mosquitto>
```

Figure 8. Mosquitto - version 2.0.15 starting

For communication between the server and the Braille dis-

[2] https://docs.wokwi.com/pt-BR/
[3] https://mosquitto.org/



```
Converts text to Braille
Start function localize_notation: 07:36:26
1684838186.656801s
1684838186656.801ms
CONVERTER text to Braille - converts the text in the file to Braille
Start function text_to_convert_notation: 07:36:26
1684838186.656801s
1684838186656.801ms
[{'hex': ['74', '65', '73', '74', '65']}, {'ascii': ['t', 'e', 's', 't', 'e']}, {'binary': ['00
1111', '010001', '000111', '001111', '010001']}, {'dots': ['2345', '15', '234', '2345', '15']},
 {'unicode': ['⠞', '⠑', '⠎', '⠞', '⠑']}]
converted: [{'hex': ['74', '65', '73', '74', '65']}, {'ascii': ['t', 'e', 's', 't', 'e']}, {'bi
nary': ['001111', '010001', '000111', '001111', '010001']}, {'dots': ['2345', '15', '234', '234
5', '15']}, {'unicode': ['⠞', '⠑', '⠎', '⠞', '⠑']}]
Call endpoint /convert_notation
Start function convert_notation: 07:36:26
1684838186.6588044s
1684838186658.8044ms
192.168.0.232 - - [23/May/2023 07:36:26] "POST /convert_notation/ HTTP/1.1" 302 -
192.168.0.232 - - [23/May/2023 07:36:26] "GET / HTTP/1.1" 200 -
192.168.0.232 - - [23/May/2023 07:36:26] "GET /static/images/2BRACELL.png HTTP/1.1" 304 -

192.168.0.232 - - [23/May/2023 07:36:26] "GET /favicon.ico HTTP/1.1" 200 -
```

Figure 9. Architecture – API: send JSON data as a message MQTT to ESP32 simulator – authors - 2022

play, the Wokwi platform, link to the project in development[4], allowed local simulation and the use of available features, such as the use of the Mosquitto broker[5] (Figure 8) for MQTT communication. However, a problem was identified regarding the size of the message in MQTT since the API prepares a JSON (JavaScript Object Notation) (Figure 9 and Figure 6), which composes a data structure with the text of the file requested by the user, whose size can exceed the limit supported by MQTT. To solve this issue, we plan to limit the number of characters stored in each text file, ensuring that the messages stay within the supported limit.

The word or phrase is converted to Braille into a structure that will allow future work in various aspects. The database where the letters, numbers, and characters are converted through their respective representations is provided by a ss-braille.csv file[6] available on the Kaggle platform[7]. The file has been carefully updated to meet the Braille conversion needs, such as starting with uppercase letters, as well as words for writing in Brazilian Portuguese, e.g., c and ç. An example below shows the structure and types of conversion for the word "test", sent by the API to the display via MQTT. The **hexadecimal** or hex represents the characters in hexadecimal according to

the ASCII table[8]. The **ascii** represents the characters according to the ASCII table[9]. The **binary** sends a set of 6 bits between 0 and 1 for each character of the word, which can be represented by 0 (zero) when the dot is down and 1 when it is raised. The **dots** represent a set of 6 numbers from 1 to 6 that indicate which dot will be raised. For example, if the set includes "12," dots 1 and 2 will be raised, while the numbers "345," absent from the set, represent the dots down. The **unicode** represents each character in their respective Braille dots used for display. Regarding display, the article is directed towards an architectural solution that can benefit visually impaired individuals as well as educators and interested parties. So, in addition to enabling communication with the API through voice commands, the architecture provides an interface for conversion. The research project aims to integrate Cloud architecture and the MQTT protocol with IoT Arduino devices, along with the use of Braille for information display. The primary focus is not on commercializing the results but rather on promoting accessibility and inclusion for visually impaired individuals in the context of the Internet of Things. The purpose is to develop a technological solution that allows visually impaired individuals to interact more independently with IoT devices and services, providing them with access to relevant information and improving their quality of life. The research aims to contribute to the advancement of knowledge and offer a practical application that benefits society, especially those with special accessibility needs.

In the scenario where Arduino is simulated through Wokwi, the processing of the MQTT message return is handled by the Arduino simulator itself. It takes care of each character present in the received text and then separates them into groups of pins that, in turn, are responsible for controlling the servomotors. These servomotors represent the dots in a Braille cell, as shown in Figure 7.

Despite the challenges faced, the Wokwi platform provides a solid foundation for testing and adjusting the "Voice o Braille" project, allowing the visualization and simulation of the system's operation before implementing it in a real environment. This approach is essential to identify and solve problems, as well as to improve user interaction, ensuring that the final solution is efficient and effective in including visually impaired individuals in social and educational forums.

V. CONSIDERATIONS

This work presents an architectural proposal to promote Braille literacy for visually impaired individuals. The architecture uses the MQTT protocol for communication between devices and enables interaction through both visual and tactile interfaces, with voice commands. The system is cloud-based, facilitating scalability and remote access. The Braille display device is structured with Arduino and components, and manufacturing is done with a 3D printer. The cloud implementation allows for data sharing and storage, with the possibility of future improvements, such as a specific converter for Brazilian Portuguese. The proposal aims to promote social and educational inclusion, democratizing access to knowledge and making education more accessible and equitable.

VI. CONCLUSION

The research presented an architecture that can benefit Braille literacy for visually impaired individuals. The related works on the context of the Braille system are recent, and there are still many discoveries that can be made regarding the developed studies. During the research, the study by Santiago

[4] https://wokwi.com/projects/348875560981627476

[5] https://mosquitto.org/

[6] https://www.kaggle.com/datasets/josimarsts/ssbraille?select=ssbraille.csv

[7] https://www.kaggle.com/

[8] https://www.ascii-code.com/

[9] https://www.ascii-code.com/

and Bengtson (2020) presented a proposal for a Braille display that contextualizes the main idea of the architecture proposed in this work. However, the authors used a system installed on physical machines and suggested, as future work, the expansion of new possibilities for communication and data sending to the display.

The current project aims to offer MQTT solutions for a Braille display, following the Braille System's standards. This implementation is performed through a specific API developed for this architecture. The API's source code is available on GitHub, allowing other programmers to collaborate and improve the project. In addition, the architecture allows for the creation of other converters according to the specific needs of each project.

The central objective is to provide an MQTT solution for the Braille display that is compatible with Braille System standards, thus ensuring an appropriate experience for users who depend on this reading method. By making the source code available on GitHub, the project becomes open and collaborative, encouraging the participation of other programmers who can contribute with new ideas and improvements.

Another crucial point is the flexibility offered by the architecture, allowing other converters to be implemented according to the specific needs of each project. This makes the solution more versatile and adaptable to different contexts and specific requirements.

In conclusion, the proposed architecture seeks to meet the demands of MQTT communication for the Braille display, following the guidelines of the Braille System and promoting collaboration from the developers' community to continuously improve the project. By implementing new converters according to each case's specific needs, it is hoped that the solution can be widely applicable in different scenarios and contribute to a more inclusive and accessible experience for visually impaired individuals.

Although the MQTT protocol presented limitations regarding the message size, it can still be used as long as it does not exceed this limit. The proposed architecture needs to be analyzed, and depending on its usage, the cloud implementation may vary. However, as cloud computing favors scalability, it is possible to start with free or basic plans and expand as needed. As an additional benefit, the architecture allows tutors/teachers to dispense the need for prior knowledge in Braille and can enable the massive use of exercises for various students, instead of individualized classes, even if the teacher is not familiar with the Braille system. The architecture will enable visual and tactile interaction through the device and voice commands. An implementation of a specific converter in Braille for words in Brazilian Portuguese can be added to the architecture, considering that some letters and accents make the conversion particular in this country. A performance comparison between other protocols, in addition to MQTT, is also considered, following the same architecture suggested in this project to solve the issue of the message size limit/JSON sending. This comparative study would evaluate the performance of these two protocols in terms of message transmission efficiency and data processing in IoT applications, considering the message/JSON size restriction. This could provide valuable insights for selecting the most suitable protocol in specific scenarios, considering each application's performance and data transmission capacity requirements. The implementation codes will be improved and made available later.

REFERENCES

[1] A. B. Possato and P. O. Monteiro, "Docentes de tecnologia da informação e comunicação," *Trabalho & educação*, vol. 29, no. 1, 2020.

[2] I. Machado Junior and J. P. Vece, "Contribuições da computação em nuvem como ferramenta pedagógica na educação superior," *The Journal of Engineering and Exact Sciences*, vol. 2, no. 3, pp. 92–106, 2016.

[3] C. D. . IBGE, "Censo demográfico 2010," https://biblioteca.ibge.gov.br/visualizacao/periodicos/94/cd_2010_religiao_deficiencia.pdf, jul 2010, (1) As pessoas incluídas em mais de um tipo de deficiência foram contadas apenas uma vez. (2) Inclusive as pessoas sem declaração dessas deficiências. (3) Inclusive a população sem qualquer tipo de deficiência.

[4] B. A. Kitchenham, D. Budgen, and O. P. Brereton, "Using mapping studies as the basis for further research - A participant-observer case study," *Inf. Softw. Technol.*, vol. 53, no. 6, pp. 638–651, 2011. [Online]. Available: https://doi.org/10.1016/j.infsof.2010.12.011

[5] V. C. Lins and A. M. De Morais, "Simulação e avaliação de desempenho de uma blockchain para aplicações iot," *Gestão.org*, vol. 19, no. 2, pp. 169–183, 2021.

[6] C. G. Calancea, C.-M. Miluţ, L. Alboaie, and A. Iftene, "iassistme - adaptable assistant for persons with eye disabilities," *Procedia Computer Science*, vol. 159, pp. 145–154, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S187705091931347X

[7] M. A. Spohn and F. D. S. Bonetti, "Análise de desempenho de uma arquitetura de fog computing para internet of things via estudos de caso," *Revista Brasileira de Computação Aplicada.*, vol. 11, no. 3, pp. 72–87, 2019.

[8] G. Santiago, C. Bengtson, D. Pino, C. Pendenza, and J. Santos, "Universidade federal de são carlos - ufscar," https://dcomp.ufscar.br/wp-content/uploads/2016/05/DComp-TR-002.pdf, 2020.

[9] S. Gandhi, B. Thakker, and S. Jha, "Braille cell actuator based teaching system for visually impaired students," in *Braille cell actuator based teaching system for visually impaired students*. Institute of Electrical and Electronics Engineers Inc., 2017, pp. 1381–1385, cited By 3; Conference of 1st IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, RTEICT 2016 ; Conference Date: 20 May 2016 Through 21 May 2016; Conference Code:125896. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85015099644&doi=10.1109%2fRTEICT.2016.7808057&partnerID=40&md5=bd3c9373d89ba72173025f1571ee4839

[10] J. Pradeepkandhasamy, A. Priya, and P. Chellappan, "Voice o braille," *Materials Today: Proceedings*, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214785320407710

[11] O. I. Ramos-García, A. A. Vuelvas-Alvarado, N. A. Osorio-Pérez, M. Ruiz-Torres, F. Estrada-González, L. S. Gaytan-Lugo, S. B. Fajardo-Flores, and P. C. Santana-Mancilla, "An iot braille display towards assisting visually impaired students in mexico," *Engineering Proceedings*, vol. 27, no. 1, 2022. [Online]. Available: https://www.mdpi.com/2673-4591/27/1/11

[12] D. Martillano, A. Chowdhury, J. Dellosa, A. Murcia, and R. Mangoma, "Pindots: An assistive six-dot braille cell keying device on basic notation writing for visually impaired students with iot technology," in *Proceedings of the 2018 2nd International Conference on education and e-learning*, ser. ICEEL 2018. ACM, 2018, pp. 41–47.

[13] Anthony Zhang (Uberi), *SpeechRecognition v.3.8,1*, , 2014. [Online]. Available: https://github.com/Uberi/speech_recognition/blob/master/reference/library-reference.rst

[14] Pierre Nicolas Durette and Contributors, *gTTS v. 2.3.0*, , 2014. [Online]. Available: http://gtts.readthedocs.org/

[15] Hubert Pham, *PyAudio 0.2.12*, , 2006. [Online]. Available: https://pypi.org/project/PyAudio/

[16] Armin Ronacher, *Flask 2.2.2*, Pallets, 2010. [Online]. Available: https://pypi.org/project/Flask/

DOI: 10.15439/2023F5773
ISSN 2300-5963 ACSIS, Vol. 37

A simple algorithm for computing a multi-dimensional the Sierpiński space-filling curve generalization.

Ewa Skubalska-Rafajłowicz, Member, IEEE
0000-0002-2795-3835
Dept. of Computer Engineering,
Wroclaw University of Science and Technology,
Wyb. Wyspianskiego 27, 50 370 Wrocław, Poland,
Email: ewa.skubalska-rafajlowicz@pwr.edu.pl

***Abstract*—Transforming multidimensional data into a one-dimensional sequence using space-filling curves, such as the Hilbert curve, has been studied extensively in many papers. This work provides a systematic presentation of the construction of an arbitrarily accurate multidimensional space-filling curve approximation which is a generalization of the Sierpiński space-filling curve. At the same time, according to the space-filling curve construction, we present a simple algorithm for determining one of the counter-images on a unit interval of a data point lying in a multidimensional cube. The computational complexity of the algorithm depends linearly on the dimension of the cube. The paper contains numerical algorithms for local generation of the curve approximation and determination of the quasi-inverse of a data point used to transform multidimensional data into the one-dimensional form.**

I. Introduction

The space-filling curve (SFC) is defined as a continuous mapping converting a unit interval $[0,1]$ onto the d-dimensional unit cube ($I_d = [0,1] \times \ldots \times [0,1]$, $d \leq \infty$) [27], [2]. This means that the space-filling curve passes at least once through each point of the I_d cube. This allows many multidimensional computational problems to be considered as one-dimensional problems without losing the essential properties of the original problems.

Space-filling curves were first described by G. Peano in 1890 [25], and then by D. Hilbert [12], and W. Sierpiński [30].

In the 1930s, space-filling curves, which are measure preserving, have been used in the theory of integration in multidimensional spaces [27], [18].

The applications of space-filling curves are pretty broad, though in many cases, they are limited to two-dimensional curves. We can mention as examples: image processing [24], image compression [22], image encryption [4], image malware classification [23], MRI image sampling [29], the cryptographic transformation scheme for spatial query processing [13], encryption technology for data privacy-preserving [16], raster tool path generation for layered manufacturing [14], among many others. Nair et al. [21] used the Hilbert space-filling curve to explore the space of the robot and detect obstacles.

The Sierpiński space-filling curve was applied to solve discrete optimization problems [41]. In particular, Bartholdi and Platzman [3], [26] applied the Sierpiński curve to find an approximate, near optimal solutions of the planar traveling salesman problem with Euclidean distances. An important property of a space-filling curve is the preservation of proximity by the points from $[0,1]^d$ transposed to $[0,1]$. This property means that points lying close to each other on the curve are also close in the multidimensional space. Points far apart on a curve can be close to each other in the multidimensional space. In order to estimate the actual distance of a pair of points in $[0,1]^d$, say (x,y) one should have determined all $2*2^d$ points on the unit interval whose images are respectively x and y. Hence the idea of using space-filling curves to find nearest neighbors of subsequent point from multi-dimensional space.

Multidimensional space-filling curves are also used in global optimization algorithms [31],[32], [33] and their applications, e.g. in experimental design [35]. Lawder et al. [15] discuss multidimensional indexing for database management systems based on space-filling curves.

On the other hand, the space-filling-based transformations retain essential statistical information. For example, it is proved that the Bayes risk is invariant under these transformations for every distribution with a bounded support [38].

Sampling of multidimensional space using one-dimensional equidistributed sequences transformed by a multi-dimensional space-filling curve was developed in [40], [11]. Data-dependent space-filling curves for non-uniform grid were proposed by[34],[43].

The Hilbert space-filling curve was used in parallel codes for numerical simulations [5], and the Sierpiński curve helped to streamline finite element calculations [1]. and for load-balancing in distributed computing [17].

The algorithms in which space-filling curves were used to reduce the dimension of data and then to analyze them concerned, among others, the determination of the box-counting fractal dimension [39], the classification of multidimensional data [38] and the data clustering [19], [42].

It is known, that the continuous mapping $F_d : I_1 \rightarrow I_d$

can not be one-to-one [27], [3], [26]. The geometric points of intersection of the curve with itself correspond to many points of the unit interval. Thus, from the viewpoint of such possible applications, it is crucial to find $t \in I_1$ such that $F_d(t) = x$ for given $x \in I_d$, i.e., to provide a quasi-inverse of $F_d(t)$.

This paper provides a fast and relatively simple algorithm for computing the approximate quasi-inverse for any dimension d. Naturally, this is intrinsic to the definition of the space-filling curve generation. One can transform each element of the multidimensional data set using quasi-inverse separately and at any moment. It does not require the construction of the entire space-filling curve. Such transformation allows us to have a linear order of data in higher dimensions. Since the curve is a closed curve, the resulting order is cyclical.

The outline of the paper is as follows. Section 2 introduces the main ideas of constructing the d-dimensional Sierpiński space-filling curve generalization. Section 3 contains two algorithms. Algorithm 1. implements the construction of the nodal point of the d-dimensional Sierpiński curve, i.e., it transforms the selected data point from the unit interval onto the d-dimensional unit cube. Algorithm 2 implements the quasi-inverse mapping connected to the generalized Sierpiński curve, i.e., it allows us to obtain a one from 2^d positions on the unit interval of a given point from the d-dimensional hypercube. The example of using the algorithm to transform 4-dimensional Iris data into unit interval is given at the end of the section. The last section summarizes the contents of the paper.

II. The method of constructing the d-dimensional space-filling curve.

The method of constructing the d-dimensional space-filling curve can be related to the sequential division of the filled multidimensional space into elementary areas, usually multidimensional sub-cubes of the same shape and the same volume [20].

Next, a one-to-one correspondence is established between the 2^{dk} elementary intervals U_k of the length 2^{-dk} and between the $(2^d)^k$ sub-cubes C_k of size $2^{-k} \times 2^{-k} \ldots \times 2^{-k}$ ($k = 1, 2, 3, \cdots$). The correspondence is such that any two adjacent sub-intervals correspond to two adjacent sub-cubes and moreover, 2^d sub-intervals U_{k+1} (of the length $2^{-d(k+1)}$) which constitute a sub-interval U_k correspond to the 2^d sub-cubes C_{k+1} associated with the corresponding sub-cube C_k. In this correspondence, adjacent sub-cubes of any level of division k are related to adjacent subintervals of the same degree of partition in the unit cube [20], [18], [27]. In this way, one can define the classical space-filling curves such as the Peano, Hilbert, and Sierpi ́nski.

Define the family W of 2^d mappings $w_i : R^d \to R^d$, $i = 0, \ldots, 2^d - 1$ of the following form:

$$w_i(x_1, x_2, \ldots, x_d) = \begin{cases} \frac{1}{2} - (\frac{1}{2} - \beta_{1,i})x_1 \\ \frac{1}{2} - (\frac{1}{2} - \beta_{2,i})x_2 \\ \cdots \\ \frac{1}{2} - (\frac{1}{2} - \beta_{d,i})x_d \end{cases} \quad (1)$$

where $\beta_{j,i} \in \{0, 1\}$, $j = 1, 2, \ldots, d$, $i = 0, 1, \ldots, 2^d - 1$. For $\beta_{j,i} = 0$ we get the transformation of the form $\frac{1}{2} - \frac{1}{2}x_j$ and for $\beta_{j,i} = 1$ we get $\frac{1}{2} + \frac{1}{2}x_j$. Mappings w_i are indexed in such a way that vectors $B_i^d = (\beta_{1,i}, \beta_{2,i}, \ldots, \beta_{d,i})^T$ determine w_i uniquely.

B_i^d forms a list of d–dimensional vectors containing only 0 and 1, where each pair of the adjacent vectors differs at exactly one position. In geometric terms, such a list of vectors describes a closed (Hamiltonian) path through all vertices of d-dimensional unit cube ($I_d = [0, 1] \times [0, 1] \times \ldots \times [0, 1]$).

Among the many different possibilities, we limit ourselves here to the order defined by the classical, reflective (reflected) binary Gray code (see, e.g., [9]). $d + 1$–dimensional code is formed from the d–dimensional one as follows:

$$B_0^{d+1}, \ldots, B_{2^d-1}^{d+1} = (B_0^d, 0), \ldots, (B_{2^d-1}^d, 0)$$

and the reverse order added

$$(B_{2^d-1}^d, 1), \ldots, (B_0^d, 1)$$

For example, for $d = 3$ we obtain a closed sequence of the vertices of the 3 dimensional unit cube in the following form closely related to the sequence of mappings $(w_0, w_1, \ldots, w_7)$ and $w_8 = w_0$:

$$(000), (001), (011), (010), (110), (111), (101), (100), (000).$$

Define a number sequence b_k, $k = 1, \ldots$ such that:
$b_1 = 1$, $b_k = 2^{k-1} - b_{k-1} + 1$, $k = 2, 3, \ldots$

In this way, we obtain a fast-growing sequence of the positive integers $(1, 2, 3, 6, 11, 22, \ldots)$, index sequences that specify the position of vertex $(1, \ldots, 1)$ in the sequence (and the corresponding mappings) scanning the vertices of I_d.

Furthermore, it is easy to verify that:

Property 1: $b_d \to \infty$ as $d \to \infty$, but $2^{-d}b_d$ ranges from $\frac{1}{2}$ to $\frac{1}{3}$ as d ranges from 2 to infinity.

The number b_d is equal to the smallest distance between two of the most distanced vertices of the cube: vertex $(1, 1, \ldots, 1)$ and vertex $(0, 0, \ldots, 0)$ and as a consequence

$$B_{2^d-b_d}^d = (1, 1, \ldots, 1).$$

$w_{2^d-b_d}$ maps vertex $(1, 1, \ldots, 1)^T$ to the same vertex $(1, 1, \ldots, 1)^T$.

The mappings w_i show how the unit cube is split onto 2^d smallest sub-cubes (of size 2^{-1},2^{-2} $\ldots$) . Successive repetition of such partitions produces a sequence of $(2^d)^k$ sub-cubes, which were obtained by consecutive mappings with indices differing by one, i.e., defined by repeated sequences of numbers $s = (0, 1, 2, \ldots 2^d - 1, 2^d)$, where 0 and 2^d symbolize two parts of the same sub-cube (associated with the vertex corresponding to the starting point node $(0, \ldots, 0)$ of the current sub-cube). At each subsequent split, the sequence $(0, 1, 2, \ldots, 2^d - 1, 2^d)$ is replaced in the following way:

$$0 \to (2^d - b_d, \ldots, 2^d - 1, 2^d),$$

$$1 \to (0, 1, 2, \ldots, 2^d - 1, 2^d),$$

$$\ldots$$

$$2^d - 1 \to (0, 1, 2, \dots, 2^d - 1, 2^d),$$

$$2^d \to (0, 1, \dots, 2^d - b_d - 1, 2^d - b_d).$$

Notably, a space-filling curve is defined as a limit of the uniformly convergent space-filling curve approximations formed by line segments with ending points in adjacent sub-cubes. The approximations could differ, but the limit curve depends only on U_k and C_k structures [18], [27]. In our case, the endpoints lie on the chosen vertices of the subsequent sub-cubes. They have been chosen, so their positions do not change in subsequent iterations (for the next k). The next approximating curve is created by adding successive points without changing the location of the previous ones. It is worth emphasizing here that the refinement of the curve approximation takes place locally, separately in each sub-interval and the corresponding sub-cube. Refining the curve approximation in one of its fragments (a given sub-cube) does not affect refining the curve in its other fragment. The first partition of the unit interval consists of $2^d + 1$ sub-intervals with all intervals of the length $1/2^d$. The only exceptions were the first and last intervals, which are $b_d/2^{2d}$ and $(2^d - b_d)/2^{2d}$ (together add up to $1/2^d$). These two shorter sub-intervals correspond to the sub-cube with the vertex $(0, \dots, 0)$, where the conventional beginning and end of the curve locate. The position of a point inside a particular sub-cube C_k indicates the sub-interval U_k where its counter-images find.

Transformations 1 are repeated without changing the scale of the cube, because the coordinates in subsequent divisions (in smaller and smaller scales) are each time scaled to the size of the unit cube by inverse transformations:

$$x_i = \begin{cases} 1 - 2x_i, \text{ if } x_i < 1/2, \\ 2x_i - 1, \text{ if } x_i \geq 1/2, \end{cases} \quad i = 1, \dots, d. \tag{2}$$

Another possibility is to assume that $1 - 2x_i$ is performed when $x_i \leq 1/2$ and complementarily, transformation $2x_i - 1$ is performed for $x_i > 1/2$. Thus, the point with a finite binary expansion of all its coordinates can be connected to one of 2^d adjacent sub-cubes. Each such combination of partitions (for every coordinate, we can have partition $[0, 1/2), [1/2, 1]$ or $[0, 1/2], (1/2, 1]$) is enough to determine successive approximations of the quasi-inverse of the space-curve (see Algorithm 2), because the scaling of the length of the respective subintervals (by multiplying by 2^{-di}, $i \leq k$) is independent of the orientation of the currently considered sub-cube of side size 2^{-i}. Let's note that each point of the cube I_d is within $2^{-kd/2-1}$ distance from one of the vertices of the approximating curve.

Usually, the 2-D Sierpiński curve is defined as a correspondence between intervals and triangles. We will here rely on the construction of the 2-D Siepiński curve, which basis on a quadruple partition of a square (see Fig. 1), i.e., squares of side size 2^{-k}, where k is the number of the subsequent divisions of the unit square. As in proposed in this paper approach, the $b_d/2^{2d}$ and $(2^d - b_d)/2^{2d}$ intervals correspond to partitioning the cube onto two triangles in two dimensional space. When d is larger, the division of the multi-cube is more complicated, and the ratio of the volumes of the two parts depends on d.

The other version of the Sierpiński space-filling curve generalization can be obtained by changing the direction of passing the cube vertices in the sub-cubes obtained as a result of the w_i transformation with the odd index i.

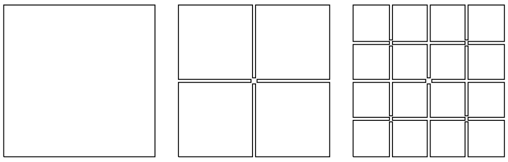


Fig. 1: Approximations of the Sierpiński space-filling curve in 2-D.

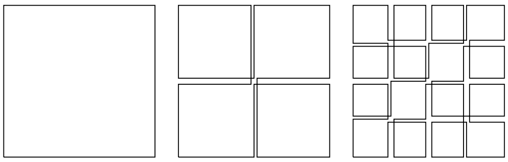


Fig. 2: Modification of the Sierpiński space-filling curve in 2-D.

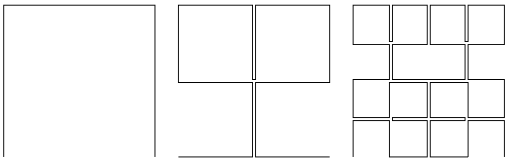


Fig. 3: Approximation of the Hilbert space-filling curve on the plane.

As a consequence, in the two-dimensional case, we obtain the agreement with the original Sierpiński curve (Fig. 1 in contrast to the 2-D space-filling curve roughly visualized in Fig. 2. Fig. 3 shows an approximation of the Hilbert curve. The Hilbert curve uses the same vertex order in the elementary cube as in the case of the Sierpiński curve [6], however, it does not form a closed cycle as in our case.

A. Properties of the proposed family of the space-filling curves

The previously defined sequence of approximating curves is uniformly convergent to the space-filling curve [20], [18].

The presented here generalization of the Sierpiński SFC treated as the map $F_d : I_1 \to I_d$, has the following properties:

a) $F_d(t)$ is a measure preserving map of I_1 onto I_d,
b) mapping F_d forms closed curve, i.e. $F_d(0) = F_d(1)$,
c) $F_d(t)$ is a Hölder continuous mapping of order $1/d$, in the following sense

$$\| F_d(t_1) - F_d(t_2) \| \leq 2(d+3)^{1/2} (\Delta(t_1, t_2))^{\frac{1}{d}} , \tag{3}$$

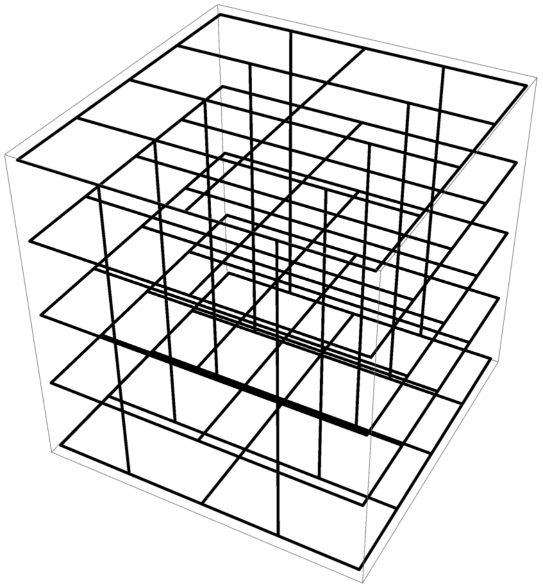


Fig. 4: Approximation of the proposed 3-D Sierpiński space-filling curve with 512 nodal points.

where $\Delta(t_1, t_2) = min\{| t_1 - t_2 |, 1- | t_1 - t_2 |\}$ for $t_1, t_2 \in [0,1]$is a metric on a circle, while $\| \cdot \|$ denotes the Euclidean norm in R^d. The constant $2(d+3)^{1/2}$ (in property c) is an upperbound. The same bound is valid for the multidimensional Hilbert curve [37], [11]. Furthermore, it is known that for the 2-D Sierpiński curve, the smallest possible constant equals 2 [26] or is close to $6^{1/2}$ for the 2-D Hilbert curve [10], [37].

III. The numerical algorithms

The first algorithm (Algorithm 1)approximates the points of the d-dimensional Sierpiński curve using the entered points from the unit interval. More precisely, it computes the image of any point $t \in [0,1]$ in $[0,1]^d$. The algorithm requires a number of iterations depending on the level of the curve approximation k, and the accuracy of the approximation of each coordinate point is 2^{-k}. The computational complexity of the algorithm is $O(kd)$. Figure 4 depicts an approximation of the 3-D Sierpiński space-filling curve.

The second algorithm (Algorithm 2) transforms a multivariate data point into a unit interval. The algorithm is iterative, and gives an approximation of the quasi-inverse of the d-dimensional Sierpiński curve, depending on the level of approximation, say k. The accuracy of determining the position on a unit interval of an image $x \in [0,1]^d$ is 2^{-dk}. The computational complexity of the algorithm is $O(kd)$.

In the case of two dimensions, the algorithms provide the curve (and its quasi-inverse) depicted on Fig. 1 b). A slight modification of the algorithms in places marked $\star$ allows us to obtain approximations of the original Sierpiński curve and its quasi-inverses.

The Fig. 5 illustrates the application of the transformation multivariate data on the example of Iris data [8], which contains 150 measurements (4 dimensional) of three different species of irises: Iris Setosa , Iris Versicolour, and Iris Virginica. It is easy to see that the Iris Setosa forms a separate cluster in the unit interval and only a small fraction of Iris Virginica is mixed with Iris Versicolor. Computations performed to generate Fig. 4 and Fig. 5 were made in Matematica 11.3 on the Intel(R) Core(TM) i7-6500U CPU, 2.50GHz.

```
Data: $d$, $k$, $t$, ($t \in [0,1]$)
Result: $x \in [0,1]^d$
$x \leftarrow (1,1,\dots,1)$ ;
$b_d \leftarrow 1$;
for $i \leftarrow 1$ to $d-1$ do
    $b_d \leftarrow 2^i - b_d + 1$ ;
end
$cd \leftarrow b_d * 2^{-d}$ ;
$KM \leftarrow [\,]$;
for $j \leftarrow 1$ to $k$ do
    $km \leftarrow \lfloor t * 2^d - 1 + cd \rfloor + 1$ ;
    $(\star) t \leftarrow t * 2^d + cd - km$ ;
    if $km == 2^d$ then
        $km \leftarrow 0$ ;
    end
    append $km$ to $KM$ ;
end
for $j \leftarrow 1$ to $k$ do
    $km \leftarrow KM[k - j + 1]$ ;
    $B \leftarrow [\,]$;
    while $i < d + 1$ do
        $be \leftarrow 1$;
        if $km < 2^{d-i}$ then
            $be \leftarrow 0$;
        end
        append $be$ to $B$ ;
        $km \leftarrow km - be * 2^{d-i}$;
        if $be == 1$ then
            $km \leftarrow 2^{d-i} - km - 1$ ;
        end
        $i \leftarrow i + 1$;
    end
    for $i \leftarrow 1$ to $d$ do
        $x_{d-i+1} \leftarrow 1/2 - (1/2 - B[i]) x_{d-i+1}$;
    end
end
modification for $d = 2$:
$(\star) t \leftarrow (cd + km - 1 + t)/2^d$ by
if $km$ is odd then
    $t \leftarrow (-cd + km + 1 - t)/2^d$ ;
else
    $t \leftarrow (cd + km - 1 + t)/2^d$ ;
end
```

Algorithm 1: Mapping of $t \in [0,1]$ into a point $x \in [0,1]^d$

Calculation of the transformation of a single point from a space with dimension $d = 4$ and accuracy 2^{-k}, $k = 10$ required approx. 0.000625 s. With the same accuracy and $d =$

```
Data: $d$, $k$, $(x_1, x_2, \dots, x_d)$, ( $x \in [0,1]^d$)
Result: $t \in [0,1]$
$b_d \leftarrow 1$;
for $i \leftarrow 1$ to $d-1$ do
    $b_d \leftarrow 2^i - b_d + 1$ ;
end
$cd \leftarrow b_d * 2^{-d}$, $t \leftarrow 1 - cd$, $KM \leftarrow [\,]$, $B \leftarrow [\,]$;
for $j \leftarrow 1$ to $k$ do
    for $i \leftarrow 1$ to $d$ do
        if $x_i < 0.5$ then
            $be \leftarrow 0$, $x_i \leftarrow 1 - 2 * x_i$, append $be$ to $B$ ;
        else
            if $x_i \geq 0.5$ then
                $be \leftarrow 1$, $x_i \leftarrow 2 * x_i - 1$, append $be$ to
                  $B$ ;
            end
        end
        $ww \leftarrow 0$, $km \leftarrow 0$ ;
        for $i \leftarrow 1$ to $d$ do
            if $be + ww == 1$ then
                $km \leftarrow km + 2^{d-i}$ ;
            end
            $ww \leftarrow |be - ww|$;
        end
    end
    if $km == 2^d$ then
        $km \leftarrow 0$ ;
    end
    append $km$ to $KM$;
end
for $j \leftarrow 1$ to $k$ do
    $km \leftarrow KM[k - j + 1]$ ;
    $(\star) t \leftarrow (cd + km - 1 + t)/2^d$ ;
    if $t < 0$ then
        $t \leftarrow 1 + t$ ;
    end
end
modification for $d = 2$:
$(\star) t \leftarrow t * 2^d + cd - km$ by
$t \leftarrow t * 2^d + cd - km$ ;
if $km$ is odd then
    $t \leftarrow 1 - t$ ;
end
```

Algorithm 2: Mapping of $x \in [0,1]^d$ into $t \in [0,1]$

2 the execution time was shortened twice.

IV. Concluding Comments

We present a simple algorithm for computing a transformation of multidimensional data points onto the unit interval using the proposed a Sierpiński type space-filling curve generalization. Specific ideas regarding the generalization of the Sierpiński curve were considered in the author's monograph [37], but the algorithms presented in this work are new and have not been published anywhere.

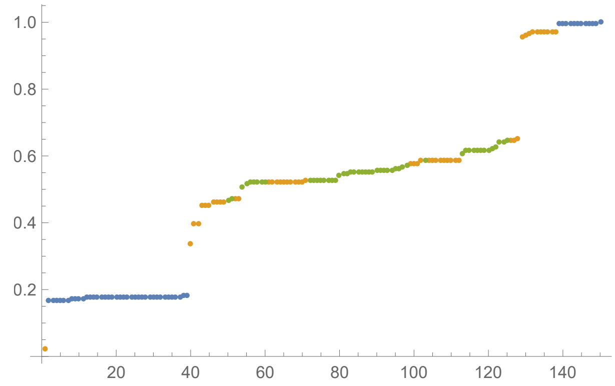


Fig. 5: Three species of the 4-dimensional Iris data: Iris Setosa (blue), Iris Versicolour (yellow), and Iris Virginica (green)-150 data points - after dimensionality reduction.

It is known that there is a close relationship between topological dimension d of the I_d cube and the maximum value of the Hölder's exponent of a space-filling curve. The value of $1/d$ is the maximum value. There is no d-dimensional space-filling curve with an exponent greater than $1/d$ [18]. Hölder's inequality results in an important property of space-filling based transformations, which is to ensure the proximity of data points whose counter-images are closely located on a unit interval. Multi-dimensional Sierpiński curves are another tool for analyzing multidimensional data.

References

[1] Bader M., Schraufstetter S., Vigh C., Behrens J.: Memory effcient adaptive mesh generation and implementation of multigrid algorithms using Sierpinski curves, International Journal of Computational Science and Engineering, 4(1), 1742–7193, (2008).

[2] Bader M.: Space-Filling Curves. An Introduction with Applications in Scientific Computing. Springer-Verlag Berlin Heidelberg (2013).

[3] Bartholdi J. J., Platzman L. K.: Heuristics based on spacefilling curves for combinatorial problems in Euclidean space, Management Sci., 34, 291–305, (1988).

[4] G. Bhatnagar, Q. M. J. Wu, and B. Raman: Image and video encryption based on dual space-filling curves, Computing Journal,55(6), 667–685, (2012).

[5] Bungartz H.-J., M. Mehl M., T. Neckel T., and T. Weinzierl T.: The PDE framework Peano applied to fluid dy- namics: an efficient implementation of a parallel multiscale fluid dynamics solver on octree-like adaptive Cartesian grids, Computational Mechanics, 46(1), 103–114, (2010).

[6] Butz A.R.: Alternative algorithm for Hilbert's space-filling curve, IEEE Trans. on Computing, 20, 424–426 (1971).

[7] Falconer K.: , Fractal Geometry: mathematical foundations and applications, John Wiley & Sons Ltd., Chichester, 3-nd ed.(2014).

[8] Fisher,R.A.: The use of multiple measurements in taxonomic problems, Annual Eugenics, 7 II, 179–188 (1936).

[9] Gilbert E. N. :, Gray codes and Paths on the n-cube, Bell System Tech. J., 37, 815–826 (1957).

[10] Gotsman C., Lindenbaum M.:, On the metric properties of discrete space-filling curves, IEEE Trans. on Image Processing, 5(5), 794–797 (1996).

[11] He Z. and Owen A. B.: Extensible grids: uniform sampling on a space filling curve, Journal of the Royal Statistical Society. Series B (Statistical Methodology) 78(4), 917–935, (2016).

[12] Hilbert D. : Ueber die stetige Abbildung einer Linie auf ein Flaechenschtueck, Mathematische Annalen 38, 459–469, (1891).

[13] Kim, H.I.; Hong, S.; Chang, J.W.: Hilbert curve-based cryptographic transformation scheme for spatial query processing on outsourced private data. Data & Knowledge Engineering 104 (C), 32–44 (2016).

[14] Kumar G.S., Pandithevan P., Ambatti A.R.: Fractal raster tool paths for layered manufacturing of porous objects, Virtual and Physical Prototyping 4(2),91–104, (2009).

[15] Lawder J. K. and King P. J. H.:Using space-filling curves for multidimensional indexing, in Proc. 17th BNCOD, in Lecture Notes in Computer Science, 1832. 20–35, (2000)

[16] Lian, H., Qiu, W., Yan, D., Guo J., Li Z., Tang, P.: Privacy-preserving spatial query protocol based on the Moore curve for location-based service, Computers & Security, 96, 101845, (2020).

[17] Melian, S., Brix, K., Müller, S., Schieffer, G. (2011). Space-Filling Curve Techniques for Parallel, Multiscale-Based Grid Adaptation: Concepts and Applications. In: Kuzmin, A. (eds) Computational Fluid Dynamics 2010. Springer, Berlin, Heidelberg

[18] Milne S.C.:, Peano Curves and Smoothness of Functions, Advances in Mathematics 35 ,129-157 (1980).

[19] Moon, B., Jagadish, H.V., Faloutsos, C., Saltz, J.H.: Analysis of the clustering properties of the hilbert space-filling curve. IEEE Trans Knowl Data Eng 13 (1), 124–141 (2001). doi:10.1109/69.908985

[20] Moore E. H.:, On certain crinkly curves, Trans. Amer. Math. Soc., 1 (1900), pp. 72–90.

[21] Nair S.H., Sinha A., Vachhani L.: Hilbert's space-filling curve for regions with holes, Proc. of IEEE Conference on Decision and Control, 313–319 (2017).

[22] Ouni, T., Lassoued, A. and Abid M.: Lossless image compression using gradient based space filling curves (G-SFC). Signal, Image and Video Processing 9, 277–293 (2015).

[23] O'Shaughnessy S., Sheridan S.:Image-based malware classification hybrid framework based on space-filling curves, Computers & Security, 116, 102660, (2022).

[24] Patrick E.D., Anderson D.R., Bechtel F.K., Mapping multidimensional space to one dimension for computer output display, IEEE Trans. on Comput. 17, 949–953, (1968).

[25] Peano G.:, Sur une courbe qui remplit toute une aire plane, Math. Ann., 36, 157–160 (1890).

[26] L. K. Platzman and J. J. Bartholdi:, Spacefilling Curves and the Planar Traveling Salesman Problem, Journal of ACM, 36, 719–737 (1989).

[27] Sagan H., Space-filling Curves, Springer-Verlag, New York, 1994.

[28] Schrack G., Stocco L. : Generation of Spatial Orders and Space-Filling Curves, IEEE Trans. on Image Processing, 24(6, 1791–1800, (2015).

[29] S. Sharma, K. Hari and G. Leus, Space filling curves for MRI sampling in IEEE ICASSP, 1115–1119, (2020).

[30] Sierpiński W.:, O pewnej krzywej wypełniającej kwadrat. Sur une nouvelle courbe continue qui remplit toute une aire plane., Bulletin de l'Acad. des Sciences de Cracovie A.,463–478, (1912).

[31] Sergeyev, Y.D.: An information global optimization algorithm with local tuning. SIAM J. Optim. 5, 858–870 (1995)

[32] Sergeyev Y.D., Strongin R.G. , Lera D.: Introduction to Global Optimization Exploiting Space-Filling Curves, Springer New York Heidelberg Dordrecht London (2012).

[33] Sergeyev Y.D., ·Maria Chiara Nasso M.C.,Lera D.: Numericalmethods using two different approximations of space-filling curves for black-box global optimization, Journal of Global Optymization, published online 08.08.2022 https://doi.org/10.1007/s10898-022-01216-1, published online 08.08.2022.

[34] Skubalska-Rafajłowicz E.:, Applications of the space-filling curves with data driven measure–preserving property, Nonlinear Analysis, Theory, Methods and Applications, 30(3), 1305–1310, (1997).

[35] Skubalska-Rafajłowicz E., Rafajłowicz E.: Searching for optimal experimental designs using space-filling curves Applied Mathematics and Computer Science. 8(3) 647–656 (1998).

[36] Skubalska-Rafajłowicz E., Rafajłowicz E.: Space-filling curves in generating equidistrubuted sequences and their properties in sampling of images W: Signal processing / ed. by Sebastian Miron. Vukovar : In-Teh, 131–149 (2010.

[37] Skubalska-Rafajłowicz E.:, Krzywe wypełniające w rozwiązywaniu wielowymiarowych problemów decyzyjnych, Wydawnictwo Politechniki Wrocławskiej, 2001.

[38] Skubalska-Rafajłowicz E.:, Pattern recognition algorithm based on space-filling curves and orthogonal expansion, IEEE Trans. on Information Theory 47(5) 1915–1927, (2001).

[39] Skubalska-Rafajłowicz E.: A new method of estimation of the box-counting dimension of multivariate objects using space-filling curves. Nonlinear Analysis, Theory, Methods & Applications. Series A, Theory and Methods. 63 (5–7), 1281–1287,(2005).

[40] Skubalska-Rafajłowicz E., Rafajłowicz E.: Sampling multidimensional signals by a new class of quasi-random sequences. Multidimensional Systems and Signal Processing. 23 (1/2) 237–253, (2012).

[41] Steele J. M.: Efficacy of Spacefilling Heuristics in Euclidean Combinatorial Optimization, Operations Research Letters 8 , 237–239, (1989).

[42] Vogiatzis D. and Tsapatsoulis N.:Clustering Microarray Data with Space Filling Curves, in Proceedings of the 7th international workshop on Fuzzy Logic and Applications: Applications of Fuzzy Sets Theory, LNAI 4578, 529–536, (2007).

[43] Zhou L., C. R. Johnson and D. Weiskopf: Data-Driven Space-Filling Curves, IEEE Transactions on Visualization and Computer Graphics, 27(2) 1591–1600, (2021).

DOI: 10.15439/2023F1379
ISSN 2300-5963 ACSIS, Vol. 37

Knowledge-Based Creation of Industrial VR Training Scenarios

Paweł Sobociński, Jakub Flotyński, Michał Śliwicki, Mikołaj Maik, Krzysztof Walczak
Department of Information Technology, Poznań University of Economics and Business
Email: [pawel.sobocinski, jakub.flotynski, michal.sliwicki, mikolaj.maik, krzysztof.walczak]@ue.poznan.pl

***Abstract*—The application of virtual reality (VR) for building training systems has grown in popularity across diverse fields. This trend is particularly prevalent in Industry 4.0, where many real-world training scenarios can be expensive or pose potential dangers to trainees. The most important aspect of professional training is domain-specific knowledge, which can be expressed using the semantic web approach. This approach facilitates complex queries and reasoning against the representation of training scenarios, which can be useful for educational purposes. However, current methods and tools for creating VR training systems do not utilize semantic knowledge representation, making it difficult for domain experts without IT expertise to create, modify, and manage training scenarios. To address this issue, we propose an ontology-based representation and a method of modeling VR training scenarios. We demonstrate our approach by modeling VR training scenarios for Industry 4.0 in the field of the production of household equipment. The domain knowledge used represents training activities, potential errors, and equipment failures in a way comprehensible to domain experts.**

I. Introduction

COMPARED to traditional training methods, VR training systems offer significant advantages. First, training in VR is more engaging and attractive to users compared to paper, audio, or video materials. Second, virtual training eliminates the need for physical infrastructure or dangerous equipment, reducing the risk posed to users. Moreover, it liberates companies from acquiring expensive or unavailable devices, especially in Industry 4.0 environments where production devices cannot be suspended. Finally, VR training can be carried out to a certain degree without the need for instructors. This makes it simpler to organize, more cost-effective, more efficient, and more flexible compared to conventional training methods.

However, creating effective VR training environments with behavior-rich scenes and objects requires expertise in programming and 3D modeling, as well as domain knowledge to prepare practical and meaningful training scenarios in a specific domain. As a result, the development process often involves collaboration between IT specialists and domain experts, who typically have limited knowledge of IT. This collaboration can make the development of VR training environments complex, time-consuming, and costly. Therefore, the availability of user-friendly tools for domain experts to design VR training with domain knowledge is crucial in reducing the required time and effort and promoting the use of VR in training.

The semantic web is a leading method for representing domain knowledge, providing a range of standards for conveying content in a manner understandable to humans and processable by software. Ontologies are the primary form of content representation in the semantic web, formulated using the Resource Description Framework (RDF), the Resource Description Framework Schema (RDFS), and the Web Ontology Language (OWL). RDF establishes a data model, whereas RDFS and OWL expand RDF terminology allowing to build ontologies. The semantic web standards rely on description logic, which enables the representation of concepts, roles, and individuals. Such representations can be subject to reasoning, leading to the inference of implicit knowledge based on explicit knowledge and precise queries, including highly complex conditions. This is highly beneficial for content creation and management by users across various domains.

To date, the semantic web has primarily been used for the representation of 3D content, including its geometry, structure, and presentation, which is insufficient for managing complex VR training, with its users, tasks, and equipment, as well as possible problems and errors. User-friendly tools are needed for domain experts to design VR training with domain knowledge, making the development of training environments less complicated, less time-consuming, and more cost-effective.

In this paper, we present a new method for creating VR training scenarios that utilize the semantic web. Our approach includes two primary components: an ontology-based representation of domain knowledge in training scenarios and a user-friendly semantic scenario editor. The ontology-based representation covers various elements such as users, tasks, equipment, and potential problems or errors that may arise during training scenarios. Using the semantic scenario editor, domain experts can easily design scenarios through an intuitive visual interface. This method allows for domain-specific descriptions of training scenarios and scenes using well-known semantic web standards. Furthermore, the process of selecting and combining appropriate objects for training scenarios, as well as verifying modeling results, can be completed using well-recognized activities on description logics such as instance checking, query answering, and consistency checking against the used ontologies.

The project discussed in this paper focuses on developing a VR training system for the production of household appliances. Hence, all examples and discussions are centered on this application domain. However, the proposed approach can be adapted for other domains if the relevant objects and actions are identified and semantically described.

The remainder of the paper is structured as follows: Section II provides an overview of the current state of VR training environments and existing approaches to semantic modeling of VR content. Section III outlines the proposed approach, while Section IV explains the ontology-based representation of training scenarios. The semantic scenario editor, which utilizes this representation, is discussed in Section V. Section VI presents an example of VR training. Section VI-F provides a discussion of the results. Finally, Section VII concludes the paper and indicates possible future research.

II. Related work

Up until now, little attention has been given to the utilization of ontologies in virtual reality training and education by the research community. A study by [1] proposed ontologies for creating VR training at various levels of abstraction, including high, medium, and low. The high-level ontology defines entities that represent physical and non-physical objects that may occur, such as avatars, tools, vehicles, roles, animals, and events that could happen in VR environments. The medium-level ontology builds on the high-level ontology by providing a classification of avatars, tools, vehicles, roles, and animals with more concrete entities. The low-level ontology describes entities that are specific to a particular VR environment.

A medical diagnosis system has been described in [2]. It leverages an ontology-based approach to represent medical knowledge, where separate ontologies are utilized to illustrate patients' physical and mental states. An avatar, which communicates with patients through voice, employs these ontologies. To make diagnoses, the system employs probabilistic reasoning with a Bayesian network.

Numerous studies have focused on representing 3D content through ontology-based approaches, which involve a range of geometrical, structural, spatial, and presentational elements. An extensive evaluation of these methods has been provided in [3], and a summary of the existing techniques can be found in Table I. Among the methods, four aim to address low-level abstraction that is specific to graphics, while six approaches support high-level abstraction that is either general or specific to a domain. Furthermore, three of these methods can be employed with different domain ontologies.

TABLE I: Comparison of semantic 3D content modeling methods

	Level of Abstraction	
Approach	Low (3D graphics)	High (application domain)
De Troyer et al. [4]	✓	general
Gutiérrez et al. [5]	✓	humanoids
Kalogerakis et al. [6]	✓	-
Spagnuolo et al. [7]	-	humanoids
Floriani et al. [8]	✓	-
Kapahnke et al. [9]	-	general
Albrecht et al. [10]	-	interior design
Latoschik et al. [11]	-	general
Drap et al. [12]	-	archaeology
Trellet et al. [13]	-	molecules
Perez-Gallardo et al. [14]	✓	-

Another example of a knowledge-based 3D design method has been described in [15]. The paper presents a collaborative method for the interactive development of aircraft cabin systems in VR based on preliminary design data. The knowledge is stored in an ontology which is linked with design rules and external parameters, which can generate missing information needed for the design of cabin systems. The design rules are based on requirements, safety regulations as well as expert knowledge for design interpretation that has been collected and formalized. The data is stored in an XML file that can be used to generate a 3D virtual cabin mockup in which users have the possibility to interact with cabin modules and system components via controllers. This VR model enables interaction with complex product data sets by visualizing metadata and analysis results along with the cabin geometry, making it even better comprehensible and processable for humans. It allows the design to be evaluated and optimized at a low cost before the concepts are validated in a real prototype.

Another example of ontologies for VR is the OntoPhaco project presented in [16]. The goal of the OntoPhaco project was to develop a new approach to the evaluation and design of ontologies in ophthalmology, specifically for cataract surgery training. The authors propose a solution on how to design a proper domain model to support VR training in ophthalmology, which includes the OntoPhaco ontology, built using OntoUML based on UML. They also introduce systematic verification and validation processes that include theoretical and hypothetical evaluation of the system and the use of feedback from domain experts to verify and revise the ontology for VR training. The conducted evaluation shows, that OntoPhaco has the potential to improve the learning experience of students and facilitate the development of VR training in the future.

There is also an example of ontology-based, general-purpose and Industry 4.0-ready architecture to use with systems supporting factory workers that use mixed reality [17]. In the paper, authors describe a general ontology, that is capable of structuring knowledge to enable interoperability and standardization between such systems. The approach enables data findability and reusability. The proposed architecture was implemented and validated in two case studies in the manufacturing sector: scheduled maintenance and alarm management, and customer order management.

III. Overview of the Approach

The review presented in Section II shows that universal, cross-domain methods and tools for creating interactive VR training scenarios are still missing. The existing ontologies for VR are limited to either 3D-specific features that focus on the properties of static 3D content or domain-specific features that focus on a single application domain. There is a lack of domain-independent conceptualization of actions and interactions, which can be utilized by non-technical users to create VR environments with minimal assistance from programmers and graphics designers. Solutions that focus on 3D content behavior, like [18], are broad in scope and do not provide the specific concepts and roles required for training scenarios.

The main contribution of this paper is a solution to the problem mentioned above: an approach to semantic representation and modeling of VR training scenarios. The approach is illustrated in Fig. 1 and consists of two key elements: the *ontology-based representation of training scenarios* and the *semantic scenario editor*. The *ontology-based representation* comprises resources based on semantic web standards such as RDF, RDFS, and OWL, which cover training scenarios, scenes, and objects in terms of both their semantics and visualization. The central component of the representation is the *scenario ontology*, consisting of a TBox and an RBox, which includes concepts (classes) and roles (properties) associated with training scenarios, scenes, infrastructure objects, and equipment. Since the classes and properties are general, the ontology can be utilized in various application domains.

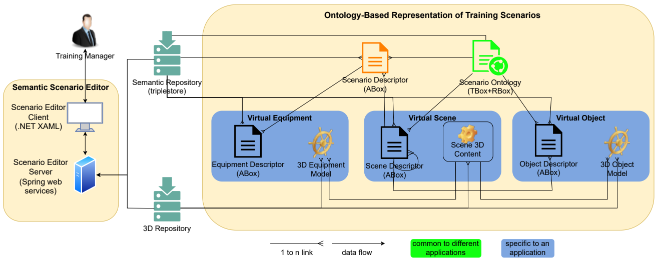

Ontology-Based Representation of Training Scenarios
Training Manager
Semantic Scenario Editor
Scenario Editor Client (.NET XAML)
Scenario Editor Server (Spring web services)
Semantic Repository (triplestore)
3D Repository
Scenario Descriptor (ABox)
Scenario Ontology (TBox+RBox)
Virtual Equipment
Equipment Descriptor (ABox)
3D Equipment Model
Virtual Scene
Scene Descriptor (ABox)
Scene 3D Content
Virtual Object
Object Descriptor (ABox)
3D Object Model
1 to n link
data flow
common to different applications
specific to an application

Fig. 1: Overview of knowledge-based representation and modeling of VR training scenarios.

Based on the scenario ontology, four kinds of *descriptors* are created: *scenario descriptors*, *scene descriptors*, *object descriptors*, and *equipment descriptors*. Each descriptor is an ABox that represents individuals linked to a specific scenario, scene, object, or piece of equipment, respectively. These individuals are characterized using classes and properties defined in the scenario ontology. Furthermore, every descriptor associated with a scene, an object, or a piece of equipment is connected to relevant synthetic content, which consists of hierarchical, interconnected and reusable 3D components, 2D graphics, as well as scripts that implement animations and interactions. The scenario ontology and descriptors are comprehensively described in Section IV. The creation of synthetic 3D content for scenes, objects, and equipment is achievable through the use of our scene editor and the Unity game engine, but it is beyond the scope of this paper [19].

The scenario ontology and descriptors are stored in a *Semantic Repository*, which is a triplestore, whereas the 3D content of scenes and objects is stored in the *Content Repository*, which is a relational database.

The *Semantic Scenario Editor* is a client-server application comprising a *client* and a *server*. The client is a desktop application designed utilizing .NET, with a GUI described in the XAML language, while the server is a Java-based application that provides RESTful web services developed using the Spring library. The client offers a user-friendly interface that enables a *Training Manager* to create and modify training scenarios by utilizing the scenario ontology, descriptors, and 3D content stored in the repositories. The editor is described in detail in Section V.

IV. Semantic Representation of VR Training Scenarios

The proposed semantic representation of VR training scenarios is based on an ontology and uses domain-specific classes and properties, which are comprehensible to domain experts. The representation comprises two primary components: the *scenario ontology* which is common to different domains and applications, and descriptors that are domain-specific and built on top of the ontology.

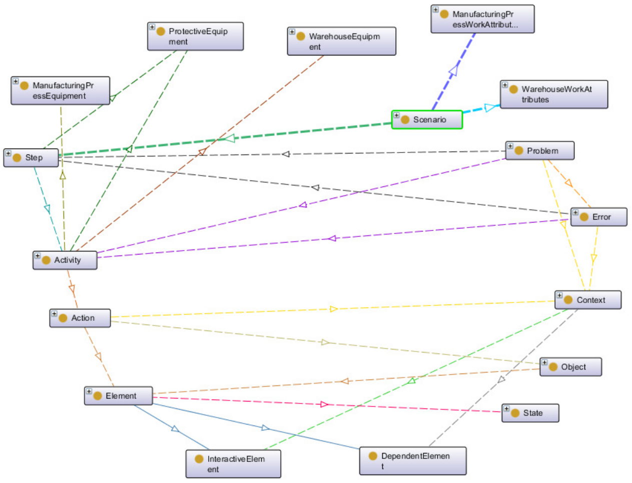


Fig. 2: Classes and properties of the scenario ontology.

The classes and properties that constitute the ontology are illustrated in Figure 2. They can be categorized into four distinct groups:

a) Training Objects: *Training objects* are the primary elements of every VR training scenario, e.g., forklifts, pallet trucks, and batteries. A virtual object is a tuple of an *object descriptor*, which semantically represents the physical object used in the training, and a 3D model, which visually represents the physical object. An object descriptor is an assertional box that describes individuals related to the physical training object using classes and properties specified in the scenario ontology. In an object descriptor, the physical object is represented by an individual of the *Object* class. An object individual comprises individuals of the *Element* class. Hence, it forms a hierarchy. Objects' elements have possible *states*. Elements' states can be changed during VR training. There are two types of *elements*: *interactive element* and *dependent element*. The state of an *interactive element* is changed by a trainee, whereas the state of a *dependent element* is changed as a consequence of a change of the state of an interactive element. Every object, element, and state has a name, description, and image intelligible to domain experts.

b) Virtual Equipment: *Virtual equipment* represents either work equipment, e.g., a toolkit from which users can select tools needed to complete the scenario, or protective equipment, e.g., helmets and gloves. A *virtual piece of equipment* is a pair of an *equipment descriptor*, which semantically represents the physical equipment, and a 3D model, which visually represents the physical equipment. An equipment descriptor is an assertional box that represents individuals related to the equipment using classes and properties specified in the scenario ontology. In an equipment descriptor, a piece of equipment is represented by an individual of a sub-class of the *Equipment* class. Different sub-classes of equipment may be specified depending on the particular domain of training. Every piece of equipment has a name and description, which are understandable to domain experts.

c) Training Scenes: Each VR training scenario is designed for a specific *VR training scene*, which comprises a *scene descriptor* and synthetic 3D content. The scene descriptor represents individuals related to the training scene using the classes and properties from the scenario ontology. A training scene is an individual of the *Scene* class, which includes individual objects of the *Object* class. Likewise, the synthetic content of the virtual scene includes the 3D models of the virtual objects. Every scene has a name and a description, which are understandable to domain experts.

In practical VR training applications, multiple virtual scenarios may be designed for slightly different virtual scenes, such as two factories with different placement of battery charging points. As a result, a scene may be a super-scene to other scenes, and each scene may inherit from another scene. The scene descriptors describe which objects are included or excluded in a scene or its sub-scenes, and every scene includes an object that is either included directly in the scene or included in its super-scene but not excluded in any scene on the inheritance path to the super-scene.

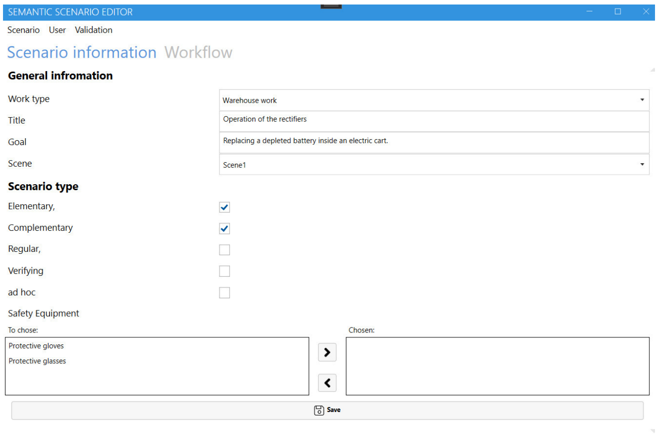
V. Training Scenario Editor

A. Architecture

The Semantic Scenario Editor is composed of a client-server system consisting of two main parts: the *Scenario Editor Server* and the *Scenario Editor Client*. The Scenario Editor Server is a Java-based program with RESTful web services using the Spring library, which accesses the *Semantic Repository* and the *Content Repository*. The system offers four services. The *Scene Service* allows for the selection of scenes that can be used to create training scenarios. Every scene can have a different set of available objects. The *Object Service* provides objects, their elements, and their respective states. The *Equipment Service* provides the available equipment for training in the application domain, which is common to all potential training scenarios. The *Workflow Service* provides information about the workflow of scenarios. The workflow consists of steps, which are divided into activities. Each activity consists of several actions as well as possible problems and errors. Such a structure allows for an easier understanding and editing of the scenario workflow. The Semantic Repository is supported by the Apache Fuseki server, which enables semantic reasoning and query processing.

The *Scenario Editor Client* is a user-friendly visual tool that training managers use. It is based on Windows Presentation Foundation. The main purpose of the tool is to permit the specification of training scenarios. The client displays scenario attributes and their possible values in different fields of visual forms (see Fig. 3). The attributes are accessed from and saved to the Semantic Repository via the Scenario Editor Server. The forms are presented in a simple layout that includes attribute names, text boxes, and drop-down lists where the manager can enter the necessary information. The drop-down lists show values obtained from the *scenario ontology*.

The general information includes the scenario title and the type of work, which may be either warehouse work or manufacturing press work. The manager also specifies whether the scenario is *elementary*, *complementary*, *regular*, *verifying*, or *ad hoc*. Finally, the manager selects the necessary pieces of protective equipment to complete the scenario from the list of all available equipment.



General infromation

Work type	Warehouse work
Title	Operation of the rectifiers
Goal	Replacing a depleted battery inside an electric cart.
Scene	Scene1

Scenario type

Elementary,	☑
Complementary	☑
Regular,	□
Verifying	□
ad hoc	□

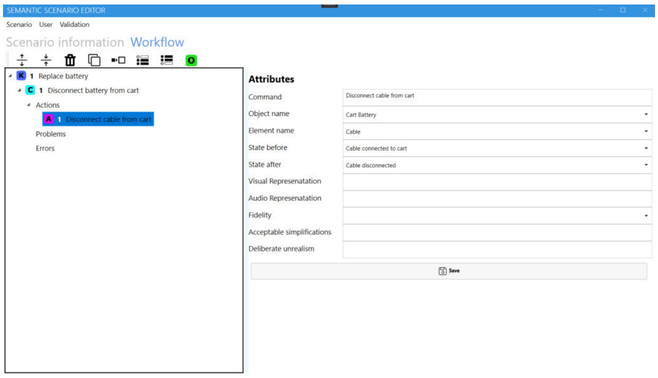
Safety Equipment

To chose:	Chosen:
Protective gloves Protective glasses	

Fig. 3: General information about a scenario.

In addition to providing general information, the author also specifies a scenario's workflow, which includes steps, activities, and actions that trainees must perform. In each scenario, there must be at least one step that contains at least one activity, which in turn contains at least one action (see Section IV). Actions are linked to interactive and dependent objects' elements, as well as potential issues and errors that may arise during the action.

In the Scenario Editor, the workflow of each scenario is presented in the form of a tree, which is a widely used and easy-to-understand method for displaying hierarchical data (see Fig. 4). The tree includes scenario steps, activities, actions, problems, errors, and objects, which are represented by distinct icons. The editing manager can expand and collapse the list of sub-items for each item in the tree. In addition, there are optional sub-items for grouping actions, errors, and problems in activity and problem items. Using the toolbar and context menu, the author can visually add, modify, and delete items in the tree. Moreover, the order of the steps, activities, and actions can be changed by dragging and dropping.



K 1 Replace battery
C 1 Disconnect battery from cart
Actions
A 1 Disconnect cable from cart
Problems
Errors

Attributes

Command: Disconnect cable from cart
Object name: Cart Battery
Element name: Cable
State before: Cable connected to cart
State after: Cable disconnected
Visual Represenatation:
Audio Represenatation:
Fidelity:
Acceptable simplifications:
Deliberate unrealism:

Fig. 4: Tree view of a scenario's workflow.

VI. Example VR Training Scenario

A. General information

We have implemented a prototype of a VR training system based on the semantic training scenario representation using the Unity game engine. To permit the completion of the training scenario without using the hardware controllers, the *Oculus Integration* plug-in has been utilized together with supplementary plug-ins. Thanks to the advanced capabilities of the equipment, which allows for direct tracking of users' hands, prospective trainees are not required to learn how to operate VR controllers. Instead, they can interact with objects in the virtual environment directly using their own hands, which increases the level of immersion.

The system has been developed utilizing resources provided by *Amica S.A.*, which is one of the largest producers of household equipment in Poland. These resources consist of in-depth information regarding the training process for the designated workstation, as well as comprehensive information regarding the relevant equipment. These resources became a base for creating scenes and objects used in training scenarios regarding the operation and maintenance of electric carts, forklifts and industrial presses. The elements of the scene are described by scenario descriptors, scene objects, and equipment descriptors. The main goal of the system is to teach the trainee, how to use industrial equipment safely.

B. Training scenario

In order to show the functionality of the system, we have created an example scenario that uses the developed semantic representation. The scenario has been formulated to illustrate knowledge about the procedure of replacing a depleted battery inside an electric cart. Trainees can visually inspect the condition of the cart, as per the official documentation of the real-life training session, and are also capable of operating the cart, provided that the battery is functional and all connections are properly established. All components have been designed and programmed to accurately replicate the simulated reality, including such factors as cable physics, electrical plug connections, battery charge level and depletion during operation, as well as a realistic electric cart driving system that permits trainees to execute the battery replacement procedure at any designated workstation.

The workflow for this scenario has been developed in collaboration with domain experts to enable a realistic training experience for trainees. The scenario has been created in accordance with the provided guidelines, and block diagrams have been used to illustrate the training process. Functionality facilitating the monitoring of trainees' progress during training, as well as the provision of pertinent feedback regarding positive or negative training outcomes has been incorporated.

C. Training scene

The virtual scene in which the training scenario takes place has been constructed using materials furnished by the company as well as photographs obtained during a site visit to the production hall (Fig. 5). Adequate lighting has been selected to accurately simulate the conditions present in the authentic production hall. Since the training scenario is conducted in only a small section of the hall, rendering is optimized for performance by limiting the display to the sector of the production hall utilized during the training, i.e., the section of the hall containing batteries and chargers, enabling charging and replacement of batteries in electric carts.

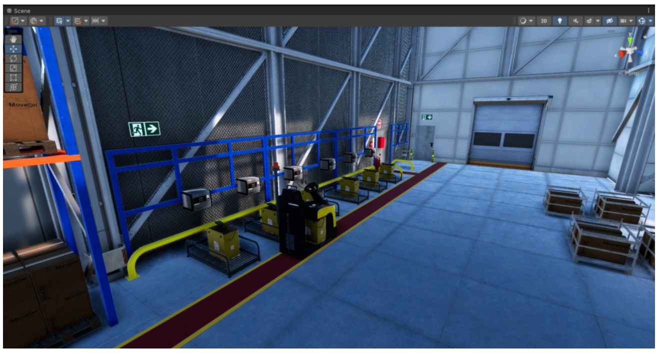


Fig. 5: Production hall in the scenario

D. Virtual objects

In industrial VR training scenarios, *virtual objects* representing real equipment are key to construction of accurate representation of a real-life training exercise. For the scenario, a number of 3D models have been developed specifically for this purpose, utilizing photographs obtained during the site visit to the production hall. The most critical virtual objects include the electric cart, battery, plugs, and charging stations (Fig. 6-9). These objects, in combination with the main *player object* representing the trainee, are employed to construct an environment that enables the execution of the scenario.

E. Scenario steps

By utilizing the aforementioned semantic representations of *Objects*, *Steps, Actions*, and the description of the scene, the initial training scenario has been created. The primary objective of the trainee is to replace a nearly depleted battery

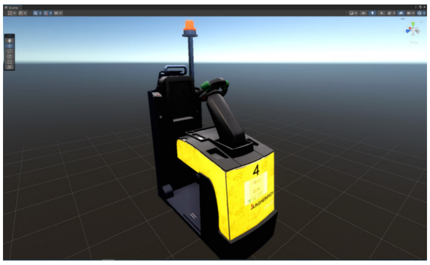


Fig. 6: Electric cart

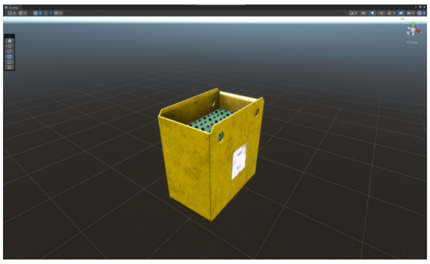


Fig. 7: Battery for the electric cart

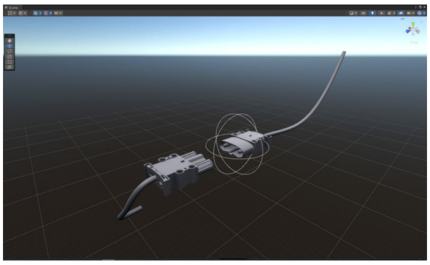


Fig. 8: Plugs of the electric cart

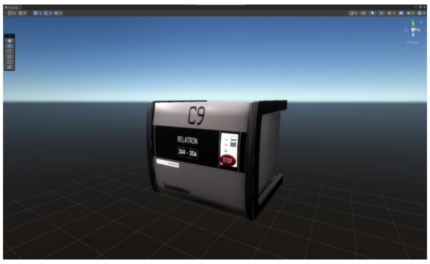


Fig. 9: Charging station for the battery

inside the electric cart with a new one from the charging station. The training scenario consists of the following steps:

1) Perform a visual inspection of the electric cart (Fig. 10).
2) Start the cart by ensuring that the battery is connected and locked and then use the electronic card to initiate the power-up sequence (Fig. 11).
3) Drive the cart into the correct position alongside the charging station (Fig. 12).
4) Replace the nearly depleted battery with a fully charged one (Fig. 13).
5) Connect the new battery unit to the cart and ensure that the blockade is properly in place (Fig. 14).
6) Conduct a visual and manual inspection to verify that the new battery unit is correctly installed and that the cart is in good working condition.
7) Drive the cart away from the charging station to finish the scenario.

At the beginning of the scenario, the trainee is situated in front of the cart and must either physically move into the correct position or use the appropriate hand gesture to teleport themselves into the desired position within the cart. The subsequent step is to verify that the cart is in good condition. The cart may appear fully functional or exhibit visual cues indicating its malfunction. This is achievable by adjusting various settings, which enable the *virtual objects* to exhibit signs of damage or malfunction. Once the visual inspection is completed, the next objective is to activate the vehicle. This step is fairly intricate and may prove challenging for inexperienced trainees, however, this is a deliberate design decision, as the scenario is intended to provide trainees with a safe and controlled environment for practice.

Activating the electric cart involves several *Steps*. Initially, the trainee must verify that the battery is properly connected and securely locked inside the cart. Subsequently, an electronic card is utilized to initiate the vehicle's power-up sequence. If the battery is not appropriately connected and/or locked, a message will be displayed in the console log, indicating an *Error*. If this situation arises, the trainee may opt to either restart the scenario by pressing the "R" key on the keyboard or attempt to correctly connect and lock the battery and then reattempt to activate the vehicle. Once the trainee completes this process, the electric cart is primed for operation.

With the activation of the cart and the trainee correctly using the steering wheel with both hands, the electric cart can be driven. A refined driving model supports precise control of the cart, paralleling the real-world scenario. Additionally, the employment of hand tracking elevates the immersive nature of the driving experience.

The subsequent objective entails driving the cart into the appropriate position, allowing for the replacement of the battery *Object*. The trainee is permitted to drive the cart to any of the charging stations where a battery is available for replacement, as indicated by the illuminated lights at the stations. The charging stations are meticulously modeled and scripted to mirror their real-life behavior, thereby enabling trainees to learn about the various color-coded markings displayed at the stations, the available actions and how to respond in an emergency.

Once the cart is positioned correctly alongside the charging station, the trainee must utilize their electronic card to power down the vehicle and remove the battery blockade to unlock the battery. To remove the battery, the trainee must execute the appropriate hand gesture to grasp and extract the battery from the cart. Next, the trainee must unlock the fully charged battery from the charging station, retrieve it, and install it into the now empty slot.

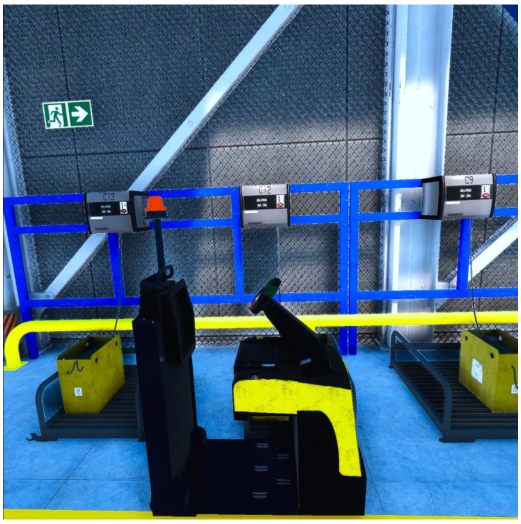


Fig. 10: Visual inspection of the electric cart

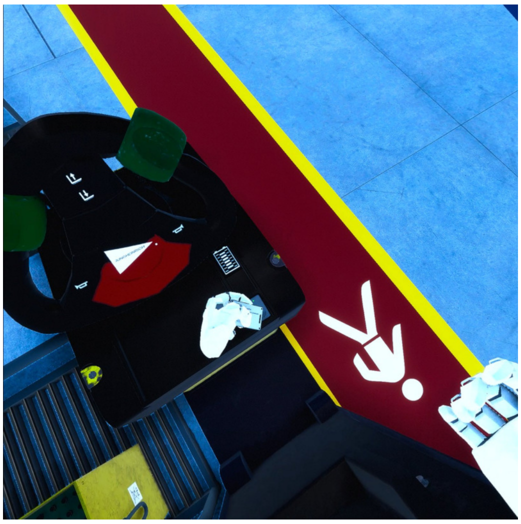


Fig. 11: Turning on the cart after visual inspection

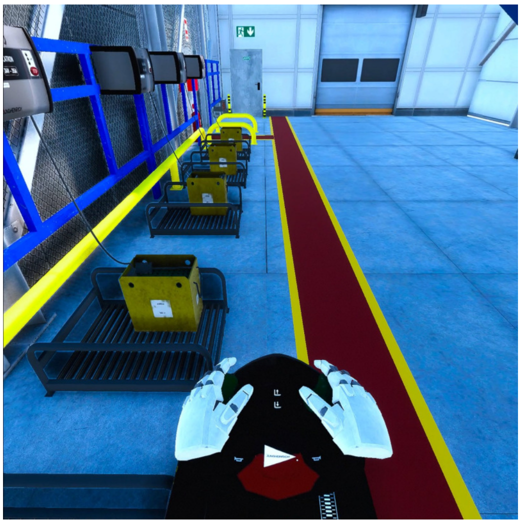


Fig. 12: Steering the cart into a station using hands

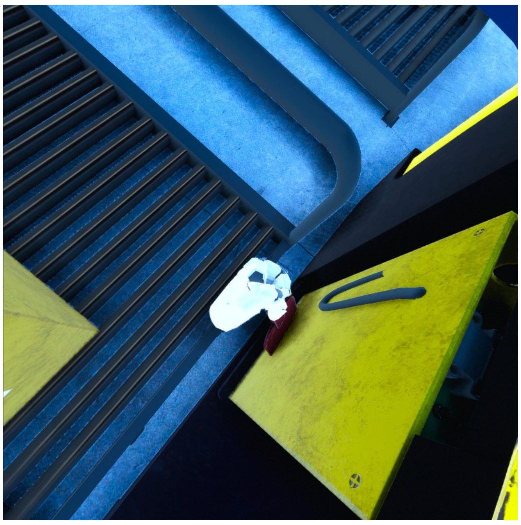


Fig. 13: Unlocking the battery using hand tracking

Following battery insertion, the trainee should confirm that the battery is correctly connected to the cart and that the blockade is securely in place. The cart may then be turned back on, and if all connections are established correctly, the trainee can resume driving the cart. This scenario is intended to train the trainees on the proper procedure for exchanging the battery, as well as teach them how to safely operate the electric cart. Trainees may learn how to drive the cart, how to adopt appropriate safety procedures while driving, and how to react in the event of an emergency.

F. Discussion

During the development of the presented training scenario, several discussions with the domain experts were conducted to ensure that the VR scenario accurately simulates real-life situations. An example of a significant alteration that has been implemented following these consultations is the need to hold the steering wheel of the electric cart with both hands during the operation. Without access to the appropriate documentation and expert knowledge, such issues could easily be overlooked.

During the battery replacement process, a trainee can encounter various challenges such as incorrectly connecting the battery, forgetting to lock it, or encountering a malfunction in the electrical system. These scenarios have been programmed to provide the trainee with a realistic learning experience and the opportunity to learn from their mistakes in a safe and controlled environment.

Overall, the scenario has been designed to provide the trainee with a realistic learning experience, replicating real-life situations as closely as possible. The use of VR technology, combined with accurate semantic representation of objects and realistic physics, creates a highly immersive and effective training environment.

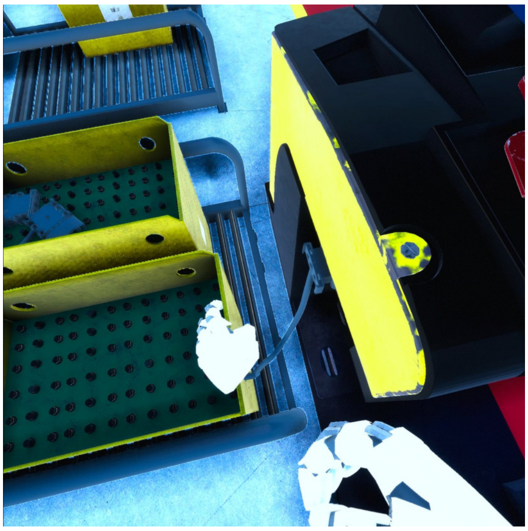


Fig. 14: Installing a new battery

VII. Conclusions

The proposed approach to the creation of VR training scenarios based on knowledge representation permits more flexible and precise modeling by utilizing domain concepts, rather than relying on low-level programming and 3D modeling. With the presented editor, trainers can easily create and modify scenarios in an efficient and intuitive manner. This makes the development of VR training environments accessible to non-technical users who can use domain terminology in the design process.

Future work includes extending the environment to allow for collaborative scenario creation by distributed users and integrating assessment of trainees' performance. Moreover, the scenario ontology will be extended to include concepts of alternative activities, which would be useful in cases when some tasks can be accomplished in multiple ways, such as in infrastructure error scenarios.

Acknowledgment

The presented research was funded by the Polish National Center for Research and Development under grant number LIDER/55/0287/L-12/20/NCBR/2021. The authors would like to thank Amica S.A. for providing training resources and insightful consultations.

References

[1] M. Dragoni, C. Ghidini, P. Busetta, M. Fruet, and M. Pedrotti, "Using ontologies for modeling virtual reality scenarios," in *The Semantic Web. Latest Advances and New Domains*, F. Gandon, M. Sabou, H. Sack, C. d'Amato, P. Cudré-Mauroux, and A. Zimmermann, Eds. Cham: Springer International Publishing, 2015, pp. 575–590.

[2] H. Fujita, M. Kurematsu, and J. Hakura, *Virtual Doctor System (VDS) and Ontology Based Reasoning for Medical Diagnosis*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 197–214. [Online]. Available: https://doi.org/10.1007/978-3-642-33959-2_11

[3] J. Flotyński and K. Walczak, "Ontology-Based Representation and Modelling of Synthetic 3D Content: A State-of-the-Art Review," *Computer Graphics Forum*, vol. 35, p. 329–353, 2017.

[4] O. De Troyer, F. Kleinermann, B. Pellens, and W. Bille, "Conceptual modeling for virtual reality," in *Tutorials, posters, panels and industrial contributions at the 26th Int. Conference on Conceptual Modeling - ER 2007*, ser. CRPIT, J. Grundy, S. Hartmann, A. H. F. Laender, L. Maciaszek, and J. F. Roddick, Eds., vol. 83. Auckland, New Zealand: ACS, 2007, pp. 3–18.

[5] M. Gutiérrez, D. Thalmann, and F. Vexo, "Semantic virtual environments with adaptive multimodal interfaces." in *MMM*, Y.-P. P. Chen, Ed. IEEE Computer Society, 2005, pp. 277–283.

[6] E. Kalogerakis, S. Christodoulakis, and N. Moumoutzis, "Coupling ontologies with graphics content for knowledge driven visualization," in *VR '06 Proceedings of the IEEE conference on Virtual Reality*, Alexandria, Virginia, USA, Mar. 2006, pp. 43–50.

[7] M. Attene, F. Robbiano, M. Spagnuolo, and B. Falcidieno, "Characterization of 3D Shape Parts for Semantic Annotation," *Comput. Aided Des.*, vol. 41, no. 10, pp. 756–763, Oct. 2009.

[8] L. De Floriani, A. Hui, L. Papaleo, M. Huang, and J. Hendler, "A semantic web environment for digital shapes understanding," in *Semantic Multimedia*. Springer, 2007, pp. 226–239.

[9] P. Kapahnke, P. Liedtke, S. Nesbigall, S. Warwas, and M. Klusch, "ISReal: An Open Platform for Semantic-Based 3D Simulations in the 3D Internet," in *International Semantic Web Conference (2)*, 2010, pp. 161–176.

[10] S. Albrecht, T. Wiemann, M. Günther, and J. Hertzberg, "Matching CAD object models in semantic mapping," in *Proceedings ICRA 2011 Workshop: Semantic Perception, Mapping and Exploration, SPME*, 2011.

[11] M. Fischbach, D. Wiebusch, A. Giebler-Schubert, M. E. Latoschik, S. Rehfeld, and H. Tramberend, "SiXton's curse - Simulator X demonstration," in *Virtual Reality Conference (VR), 2011 IEEE*, M. Hirose, B. Lok, A. Majumder, and D. Schmalstieg, Eds., 2011, pp. 255–256. [Online]. Available: http://dx.doi.org/10.1109/VR.2011.5759495

[12] P. Drap, O. Papini, J.-C. Sourisseau, and T. Gambin, "Ontology-based photogrammetric survey in underwater archaeology," in *European Semantic Web Conference*. Springer, 2017, pp. 3–6.

[13] Trellet, M., Férey, N., Flotyński, J., Baaden, M., Bourdot, P., "Semantics for an integrative and immersive pipeline combining visualization and analysis of molecular data," *Journal of Integrative Bioinformatics*, vol. 15 (2), pp. 1–19, 2018.

[14] Y. Perez-Gallardo, J. L. L. Cuadrado, Á. G. Crespo, and C. G. de Jesús, "GEODIM: A Semantic Model-Based System for 3D Recognition of Industrial Scenes," in *Current Trends on Knowledge-Based Systems*. Springer, 2017, pp. 137–159.

[15] M. Fuchs, F. Beckert, J. Biedermann, and B. Nagel, "A collaborative knowledge-based method for the interactive development of cabin systems in virtual reality," *Computers in Industry*, vol. 136, p. 103590, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0166361521001974

[16] B. Youcef, M. Ahmad, and M. Mustapha, "Ontophaco: An ontology for virtual reality training in ophthalmology domain – a case study of cataract surgery," *IEEE Access*, vol. PP, pp. 1–1, 11 2021.

[17] F. Longo, G. Mirabelli, L. Nicoletti, and V. Solina, "An ontology-based, general-purpose and industry 4.0-ready architecture for supporting the smart operator (part i – mixed reality case)," *Journal of Manufacturing Systems*, vol. 64, pp. 594–612, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0278612522001303

[18] J. Flotyński, M. Krzyszkowski, and K. Walczak, "Semantic Composition of 3D Content Behavior for Explorable Virtual Reality Applications," in *Proceedings of EuroVR 2017, Lecture Notes in Computer Science*, J. Barbic, M. D'Cruz, M. E. Latoschik, M. Slater, and P. Bourdot, Eds. Springer, 2017, pp. 3–23.

[19] M. Maik, P. Sobociński, K. Walczak, D. Strugała, F. Górski, and P. Zawadzki, "Flexible photorealistic vr training system for electrical operators," in *Proceedings of the 27th International Conference on 3D Web Technology, New York, NY, USA*, ser. Web3D '22. Association for Computing Machinery, 2022. [Online]. Available: https://doi.org/10.1145/3564533.3564561

DOI: 10.15439/2023F1683
ISSN 2300-5963 ACSIS, Vol. 37

Interactive, Personalized Decision Support in Analyzing Women's Menstrual Disorders

Łukasz Sosnowski
(0000-0003-2388-4008)
Systems Research Institute,
Polish Academy of Sciences
Newelska 6, Warsaw, Poland
sosnowsl@ibspan.waw.pl

Soma Dutta
(0000-0002-7670-3154)
University of Warmia and Mazury
in Olsztyn
Słoneczna 54, Olsztyn, Poland
soma.dutta@matman.uwm.edu.pl

Iwona Szymusik
(0000-0001-8106-5428)
Dep. of Obstetrics, Perinatology and Neonatology,
Center of Postgraduate Medical Education,
Ceglowska 80, Warsaw, Poland
iwona.szymusik@gmail.com

Wojciech Chaber
OvuFriend Sp. z o.o.,
Zlota 61/100, 00-819 Warsaw, Poland
wojciech.chaber@ovufriend.com

Paulina Kasprowicz
OvuFriend Sp. z o.o.,
Zlota 61/100, 00-819 Warsaw, Poland
paulina.kasprowicz@ovufriend.com

***Abstract*—This paper is in continuation to the paper published in FedCSIS 2022. In the earlier paper we presented the general scheme behind the AI based model for determining the possible ovulation dates as well as the possibility of some health risks. Here apart from the already discussed schemes for Premenstrual Syndrome (PMS), Luteal Phase Defect (LPD), and polyp and fibroids, a few additional schemes like hypothyroidism, polycystic ovary syndrom (PCOS) are included. Moreover, we attempt to throw light on the novelty of this AI based scheme from the perspective personalized, case sensitive, interactive medical support which does not depend only on a preset rule based system for diagnosing diseases.**

***Index Terms*—Medical decision support, Interactive AI, Explainable AI**

I. Introduction

FROM *the emergence of Artificial Intelligence, many researchers from different aspects contributed towards a broader goal of an artificially intelligent agent. However the progress is still far from the level of reaching close to human-like reasoning. This may be the reason that the current literature on AI showcases examples where the researchers specify the need by introducing terms like 'human-centered AI' [1], 'human-in-the-loop of machine learning' [2] etc.* explainable Reflection of similar thoughts can be noticed also in the context of decision support for different health care systems, such as IBM's dream project Watson, which was supposed to revolutionize everything from diagnosing patients and recommending treatment options to finding candidates for clinical trials; however, it failed as instead of trained with real data it was trained with hypothetical cases provided by a small group of doctors. The reason is quite understandable as expecting that the model's success with test data will directly translate to the real world does not really meet in reality. So, one way to improve the performance of an AI system is to engage users in providing feedback in order to continually improve the model. Thus, the terms like personalized medicine [3], evidence based medicine [4] are becoming prevalent in the literature of AI.

The main feature in both [3], [4] is to create such protocols for medical care that combine the knowledge from the existing literature of medicine, experience of the professionals, as well as the input parameters, habits, life style, and preferences of the individual patients. Moreover, in order to be sure that such data are properly gathered as well as to guide a patient during the intermediate steps of performing tests required for the diagnosis, an interactive interface among different stakeholders of the AI system is also required. So, it is quite clear that such a paradigm combines together different kinds of physical entities, information associated to them (e.g., a patient and her informational base, a team of experts and their informational bases) and their interactions, which as a whole indicates a real physical process of computation on a complex granule [5], [6].

In continuation to a series of papers [7]–[10], here we present the developments made in the platform of OvuFriend[1] focusing on introducing the above mentioned aspects in an AI system for helping women in determining the possibility of conceiving and understanding the hidden risk of health problems based on their input. The platform of OvuFriend 1.0 was developed as a part of R&D project where through a mobile app an user can put the data related to her physical and mental states during a specific menstrual cycle, and the underlying algorithm of the app helps to get an analysis of the possibility of conceiving or not conceiving. The second stage of OvuFriend's project, namely OvuFriend 2.0, focuses on the

Co-financed by the EU Smart Growth Operational Program 2014-2020 under the project "Developing innovative solutions in the domain of detection of frequent intimate and hormonal health disorders in women of procreative age based on artificial intelligence and machine learning - OvuFriend 2.0", POIR.01.01.01-00-0826/20.

[1] www.ovufriend.pl

analysis of whether a particular user has the possibility of having the risk of Premenstrual Syndrome (PMS) Luteal Phase Defect (LPD), benign growths like polyps, fibroids in the uterus, Polycystic Ovary Syndrome (PCOS) or hypothyroidism Among them in [11] a detailed discussion regarding the schemes of PMS, LPD, and the risk from polyp and fibroids are discussed. Here we add the schemes corresponding to hypothyroidism and PCOS.

Apart from discussing the two new schemes, namely hypothyroidism and PCOS, in this paper one of our main targets is to present in what sense the AI model of OvuFriend incorporates the features like (i) learning and updating based on real data (ii) adaptation of diagnosing strategies based on interactions with users and medical as well as analytical experts, and (iii) visual as well as linguistic explainability of the relationship between the gathered data and their labelling. Inclusion of these features makes the OvuFriend platform for women's healthcare more close to the above mentioned AI paradigm of interactive, personalized, evidence based healthcare support systems keeping human in the loop.

The paper is organized as follows. Section II presents a brief general description of the scheme running behind the OvuFriend app as well as the schemes analyzing certain health risks based on a complete cycle data of a user. The schemes for hypothyroidism and PCOS, requiring a sequence of cycles data, are presented in Section III. Section IV presents the process of building the reference set based on which the app can decide effectively over new cases. In particular we would emphasize on the novelty of the process of selecting reference set which allows a team based interactive environment among the experts, the user and the consultant in the process of deciding how, when and for whom which strategy of treating and diagnosing should be selected so that the data gathered from them can be used in the reference set with certain reliability. Section V presents concluding remarks.

II. General Scheme of Ovulation and health risks

The general scheme in OvuFriend 2.0 for having an AI based app determining the possible days of ovulation as well as the possibility of different health risks is developed based on three hierarchical levels, known as *Detector level*, *Cycle level*, and *User level*. In the detector level the user can put information related to her mental and physical health over (at least) one complete cycle. Relative to the need a set of attributes is set by the medical experts. Based on the input of a particular user the values for those attributes are determined by a team of medical experts. From the values of the attributes from a completed cycle, the cycle level concepts such as *ovulation happened*, *days of ovulation*, *follicular phase interval*, *luteal phase interval*, *PMS score* etc are determined. In the user level the system aggregates the data related to the detector level as well as the cycle level concepts of a particular user for a finitely many cycles. *Risk of PMS*, *risk of LPD*, *risk of infertility* etc are a few examples of the user level concepts. For a cycle level concept, the system calculates the probabilistic ratio of the concerned cycle level concept over the total number of cycles considered for a particular user. Moreover, the system is also fed with a threshold value for each such concept. The threshold value for a particular concept is learned and with time this is updated based on the opinions of the medical experts and the histories of already recorded and analysed cases. If the respective ratio for a particular user level concept is greater than the prefixed threshold for that concept the user is notified about the possibility of such health risk.

As prerequisite the data related to the physical and mental health of a woman before, during, and after a complete menstrual cycle is collected. After gathering data over a complete cycle (or a few consecutive cycles) the analysis for different health risks starts. Initially, the data is processed to investigate whether the ovulation has occurred and then based on that to find the possible days of its occurrence. At this stage all the detector level concepts are analysed. If through the primary analysis it is determinned that ovulation has been occurred, then an attempt is made to indicate two intervals of equal length falling into the follicular phase and the luteal phase of the concerned cycle respectively [11].

A. Summary of schemes requiring one complete cycle data

The schemes requiring a complete cycle data, namely PMS, LPD, polyp and fibroids, are already discussed in [11]. Though the basic formulas for calculating these health risks are different, the general form of the underlying algorithms is similar.

To analyze the risk of PMS, which is a combination of symptoms that many women get about a week or two before their period, the coefficients of occurrence of the physical symptoms and mood symptoms are calculated (see [11]). The set of symptoms and formulas for calculating the coefficients based on them are defined with the help of a team of medical experts. From the user's input all physical and mood symptoms are counted for both the phases P_1 and P_2. Aggregating the number of physical and mood symptoms in a phase the coefficients are calculated according to the following formulas.

$$P_iMoodFeelCoeff = \frac{(SumOfOccurrenceP_iMood)}{K_1 \times PhaseLength} \times \alpha + (1-\alpha) \quad (1)$$

where i = 1, 2 and $\alpha \in (0,1)$,

$$P_2PhysFeelCoeff = \frac{(SumOfOccurrenceP_2Phys)}{K_2 \times PhaseLength} \times \beta + (1-\beta) \quad (2)$$

where $\beta \in (0,1)$.

The symbols $SumOfOccurrenceP_iMood$ and $SumOfOccurrenceP_iPhys$ respectively indicate the number of mood and the number of physical symptoms occurred in a particular phase P_i, and K_1 and K_2 represent respectively the total number of all moods and physical symptoms listed in the system. The factors α and β represent the significance of the given components in the respective coefficients. Using the above coefficients *PMS score*, denoted as PMS_{score}, is calculated by the following formula.

$$PMS_{score} = \frac{P_2MoodFeelCoeff}{P_1MoodFeelCoeff} + \frac{P_2PhysFeelCoeff}{w_1} \quad (3)$$

where w_1 is the weight chosen by a team of medical experts.

Luteal Phase Defect (LPD) is a health condition that may play a role in infertility. The general prerequisite for determining the risk of LPD [12] is same as what is discussed above. The specific formula that is fed to the algorithm in order to calculate the susceptibility of LPD (Equation 4) is as follows.

$$LPD_{score} = w_1 * LutParameters + w_2 * DecFer \tag{4}$$

The parameters $LutParameters, DecFer \in [0, 1]$ respectively denote the values for *Luteal Phase Parameters* and *Decreased Fertility*. The *Luteal Phase Parameters* are determined based on the luteal phase length and various other factors related to the analysis of bleeding during the luteal phase. The *Decreased Fertility* depends on the period of time in which the attempts are made for conceiving a child, the number of miscarriages etc. The values for $LutParameters, DecFer$ are obtained based on the input data of a particular user, and w_1, w_2 are some weights that are chosen by the team of experts based on their collective knowledge regarding the significance of $LutParameters$ and $DecFer$ in indicating LPD.

Presence of fibroids and polyps too may cause infertility and recurrent pregnancy loss. The algorithm starts with checking whether ovulation has occurred. The primary analysis focuses on the data related to inter-menstrual bleeding or spots. Then examining the cycle level concepts and associated symptoms characterizing polyp or fibroids starts. The values for disordered menstruation ($DisMens$), decreased fertility ($DecFer$), and the values for physical symptoms related to such diseases ($PhysSymp$) are obtained from the input data of the user. Then the following score is calculated.

$$Score = w_1 * DisMens + w_2 * DecFer + w_3 * PhysSymp \tag{5}$$

The weights w_1, w_2, w_3 are chosen by the team of experts. All these values are scaled in the interval $[0, 1]$ based on the information related to inter-menstrual bleeding, long-lasting menstruation, intensity of menstruation, miscarriage, long trying time for conceiving, pelvis pain, polyuria etc.

In each of the above contexts, for a given user the grade of the susceptibility of a particular disease is calculated by considering $\frac{k}{n}$ if in k such cycles, out of n cycles, the susceptibility of the respective disease is detected.

III. Schemes requiring consecutive cycles' data

In this section we present two newly analyzed health risks, namyly PCOS and hypothyroidism, which require a sequence of consecutive cycles' data of a user.

A. Scheme for PCOS

PCOS creates a condition where the ovaries produce an abnormal amount of androgens, that are usually present in women in small amounts [13]. Contrary to the above mentioned schemes, to analyse the risk of PCOS the algorithm needs the data of the user for a few months. Based on the detector level parameters such as stress, appetite, depression, hypersensitivity, insomnia, problem in concentration, BMI, length of cycle etc., relevant cycle level concepts such as *increasing level of anxiety*, *lower self-esteem*, *family history of PCOS*, *long cycle*, *extended trying time for conceiving* etc are determined. Some of the above mentioned cycle level concepts are marked with binary values and some with fuzzy values, on a scale of $0 \leq 0.33 \leq 0.66 \leq 1$; these values are marked over a span of time. After completion of a cycle, all the relevant cycle level concepts are determined. Each of the considered cycles is then characterized with the help of these concepts described on a multidimensional time series.

Some groups of symptoms are analyzed by qualitative as well as quantitative indicators. For example, it is checked whether any of the symptoms belonging to the group occurred at least once on a given day (qualitative), as well as how many symptoms (quantitative) from the group occurred on a day. The frequency of occurrence of a symptom usually is analyzed based on the selected time period. For instance, the occurrence of the symptom 'fatigue' 4 times in a 45-day cycle may indicate the greater possibility for anxiety than occurrence of the same symptom 4 times during half of the time span of the cycle. Compare to the above schemes here the algorithm chooses the next plan of actions based on an interaction with the user. There are different forms available for deeper analysis of some of the above mentioned detector or cycle level concepts. If a user meets the PCOS boundary conditions, she is asked to provide some specific parameters in the follicular phase of the cycle for consecutive 3 days. If the user rates them three times negatively, the label for *low self esteem* is activated. Then further the user is led to complete a more detailed low self-esteem survey.

The analysis for PCOS also starts with checking the possibility of ovulation and determining respective intervals. To enable PCOS susceptibility analysis, the input for the cycle must be completed for at least 10 days; the same data for previous two cycles must also be available meeting the same conditions. Symptoms for PCOS persist for a long time. So, one cycle may not reliably assess the presence of PCOS. Moreover, exploring three consecutive cycles increases the likelihood of the observations of the user. For each of these series of cycles, possible ovulation is determined.

The coefficient $cycle_nScore$ for the nth cycle is calculated based on the following formula.

$$\begin{split} cycle_nScore = X_{1n} * w_1 + X_{2n} * w_2 + X_{3n} * w_3 + \\ X_{4n} * w_4 + X_{5n} * w_5 + X_{6n} * w_6 + X_{7n} * w_7 \end{split} \tag{6}$$

where X_{in} is calculated based on the number points obtained for the i-th group of concepts that have appeared in the n-th cycle. For example, $X_{5n} = \frac{increased\_anxiety+depressive\_mood}{2}$ indicates that the two operands in the numerator represent the number of points obtained for those two parameters from the 5-th group of concepts in the n-th cycle. The weights w_i, $1 \leq i \leq n$ are selected based on the significance of a group of symptoms over other. Then the sum of the points of each cycle from the sequence is added and normalized according to the formula below.

$$normScore = \frac{\Sigma_{i=1}^{3} cycle_iScore}{3 * \Sigma_{j=1}^{7} w_j} \tag{7}$$

Based on the values for $nomScore$, different possible sequences of cycles, recorded over a time period, are ranked in the descending order. The two chosen sequences of three cycles can be such that one of the cycles can be the first in one sequence and middle in another sequence. So, the sequence with highest $normScore$ is selected for the analysis of PCOS.

The scheme of PCOS is presented in Fig. 1. To determine the PCOS susceptibility one sequence of cycles, which is completed in last six months, is selected from the history of a user. If among a series of cycles at least two are detected with a vulnerability of PCOS, the respective user is assigned to PCOS risk. Then, at user level the degree of risk is calculated based on the ratio of the number of PCOS-susceptible cycles to the number of months over which the observation is made.

B. Scheme for Hypothyroidism

In hypothyroidism the thyroid gland does not produce enough thyroid hormones, leading to changes in the menstrual cycle. The scheme for hypothyroidism is quite similar to the scheme for PCOS. Here also the algorithm requires data for three consecutive cycles. Data for all the cycles are processed to test determine the date of onset of ovulation as well as the detector and the cycle level concepts. If, in each of the cycles from the sequence, enough data is marked for the algorithm to determine the occurrence of ovulation, the algorithm proceeds to the next stage. The cycle level concepts and the symptoms, such as feeling cold, feeling sleepy, concentration problems, decreased appetite, constipation, swelling, decreased libido, memory problems, etc., which are relevant to hypothyroidism, are selected. Then for each cycle a score, denoted as Sc_n, is determined from the sequence using the following formula.

$$\begin{aligned} Sc_n = w_1 * PhySym_n + w_2 * Len_n + w_3 * Ov_n \\ + w_4 * DecFer_n + w_5 * PsySym_n \end{aligned} \tag{8}$$

The weights $w_1, \ldots, w_5$ are chosen by the experts, and the values of the parameters are computed from the input of user. The symbol $PhySym_n$ denotes the value corresponding to the physical symptoms during the n-th cycle, Len_n indicates the length of the n-th cycle, Ov_n corresponds to the number of ovulations occurred in the n-th cycle, $DecFer_n$ stands for the value of the decreased fertility in the n-th cycle, and $PsySym_n$ represents the value corresponding to the psychological symptoms in the n-th cycle. From the scores of three consecutive cycles the score for the risk of hypothyroidism is calculated for the whole sequence using the following formula.

$$Score_{Hypth} = Sc_1 + Sc_2 + Sc_3 \tag{9}$$

where 1, 2, 3 denote the numbers of the cycles in the sequence.

If the score is greater than or equal to the preset threshold, the most recent cycle in the sequence is assigned a hypothyroidism susceptibility at the cycle level and the score is then calculated just by adding the score obtained in three consecutive cycles. The score obtained for each such single cycle from a chain of three consecutive cycles is used to assess the risk of developing hypothyroidism at the level of the user. In this process all completed cycles, that have occurred during the last n months, are selected. Then all possible sequence combinations of three consecutive cycles are created from them, and the sum of the scores is calculated for each sequence. If the sum of the scores for any of the sequences is greater than or equal to the pre-fixed cut-off value, a risk of hypothyroidism is assigned to the user, and a grade is calculated in the range of $[0, 1]$. After the analysis of a user's risk for hypothyroidism the data and analysis, obtained from the sequences, are again assessed by medical experts. Based on such history of sequences the cut-off point is updated.

IV. Interactively adapting treatment and diagnosis strategy based on users' perception

We now attempt to illustrate the key features of OvuFriend's application which allow to create an interface for telemedicine consultation and choose appropriate course of actions based on analyzed data of a user. Through the interactive interface a user, a team of experts (medical and analytic), and a consultant together may share a platform for interacting with queries and respective answers, uploading and scanning documents/results, presenting an illustrative graphical representation of causes and outcomes related to a concept, and choosing labels for certain values of parameters based on consensus. In this regard, we present the design of some components and their roles contributing towards the working strategy of the app.

A. Building reference set incorporating real data through interactions

In Introduction we have discussed about failure of different decision support systems trained based on hypothetical data. Here, the reference set, for training the model of the app, has been chosen from three different populations of users.

One population pertains to the already registered users of the app for whom certain vulnerabilities are detected on the basis of physical symptoms, mood symptoms and parameters of menstrual cycles declared in the system. Based on the data recorded in the cycles of the users further medical examinations are suggested. Then based on context, indicated by a precise flowchart of the algorithm, the users are selected to be included in the reference set when some specific results are confirmed by blood tests, TSH, Testosterone, progesterone, ultrasound examination of the reproductive organs etc., or on the basis of a questionnaire completed in the app, serving as a medical teleconsultation. The second population pertains to women who have participated in a questionnaire survey conducted on the OvuFriend's platform and have declared certain diseases voluntarily. For such users based on the results of survey uploaded to the system they are selected for inclusion in the reference set. The third population pertains to the women who as a part of marketing activities of the app are envouraged to declare problems of having certain diseases or noticing symptoms from a given set of relevant symptoms on the OvuFriend's platform. In response to their willingness to take part in the project, they are offered free medical examinations, and in case of positive result for some

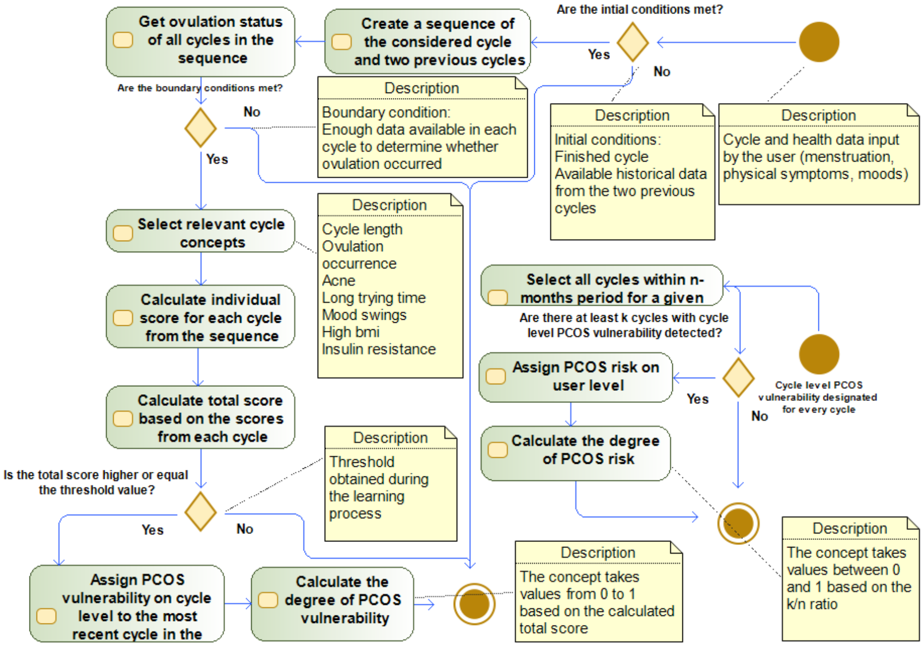

Get ovulation status of all cycles in the sequence
Create a sequence of the considered cycle and two previous cycles
Are the intial conditions met?
Yes
No
Are the boundary conditions met?
No
Yes
Description
Boundary condition:
Enough data available in each cycle to determine whether ovulation occurred
Description
Initial conditions:
Finished cycle
Available historical data from the two previous cycles
Description
Cycle and health data input by the user (menstruation, physical symptoms, moods)
Select relevant cycle concepts
Description
Cycle length
Ovulation occurrence
Acne
Long trying time
Mood swings
High bmi
Insulin resistance
Calculate individual score for each cycle from the sequence
Calculate total score based on the scores from each cycle
Select all cycles within n-months period for a given
Are there at least k cycles with cycle level PCOS vulnerability detected?
Assign PCOS risk on user level
Yes
No
Cycle level PCOS vulnerability designated for every cycle
Calculate the degree of PCOS risk
Description
Threshold obtained during the learning process
Is the total score higher or equal the threshold value?
Yes
No
Assign PCOS vulnerability on cycle level to the most recent cycle in the
Calculate the degree of PCOS vulnerability
Description
The concept takes values from 0 to 1 based on the calculated total score
Description
The concept takes values between 0 and 1 based on the k/n ratio

Fig. 1. Complete scheme for determining risk vulnerability of PCOS (2 diagrams)

tests they are given access to the Ovufriend's application to upload their results. Then further data on historical cycles are collected during a survey conducted by a consultant.

The consultant contacts using the data provided by the application or the OvuFriend's platform. If some conditions are met the consultant asks questions about the symptoms, relative to the analyzed disease. The questionnaire is designed by a medical expert. So, though all the questionnaire surveys and teleconsultation processes are conducted within the algorithmic set up of the app, it includes both human in the loop as well as real physical interactions.

Moreover, the users after completion of the tests the scans of medical examinations upload to the system or sent to OvuFriend's platform. The exchange of information between the consultant and the team of experts is carried out using spreadsheets saved on a cloud drive, due to which it is possible to track the editions by all team members. The information obtained by the user, including medical tests, and the answers given during surveys, are then checked by the analytical team, in cooperation with a medical expert.

B. Explainable model storing and labeling reference set data

An explainable AI model is another great challenge on which the present days AI development is still struggling. OvuFriend's model is capable to address the above mentioned challenge to some extent. In particular, it refers to the part of the model where each user's data along with the scans of test results are stored against an uniquely generated user-id.

The data obtained in the survey along with the test results are uploaded by the consultant in the cloud environment for review of the analytical team. On a regular interval all

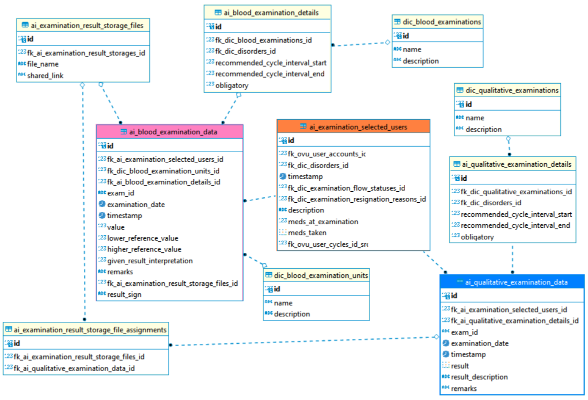

ai_examination_result_storage_files
id
fk_ai_examination_result_storages_id
file_name
shared_link
ai_blood_examination_details
id
fk_dic_blood_examinations_id
fk_dic_disorders_id
recommended_cycle_interval_start
recommended_cycle_interval_end
obligatory
dic_blood_examinations
id
name
description
dic_qualitative_examinations
id
name
description
ai_blood_examination_data
id
fk_ai_examination_selected_users_id
fk_dic_blood_examination_units_id
fk_ai_blood_examination_details_id
exam_id
examination_date
timestamp
value
lower_reference_value
higher_reference_value
given_result_interpretation
remarks
fk_ai_examination_result_storage_files_id
result_sign
ai_examination_selected_users
id
fk_ovu_user_accounts_ic
fk_dic_disorders_id
timestamp
fk_dic_examination_flow_statuses_id
fk_dic_examination_resignation_reasons_id
description
meds_at_examination
meds_taken
fk_ovu_user_cycles_id_src
ai_qualitative_examination_details
id
fk_dic_qualitative_examinations_id
fk_dic_disorders_id
recommended_cycle_interval_start
recommended_cycle_interval_end
obligatory
dic_blood_examination_units
id
name
description
ai_examination_result_storage_file_assignments
id
fk_ai_examination_result_storage_files_id
fk_ai_qualitative_examination_data_id
ai_qualitative_examination_data
id
fk_ai_examination_selected_users_id
fk_ai_qualitative_examination_details_id
exam_id
examination_date
timestamp
result
result_description
remarks

Fig. 2. Relational database for storing information related to a user and creating links for navigating between relevant information

the uploaded information of the users are analyzed by the analytical team and a medical expert. If based on the initial information a user is considered suitable for the project, a user-id is created and the information details is saved in the database. Fig. 2 shows a fragment of a relational database, presenting directory containing files with records of medical examinations of different project users.

Medical examinations stored in the database are presented to the doctor by means of a visualization that combines the information provided by the users, in the form of scans of tests and the data registered in the app for individual cycles. More specifically, the visual representation related to the measurement details and symptoms of a user also includes additional information obtained from the database, e.g. cycle type, cycle status, registered drugs, anomaly detector result and descriptive information regarding the analysis of the expert tagged with a comment and information obtained from the user during the consultation before performing tests. The visualizations in this project are based on the experience gathered in previous projects based on a well-known approach from other areas using fuzzy linguistic summaries widely described in [14], [15]. This type of visualization allows for a comprehensive presentation of all archival information registered by the user in the app including information from the labeling stage, through the period in which the test is performed, to the current cycle at the time of presentation of the final results. At the next stage based on the visualizations presented to the team of experts the selection of the final labeling of the reference set is performed.

The model of creating such visualization serves two aspects of explainability. In one hand, it presents a comprehensive visual representation of the data with all notes and comments from the user and the medical expert, and on the other hand it presents a visual relationship between the results of the analysis made by the app and the measurement data registered in the app depending on particular disease. For each of the anomalies, i.e, thyroid diseases, PCOS, NFL etc., the respective visual representation is created during consultations with a medical expert. That is, based on the data saved on cloud against each user-id, the medical expert can create some cause-effect relations among the measurement data and the concerned diseases, and that information gets translated to the system creating a visual representation of the selected relations using some software packages for time series analysis.

Fig. 3 presents how through a spreadsheet visualization of all data relative to a particular patient is presented in a compact and comprehensive way to the analytical team as well as to the user. In Fig. 3 the information presents values of different parameters over three consecutive cycles, length of each of which is presented in the header. In the left hand side using a sliding option for going up and down one can check information concerning a particular day over this sequence of cycles, and in the bottom the labels are chosen automatically by the algorithm based on values of the parameters entered from user's input.

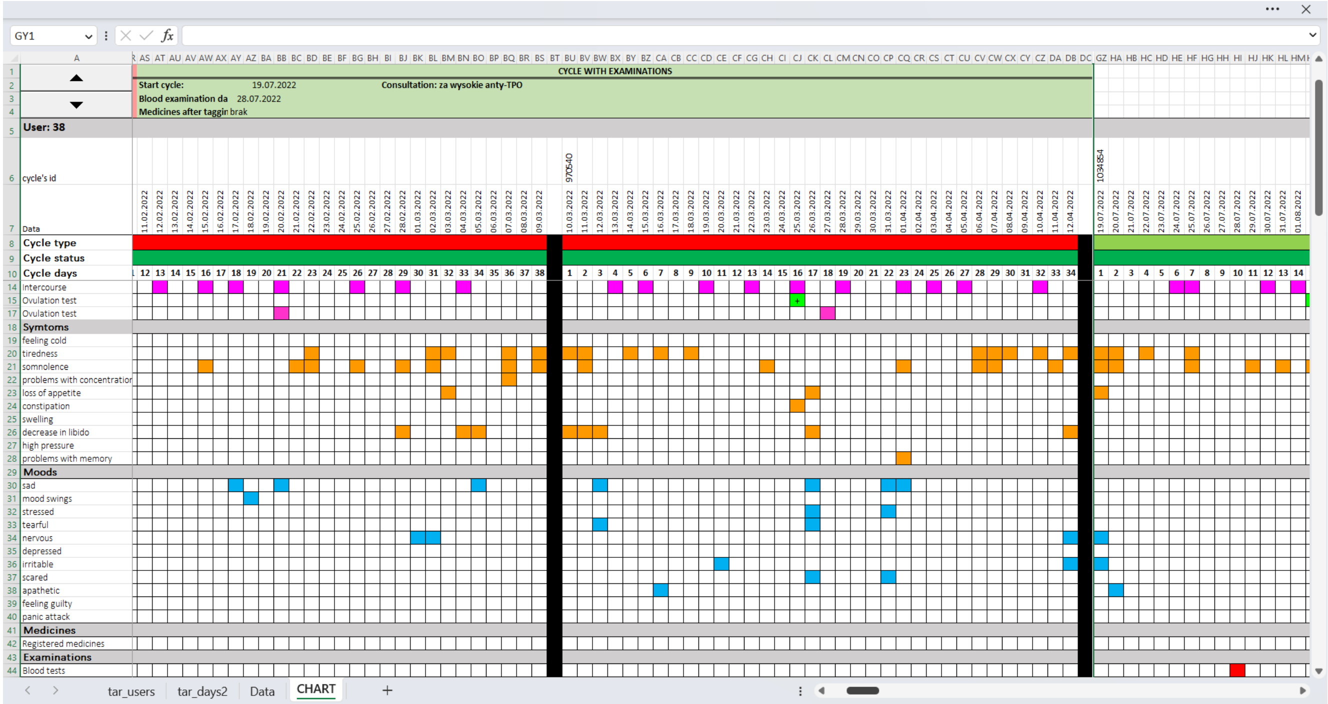
The visualization for each disease consists of two files.

(i) One is a sql file in which data is generated for each user included in the app. Here the data presents a user and her cycles divided into days. The range of days selected for visualization depends on the number of cycles recorded in the app and varies depending on the number of cycles entered and their lengths for each user. A rule is fed to the app to create the visualization; the time axis is defined by the initial data related to one cycle or three cycles used in the labeling process. The cycle, in which the tests are performed, are marked as the anchor points. From the tagged cycles, a maximum of three cycles are searched back. From here the data is supposed to be represented by visualization. All the cycles (or cycle) included in the app for labeling, including the cycle in which tests have been performed, are presented. As the cycle, in which tests are performed, is marked on the chart, the performed transformations saved in this file lead to two main tables: the users table and the days table. The user table contains information such as: chart number, user_id, information about the tagged cycle (cycle_id, length, start date), information about tagging by the expert (shipped package number, order in the package, expert tagging result, comment), information about medications taken, date of the test, test result, link to the test file, type of test, comment obtained after contacting the user etc. The days table contains data for visualized cycles for each user in a specific package, including: user_id, cycle_id, cycle order on the chart, date, cycle day, information about mucus, cervix, bleeding, intercourse, ovulation tests, pregnancy, symptoms, moods, concepts, as well as detector indications, cycle type and information about cycle status.

(ii) The other one is a xlsx file in which data prepared with the use of SQL code are read. Using the ODBC connection to the PostgreSQL database, previously prepared tables with users and days are uploaded (saved in the .sql file). Next, the data is transformed in order to visualize them on the timeline, the length of which varies based on the number of cycles registered by the user in the app. The tab with the chart shows the graphical form of the automatically transformed data, depending on the refresh of the data in the .sql file, by defining the appropriate package number. Switching between the users is possible using the user selection control in the form of arrows, a vertical slider scrolling between visualization sections and a horizontal slider scrolling between user cycles.

The presence of information in a line is conditioned by its color, depending on the day of occurrence. The vertical black bars are used to separate the cycles from each other. In the right of a black marker a new cycle starts with counting of the days in the cycle and the intensity of bleeding over days is represented on the graph. Cycles are presented on a timeline from the oldest to the newest. The users whose cycles already have been labeled by a medical expert are selected for the visualization of cycles after the tests; that is, sequentially first they participate in the tagging process, receive a referral for tests from a given medical package depending on the disease, perform the test and send the scans of the results, which are saved in the appropriate folder in the google drive, which can be accessed by the team of experts and the consultant responsible for contacting the users. The user-ids, corresponding to the selected cycles for presentation, are fed to the packages by which visualizations are created.

The visualizations are presented in such a way that the medical expert can have an insight into the widest possible range of information of the patients. The whole presentation is realized in an interactive way. A medical expert, using the buttons in the upper left corner of each of the presented visualizations, can switch tabs and obtain different information related to a chosen user. On the other hand, using the horizontal scroll bar


CYCLE WITH EXAMINATIONS

Start cycle: 19.07.2022

Consultation: za wysokie anty-TPO

Blood examination da 28.07.2022

Medicines after taggin brak

User: 38

Fig. 3. Hypothyroidism labelling form prepared for medical experts to evaluate susceptibility of selected cycles. Form is based on such attributes as: bleeding, mucus, bbt, cervix, intercourse, feeling cold, tiredness, somnolence, sad, mood swings, tearful, stressed, nervous, depressed, irritable, scared, problem with the concentration, sleep disturbance, constipation, etc.

it is possible to navigate from one cycle to other and obtain a view of the complete history of the user's recorded cycles.

V. Experimental Results: Concluding Remarks

In Section IV, we presented the design of reference set as one of the unique selling points of OvuFriend's application. In this section we would present a brief summary of the experimental results obtained based on the chosen reference set. In contrary to the experimental results obtained in the earlier stages [11], here we present the experimental results based on the actual users whose health risks or anomalies have been analyzed by OvuFriend's schemes.

The reference set consists of a list of users assigned to the selected anomaly with the actual class specified by the physician. Subsets for individual anomalies are balanced in terms of the number of positive and negative classes, so as to obtain a similar number of elements in both the classes. As the different methods of data processing depend on the anomaly, the users have been grouped by anomaly, not by a group of diseases. Later the final evaluation is calculated based on the average results of evaluations performed for different anomalies falling within a group of diseases. For example, in case of LPD 57 cases from each of positive and negative classes are selected; while in case of PCOS the reference set contains 94 cases from each of positive and negative classes.

For evaluating the effectiveness of each of the algorithms four experiments have been conducted for each of the disorders. Using ReSample evaluation [16] each of the experiments is conducted such as 1000, 500, 100, and 10 times respectively. Two disjoint subsets are designated as the training set and test set where the former contains 33% of the reference set and the later consists of remaining 67%. Evolutionary algorithms with a fitting function based on a combination of the accuracy measure and the F1Score measure are used to train the respective thresholds for all the disorders and these values are learned on each iteration of the training set containing 33% of the tagged cycles sample in particular disorders. The test procedure is performed on remaining cycles in given disorders which accounted for 67%. For each iteration results are stored in the contingency table. Then all True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) are calculated to find the effectiveness measures of the algorithms. Due to page limitation here only the result for 1000 repetitions is presented (cf. Table I). presents .

The label TP means that cycle is tagged with at least 0.6 by the medical experts and is classified as a positive case of the disease by the algorithm; whereas, the case for FP is determined when the experts have given mark below 0.6 but the algorithm has classified the case into positive class. The case for TN is obtained when the experts have assigned less than 0.6 and the algorithm has classified as negative as well. Finally the cases for FN is indicated when the algorithm has classified as negative but the expert evaluated as greater or equal 0.6.

From Table I it is visible that the obtained results are quite satisfactory; especially comparing to the experimental

TABLE I
RESULTS AVERAGED OVER 1000 ITERATIONS OF THE RESAMPLE ROUTINE. ABBREVIATIONS: # - SAMPLE, TP - TRUE POSITIVES, TN - TRUE NEGATIVES, FP - FALSE POSITIVES, FN - FALSE NEGATIVES, PR - PRECISION, RE - RECALL, F1 - F1 SCORE, $_{mn}$- MIN, $_{mx}$- MAX, AC - ACCURACY, PCOS - POLYCYSTIC OVARY SYNDROME, HYP - HYPOTHYROIDISM, LPD - LUTEAL PHASE DEFICIENCY

Type	Sample	TP	TN	FP	FN	PR	RE	F1	AC	F1_mn	F1_mx	AC_mn	AC_mx
HYP	126000	61472	39922	23096	1510	0.727	0.976	0.833	0.805	0.757	0.902	0.730	0.881
PMS	68000	37448	13668	7783	9101	0.828	0.804	0.816	0.752	0.575	0.905	0.544	0.868
PCOS	126000	60643	42814	20065	2478	0.751	0.961	0.843	0.821	0.358	0.903	0.516	0.889
LPD	76000	35526	24555	13342	2577	0.727	0.932	0.817	0.791	0.089	0.899	0.461	0.882

results obtained in the earlier stage [11] based on hypothetical data, here the experimental results based on real data is commendably good. Moreover, apart from the quantitative values showing satisfactory experimental results, in this paper, our main emphasize has been on the qualitative worth of the AI model which attempts to address a few important at the same time challenging aspects of AI. Precisely the novelty of the approach includes building of an AI model which is (i) trained on real data, (ii) sensitive to user's perceptions, (iii) able to learn and revise through interactions among the stakeholders (such as a user and the (medical) experts), (iv) designed to adapt suitable strategies for the required course of actions (e.g., suggesting further tests or filling up additional questionnaire etc.) in the process of decision making, and (v) possessing the ability of explainability of its decision to the stakeholders through a process of visualization. As a whole the proposed model is a good attempt towards the goals envisaged by the paradigms like personalized, evidence based medicine [3], [4], human-centered AI [1], and IGrC [5], [6].

However, there are a few aspects where the model has limitations. Firstly, it is difficult to collect a very large collection reference data as use of the application and participation in the project is voluntary. Moreover, usually users, who are trying to get pregnant, are interested in using the app; whereas for reference data only historical data that was recorded before pregnancy can be used. Another limitation is related to full implementation of the developed algorithms and preparing them to work in a real environment. Simultaneously in many places replacing manual process of setting parameters and weights (by experts) by ML, so that all relevant parameters are learned from reference data sets, is also required.

REFERENCES

[1] Ben Shneiderman. *Human – Centered AI.* Oxford University Press, Oxford, UK, 2022.

[2] Robert (Munro) Monarch. *Human-in-the-Loop Machine Learning. Active Learning and Annotation for Human-Centered AI.* MANNING, Shelter Island, NY, 2021.

[3] Carlos Fernández-Llatas and Jorge Munoz-Gama *et al*. *"Process Mining in Healthcare"*, pages 41–52. Springer, 2020.

[4] Margaret A. Hamburg and Francis S. Collins. "The Path to Personalized Medicine". *New England Journal of Medicine*, 363(4):301–304, 2010.

[5] Soma Dutta and Andrzej Skowron. Interactive granular computing model for intelligent systems. In Z. Shi, M. Chakraborty, and S. Kar, editors, *Intelligence Science III. 4th IFIP TC 12 International Conference (ICIS 2020), Durgapur, India, February 24-27, 2021, Revised Selected Papers*, volume 623 of *IFIP Advances in Information and Communication Technology (IFIPAICT) book series*, pages 37–48. Springer, Cham, Switzerland, 2021.

[6] Soma Dutta and Andrzej Skowron. Interactive Granular Computing Connecting Abstract and Physical Worlds: An Example. In Holger Schlingloff and Thomas Vogel, editors, *Proceedings of the 29th International Workshop on Concurrency, Specification and Programming (CS&P 2021), Berlin, Germany, September 27-28, 2021*, volume 2951 of *CEUR Workshop Proceedings*, pages 46–59. CEUR-WS.org, 2021.

[7] Lukasz Sosnowski and Tomasz Penza. "Generating Fuzzy Linguistic Summaries for Menstrual Cycles". volume 21 of *Annals of Computer Science and Information Systems*, pages 119–128, 2020.

[8] Joanna Fedorowicz, Lukasz Sosnowski, Dominik Slezak, Iwona Szymusik, Wojciech Chaber, Lukasz Milobedzki, Tomasz Penza, Jadwiga Sosnowska, Katarzyna Wójcicka, and Karol Zaleski. "Multivariate Ovulation Window Detection at OvuFriend". In Tamás Mihálydeák, Fan Min, Guoyin Wang, Mohua Banerjee, Ivo Düntsch, Zbigniew Suraj, and Davide Ciucci, editors, *Rough Sets - International Joint Conference, IJCRS 2019, Debrecen, Hungary, June 17-21, 2019, Proceedings*, volume 11499 of *Lecture Notes in Computer Science*, pages 395–408. Springer, 2019.

[9] Lukasz Sosnowski, Iwona Szymusik, and Tomasz Penza. "Network of Fuzzy Comparators for Ovulation Window Prediction". volume 1239 of *Communications in Computer and Information Science*, pages 800–813. Springer, 2020.

[10] Lukasz Sosnowski and Jakub Wróblewski. "Toward automatic assessment of a risk of women's health disorders based on ontology decision models and menstrual cycle analysis". In Yixin Chen, Heiko Ludwig, Yicheng Tu, Usama M. Fayyad, Xingquan Zhu, Xiaohua Hu, Suren Byna, Xiong Liu, Jianping Zhang, Shirui Pan, Vagelis Papalexakis, Jianwu Wang, Alfredo Cuzzocrea, and Carlos Ordonez, editors, *2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, December 15-18, 2021*, pages 5544–5552. IEEE, 2021.

[11] Lukasz Sosnowski, Joanna Zulawinska, Soma Dutta, Iwona Szymusik, Aleksandra Zygula, and Elzbieta Bambul-Mazurek. Artificial intelligence in personalized healthcare analysis for womens' menstrual health disorders. In Maria Ganzha, Leszek A. Maciaszek, Marcin Paprzycki, and Dominik Slezak, editors, *Proceedings of the 17th Conference on Computer Science and Intelligence Systems, FedCSIS 2022, Sofia, Bulgaria, September 4-7, 2022*, volume 30 of *Annals of Computer Science and Information Systems*, pages 751–760, 2022.

[12] Kenneth A. Ginsburg. "Luteal Phase Defect: Etiology, Diagnosis, and Management". *Endocrinology and Metabolism Clinics of North America*, 21(1):85–104, 1992. Reproductive Endocrinology.

[13] Neil F. Goodman, Rhoda H. Cobin, Walter Futterweit, Jennifer S. Glueck, Richard S. Legro, and Enrico Carmina. "American Association of Clinical Endocrinologists, American College of Endocrinology, and Androgen Excess and PCOS Society Disease State Clinical Review: Guide to the Best Practices in the Evaluation and Treatment of Polycystic Ovary Syndrome - Part 1". *Endocrine Practice*, 21(11):1291–1300, 2015.

[14] Janusz Kacprzyk and Ronald R. Yager. "Linguistic summaries of data using fuzzy logic". *International Journal of General Systems*, 30(2):133–154, 2001.

[15] Janusz Kacprzyk and Slawomir Zadrozny. "Fuzzy logic-based linguistic summaries of time series: a powerful tool for discovering knowledge on time varying processes and systems under imprecision". *Wiley Interdiscip. Rev. Date Min Knowl. Discov.*, 6(1):37–46, 2016.

[16] P.I. Good. *"Resampling Methods: A Practical Guide to Data Analysis"*. Birkhäuser Boston, 2005.

DOI: 10.15439/2023F5207
ISSN 2300-5963 ACSIS, Vol. 37

Data science to identify crimes against public administration

Luan Bruno Souza
Postgraduate Program in Computer Science
(PROCC/UFS) – Aracaju, Brazil
Laboratory of Technological Innovation in Health
(LAIS) – Natal, Brazil
Prosecution Office of Sergipe (MPSE) – Aracaju,
Brazil
luanbrunos@gmail.com

Methanias Colaço Júnior
Postgraduate Program in Computer
Science (PROCC/UFS) – Aracaju, Brazil
Laboratory of Technological Innovation in
Health (LAIS) – Natal, Brazil
Advanced Center for Technological
Innovation (NAVI) – Natal, Brazil
mjrse@hotmail.com

Rodrigo Silva
Ministry of Health – Brasília, Brazil
Laboratory of Technological Innovation
in Health (LAIS) – Natal, Brazil
rodrigo.silva@lais.huol.ufrn.br

Raphael Fontes, Caldeira Silva, Jailton Paiva,
Ricardo Valentim, Gabriel Lins
Laboratory of Technological Innovation in
Health (LAIS) – Natal, Brazil
Advanced Center for Technological
Innovation (NAVI) – Natal, Brazil
raphaelf.ti@gmail.com,{caldeira.silva, jailton.paiva,
ricardo.valentim, gabriel.lins}@lais.huol.ufrn.br

***Abstract*—Context: The management of public resources is subject to illegal acts and the automatic identification of such acts depends on the analysis of a lot of data. Objective: The object of this work is the analysis of scientific publications through a study based on systematic mapping with the purpose of evaluating them in relation to the use of automated tools to identify crimes against public administration in databases from the perspective of researchers in the data science context. Method: Using PICO strategy (Population, In- tervention, Comparison, and Outcome), a systematic mapping was conducted to find the primary studies in the literature and collect evidence for directing future research. Results: Nineteen works were found that fit the proposed cri- teria. Almost 80% of the studies found seek to identify some type of fraud in bidding processes, obtaining accuracies between 72% and 99%. The research also revealed different techniques for approaching the problem. Considering all the works, the most used databases are bidding bases, lawsuits, public notices and corporate structure of companies, respectively. Conclusions: The work has shown a recent increase in interest in analyzing public data for irregularities. It is expected that this analysis will help control bodies elucidating different ways of detecting crimes against the public administration in an automated way.**

***Index Terms*—Crime, Corruption, Public Administration, Data Science**

I. Introduction

PUBLIC resource management in many countries, as well as in Brazil, is unfortunately subject to illicit acts, which aim at the subtraction usage of the same resources for the public benefit. Among the most common crimes against public administration, according to Brazilian law, are Corruption, Embezzlement, Prevarication, and Concussion. In the Brazilian context, a study carried out by the Department of Competitiveness and Technology (Decomtec) of Fiesp (Federation of Industries of São Paulo) revealed that the economic and social damage caused by corruption in the country reaches R$ 69 billion reais per year [8]. At the same time, the Anti-Corruption Capacity Index (CCC), which is prepared by the American business entity Americas Society/Council of the Americas (AS/COA) and the British consultancy Control Risks, indicates that, since 2019, Brazil has been falling in the ranking that measures each nation ability to fight corruption [24]. In addition, Brazil ranks 96th in the Corruption Perceptions Index, organized by Transparency International, which order countries' ranks according to the degree to which corruption is perceived to exist among public and political officials, in a total of 180 countries [10].

This difficulty in combating crimes against the public administration involves the difficulty in analyzing a large data volume referring to the public asset movement, often dispersed in different databases. As a result, a good part of the investigative processes about damages to public funds originated in complaints made by the citizens themselves [25]. However, despite the difficulty imposed by the large information volume, it is precise that a good part of government services are stored (and to some extent available) in digital format that makes their analysis through Data Science and Data Analytics usage.

Given this scenario, it is necessary to use and improve techniques and tools which aim to detect, identify or predict the potential crime existence against the public administration. In many cases, these deductions can only be extracted from the unified analysis of distinct databases. The information collected in heterogeneous databases, in order to assist in the decision-making process, is already widely used in the private sector worldwide, for example, in training Credit Score - an index that determines how safe it is to provide credit to a given consumer [11].

The present work objective, therefore, is, through a systematic mapping accomplishment, to carry out a survey on the studies that aim at the development and improvement of crime detection techniques against the public administration in databases, which techniques are most used, which crime types are most addressed, and which databases are most used. The work also aims to observe the countries with the greatest interest in exploring the problem and whether it is possible to establish a correlation between this interest and the corruption perception indices, according to [10], as well as the interest evolution over time.

The rest of the work is organized as follows: section 2 describes the work methodology, the research questions raised, and the search strategies. Section 3 presents the results obtained after the search, as well as the answers to the research questions. A work narrative synthesis is described in section 4. Section 5 looks at threats to the work's validity. Conclusions and final considerations are presented in section 6.

II. Methodology

The following section describes the methodology used to carry out the work. To guide the research question formulation and the bibliographic search, the PICO strategy was used [23]. The PICO strategy guides the research question construction and the bibliographic search, and allows the researcher, when having doubt or question, to locate, accurately and quickly, the best scientific information available. It presents four fundamental research elements: Population, Intervention, Control, and Outcome, which the authors used to describe all components related to the identified problem and structure the research questions.

A. Research questions

Following are the research questions:

QP1. What crime types against public administration are most commonly identified in these works?

QP2. What are the most widely used data science approaches to detect them?

QP3. What are the approach performance metrics?

QP4. What are the most used databases for the approach application?

QP5. What are the main journals and conferences on the topic?

QP6. In which years were more articles published in this area?

QP7. Which countries have the most publications in this area?.

B. Search Strategy

The research was designed according to the PICO strategy [23], and the result is illustrated in Table 1. Therefore, keywords were established for each category. The resulting set is described in Table 2. The first keywords were selected from some control articles, similar to solution sought in this work. In addition, other keywords were included based on criteria such as related works, similarity and synonyms. The keyword set was then refined, removing redundant words and identifying word stems. The process result is illustrated in Table 3.

Table 4 shows the string used for searches in the databases. The population keywords were subdivided into three blocks, the first being related to the action (detection and its correspondences), the second related to the object sought (crime, corruption, and its correspondences), and the third block related to where to find the objects sought

Table I. PICO Strategy Categories

Category	Description
Population	Publications that directly address the crime identification against the public administration.
Intervention	Context of applications that use automated approaches to identify crimes against public administration.
Control	Applications that do not use automated approaches to identify crimes against the public administration.
Result	Automated approaches to identify crimes against the public administration through computing usage.

Table II. Keywords by category

Category	Description
Population	crime detection against public administration, corruption detection, collusion detection, fraud detection, corruption in public sector, fraud detection in public procurement, risk pattern in public sector, cartel detection, corruption risk assessment, offences against public administration, public ghost employee, public ghost payroll, organized crime, prevarication, public treasury, public procurement, public bidding, government purchasing, bid rigging, public fund, money laundering
Intervention	data mining, data science, text mining, artificial intelligence, a.i, data crossing, crossing technologies, data combination, data manipulation, machine learning, neural network, deep learning, cluster analysis, algorithm
Control	-
Result	decision support system, dss, knowledge discovery, automated system, automated information system, prototype, online analytical processing, olap, intelligent agent, corruption indicator, predictive, model, predictive analytics, model

Table III. Keywords Refined by Category

Population	**crime detect*, collusion detect* corruption detect*, fraud detect* offences detect*, cartel detect* prevarication detect*, ghost payroll detect*, ghost employee detect*, bid rigging, money laundering, corruption risk, public administration, public sector, public procurement public treasury, public bidding, public employ*, government* purchas* government* treasury, public fund, risk pattern**
Intervention	data mining, data science, text mining, data crossing, artificial intelligence, crossing technologies, data combination, data manipulation, machine learning, neural network, deep learning, cluster analysis, algorithm
Result	decision support system, dss, knowledge discovery, automated system, prototype, automated information system, online analytical processing, olap, intelligent agent, corruption indicator, approach, predictive model, predictive analytics, model

Table IV. Generic Search String

("detect*" OR "search*" OR "find*" OR "look* for" OR "predict*") **AND ("crime" OR "corruption" OR "clue" OR "fraud*" OR "collusion" OR "offence" OR "cartel" OR "malfeasance" OR "prevarication" OR "ghost payroll" OR "ghost employee" OR "bid rigging" OR "irregularity" OR "money laundering"OR "anomaly" OR "suspicious")** **AND ("public administration" OR "public sector" OR "public procurement" OR "government* procurement" OR "public treasury" OR ("bidding" AND ("public" OR "government*")) OR "public employ*"OR "government* purchas*" OR "government* treasury" OR "public fund")**	Population
AND ("data mining" OR "data science" OR "text mining" OR "artificial intelligence" OR "data crossing" OR "crossing technologies" OR "data combination" OR "data manipulation" OR "machine learning" OR "neural network" OR "deep learning" OR "cluster analysis" OR "algorithm")	Intervention
AND ("decision support system" OR "dss" OR "knowledge discovery" OR "automated system" OR "automated information system" OR "prototype" OR "online analytical processing" OR "olap" OR "intelligent agent" OR "predictive model*" OR "predictive analytics" OR "model" OR "corruption indicator" OR "approach*")	Result

(public sector, bids, and their correspondence). Searches in titles, abstracts, and keywords were used in the Scopus, IEEE Xplore Digital Library, Web of Science, and ACM Digital Library search bases.

Following are the Inclusion Criteria: (1) Short and complete works published and available in full in scientific databases, with title, abstract, and keywords available in the English language; (2) Recent works (published from 2010 onwards), however, they have already been approved by the scientific community. (3) Works that propose a method, tool, or application for the detection, selection, or fraud or crime prediction against public administration in databases through Data Science usage. The 2010 limit year was determined to be immediately prior to the Law implementation on Information Access [1], which regulated the citizens' constitutional right to access public information.

The following are Exclusion Criteria: (1) Duplicate works; (2) Restricted works; (3) Revision works; (4) Works that do not seek to detect crimes; (5) Works that seek to detect or predict other crimes outside the context of this work.

C. *Information Extraction Strategy*

To assess the work quality and answer the exposed research questions in section 2.1, a form was designed to be answered for each article read completely. According to [12], data extraction forms should be designed to collect all the information necessary to address the issues and quality criteria of the study. Table 5 presents the extraction form used in this research. For the attributes Crime Types, Approaches, Performance Metrics, and Databases, the results are multivalued, that is, there is the possibility of more than one answer of the same attribute for each article.

Table V. Extraction Form

Attribute	Description
Crime Type	Identification of the crime type against the public administration which the work aims to identify. Part of this task was already carried out in exclusion criterion 5, which sought to remove crime identification work outside the public administration context.
Approach	The Data Science identification approach used in the crime identification
Performance Metric	The evaluation criteria identification of the approach according to the authors' experiment, if there is any.
Database	The databases identification, structured or not, analyzed by the approaches.

III. Results and Discussion

The following subsections describe the search process and discuss the results. In subsection 3.1 the resulting treatment and the exclusion criteria application until the analysis base formation is described. Subsection 3.2 runs briefly over each selected job. From subsection 3.3 onwards, the research questions are answered based on the results.

A. *Results*

Once the works were searched in the specified databases using the keywords, the first step was the duplicate work removal since they were found in more than one database. Figure 1 presents a flow describing the article extraction process from this phase to analysis. The search sum in the databases returned a total of 251 works, a number that was reduced to 223 after the duplicate article removal.

Then the other exclusion criteria were applied. Two articles were removed for being of a restricted domain. Afterward, the work title was read to identify review articles. Along with the title, the work abstract was also observed, which allowed us to remove those that were not intended to detect fraud and crimes. These three criteria allowed us to reduce the number of works to a total of 100 articles.

After the exclusion according to these criteria, we were left with 100 works that aimed to search for techniques and tools to detect crimes or fraud automatically. Yet, many of these works did not aim at crime identification in the public administration sphere. Among the events sought by these articles were common crimes, hacking invasions, health insurance fraud, and even illegal immigration. Frequently reading the title and abstract were sufficient for this discernment, but often the article introduction needed to be read for more precision. Finally, after the last step in applying the removal criteria, we reached 19 articles. All have been read completely and a brief commentary is described in the following sections.

B. *Work Abstracts*

The works of [14] and [20] present an approach to crime detection based on users' perceptions. The first is based on the post content on the social network Twitter, while the authors of the second created a survey to be applied by public

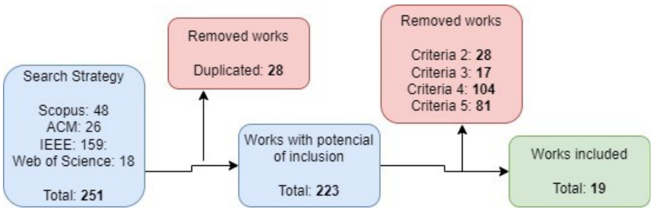


Figure 1. Prism chart with data extraction

service users. The article by [13] seeks to analyze financial transactions in order to find suspicious transactions that lead to money laundering, while [3] propose an ontology to, applied to a data warehouse, identify inconsistencies in payroll.

From here, the articles focus on fraud detection in the public purchase sphere. [7], [16], and [21] seek to identify potential signs of fraud already in terms of the bids using opening, among other devices, text mining techniques. The work of [5] proposes to use Bayesian networks to identify fractional purchases, where the bidding process is suppressed if each purchase value does not exceed a maximum value defined by Brazilian legislation.

The article by [17] added, to the bidding database, a bid list against those for which there are legal proceedings or formal complaints. The objective is to detect patterns of attributes of problematic processes in order to identify problems in new bidding processes using random forest. [9] used a similar approach, in addition to using other available process data, such as budget, duration, delays, time before electoral processes, and geographic patterns.

Other works seek to detect bidding processes with potential collusion through the association network analysis of other purchases involving the same buyers or suppliers. They are [22], [6], [19] and [4]. For this, they use techniques such as association rules and random forest. Articles such as [2], [18], and [26] use clustering algorithms to group competitive and non-competitive bidding processes based on data such as the ratio between the bid values offered by companies and initial value of the contract.

Finally, [25] and [15] propose the veracious data analysis suite creation of bidding processes, precisely with the addition of information available in other databases. It allows the fraudulent schemes detection that could only be elucidated from this distributed information joining. Auxiliary databases include corporate structure data of companies, income transfer programs, and electoral data

C. QP1: What crime types against public administration are most commonly identified in these works?

The vast majority (78.9%) of the work authors focused their efforts on automated techniques to detect fraud in bids, as illustrated in Figure 2. However, the works differ on when the detection attempt is performed. Some works, such as [7], [16] and [21], seek to identify potential fraud signs in terms of the bid opening. Other works, such as [2] and [18], use variables found during the bidding process to find collusions, such as bid values and time intervals. Finally, works such as [22] and [4] based on the compilation of different bidding processes already carried out in search of participation patterns and winners. There are still other works, such as [25] and [15], which use multiple approaches to detection.

Two other works ([14] and [20]) did not define a specific crime type but were concerned with detecting fraud in a more comprehensive way through opinion collection and user perception. There are also works aimed at finding fraud in the government employee payroll [3] and money laundering [13].

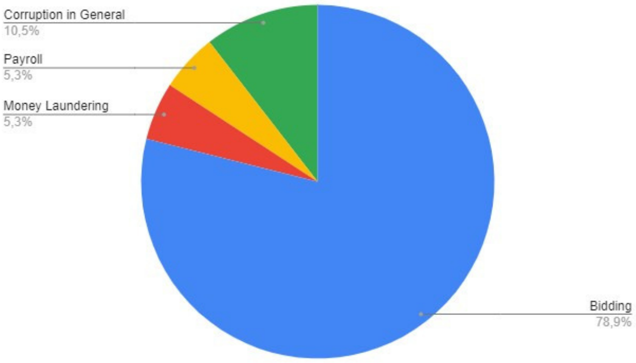


Figure 2. Crimes or Frauds identified in the approaches

D. QP2: What are the most widely used data science approaches to detect them?

As seen in the previous item, most works focus on fraud detection in government purchases through bidding processes. Some works, such as [7], [16] and [21], seek to identify signs already in terms of opening the process. For this, text mining tools are used to analyze specific term elements. Once found, they apply logistic regression or deep learning algorithms to detect a competitiveness lack in bidding terms, which could point to a possible collusion between the bidder and interested companies. On the other hand, works aimed at identifying fraud in the same processes using data generated during the bidding process, such as [2], [26] and [18], using data as the ratio between bid value and initial bid value, through clustering algorithms to differentiate competitive and non-competitive processes. Finally, clustering algorithms, as well as association rules used by works such as [22] and [4] to identify collusions between companies and suppliers through several bidding process analysis. Other works, such as [25] and [15], combine other techniques for this detection, in addition to the assigning score possibility to certain companies that participate in bidding processes. For this, they use other data sources in addition to the basis of contracts and public bids generally used in other approaches, as detailed in the following section.

The works [14] and [20], which did not define a specific crime type because they are concerned with detecting fraud in a more comprehensive way, making use of reports and impressions of public service users. While [14] use machine learning techniques to detect fraud evidence in public services through posts on Twitter, [20] applied forms to users of different services in order to search for inefficiency signs based on the responses to these forms using clustering algorithms.

To detect money laundering crimes, [13] used a Bayesian classifier based on a bank operation set. As for looking for inconsistencies in payrolls (not necessarily fraud) [3] de-

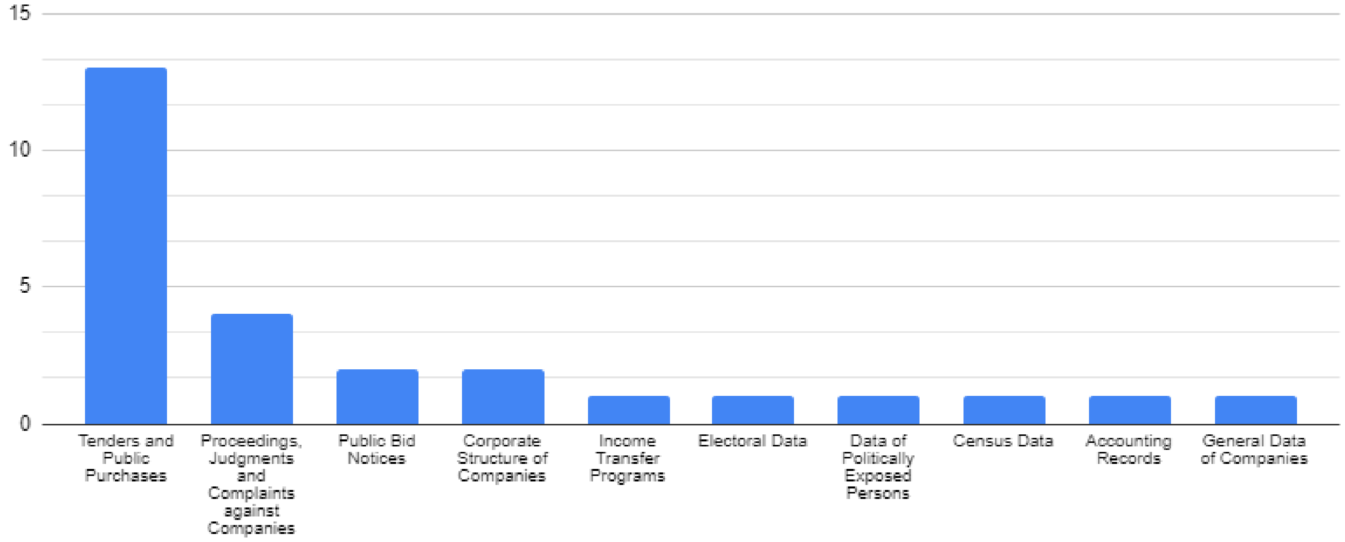


Figure 3. Databases used to detect bid fraud

fined an ontology indexing process through concept maps and audit indicators as a tool for documenting evidence.

E. QP3: What are the approach performance metrics?

In general, the authors used accuracy as the predominant form of statistical evaluation of the proposed models, with the exception of [22] who obtained an assertiveness of 90% according to its own evaluation index, called RQ, when trying to identify cartel formation and [7] who obtained a Mean Square Error (MSE) of approximately 0.0013 when trying to predict fraudulent bids from their opening terms.

Also, when analyzing the bid opening terms [16] obtained an accuracy of 76% using SVM, while [21] obtained accuracies between 72% and 85%, depending on the product group of the used bidding process utilizing Logistical Regression and Bayesian Networks. [6] obtained an accuracy of 67% in identifying cartels. Through the attribute analysis of the bidding process, [26] reached an accuracy of 99%, while in the work of [17] the same rate was 90% using similar data, including data from known problematic bids. The work of [5] reached an accuracy of 99.9% analyzing fractional bids where the global value is divided into bids with lower values to circumvent some legal requirements.

Outside the bidding process context, [20] obtained an accuracy of 87.5% in the irregularity discovery when applying a questionnaire to public service users. [13] reached an accuracy of 81% when searching for suspicious financial transactions in order to find money laundering evidence. The other works found proposed data analysis models without presenting statistical validations regarding these models' assertiveness.

F. QP4: What are the most used databases for the approach application?

The answer to this question must take into account the fraud or crime type that the work aims to detect. [13], for example, used bank transaction databases to look for fraud evidence. [3] used a payroll database to build a data warehouse and define its ontology. In turn, [14] and [20] used posts on the social network Twitter and data from an applied survey, respectively, to identify fraud in the public sector through the perception of users.

Figure 3 counts the databases used to help detect fraud in bidding processes. Note that one approach can make use of more than one database simultaneously. Altogether, 13 of the 15 studies found that proposed to detect anomalies in bidding processes utilizing public bidding and procurement bases, while the other two analyzed only opening documents and the process definition. In order to negatively consider processes involving companies against which there was a history of lawsuits, some works made use of procedural bases, judicial sentences, and complaints. Other databases used were those that included the corporate structure of companies, income transfer programs, electoral data, data on politically exposed persons, census data, accounting records, and company registration data.

G. QP5: What are the main journals and conferences on the topic?

Among the results found, all of them were published in different Magazines, Journals, or Conferences. Thus there isn't a periodical or conference that stood out from the others.

H. In which years were more articles published in this area?

As shown in Figure 4, it is possible to notice an increase in the publication of works that address the researched topic from 2019, with four papers published. The year 2020 was,

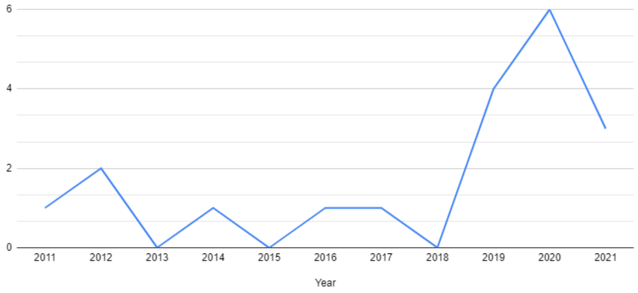


Figure 4. Databases used to detect bid fraud

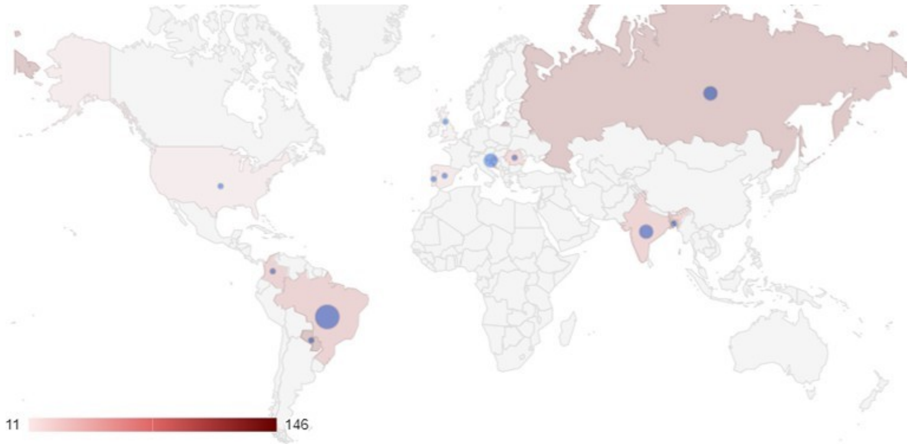


Figure 5. Publication number by country and position in the IPC-2020

until then, the one with the highest number of publications, accounting for six papers.

I. QP7: Which countries have the most publications in this area?

Figure 5 shows the distribution of published works around the world, where the size of the blue circle represents the publication number. It is possible to notice that the country with the largest publication number is Brazil (5), followed by Russia (2), India (2), and Croatia (2). Colombia, Spain, Bangladesh, the United Kingdom, the United States, Romania, Paraguay, and Portugal complete the list with one work each. Figure 5 also plots, in red, the position of that country (only where works was found) in the Corruption Perceptions Index prepared by International Transparency for the year 2020 [10], where the darker the red color, the greater the corruption perception. Through the results, it was not possible to establish a relationship between the number of published works and the corruption perception in the country.

IV. Narrative Synthesis

Quantitatively, the result observation allowed us to observe that the search for automated ways to detect fraud and crimes is relatively recent in the scientific context. For many authors, this is often due to the late digitization process of governments in relation to the private sector, especially in underdeveloped countries. With no government data available in digital format, there is no means to perform such a task.

The quantitative analysis also placed Brazil as a major contributor to this approach type, despite being in an intermediate position in the Corruption Perception Index (CPI) in 2020 provided by International Transparency. It was observed that the publication number per country cannot be directly related to the countries' perceptions of corruption, according to the same index.

The bidding processes, as analyzed, are the main target of automated fraud detection processes in the public context. In general, the authors justified the interest in this collusion type due to the large financial volumes involved in government purchases that carry out the bidding processes. Furthermore, the amount of money involved in these transactions inevitably ends up arousing malicious people's interest.

For such detections, the works take turns using predictive and deductive models. Deductive models are generally based on local legislation and prior knowledge about the fraudulent scheme typologies, which is often a disadvantage because this approach type is not able to predict new scheme formats. On the other hand, predictive models are more difficult to apply due to the absence of training bases, considering that the number of proven frauds is often insufficient for modeling this approach type.

V. Threats to Validity

The great difficulty of the current work concerns the keyword selection to search in the databases. As much as the search context is well defined, the expressions used to describe crimes, frauds, or anomalies are diverse and are subject to different regionalities and descriptions depending on the country where the laws are written making it difficult to select terms used as population keywords, according to the PICO model. This characteristic threatens above all the excessive volume of works from Brazil. Another similar difficulty is the wide term variety used to describe the methods used for detection, described in the intervention keywords. An incomplete keyword selection can considerably limit the number of results returned.

As for the exclusion criteria, the heterogeneity of different laws and policies in different countries can compromise the researcher's interpretation, regarding the often subjective analysis of these criteria. For example, in the current work, fraud against health plans was not considered, given that Brazil has a single public health system that is not very intertwined with the private system so financial fraud against health plans in Brazil generally does not involve public administration. But it is not possible to infer that this does not occur in other countries.

VI. Final Consideration

All over the world, to a greater or lesser extent, public money management deprives the population of the right to fully take advantage of the resources provided by them through taxes. This mismanagement is often intentional, resulting from criminal actions that seek to subtract or misuse public goods for their own benefit. Fortunately, the recent governmental service digitization, allied to the principle that part of this information load is in the collective domain, allows the organizational or popular initiative emergence, aimed at these illicit act identification. The information sheer volume, however, requires an automated process.

The current work described a systematic mapping with the objective of elucidating scientific works aimed at the automated tools development or improvement for the fraud detection or crimes against public administration. The research questions were raised and, based on the PICO strategy, the search keywords were selected. Once searched, the works were selected based on pre-defined inclusion and exclusion criteria.

The result analysis shows that this concern, fortunately, is growing. Several studies were found with this objective in mind, and they do so by approaching different strategies. Due to the financial resource volume involved, bidding fraud is the main target of this initiative type. Some works even look for the association of different databases, seeking the fragmented information discovery. The computational resources for this range from text mining to machine learning algorithms.

It is hoped that this work can provide guidance to entities that seek to develop initiatives and develop tools that allow a better public expenditure monitoring. As noted in this work, part of the information available for this task is in the public domain, allowing non-governmental entities to participate directly in these initiatives. However, it is the control bodies that have exclusive control over part of the data identified as a source for the detecting crime work, in addition to having civil liability for such.

It is recommended the existence of periodical works in this sense, in order to maintain the population and control institutions always updated on the best practices to achieve the final objective, which is the fight against fraud in public administration. In future works, it is recommended a better understanding of the terms used to define illegal or suspicious acts which will be used in the search string, in order to avoid the existence of false negatives in the process. In addition, a more in-depth analysis of the results offered by the applications found is also recommended, comparing them and indicating the best approach for each situation.

References

[1] Brasil (2011). Lei nº 12.527, de 18 de novembro de 2011. Dia´rio Oficial da Repu´blica Federa- tiva do Brasil.

[2] Busu, M. and Busu, C. (2021). Detecting bid-rigging in public procurement. a cluster analysis approach. Administrative Sciences, 11(1):13.

[3] Campos, S. R., Fernandes, A. A., De Souza, R. T., De Freitas, E. P., da Costa, J. P. C. L., Serrano, A. M. R., and Rodrigues, D. d. C. (2012). Ontologic audit trails mapping for detection of irregularities in payrolls. In 2012 Fourth International Conference on Computational Aspects of Social Networks (CASoN), pages 339-344. IEEE>

[4] Carneiro, D., Veloso, P., Ventura, A., Palumbo, G., and Costa, J. (2020). Network analysis for fraud detection in portuguese public procurement. In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 390–401. Springer.

[5] Carvalho, R. N., Sales, L., Da Rocha, H. A., and Mendes, G. L. (2014). Using bayesian networks to identify and prevent split purchases in brazil. In *BMA@ UAI*, pages 70–78.

[6] Domashova, J. and Kripak, E. (2021). Application of machine learning methods for risk analysis of unfavorable outcome of government procurement procedure in building and grounds maintenance domain. *Procedia Computer Science*, 190:171–177.

[7] Domingos, S. L., Carvalho, R. N., Carvalho, R. S., and Ramos, G. N. (2016). Iden- tifying it purchases anomalies in the brazilian government procurement system using deep learning. In *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 722–727. IEEE.

[8] FIESP (2010). Relato´rio corrupc¸a˜o: custos econoˆmicos e propostas de combate. In *DECOM- TEC*. FIESP - Federac¸a˜o das Indu´strias do Estado de Sa˜o Paulo.

[9] Gallego, J., Rivero, G., and Mart´ınez, J. (2021). Preventing rather than punishing: An early warning model of malfeasance in public procurement. *International journal of forecasting*, 37(1):360–377.

[10] International, T. (2021). Corruption perceptions index.

[11] Investopedia (2021). Credit score. https://www.investopedia.com/terms/ c/credit_score.asp. Accessed: 2021-10-16.

[12] Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University*, 33(2004):1–26.

[13] Kumar, A., Das, S., and Tyagi, V. (2020). Anti money laundering detection using na¨ıve bayes classifier. In *2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON)*, pages 568–572. IEEE.

[14] Li, J., Chen, W.-H., Xu, Q., Shah, N., and Mackey, T. (2019). Leveraging big data to identify corruption as an sdg goal 16 humanitarian technology. In *2019 IEEE Global Humanitarian Technology Conference (GHTC)*, pages 1–4. IEEE.

[15] Mart´ınez-Plumed, F., Casamayor, J. C., Ferri, C., Go´mez, J. A., and Vidal, E. V. (2018). Saler: a data science solution to detect and prevent corruption in public administration. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 103–117. Springer.

[16] Modrusan, N., Rabuzin, K., and Mrsic, L. (2020). Improving public sector efficiency using advanced text mining in the procurement process. In *DATA*, pages 200–206.

[17] Niessen, M. E. K., Paciello, J. M., and Fernandez, J. I. P. (2020). Anomaly detection in public procurements using the open contracting data standard. In *2020 Seventh International Conference on eDemocracy & eGovernment (ICEDEG)*, pages 127–134. IEEE.

[18] Padhi, S. S. and Mohapatra, P. K. (2011). Detection of collusion in government procurement auctions. *Journal of Purchasing and Supply Management*, 17(4):207–221.

[19] Popa, M. (2019). Uncovering the structure of public procurement transactions. *Business and Politics*, 21(3):351–384.

[20] Pramanik, A., Sarker, A., Islam, Z., and Hashem, M. (2020). Public sector corruption analysis with modified k-means algorithm using perception data. In *2020 11th International Conference on Electrical and Computer Engineering (ICECE)*, pages 198–201. IEEE.

[21] Rabuzin, K. and Modrusan, N. (2019). Prediction of public procurement cor- ruption indices using machine learning methods. In *KMIS*, pages 333–340.

[22] Ralha, C. G. and Silva, C. V. S. (2012). A multi-agent data mining system for cartel detection in brazilian government procurement. *Expert Systems with Applications*, 39(14):11642–11656.

[23] Santos, C. M. d. C., Pimenta, C. A. d. M., and Nobre, M. R. C. (2007). A estrate´gia pico para a construc¸a˜o da pergunta de pesquisa e busca de evideˆncias. *Revista Latino-Americana de Enfermagem*, 15:508–511.

[24] Simon, R. and Aalbers, G. (2019). The capacity to combat corruption (ccc) index. *AS/COA*, page 15.

[25] Velasco, R. B., Carpanese, I., Interian, R., Paulo Neto, O. C., and Ribeiro, C. C. (2021). A decision support system for fraud detection in public procurement. *International Transactions in Operational Research*, 28(1):27–47.

[26] V.I., D., N.V., M., and M.I., B. (2017). Adaptation of cluster analysis methods in respect to vector space of social network analysis indicators for revealing suspicious government contracts. In *2017 5th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*, pages 57–62.

DOI: 10.15439/2023F9176
ISSN 2300-5963 ACSIS, Vol. 37

Evolutionary k-means Graph Clustering Method to Obtain the hub&spoke Structure for Warsaw Communication System

Barbara Mażbic-Kulma
0000-0002-7668-4694
Warsaw School of Information Technology
ul. Newelska 6
01-447 Warsaw, Poland
Email: kulma@wit.edu.pl

Jan W. Owsiński, Jarosław Stańczak, Aleksy Barski, Krzysztof Sęp
0000-0002-2750-6584, 0000-0003-3722-0085
0000-0003-3746-1778, 0000-0002-1453-4148
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6 01-447 Warsaw, Poland
Email: {owsinski, stanczak, aleksy.barski, sep}@ibspan.waw.pl}
Warsaw School of Information Technology
ul. Newelska 6
01-447 Warsaw, Poland

***Abstract*—The k-means method is one of the most frequently used clustering methods due to its efficiency and ease of modification and adaptation to the problem being solved. This paper presents modification of k-means method used for clustering in graphs. The method is presented on the example of generating the hub&spoke structure in the graph of public transport connections in Warsaw. Optimization of the public transport is one of the most important tasks for large cities. An efficient transport system is very important for its inhabitants. One of possible solutions is introducing the idea of hub&spoke to transport system. In this approach it is important to detect main stations, called hubs, which will create axes of high-speed connections (city trains, metro, high-speed trams), from which passengers can transfer to slower local connections to get to their rather close destinations. In the presented approach we propose to find locations of such main changeover stations using an evolutionary k-means algorithm.**

***Index Terms*—hub&spoke, evolutionary k-means algorithm, city transport system.**

I. Introduction

THE BASIC k-means algorithm became the starting point for the construction of many of its modifications adapted to different needs. In this paper we present its modification used for clustering in graphs, applied to obtain the hub&spoke structure ([2], [11]) of public transport system in Warsaw.

Obtaining high efficiency of urban transport is a very big challenge. This can be achieved by high financial outlays for building new fast connections (metro lines, fast trains or fast trams) or to some extent by optimizing the existing system. In this work we propose significantly cheaper approach, which requires (probably slow and well thought out) rearranging the transport in the city using the hub&spoke structure. The hub&spoke structure was successfully used in the 1970s to reorganize air traffic [6]. Currently, the air transport is so developed (or there are so many possibilities now) that this paradigm of connections is often abandoned (an example of which is the announced cessation of production of the A380 aircraft, designed mainly for the mass transportation of passengers between hub airports), which does not mean that the method will not be useful in other areas of transport. It seems that public transport in big cities can be such an application of the hub&spoke structure.

The properties and definitions of the hub&spoke structure were presented in works: [1], [7], [11], [12], [13] [14] and [16]. The basis of the idea of arranging urban transport as a structure hub&spoke is that individual parts of the city are connected by a network of fast means of transport connections, mainly rail, metro and fast trams. Selected stops of these fast means of transport, can become communication hubs where one can change to slower, local means of transport (buses, ordinary trams or even bicycles,...), but the whole journey usually lasts shorter, because the main burden of transport has been transferred to fast and high-capacity means of transport. Of course considered city should have such a fast means of transport.

The proposed ideas for changing the concept of urban transport do not assume a revolutionary removal of stops and connections (which may cause passenger protests), but rather a slow reorganization of the system so that it evolves towards a more effective hub&spoke structure, while maintaining many connections that break this structure due to the habits of users.

Our new approach to this problem is based on the described further evolutionary k-means method for graphs clustering (EKMG), which finds groups of strongly connected stops and designates a central one as a communication hub. The data for the EKMG algorithm come from the preprocessed timetable for the city transport system.

II. Basic Notions, Definitions and Algorithms

A. Clustering methods

Clustering is a disjunctive partitioning of set of data X, containing n-dimensional elements x into p nonempty subsets called clusters, containing elements similar in some sense or measure, while the elements belonging to different clusters should be highly dissimilar in the same sense or measure.

This aim can be obtained using many methods, one of the most commonly used is the k-means method ([4], [5] and [20]), which is presented in the Algorithm 1:

1. Choose the number of sought clusters.
2. Generate starting positions of cluster centroids
3. Calculate distances of all clustered objects to all cluster centroids.
4. Assign objects to clusters with the closest centroids.
5. Update cluster centroids as geometric centers of their clusters.
6. If the assignment of objects to clusters in the subsequent two steps does not change, then go to 7, else go to 3.
7. End.

Algorithm 1. The basic k-means method algorithm.

The properties of this algorithm strongly depend on the accepted minimized distance measure:

$$C_D = \Sigma_q \Sigma_{i \in Aq} d(x_i, x_q), \qquad (1)$$

where:

$d(.,.)$ - denotes the Euclidean (or different) distance;

x_i – clustered data items;

x_q – centroids (center or mean points) of clusters A_q, q=1,... p.

Very important parameter of this method is p – the number of clusters which is imposed by algorithm users but is not tuned by the method. Mentioned earlier and described more precisely in section 4 the EKMG method can deal with this problem.

B. Evolutionary algorithms

The standard evolutionary algorithm (EA) works as this is shown below ([2]):

1. Random initialization of the population of solutions.
2. Reproduction and modification of solutions using genetic operators.
3. Evaluation of obtained solutions.
4. Selection of individuals for the next generation.
5. If stop condition is not satisfied go to 2, else go to 6.
6. End.

Algorithm 2. The standard evolutionary algorithm scheme.

As it is known from further works ([9], [10]) this simple algorithm requires several improvements in order to work efficiently:

- the invention of a proper encoding of solutions,
- development of specialized genetic operators, appro priate for the accepted solution encoding (if standard ones are not proper),
- formulation of the fitness function to be optimized by the algorithm.

The stop condition is usually described by a certain number of iterations.

C. Basic graph notions

We treat the city transport system as a graph with stations as graph nodes and transport lines as edges, thus some basic notions from graph theory are presented here, following [21].

A graph is a pair $G = (V, E)$, where V is a non-empty set of vertices and E is a set of edges. Each edge is a pair of vertices $\{v_1, v_2\}$ with $v_1 \neq v_2$.

In our problem we can consider also a directed graph, which is an ordered pair $G = (V, A)$ where

- V is a set whose elements are called vertices or nodes;

- A is a set of ordered pairs of vertices, called arcs (directed edges).

A simple non-weighted graph can be described using a neighborhood matrix with elements a_{ij}, which describe the connection between vertices i and j of the graph, $a_{ij} \in \{0, 1\}$, 0 - no connection, 1 - presence of connection.

In our work we consider mainly generalization of the neighborhood matrix for weighted graphs, where elements a_{ij} describe not only the presence or no of the connection, but also its strength (for instance the capacity or travel time of connection).

A hub and spoke structure (proposed in [11] and [12]) is a graph $H_s = (G_h \cup G_s, E)$ where the subset G_h corresponds to at least a connected graph (of hubs) with the relevant subset of set E, each vertex of subset G_s (of spokes) has degree 1 and is connected exactly with one vertex from subset G_h.

III. Description of Warsaw's Transport System

The timetable describing the urban transport operation in Warsaw can be downloaded from https://www.wtp.waw.pl/rozklady-jazdy/ and ftp://rozklady.ztm.waw.pl. The public transportation system is presented in Fig. 1.

As it can be seen, this network is quite well developed and consists of:

- metro - 2 lines,
- high-speed city rail (SKM), suburban railway (WKD) and rail (KM) - 12 lines,
- trams - 26 lines,
- city, suburban and night buses - 303 lines,
- and over 10,000 stops for all mentioned means of transport.

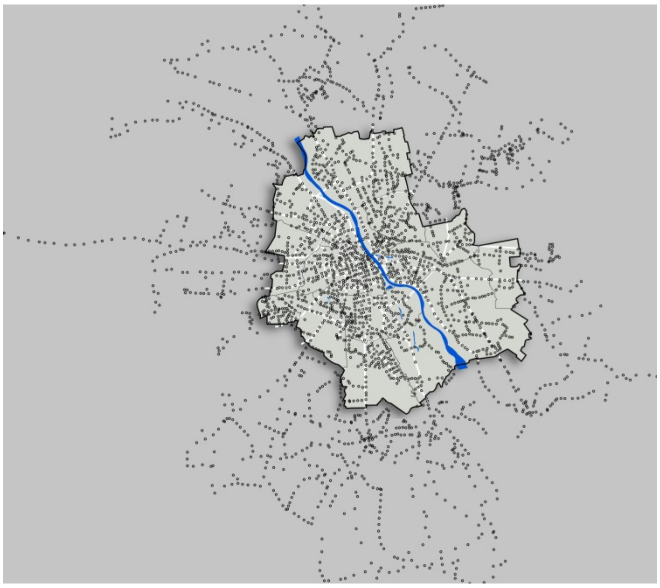


Fig. 1. Warsaw passenger transport system – 2840 unified stops (the state of the timetable as of November 25, 2021).

WTP (Warszawski Transport Publiczny, Warsaw Public Transport, the former abbreviation ZTM is still often used) conducts transport on most of the public transport lines in Warsaw. Other means of transport like private carriers and taxis were not included here due to the lack of possibility to know and to influence their transportation systems, also long-distance buses and trains were not considered to prepare the data for computations, because they are rarely used as urban mean of transport.

The public urban transport system has a large amount of about 10,000 stops, but for calculation purposes it has been reduced to 2,840 stops (graph nodes) as a result of combining into one stop stops in opposite directions and stops divided into "substops" in places with heavy passenger traffic (railway stations, bus terminals) or stops where different means of transport meet (for instance metro and bus or tram).

Our graph model of Warsaw transport system was built only on the basis of the timetable data taken from WTP/ZTM. In the constructed simplified graph of connections, we assumed that its vertices (communication stops), have a direct connection, as long as there is at least one running communication line that connects them. Thus the graph consists of overlapping blocks, because different communication lines have common stops.

The processed data obtained from the timetable may present several properties of the transportation system:

- presence or not the direct connections between the stops,
- frequency or the number of courses in a certain unit of time,
- travel time,
- potential capacity of means of transport in a certain unit of time (data about capacity of vehicles serving particular connections can be found in WTP/ZTM websites).

In the case of stops and connections common to many communication lines, the final values taken for computations are, in our case, appropriately modified (aggregated), so tha e.g. frequencies or capacity are added and the travel time is averaged in order to consider of connections from more lines at the same destination.

IV. Evolutionary k-means Method for Graphs Clustering (EKMG)

The EKMG method is based on evolutionary k-means method (EKM), which is described in detail in work [17]. In short words it can be summarized as follows:

1. Random initialization of the population of solutions (different centroids and numbers of clusters in solutions).

2. Reproduction and modification of solutions using genetic operators.

3. Evaluation of obtained solutions:

a) total minimized distance (2) is equal to infinity, the number of steps is equal to 0

b) take the number and centers of sought clusters from evaluated solution,

c) calculate distances (meant as in formula (3)) of all clustered objects to all cluster centroids,

d) assign objects to clusters with the closest centroids,

e) update cluster centroids as geometric centers of their clusters,

f) if calculated total distance for new data clustering (2) is less than calculated in previous step and number of steps is less than 5, then go to b).

g) the last value computed of the criterion (2) is the value of fitness function of the evaluated solution.

4. Selection of individuals for the next generation.

5. If stop condition of EA is not satisfied go to 2, else go to 6.

6. End.

Algorithm 3. The evolutionary k-means algorithm.

In this approach “solutions” are different instances of k-means algorithm with different numbers of clusters. Numbers of clusters and locations of their centroids can be modified by genetic operators of EA.

The minimized fitness function is similar to (1):

$$C_{Dr} = \Sigma_q \Sigma_{i \in Aq} d_r(x_i, x_q), \qquad (2)$$

where:

$d_r(.,.)$ - denotes a modified distance (Euclidean or different), described further by equations (3) and (4),

x_i– clustered data,

x_q – centroids of clusters A_q, q = 1,... p_t, the value of p_t (the number of clusters) is variable.

The modified distance $d_r(.,.)$, as used in (2), is calculated as follows:

$$d_r(x_i, x^q) = \begin{cases} d(x_i, x^q) & \text{if } d(x_i, x^q) \geqslant R \\ R & \text{if } d(x_i, x^q) < R \end{cases} \qquad (3)$$

and

$$R = (1 - r) \cdot d_{min}(x_i, x_j) + r \cdot d_{max}(x_i, x_j) \qquad (4)$$

where:

R – is the threshold value computed used threshold parameter r,

$d_{min}(x_i, x_j)$ - is the minimum value (but bigger than zero) of the accepted distance measure method among grouped different data items x_i, x_j,

$d_{max}(x_i, x_j)$ - is the maximum value of the accepted distance measure method among grouped data items x_i, x_j.

As it can be seen, the value of the threshold R is calculated on the basis of the properties of the grouped data and the given threshold parameter r, $r \in [0,1]$, which is meant to control the degree of detail of the clustering and indirectly the number of detected clusters. The threshold value also prevents the algorithm to find the trivial solution, where equation (2) is equal 0 and all data become centroids of their own one-data clusters.

The EKMG method is an application of EKM method to find clusters in graphs. The adjacency matrix of the graph is treated as a set of data about the attribute values of the nodes of the graph.

The algorithm of this method is presented in Algorithm 4:

1. Random initialization of the population of solutions: numbers of clusters and as centroids of clusters are randomly selected existing nodes of the graph.

2. Reproduction and modification of solutions (number and position of centroids) using genetic operators.

3. Evaluation of obtained solutions:

a) total minimized distance (2) is equal to infinity, the number of steps $s = 0$

b) take the number and centers of sought clusters from evaluated solution,

c) calculate distances (meant as $d(x_i,x_q)$) of all clustered objects to all cluster centroids,

d) assign objects to clusters with the closest centroids,

e) if $s < k$ update cluster centroids as graph nodes closest to computed geometric centers of their clusters,

f) if calculated total distance for new data clustering (2) is less than calculated in previous step and number of steps is less than k (k – the number of repetitions of k-means procedure, $k = 0, 1, 2$, bigger values too much slow down computations), then go to b),

g) the last value computed of the criterion (2) is the value of fitness function of the evaluated solution.

4. Selection of individuals for the next generation.

5. If stop condition of EA is not satisfied then go to 2, else go to 6.

6. End.

Algorithm 4. The evolutionary k-means algorithm for graph clustering.

The specialized evolutionary algorithm has in this case 4 genetic operators that modify solutions:

• the number of clusters – q (mutation like operator);

• values of cluster centers (random selection of new centroid among the nodes of the cluster) – A_q (mutation like operator);

• values of cluster centers (random selection of new cen troid among the nodes of the graph) – x_i (mutation like operator);

• uniform crossover (exchange of parameters between solutions).

The mechanism described in [15] was used to manage the genetic operators and select them to modify the solutions.

V. Results of Computer Simulations

New method of graph clustering was tested on Warsaw transport system data, using the time of travel and the capacity of connections as attributes of graph nodes. Simulations were conducted for different values of r parameter, equal 0.01, 0.05, 0.1, 0.3, 0.5, 0.7 and 0.9. Results with different numbers of detected hubs are presented in consecutive Table I, Fig. 2 and Fig. 3.

As you can see in Table I, the method usually selects about 24 hubs. For higher values of r imposed, the number of detected hubs starts to decrease, which is in line with the way the clustering method works: for bigger values of r, the method finds smaller number of more general clusters, for smaller values of r, the method finds bigger number of more detailed clusters. The function of the threshold parameter value r can be compared to the zoom function in a camera lens. Of course, there is no perfect proportion here, because the data parameters of the considered problem are also important and they affect the number and distribution of the clusters found.

Table I.
Numbers of clusters detected using the EKMG method depending on imposed r value

	Number of clusters detected	
r	Criterion: time of connections	Criterion: capacity of connections
0.01	27	24
0.05	25	24
0.10	24	23
0.30	24	25
0.50	25	23
0.70	20	12
0.90	2	4

In the domain of communication hubs the properties described earlier mean that hubs are stronger or weaker connected with their hubs.

Figures 2 - 5 show the results of computations on the map of Warsaw: larger points - hubs and smaller – spokes (ordinary stops) belonging to them, marked with the same color. Presented results are obtained for value of $r = 0.3$ and $r = 0.7$ respectively for the criterion of time and capacity. As it can be seen in the pictures, hub stations are mainly located in central, important communication points of the city.

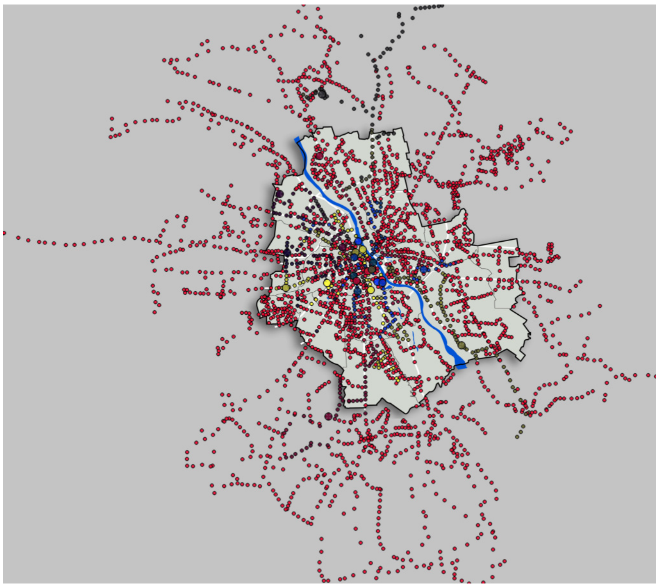


Fig. 2. Warsaw transport system with 24 hubs computed on the basis of time of connections for $r = 0.3$ (the state of the timetable as of November 25, 2021).

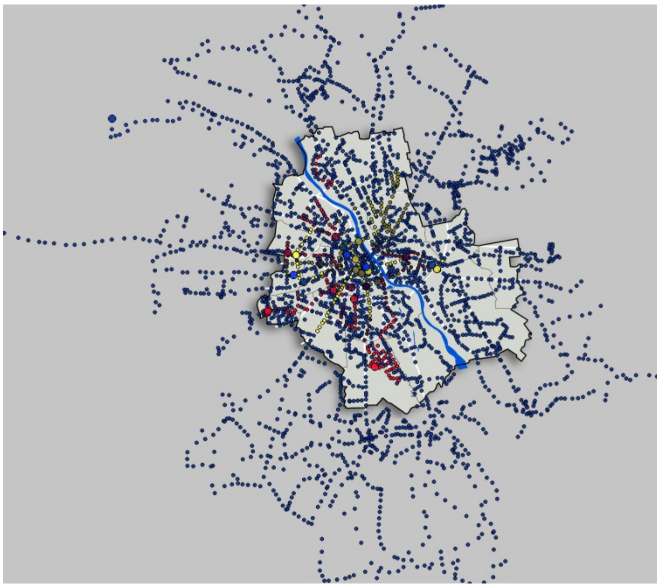


Fig. 3. Warsaw communication system with 25 hubs computed on the basis of capacity of connections for $r = 0.3$ (the state of the timetable as of November 25, 2021).

The method of finding transport hubs in the local transport system presented here is one of many possible ones, more information on other possible methods and results obtained for Warsaw can be found in the works: [13], [18] and [19]. Certainly, the stops indicated as potential hubs by several methods are definitely the best candidates for giving them such a function in reality.

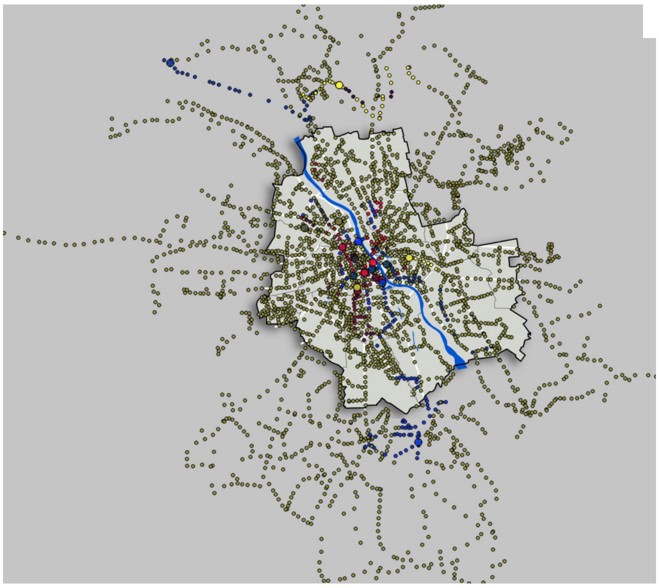


Fig. 4. Warsaw transport system with 20 hubs computed on the basis of time of connections for $r = 0.7$ (the state of the timetable as of November 25, 2021).

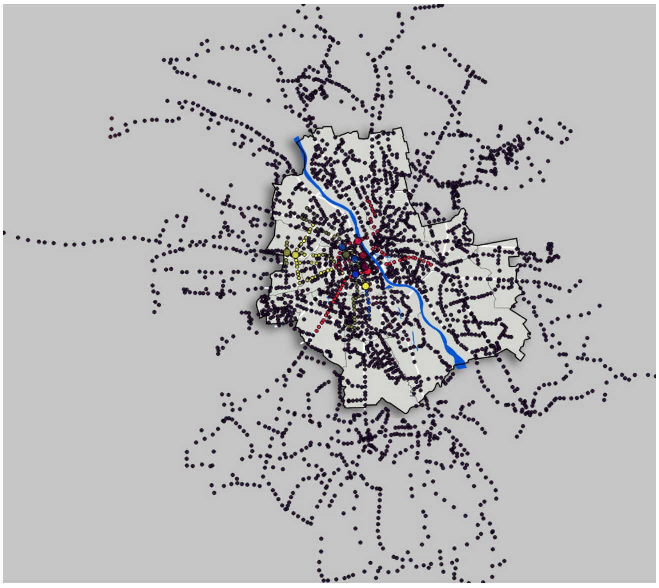


Fig. 5. Warsaw communication system with 12 hubs computed on the basis of capacity of connections for $r = 0.7$ (the state of the timetable as of November 25, 2021).

VI. Conclusions

This work deals with the possibility of applying the well-known k-means clustering algorithm in the problem of graph clustering with the possibility of improving the public transport system by using elements of the hub&spoke idea. The proposed specialized evolutionary method of the communication data processing is quite efficient and can deal with large sets of stops, characterizing big cities. We showed this feature on the example of Warsaw. As a result we obtained several solutions for different values of the threshold parameter r for the evolutionary k-means method for graphs clustering. The calculated transfer points are, of course, indicative proposals and actual transfer hubs may be created in slightly different places due to the influence of many other factors (e.g. existing buildings, land ownership), which the presented algorithm does not take into account, as only data on transport connections have been considered.

References

[1] J. Coyle, E. Bardi, and. R. Novack, "Transportation", Fourth Edition, New York: West Publishing Company, 1994.

[2] J. F. Campbell, and M O'Kelly, Twenty-Five Years of Hub Location Research, Transportation Science, 46(2), 2012, pp. 153-169.

[3] D. Goldberg, "Genetic Algorithms in Search, Optimization and Machine Learning", Addison-Wesley, Massachusetts, USA, 1989.

[4] S. Lloyd, "Least squares quantization in PCM" In: Bell Telephone Labs Memorandum, Murray Hill NJ, reprinted in: IEEE Trans. Information Theory IT-28, Vol. 2, 1982, pp. 129-137

[5] J. MacQueen, "Some methods for classification and analysis of multivariate observations", in: Proc. 5th Berkeley Symp. Math. Statistics and Probability. Vol. 1. 1967. pp. 281-297.

[6] T. Matisziw, Ch. Lee., and T. Grubesic, "An analysis of essential air service structure and performance", Journal of Air Transport Management 18, 1, January 2012, pp. 5–11.

[7] B. Mażbic-Kulma, H. Potrzebowski, J. Stańczak, and K. Sęp, "Evolutionary approach to solve hub-and-spoke problem using α-cliques", Evolutionary Computation and Global Optimization, Prace naukowe PW, Warszawa, 2008, pp. 121-130.

[8] B. Mażbic-Kulma, J. Owsiński, J. Stańczak, A. Barski and K. Sęp, Mathematical Conditions for Profitability of Simple Logistic System Transformation to the Hub and Spoke Structure, in: Atanassov, K., *et al.* Uncertainty and Imprecision in Decision Making and Decision Support: New Challenges, Solutions and Perspectives. IWIFSGN 2018. Advances in Intelligent Systems and Computing, vol. 1081. Springer, Cham., 2021, pp. 398-408.

[9] Z. Michalewicz, Genetic Algorithms + Data Structures = Evolution Programs, Springer Verlag, Berlin Heidelberg, 1996.

[10] Z. Michalewicz, and B. Fogel, How to Solve It: Modern Heuristics, Springer-Verlag, Berlin Heidelberg, 2004.

[11] M. O'Kelly, and D. Bryan, Interfacility interaction in models of hubs and spoke networks, Journal of Regional Science, 42 (1), 2002, pp. 145-165.

[12] M. O'Kelly, A quadratic integer program for the location of interacting hub facilities, European Journal of Operational Research, V. 32, 1987, pp. 392-404.

[13] J. Owsiński, J. Stańczak, A. Barski, and K. Sęp, Identifying main center access hubs in a city using capacity and time criteria. The evolutionary approach, Control and Cybernetics, 45(2), 2016, pp. 207-223.

[14] J.-P. Rodrigue The Geography of Transport Systems, New York: Routledge, 2020

[15] J. Stańczak, Biologically inspired methods for control of evolutionary algorithms, Control and Cybernetics, 32(2), 2003, pp. 411-433.

[16] J. Stańczak, H. Potrzebowski, and K. Sęp, Evolutionary approach to obtain graph covering by densely connected subgraphs, Control and Cybernetics, vol. 41, No. 3, 2011, pp. 80-107.

[17] J. Stańczak, and J. Owsiński, Evolutionary k-Means Clustering Method with Controlled Number of Clusters Applied to Determine the Typology of Polish Municipalities. In: Uncertainty and Imprecision in Decision Making and Decision Support: New Advances, Challenges, and Perspectives. IWIFSGN BOS/SOR 2020. Lecture Notes in Networks and Systems, vol. 338, 33. Springer, Cham, 2022, pp. 436-446.

[18] J. Stańczak, A. Barski, K. Sęp, and J. Owsiński, The problem of distribution of Park-And-Ride car parks in Warsaw, International Journal of Information and Management Sciences, 27(2), 2016, pp. 179-190, http://dx.doi.org/10.6186/1JIMS.2016.27.2.6.

[19] J. Stańczak, K. Sęp, and J. Owsiński, "Evolutionary methods for finding kernel & shell structures in a graph of connections" (in Polish: "Ewolucyjne metody znajdowania struktur typu "kernel & shell" w grafie połączeń"), Instytut Badań Systemowych PAN, Warszawa, 2023.

[20] H. Steinhaus, Sur la division des corps matériels en parties. Bulletin de l'Académie Polonaise des Sciences, Classe 3, 1956, 12, pp. 801-804.

[21] R. Wilson, Introduction to graph theory, Addison Wesley Longman, 1996.

Communication Papers of the 18th Conference on Computer Science and Intelligence Systems pp. 295–302
DOI: 10.15439/2023F3141
ISSN 2300-5963 ACSIS, Vol. 37

Forecasting migration of EU citizens to Germany using Google Trends

Nicholas Steinbrink
0000-0002-6163-0212
Bertelsmann Stiftung
Carl-Bertelsmann-Str. 256, 33335 Gütersloh, Germany
Email: nicholas.steinbrink@bertelsmann-stiftung.de

***Abstract*—The study examines the potential of Google Trends data as an additional data source for forecasting EU migration to Germany. For that aim, candidate search queries with relation to migration intent are proposed. The resulting Google Trends Indices (GTI) are combined with macroeconomic and past migration data and used to build a machine learning regression model. It is shown that GTI predictors can moderately reduce the forecast error and enable a slight expansion of the forecast horizon. However, the presence of outliers emphasizes the need for continuous improvement in data quality to increase the robustness of the approach.**

I. Introduction

MIGRATION policy plays a pivotal role in shaping the labor market policies of OECD countries, offering potential solutions to mitigate labor market rigidities. Notably, in the context of Germany, the labor market has greatly benefited from the free internal mobility of EU nationals, which serves as a crucial source of skilled labor migration. However, the effectiveness of migration policy faces a central challenge arising from the uncertainty surrounding the goals and scale of future migration. This uncertainty is driven by a diverse array of political and socio-economic push and pull factors. Although intra-EU labor mobility is subject to regular monitoring [1], there is currently a lack of substantial efforts towards forecasting, despite the occurrence of significant mobility shifts in the past, such as those witnessed in the aftermath of the Eurozone financial crisis.

Regarding external migration, forecasts primarily rely on three methodologies: (a) extrapolation of past migration patterns using time series methods such as ARIMA models, (b) explanatory econometric models incorporating variables like GDP and unemployment, which are presumed to be linked to migration, or (c) spatial interaction models like gravity models, connecting origins and destinations. [2] These models often integrate expert opinions within a Bayesian framework. Despite their methodological sophistication, these approaches often exhibit considerable forecast uncertainties, leading to potential over- or underestimation of actual migration.

An innovative approach to migration forecasting involves focusing on the individual planning behaviors of individuals who have made the decision to migrate, as opposed to relying solely on macroscopic factors. This can be achieved, for example, by incorporating data from migration intention surveys into the forecasting process [3]. A promising alternative lies in leveraging digital trace data, which has the potential to identify individual migration intentions earlier by capturing active behaviors of individuals seeking information about emigration and migration planning. One suitable data source for this purpose is Google Trends, which measures the temporal and regional search intensity associated with specific keywords in the Google search engine, thanks to Google's high market share.

The aim of this project is to explore the applicability of a Google Trends as a predictor in a novel forecast of migration for EU nationals to Germany. Specifically, the research seeks to determine whether Google Trends data can enhance the accuracy of a forecast method based purely on past migration patterns and macroeconomic variables, particularly within a short- to medium-term timeframe (3 to 12 months).

II. Related work

Although no specific work regarding intra-EU mobility can be found, various attempts have been made to utilize digital trace data for migration forecasts. Data sources include for example Facebook's advertising platform [4], geolocalized IP addresses from e-mails [5], as well as Twitter messages [6], [7]. Moreover, the potential of Google Trends data as a predictive data source for migration has been examined in previous studies. Boehme et al. [8] demonstrated the correlation between search activity related to migration and migration intention, as well as between migration intention and successful migration, establishing the viability of using search engine data for predictive purposes. Carammia et al. [9] have developed an early warning and forecast system based on data from Google Trends, applied to monthly asylum data for EU destination countries. Closely related to that, ongoing research by Boss et al. [10] highlights the particular usefulness of Google Trends data for forecasting bilateral refugee flows at scale across multiple corridors. In contrast, Wanner [11] presented mixed results using a minimal model based on a single Google Trends keyword as predictor for work-related regular migration from EU countries to Switzerland, emphasizing the need for further validation and research.

Thematic track: International Workshop on AI in Digital Humanities, Computational Social Sciences and Economics Research

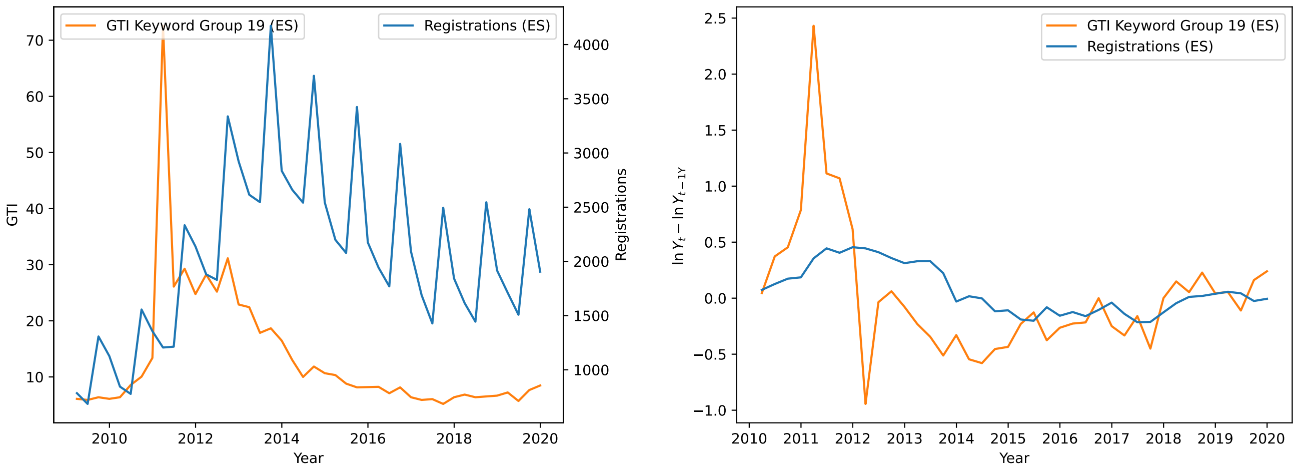


Fig. 1. Left panel: Number of registrations (blue) and GTI for keyword group 19 (related to work and jobs in Germany) for the example of Spain. The peak during 2011 coincides with a period of high unemployment in Spain during the Eurozone financial crisis and the search for jobs in Germany has possibly gone "viral" for a short period of time. Right panel: transformed values of registrations and GTI according to (1).

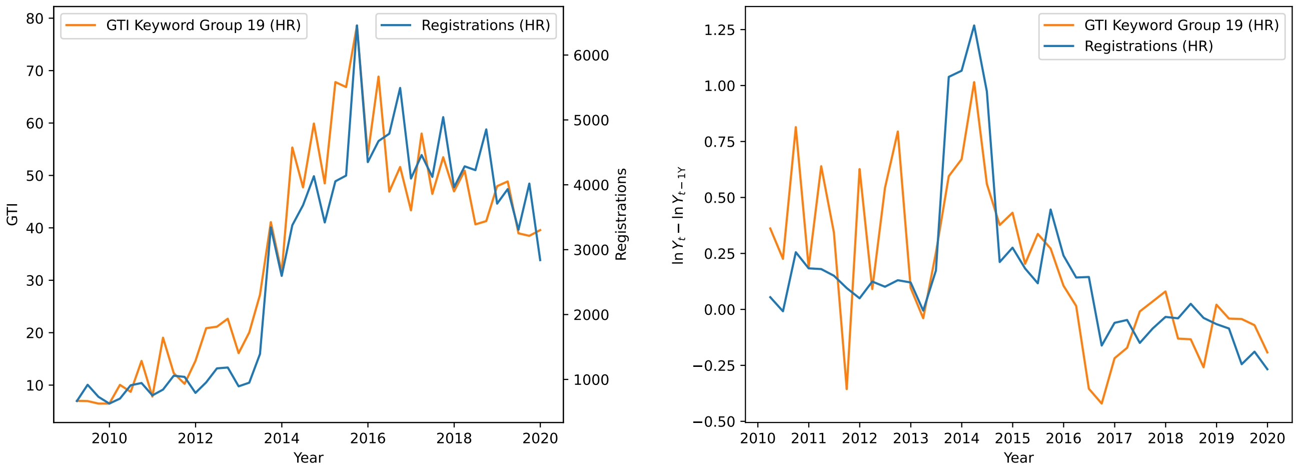


Fig. 2. Left panel: Number of registrations (blue) and GTI for keyword group 19 (related to work and jobs in Germany) for the example of Croatia. Registrations increase continuously after the admission of Croatia to the EU in July 2013 and the full access to free movement to Germany in July 2015, with the GTI following the trend. Right panel: transformed values of registrations and GTI according to (1).

III. Methods

A. Prediction target

Since EU nationals are required to register within 3 months after their arrival in Germany, the official number of registrations by country of origin, which is provided by the federal office of statistics (DESTATIS), is taken as a proxy of the number of arrivals. The number of registrations exhibits a clear seasonal pattern with a peak during the summer months and a drop during the winter months consistently for all EU origin countries (for an example, see figs. 1 and 2, left panels, blue line). A naive forecast could in most situations produce reasonable results by taking the number of registrations from the same period of the previous year as prediction. Therefore, the interesting quantity to predict is not the absolute number of registrations, but the change compared the previous year. The target variable is then stipulated as:

$$Y_{t,c} = \ln R_{t,c} - \ln R_{t-1\mathrm{y},c}, \qquad (1)$$

where $R_{t,c}$ is the number of registrations at time t of nationals with country of origin c and $R_{t-1\mathrm{y},c}$ the corresponding number lagged by one year. The log transformation ensures that all countries are given equal weight regardless of the absolute number of registrations if the forecast accuracy is determined by conventional metrics. For an example, see figs. 1 and 2 (right panels, blue line).

B. Countries of origin

The selected countries of origin encompass the 28 member countries of the EU prior to the Brexit, except Malta, Cyprus and Germany, which is the destination country. Additionally, Switzerland has been added due to the shared border with Germany and the membership in the Schengen area. Countries with only small number of registrations have been grouped according to similar migration behavior, similar size and regional proximity. These groups comprise: Austria and Switzerland; Belgium, the Netherlands and Luxembourg; the Czech Republic and Slovakia; Lithuania, Latvia and Estonia; Sweden, Finland and Denmark.

C. Google Trends index

The Google Trends index (GTI) measures the relative search interest, given a certain keyword, over a specified amount of time. The index is based on a sample of total search queries, and is normalized in such a way that the maximum value for a given index time series is set to 100. If the sample becomes too small, no data is returned by Google, which results in a flat zero response when using the API. No detailed information is specified by Google regarding the sampling process or the cutoff.

The keywords of interest have been taken from [8] and been professionally translated to the languages of the countries of origin. A crucial challenge in keyword selection is to achieve an optimal balance between specificity and generality. The keywords must be specific enough to avoid any confusion with search activity unrelated to migration. However, they should not be excessively specific, as this is essential to ensure sufficient data quality and comprehensive coverage of relevant search activity. Especially for smaller countries, the Google Trends API requests often resulted in either a flat zero response when a keyword lacked sufficient search frequency, or an overly noisy and unusable response. To balance both aims, the Google Trends queries have been enriched in the following way:

- Semantically related keywords have been grouped.
- Each keyword is considered in the language(s) of the country of origin, as well as German and English.
- To every keyword, the postfix "Germany" is added in the corresponding language.
- Multiple spellings (for instance with and without accents) have been considered.
- The keywords in one semantic group have been connected with a logical "or" (+).

An exemplary query is given in appendix V-B, while an example for a resulting GTI time series can be found in fig. 1 (left panel, orange line). Additionally, a range of keywords without the postfix "Germany" has been included as well to consider push-effects. In total, GTI time series of 48 keyword group candidates have been generated that way (see appendix V-A, table II for a complete list). To mitigate statistical variance, each query has been performed ten times and averaged.[1]

D. Data preparation and modeling

All time series have been discretized by 3-month intervals.[2] Data have been taken from 2010 to 2019. The lower limit has been set due to insufficient data quality of Google Trends prior to 2010 while the upper limit was set to avoid effects of the COVID-19 pandemic.[3] In addition to Google Trends, GDP per capita and unemployment of the countries of origins have been selected as explanatory variables. Google Trends indices, GDP and unemployment have as well been transformed according to (1) (for an example of a transformed GTI, see fig. 1, right panel, orange line). The full set of features is then given by lagged values of Google Trends indices, GDP, unemployment and autoregressive lags of the numbers of registrations themselves. Only for Google Trends a minimum lag of 3 months has been used, while for the other variables the minimum lag was 6 months due to the publication delay, which forbids smaller lags in a forecast situation.

An array of both linear and ensemble-based machine learning models has been tested with different feature-sets.[4] The choice of models is guided by similar motivations as in [10]: no prior theoretical knowledge is utilized, and the algorithms are suitable for a combined model across a multitude of origin countries with a relatively large number of features. To reduce dimensionality and mitigate multicollinearity, a feature selection step has been added, which is described in section III-E.

For each configuration, a mean out-of-sample R^2, given by $R^2_{OOS} = 1 - \sum(y_i - \hat{y}_i)^2 / \sum y_i^2$ [12], and MAE have been determined via n-fold cross-validation (CV). Each year corresponds to a CV fold, while the year 2019 has been additionally set apart as hold-out set for a sanity-check of the CV results. While in principle future information is used as training data in this CV scheme, it has been shown [13] that such a method is valid as long as residuals are uncorrelated, which is typically only the case in severe underfitting. The advantage, on the other hand, is a maximum use of the available training data and comparability across folds in contrast to time-series CV methods.

As there are no known comparable forecasts for intra-EU mobility, a range of benchmarks has been chosen to assess the forecast accuracy. The const(0) benchmark is the simplest baseline, setting the target variable constantly zero, $\hat{Y}_{t,c} = 0$, corresponding to no annual change in the registration rate. A model performing below the const(0) benchmark corresponds to $R^2_{OOS} < 0$.

[1] To force Google Trends to draw a new sample for each request, a random, sufficiently long character string can be added to each query.

[2] A finer discretization has been tested as well, but did not lead to any significant improvement of the forecast accuracy in the modeling stage.

[3] While the German border has been officially closed only for a few months, it is reasonable to assume that individual mobility has been reduced for a longer period of time during the pandemic. As a stable relationship between online search activity and registrations is a necessary assumption of the methodology, the cutoff has been set to 2020 for this principle study.

[4] The tested ensemble models include: Random Forest, XGBoost and AdaBoost. The tested linear models include: OLS, ElasticNet, Bayesian Ridge Regression, Automatic Relevance Determination, Huber Regression and Theil-Sen Regression.

The benchmark previous(1) corresponds to a random walk, where the previous lag of the target variable is set as predictor, $\hat{Y}_{t,c} = Y_{t-1,c}$. This is not a realistic scenario due to the aforementioned publication delay of registration data. Therefore, the benchmark previous(2), $\hat{Y}_{t,c} = Y_{t-2,c}$, provides a realistic benchmark based on previous lags of the target variable.

Eventually, a non-naive realistic benchmark is given by a comparison of the best models with and without GTI variables, which accounts for the benefit of adding Google Trends data itself.

E. Feature selection

In total, a maximum number of features of roughly $P \sim 51 \cdot L$, where L denotes the number of lags used for the prediction, are available. The maximum lag has been set to $L = 8$ to account for possible correlations between search activity and immigration for up to 24 months. In effect, we end up with a relatively large number of features compared to the number of data points of $N = 576$. While some models with internal variable selection mechanisms are in principle able to handle large-P-small-N-problems, a reduction of the feature dimensionality is nonetheless advisable to maximize model performance and increase interpretability.

In a pre-selection step, the complete set of features is filtered to ensure that only features which are sufficiently correlated with the target are included. To that aim, the p value of the Spearman lag-correlation of all variables with the target is determined. Only those variables with a minimum p value below a cut-off given by a conservative Bonferroni correction, $p < 0.05/P$, are kept. This conservative limit both minimizes the likelihood of spurious correlation and maximizes the robustness of the subsequent feature selection step. The remaining 10 GTI variables include both pull- and push-related queries and are indicated in table II.

Finally, the selection of the input features for the regression model has been performed separately for the linear and ensemble models, respectively, as well as for feature configurations both including and excluding the GTI and autoregressive lags of the prediction target. To estimate the optimum feature sets for the linear models, a Sequential Forward Floating Selection (SFFS) has been performed with an ordinary least squares linear regression model, which has shown to be an efficient search technique [14]. The selected features have been checked for multicollinearity using the variable inflation factor (VIF). It could be shown that by the selection procedure multicollinearity could be reduced to a moderate level of VIF < 10. Since SFFS is a rather costly greedy method, for the ensemble models Recursive Feature Elimination (RFE) with a tuned Random Forest regressor has been chosen. For both methods, the CV scheme as outlined in section III-D with MAE as optimization metric has been used to determine the optimum number of features. For all feature configurations, a stable optimum could be found. An illustration of the variables and the selection procedure is shown in fig. 3.

IV. Results

A. Model performance

Table I shows a performance comparison of the best optimized models of each class (linear and ensemble) after hyperparameter tuning with different feature configurations (all features, without GTI, without autoregressive lags of the target variable), compared to naive benchmarks, as outlined in the previous section. For the ensemble models, a Random Forest has been found to perform best, while for the linear models, a linear regression with Huber loss has shown the optimum performance.

If the cross validation results of the ML models are compared to the benchmarks, it can be observed that most models to surpass the const(0) and the previous(2) benchmark, but only the linear model with all features is on par with the previous(1) benchmark. Since the latter is not realistic due to the publication delay of registration data, this observation implies that a real-time monitoring of registrations alone would be sufficient to provide a competitive short-term forecast.

If the models are compared among themselves, it can be seen that the linear model has a clear advantage over the Random Forest model, which is prone to overfitting due to the small number of training examples. For both model classes, it can be observed that the models with GTI exhibit slight but clear reduction of the MAE by up to 10 %, compared to the models without GTI. A mean R^2_{OOS} of up to 0.54 can be achieved. The models without autoregressive lags provide, while not being competitive, still a reasonable forecast accuracy on average and could in principle be used as an alternative if past registration data are not available.

Compared to the mean CV scores, the MAE for the 2019 holdout set are smaller. However, only for the linear models with included autoregressive lags, R^2 is noticeably above zero. A plausible explanation is that the trend in registrations (1) is largely flat in 2019, suggesting that Google Trends data provide additional predictive power only if there are shifts in registrations which can not be predicted by other variables.

B. Comparison by country of origin

Fig. 4 (left panel) shows strong heterogeneity regarding the model performance for the linear model by country of origin. It can be observed that the model produces largely reasonable forecasts in terms of R^2 especially for Southern European and some Eastern European countries. Coincidentally, most EU citizens moving to Germany are native to these regions and registrations from these countries have been subject to greater variability over the last decade. However, even for countries which perform well on average, some outliers are present, corresponding to periods for which the forecast performs poorly. If the forecast errors of the best model with and without GTI, respectively, are compared (right panel), it can be seen that for some countries the model benefits more clearly from the GTI predictors. These are especially Spain (see example, fig. 5, left panel), Portugal, Greece and Italy, which have been particularly hit by the Eurozone financial crisis of the early

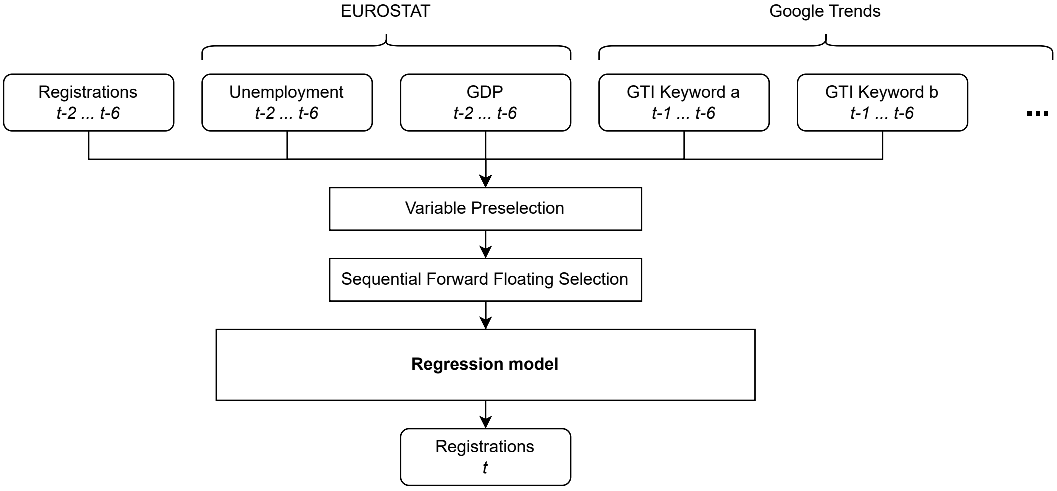


Fig. 3. Illustration of the input features and selection procedure.

TABLE I
COMPARISON OF MODEL PERFORMANCE METRICS

Model	Feature Sets	Cross Validation		Holdout	
		MAE	R^2_{OOS}	MAE	R^2_{OOS}
Benchmark cost(0)	-	0.127 (0.025)	0.00 (0.00)	0.070	0.00
Benchmark previous(1)	-	0.070 (0.005)	0.51 (0.10)	0.061	0.22
Benchmark previous(2)	-	0.092 (0.007)	0.23 (0.11)	0.068	0.07
Random Forest	all	0.082 (0.025)	0.44 (0.07)	0.068	0.07
Random Forest	without autoregression	0.092 (0.011)	0.23 (0.17)	0.072	0.05
Random Forest	without GTI	0.089 (0.010)	0.38 (0.07)	0.068	-0.05
Linear Regression	all	0.071 (0.007)	0.54 (0.08)	0.063	0.30
Linear Regression	without autoregression	0.088 (0.010)	0.34 (0.11)	0.062	0.30
Linear Regression	without GTI	0.078 (0.009)	0.47 (0.09)	0.070	-0.00

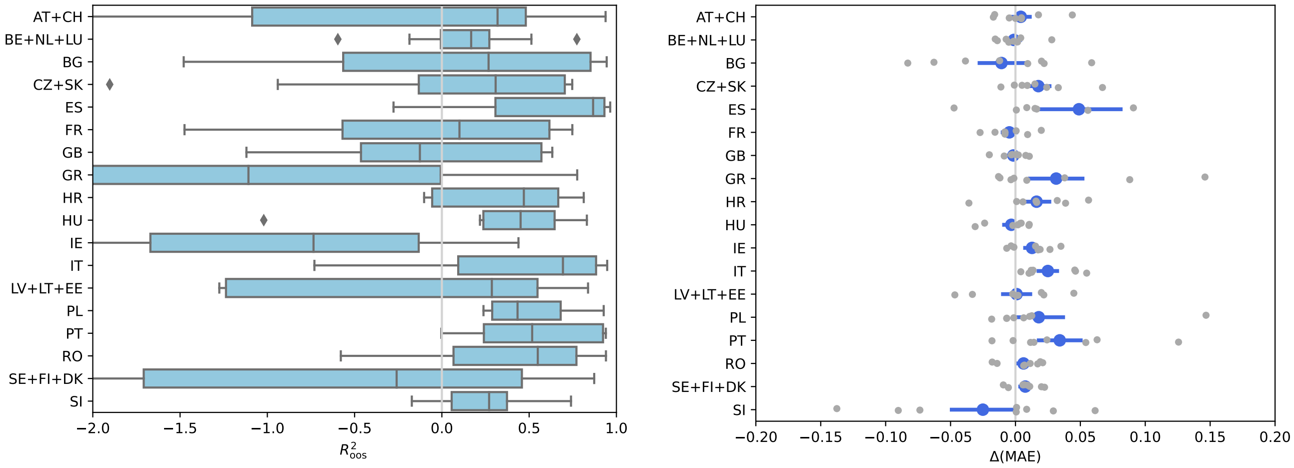


Fig. 4. Left panel: Distribution of cross validation out-of-sample R^2 by country of origin for linear model with autoregression and GTI. Right panel: Cross validation distribution of difference of mean absolute error between linear autoregressive model with and without GTI by country of origin. Positive values indicate lower prediction errors with GTI. Blue circles correspond to mean with standard errors.

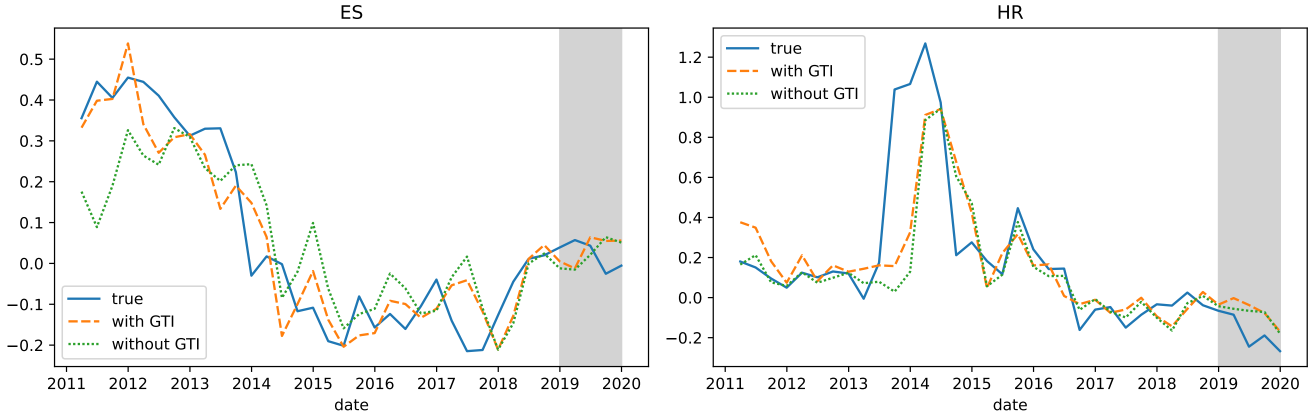


Fig. 5. Transformed number of registrations (1) (blue solid line), as well as prediction for the best linear model with GTI (orange dashed line) and without GTI (green dotted line) for Spain (left panel) and Croatia (right panel). The prediction is composed of the individual test sets of the cross validation folds. The shaded area corresponds to the holdout set of 2019.

2010s. In addition, a very subtle performance improvement can be observed for Poland, Croatia (see example, fig. 5, right panel), the Czech Republic, Slovakia and Ireland. For other countries, the benefit is negligible or even negative, such as in case of Bulgaria and Slovenia.

The existence of some forecast errors can partly be attributed to the data quality of the GTI predictors. In general, the lag correlation between GTI predictors and target is not stable in time, which can be observed in the examples in figs. 1 and 2. Especially for smaller countries, the GTI predictors can be noisy, which becomes even more pronounced in terms of relative changes, as after transformation (1). Moreover, while noise and outliers can be accommodated by using a diverse array of search queries, many of the corresponding GTI variables are unusable or missing (zero) for smaller countries. In the example of Croatia (fig. 5, right panel) this causes the forecast, for instance, to predict the sharp increase of registrations during 2013 and 2014 too late and to erroneously predict an increase in early 2011.

C. Forecast Horizons

Fig. 6 compares the best linear model with and without GTI, respectively, for different forecast horizons. Due to the machine learning setup, the results have been simulated by shifting the lags of all features $n-1$ periods to the past, with n denoting the number of forecast periods ahead, except for those features which would still be available at $t = t_0 - n$. For all forecast horizons from 3 to 12 months, the model with GTI predictors consistently outperforms the model without GTI moderately. Whereas for $n = 3$ and $n = 4$ the performance of the model without GTI becomes nearly indistinguishable from that of the const(0) baseline, the model with GTI exhibits at least some predictive power.

V. Conclusion

Google Trends can in principle be seen as a viable additional data source for a forecast of EU migration to Germany, especially as long as a real-time monitoring of registrations is not available yet. For an existing forecast based on past registrations and macroeconomic variables, the addition of Google Trends data can on average reduce the forecast error moderately and enable some expansion of the forecast horizon.

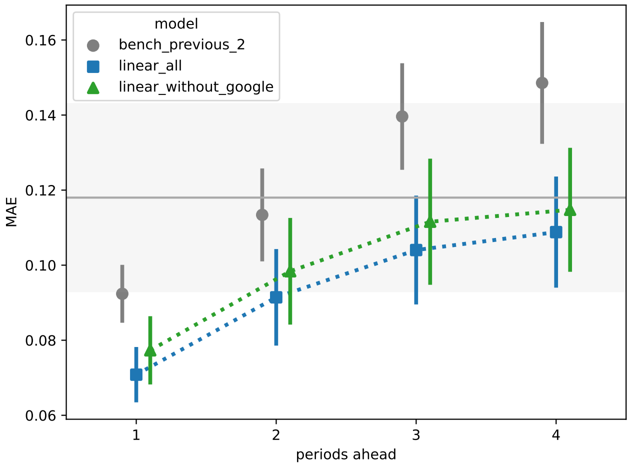


Fig. 6. Mean absolute error of the best model with GTI (blue) and without GTI (green), respectively, for forecast horizons of 1 to 4 periods corresponding to 3 to 12 months. The horizontal line represents the mean score for the const(0) benchmark with the shared area representing the standard error, while the grey circles represent the mean score for the previous(2) benchmark. All error bars are standard errors.

The benefit of Google Trends data is especially given for scenarios and origin countries with greater shifts in migration behavior, where these can not always reliably predicted by other variables. For countries where the seasonality-adjusted migration to Germany is largely stationary, no improvement is gained by Google Trends, as naive forecasts are sufficient in these cases. For a small set of origin countries, however, the Google Trends based forecast performs weakly. Even for bigger countries the forecast occasionally produces outliers. The lack of robustness can be partly attributed to insufficient data quality of the GTI predictors. It is unclear if that is a purely statistical effect. Even if multiple samples are drawn from Google Trends and averaged, the resulting time series are still noisy in many cases. Unfortunately, details about the sampling procedure are not made public by Google. As long as data quality is still an issue, a possible workaround might be a more sophisticated modeling of the relationship between GTI variables and migration intent using Bayesian inference techniques or microsimulations.

Nonetheless, it is reasonable to assume that role of Google Trends in migration forecasts will become more prominent in the future, given that Google claims to continuously improve the data quality, that usage of the Google search engine grows in some countries and that simply more data will become available.

Appendix

A. List of keyword groups

The complete list of candidate keywords groups can be found in table II.

B. Examplary query

Below, the full Google Trends search query corresponding to keyword group 19 for Spain is given as an example.

> contrato de trabajo alemania + contrato laboral alemania + contrato de empleo alemania + trabajo alemania + empleo alemania + ocupación alemania + ocupacion alemania + trabajar alemania + empleo alemania + empleos alemania + trabajo alemania + trabajos alemania + arbeitsvertrag deutschland + arbeit deutschland + arbeiten deutschland + job deutschland + jobs deutschland + work contract germany + employment germany + working germany + job germany + jobs germany

C. Source code

The source code for the study, including the data and queries, can be found online at https://github.com/bertelsmannstift/eu-migration-forecast.

Acknowledgment

This study has been part of the project "Migration Forecast EU" of the Bertelsmann Stiftung. Special thanks goes to Orkan Kösemen, who initiated and managed the project. I would like to thank Ingmar Weber, André Gröger, Tobias Heidland, Johannes Müller and Leonard Mandtler for important and valuable input to the project, as well as Carsten Große Starmann, Ulrich Kober and Jörg Dräger for making the project possible.

References

[1] European Commission (2022). "Annual report on intra-EU labour mobility." https://ec.europa.eu/social/BlobServlet?docId=26778&langId=en.

[2] Sohst, R., Tjaden, J., de Valk, H., & Melde, S. (2020). "The future of migration to Europe: A systematic review of the literature on migration scenarios and forecasts." International Organization for Migration, Geneva, and the Netherlands Interdisciplinary Demographic Institute, the Hague, https://publications.iom.int/system/files/pdf/the-future-of-migration-to-europe.pdf.

[3] Tjaden, J., Auer, D., & Laczko, F. (2018). "Linking migration intentions with flows: Evidence and potential use." *International Migration*, 57(1), 36–57, https://doi.org/10.1111/imig.12502.

[4] Zagheni, E., Weber, I., & Gummadi, K. (2017). "Leveraging Facebook's Advertising Platform to Monitor Stocks of Migrants." *Population and Development Review*, 43(4), 721–734, https://doi.org/10.1111/padr.12102.

[5] Zagheni, E., & Weber, I. (2012). "You are where you e-mail: using e-mail data to estimate international migration rates" *Proceedings of the 4th Annual ACM Web Science Conference*, 348–351, https://doi.org/10.1145/2380718.2380764.

[6] Zagheni, E., Garimella, V.K.R., Weber, I., & State, B. (2014). "Inferring international and internal migration patterns from Twitter data" *Proceedings of the 23rd International Conference on World Wide Web*, 439–444, https://doi.org/10.1145/2567948.2576930.

[7] Hawelka, B. et al (2014). "Geo-located Twitter as proxy for global mobility patterns" *Cartography and Geographic Information Science*, 41(3), 260–271, https://doi.org/10.1080/15230406.2014.890072.

[8] Böhme, M.H., Gröger, A., & Stöhr, T. (2020). "Searching for a better life: Predicting international migration with online search keywords." Journal of Development Economics, 142, 102347, https://doi.org/10.1016/j.jdeveco.2019.04.002.

[9] Carammia, M., Iacus, S.M., & Wilkin, T. (2022). Forecasting asylum-related migration flows with machine learning and data at scale. *Sci Rep* 12, 1457. https://doi.org/10.1038/s41598-022-05241-8.

[10] Boss, K., Gröger, A., Heidland, T., Krüger, F., & Zheng, C. (2023). "Forecasting Bilateral Refugee Flows with High-dimensional Data and Machine Learning Techniques." BSE Working Paper 1387, https://www.itflows.eu/wp-content/uploads/2023/03/1387.pdf.

[11] Wanner, P. (2021). "How well can we estimate immigration trends using Google data?" *Qual Quant* 55, 1181–1202, https://doi.org/10.1007/s11135-020-01047-w.

[12] S. Hawinkel, W. Waegeman, & S. Maere (2023). "Out-of-Sample R2: Estimation and Inference." *The American Statistician*, https://doi.org/10.1080/00031305.2023.2216252.

[13] Bergmeir, C., Hyndman, R.J., & Koo, B. (2018). "A note on the validity of cross-validation for evaluating autoregressive time series prediction." *Computational Statistics & Data Analysis*, 120, 70–83, https://doi.org/10.1016/j.csda.2017.11.003.

[14] Ferri, F. J., Pudil P., Hatef, M., $ Kittler, J. (1994). "Comparative study of techniques for large-scale feature selection." *Machine Intelligence and Pattern Recognition*, 16, 403–413, https://doi.org/10.1016/B978-0-444-81892-8.50040-7.

TABLE II
LIST OF CANDIDATE KEYWORDS, BASED ON [8]

ID	Keywords	With postfix "Germany"	Included in forecast
2	passport, passport office	yes	
10	immigrant, emigrant, immigrate, emigrate, immigration, emigration	yes	
11	visa, entry requirements, required documents	yes	
12	minimum wage	yes	
14	pension	yes	
15	unemployment	yes	
16	internship	yes	
17	inflation, living expenses	yes	
18	social benefits, unemployment benefits	yes	
19	work contract, employment, working, job, jobs	yes	x
20	employment agency, employer, hiring, recruitment	yes	
21	income, tax	yes	
22	GDP, prosperity	yes	
24	wage, salary	yes	x
26	economy, German economy	partially	
28	vacancies, job offers	yes	x
32	job application, application letter, job interview, resume	yes	
33	insurance premium, health insurance, social insurance	yes	
37	university qualification, university	yes	
38	credentials, diploma, certificate	yes	
39	language school, German language school, Goethe Institut	partially	x
41	language test, German language test, German certificate	partially	
42	studies, study, Bachelor, Master, phd	yes	
44	trainee, vocational training, apprenticeship, German apprenticeship	partially	
48	bank account, account	yes	
49	apartment, flat, room	yes	
51	spouse, marry, marriage, intermarriage	yes	
52	rent, utilities, rent deposit	yes	
54	move, moving, relocation	yes	
55	Germany	no	
56	customs	yes	
57	business	yes	
58	migrant, migration, foreigner	yes	
59	nationality, citizenship	yes	
60	arrival, tourist, visit	yes	
112	minimum wage	no	
113	welfare	no	
114	pension	no	
115	unemployment	no	x
117	inflation, living expenses	no	x
118	social benefits, unemployment benefits	on	x
119	work contract, employment, working	no	x
121	income, tax	no	
122	GDP, prosperity	no	
123	job, jobs	no	x
124	wage, salary	no	x
125	gross net, allowances	no	

Communication Papers of the 18th Conference on Computer Science and Intelligence Systems pp. 303–311
DOI: 10.15439/2023F7698
ISSN 2300-5963 ACSIS, Vol. 37

The Grammar and Syntax Based Corpus Analysis Tool for the Ukrainian Language

Daria Stetsenko
0000-0002-3698-4414
NASK National Research Institute
Kolska 12, 01-045 Warsaw, Poland
Email: {daria.stetsenko@nask.pl}

Inez Okulska
0000-0002-1452-9840
NASK National Research Institute
Kolska 12, 01-045 Warsaw, Poland
Email:{inez.okulska@nask.pl}

□

***Abstract*—This paper provides an overview of a corpus analysis tool - the StyloMetrix for the Ukrainian language. The StyloMetrix incorporates 104 metrics that cover grammatical, stylistic, and syntactic patterns.**

The idea of constructing the statistical evaluation of syntactic and grammar features is straightforward and familiar for the languages like English, Spanish, German, and others; it is yet to be developed for low-resource languages like Ukrainian. We describe the StyloMetrix pipeline and provide some experiments with this tool for the text classification task. We also describe our package's main limitations and the metrics' evaluation procedure.

***Index Terms*—stylometric analysis, Ukrainian linguistics metrics, text analysis, supervised learning.**

I. Introduction

Ukrainian remains one of the low-resource languages with few practical applications in machine learning and deep learning. Many studies on the Ukrainian language are conducted in terms of multilingual settings, such as training the multilingual large language models [14], [18], transformers [23], [6], or abstractive summarization [10]. We offer a corpus analysis tool for the Ukrainian language – the StyloMetrix. The underlying idea is not new in the NLP community but is recent in the Ukrainian language.

This paper provides an overview of an open-source Python package – the StyloMetrix developed initially for the Polish language and further extended for English and recently for Ukrainian. The StyloMetrix is built upon a range of metrics crafted manually by computational linguists and researchers from literary studies to analyze stylometric features of texts from different genres. The principal purport of this package is to provide high-quality statistical evaluations of the general grammatical, lexical, and syntactic features of the text, regardless of its length, genre, or author.

We organize our paper in the following way:

- we provide an overview of similar tools for text analysis and a general idea of the corpus linguistics based on the syntactic and grammar representations;
- give an exhaustive characteristic of existing metrics for the Ukrainian language, their evaluation, and limitations;
- describe a case study with the StyloMetrix as the baseline model for the text classification task, providing the metrics analysis and feature importance of the classification model.

II. Related Studies

The idea to measure specific textual features to determine a text's register or an author is not new. In 1998, D. Biber, S. Conrad and R. Reppen have developed a comprehensive methodological approach for corpus analysis based only on grammatical characteristics. D. Biber argues that, although, semantic evaluations and descriptive analysis can provide a valuable insight about the narrative, it is not enough if one needs to discern the genre of the text or to assess whether it belongs to a particular author and an epoch [4]. On the other hand, grammatical/syntactic characteristics and figures of speech may come in handy and be less decisive and more exhaustive when it comes to genre, author or style estimation. M.A.K. Halliday supports this view and emphasizes the general importance of corpus studies as a source of insight into the nature of language. He points out that *a language is inherently probabilistic and we need to extract the frequencies in the texts to establish probabilities in the grammatical system – not for the purpose of tagging and parsing, but to discover the interactions between different subsystems* [2].

The development of corpus-based grammar and syntactic tools for text mining has started in 1990s and is still an ongoing field of investigation. Some of the corpus-based techniques aim to manually study the English grammar and discourse. For instance, [1] and [16] provide introductions on how to identify and extract syntactic and grammatical constructions in corpora to build tagging and parsing algorithms. They cover various aspects, limitations and boundaries related to grammar and syntax. Other researchers concentrate on specific incarnations of the language use. For example, [29] on negation and lexical diffusion in syntactic change; [8] on prosody and pragmatics based on it-clefts and wh-clefts; [12] on automated retrieval of passives; [16] on infinitival complement clauses; and [7] has conducted the most valuable

https://github.com/ZILiAT-NASK/StyloMetrix

Thematic track: International Workshop on AI in Digital Humanities, Computational Social Sciences and Economics Research

study on generative grammar that has served as a scaffold for contemporary natural language processing. Those techniques are the basis of modern tools and web-based services for text analysis.

We follow the assumption that grammar and syntax can be enough for the tasks connected with style and author classification which are unified under the term stylometry [20].

The most popular applications for the stylometric analysis are the "Stylometry with R" (stylo) [9] (for English and Polish), WebSty [15] and CohMetrix [11]. The stylo is a flexible R package for the high-level analysis of writing style in stylometry. The package can be applied at the supervised learning for the text classification [9]. WebSty [15] is an accessible open-sourced library that encompasses grammatical, lexical, and thematic parameters which can be manually selected by the user. The tool covers the Polish, English, German and Hungarian languages. Coh-Metrix is a web-based platform that offers a wider range of descriptive statistic measurements. For example, low-level metrics counting pronouns per sentence, Text Easability Principal Component Scores, Referential Cohesion, LSA, Lexical Diversity, Connectives, Situation Model, Syntactic Complexity, Syntactic Pattern Density, Word Information, Readability, etc. [17]. The documented versions of Coh-Metrix exist for Spanish [24], Portuguese [25], and Chinese [22] (however, they are developed independently and not supported by the initial authors).

There are many tools for corpus analysis that look at concordances, n-grams, co-locations, key words and numerous frequency analysis which can be applied for the stylometric classification tasks, but most of them are quite primitive and basic with respect to the intricacy of grammar structures like tenses or syntactic phrases (the comprehensive list of tools can be accessed via the link `https://corpus-analysis.com/`

Therefore, inspired by the powerful image of grammatical patterns and syntactic clauses we build the first (to our knowledge) corpus-analysis tool for the Ukrainian language that presents a thorough statistical evaluation of the Ukrainian grammar, syntactic patterns, and some descriptive lexical assessment.

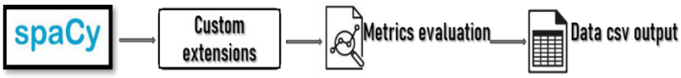


Fig. I The pipeline of the StyloMetrix.

III. Grammatical Vector Representations

A. General outline

The general pipeline of the tool is presented in the Fig. I. First, we utilize the standard spaCy pipeline of the transformer model for the Ukrainian language. The primary purpose of our package is not to build new tagger or parser algorithms but to add a higher level of grammatical and syntactic language characteristics and provide descriptive statistical measurements for each of them. Ukrainian is a fusional language, and the basic spaCy pipeline `https://spacy.io/models/uk` can trace only primary morphological features such as animacy, gender, case, number, aspect, degree, name type, verb form, and others. Nonetheless, these attributes do not cover all aspects of Ukrainian morphology and grammar, such as two types of conjugation, four types of declension, and present, past, and future tenses. Therefore, we leverage the last spaCy component of the pipeline and create custom extensions for each case. Further, the tokens that fall under specific rules are calculated by the discussed formula at the stage of the Metric evaluation and are stored in the data frame that is further available for a user in the .csv format. As for the input – the StyloMetrix can be applied to any text length starting from a single sentence.

The StyloMetrix is a tool designed to calculate the mean value of a distinct grammar rule, a lexical component or a stylistic phenomenon. The statistical measurement is derived by the standard formula:

$$\frac{\sum_0^n w}{\mathrm{N}}$$

Where $\sum_0^n w$ is the sum of all tokens that fall under the particular rule, and N – is the total amount of tokens in the text. This evaluation holds for all metrics. Hence the output is acquired as a matrix, where the text instances are at the y-axis and the x-axis is the vectors of real numbers that stand for the specific metric. The obtained matrices can be utilized for various machine-learning tasks.

Primary developed for the Polish language, which is also fusional, the package has been used for stylometric analysis and text classification. For example, [21] presents a study on erotic vs. neutral text classification using the StyloMetrix vectors as the input to the RandomForest Classifier. The general accuracy has yielded around 0.90 score, giving us an impetus to deliver the primary metrics for the Ukrainian language and test their performance on the existing annotated datasets.

B. Spacy Limitations

Before developing the rules for custom extensions and metrics, we verified the spaCy tags' correctness. Table I presents the incongruencies which have been discerned. Among the most frequent mis assignments are morphological features such as case, animacy, aspect, and gender. For example, "Ілля" is a typical male Ukrainian name that is

TABLE I.
SPACY TAGS INCONGRUENCIES.

word	spacy	Correct tag	Sentence
закрапало	Aspect=Imp	Aspect=Perf	Із стріх закрапало, а з гір струмочки покотилися.
веснянки	ADV	NOUN, Plural	Вже веснянки заспівали.
замазалося	Aspect=Imp	Aspect=Perf	Високе небо замазалося зеленобурими хмарами, припало до землі, наче нагнітило на неї.
крук	Animacy=Inan	Animacy=Animate	Тільки чорний крук надувся, жалібно закрякав з високої могили серед пустельного поля.
завдання	Case=Acc	Case-Nom	Завдання буде зроблено.
листа	Animacy=Anim	Animacy=Inan	Я напушу листа.
осінню	ADJ Case=Acc	NOUN Case=Ins	Повіває молодою осінню холодна річка з низів.
низів	Case=Gen	Case=Loc	Повіває молодою осінню холодна річка з низів.
закрапало	Aspect=Imp	Aspect=Perf	Із стріх закрапало, а з гір струмочки покотилися.

tagged as feminine by the spaCy parser. Other inconsistencies are found in the part-of-speech annotation.

We intentionally highlight this part as it directly influences the quality of our metrics. Due to the probability of tags' incorrectness, some tokens can be missing from the set; therefore, the final evaluation of the tool may be less precise. At the lexical level, we try to avoid this drawback by checking some explicit morphological characteristics through affixes or the position of a token in the sentence based on a dependency tree. The dependency tags have proven to be the most precise and robust. Hence we tend to rely on them more while implementing grammar and syntactic rules.

C. Metrics assessment

The Ukrainian version of the StyloMetrix incorporates 104 metrics subdivided into lexical forms, parts-of-speech incidence, and syntactic and grammatical structures. The complete list of metrics can be found in Appendix A. In this subsection, we strive to provide general descriptive characteristics and validation criteria for each group.

Table II describes the number of metrics per category. With the StyloMetrix, academics can extract both conventional statistics of the text and features intrinsic to the Ukrainian language. For instance, the universal metrics are the type-token ratio, functional and content words, punctuation, and parts-of-speech statistics. A few examples are presented below.

TABLE III.
TOTAL NUMBER OF METRICS PER GROUP.

Group	Number
Lexical	56 metrics
Grammar	23 metrics
Syntax	14 metrics
Part-Of-Speech	12 metrics

- **L_ADV_POS:** [потрібно, відверто] – positive adverbs [needed, sincerely]
- **L_ANIM_NOUN:** [Президент, агресор, людей] – animated nouns [President, aggressor, people]
- **L_DIRECT_OBJ:** [час, нам, армію, потенціал, альтернативи, режим] – direct object [time, us, army, potential, alternatives, regime]
- **L_INDIRECT_OBJ:** [світом, року, конференції, Україні, режимом] – indirect object (in Ukrainian denoted by case; in English translation we add prepositions) [(by) world, (during) a year, (at) a conference, (to) Ukraine, (in) regime]

Albeit the commonness of these measurements, it has been demonstrated by many researchers, e.g., the Coh-Metrix study, that these scores may provide valuable insight into the idiosyncratic characteristics of a text.

The forms prominent in the Ukrainian language belong to syntactic constructions such as parataxis, ellipses, and positioning (прикладка). Grammatical forms such as two types of the future tense, passive and active participles (дієприсливний доконаного \ недоконаного виду), adverbial perfect \ imperfect participles (дієприкметник доконаного \ недоконаного виду), four types of declensions, and seven cases. For instance:

- **SY_PARATAXIS:** [Я, хотів, чути, від, світу, ", Україна, ,, ми, будемо, з, тобою, "]. – parataxis [I wanted to hear for the world: "Ukraine, we will stand with you".]
- **VF_FIRST_CONJ:** [затримка, підтримкою, помилкою, країна] – first declension [delay, support, mistake, country]
- **L_GEN_CASE:** [виступу, безпеки, лютого, життів, домовленостей] – genitive case (in Ukrainian denoted by suffix) [performance, safety, February, lives, agreements]

The examples are the raw outputs from the metrics, with added translation into English. We evaluate metrics' performance based on the accuracy score assessed by the trained linguist. The best weighted accuracy has been achieved in the part-of-speech metrics – 0.957, due to their reliance on the spaCy tagger. The lexical metrics have obtained a weighted accuracy of - 0.934. Some discrepancies have occurred at rel-

TABLE III
RESULTS OF THE EXPERIMENTS CONDUCTED IN THE PAPER COMPARED TO THE PAPER BY PANCHENKO ET AL.

Model	**Large training set**	
	Panchenko et al.	This paper
NB-SVM	0.64	-
SM-Voting Classifier	-	**0.66**
Ukr-RoBERTa	0.75	0.82
Ukr-ELECTRA	0.72	0.89

ative and superlative adjectives, adverbs, and case misalignment because of the tagger performance. The grammar group scored 0.912; the inconsistency has occurred in declensions metrics. The syntactic group has got 0.886 in light of the complex constructions, such as parataxis and positioning, that may produce incongruencies.

The accuracy scores indicate that the metrics perform well overall but have some limitations in dealing with complex structures. As the StyloMetrix provides each metric's mean value, a researcher can skip looking into Ukrainian texts to extract the necessary features and conduct further analysis based on the obtained statistics. The descriptions are available for every metric, some with external links to the Universal Dependencies project `https://universaldependencies.org/`.

IV. EXPERIMENTS

This section attempts to represent the StyloMetrix as a baseline for text classification tasks. We further illustrate how to analyze the StyloMetrix baseline model and the possibility of making beneficial inferences about the data based solely on its output.

Conducting a supervised text classification in the Ukrainian language remains challenging due to the scarcity of labeled datasets. There exist a few open-source corpora which can be relevant to this task. For instance, the largest and most popular corpus known by now is UberText 2.0 [5]. The data is subdivided into five smaller datasets: the news dataset, which incorporates short news, longer articles, interviews, opinions, and blogs scraped from 38 news websites; the fiction dataset, with novels, prose, and poetry; the social dataset, covers 264 public telegram channels; the Wikipedia corpus; and the court dataset with decisions of the Supreme Court of Ukraine. The UA news corpus `https://github.com/fido-ai/ua datasets/blob/main/ua_datasets/src/text_classification/README.md` is a collection of over 150 thousand news articles from more than 20 news resources. Dataset samples are divided into five categories: politics, sport, news, business, and technologies. UA-SQuAD is a Ukrainian version of Stanford Question Answering Dataset, and UA-GEC: Grammatical Error Correction and Fluency Corpus for the Ukrainian language [27].The list with all state-of-the-art datasets can be found via the link `https://github.com/asivokon/awesome-ukrainian-nlp/blob/master/README.md`.

We ground our experiments on the well-established benchmark public dataset `https://www.kaggle.com/competitions/ukrainian-news-classification/data` provided by Kaggle project. The corpus has been scrapped from seven Ukrainian news websites: BBC News Ukraine, NV (New Voice Ukraine), Ukrainian Pravda, Economic Pravda, European Pravda, Life Pravda, and Unian. Ukrainian computer scientists [23] have developed the described corpus. The researchers give an exhaustive outlook on the data preprocessing steps and the number of texts in the train/test split (57789/ 24765, respectively). The Kaggle platform offers two training splits from the existing sample: large (57460) and small (9299). In their paper, the academics demonstrate their models' performance scores on the two training splits discussed [23]. We are left with the training splits because we cannot use the initial train and test split as it is unavailable to the public.

The large training data partially incorporates the small training sample; hence we leverage the larger corpus, subdividing it into 80/20% training and testing samples, with 15\% for validation. The obtained results are evaluated with macro-averaged F1-score, the same criterion as in the paper. The baseline model leveraged in the study by [23] was Naïve Bayes with SVM; we have added the StyloMetrix with Voting Classifier as our baseline. The Voting Classifier is composed of RandomForest, AdaBoost, and Logistic Regression. As for the main models, we keep the ones utilized in the paper: ukr-RoBERTa [19] and ukr-ELECTRA[26].

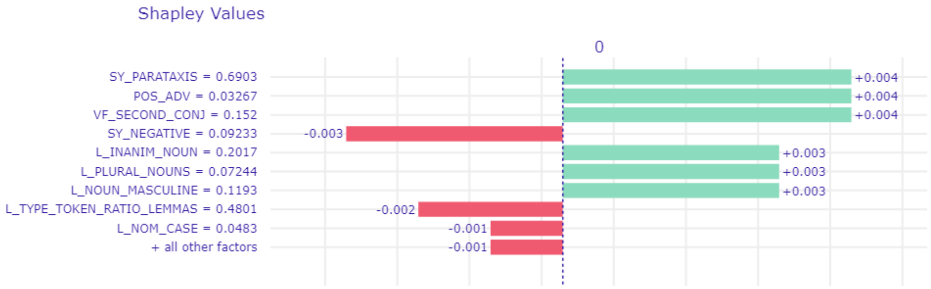
As can be inferred from Table III, the StyloMetrix-Voting Classifier has scored higher than the Naïve Bayes – SVM model but not much, which allows it to serve as a baseline for other more advanced algorithms.

Model Explanation

Unlike the Naïve Bayes – SVM model, the StyloMetrix offers the possibility to extract the descriptive statistics for each group, looking at the most discriminative metrics. For example, we have arbitrary chosen article from class 0 - BBC News Ukraine, to describe the possible data analysis approaches with the StyloMetrix.

One of the wide-used methods of explainable AI is Shapley values [13], which shows the average marginal contributions of features. To make the most of this type of explanation, Shapley values are usually applied to features that can be reversely interpreted, such as categorical values, or to anomaly detection [28]. The most common text representation offering static or dynamic embeddings like GloVe, Word2Vec, or BERT-based vectors does not allow human interpretation of

Fig IVI.
The Shapley values for class 0.



such explanations. The Shapley values indicate the most essential features locally and globally, but the features themselves remain some random columns. With StyloMetrix, on the other hand, the text vector representation consists of interpretable values: each element of the vector translates directly into a given linguistic metric. In this case, indicating the local or global contribution of top features allows for linguistic analysis of grammar or lexical patterns that impact the model's decision when predicting the class.

To implement this, we leverage an open-source library – DALEX [3]. As shown in Fig. II the metrics' significance for class 0 based on their contributions to the model's decision is described. Ultimately, the syntactic metric for parataxis, adverbs, second declension, inanimate nouns, plural nouns, and masculine nouns are prominent in texts that belong to class 0. Vice versa, negative sentences, type-token ratio, and nominative case lower the likelihood of a text falling under this category.

We can dive even deeper into the text statistics and extract the aggregated mean values of the metrics from the StyloMetrix output. As the final vectors are saved in the .csv file, it is easy to find the needed metric and estimate the average mean value for the class. For instance, based on the obtained Shapley, we provide the metrics description and average mean of all texts under class 0 (Table IV). This, in turn, serves the linguistic analysis offering a statistical baseline for a given text genre, including a wide range of metrics. Especially in a multi-class classification, it is vital to compare the baseline against other genres (classes) and draw conclusions about local and global distinctive features.

Therefore, we can conclude that the StyloMetrix as the baseline model can bring some beneficial insights about the texts and the significance of the metrics for a particular classification model. We have presented only one approach to data evaluation with the XAI tool. There are other possibilities to research this area and expand the horizon of the StyloMetrix application and existing metrics.

V. Conclusion

Albeit the idea of constructing the statistical measures of syntactic and grammar features of the text is not new, the experiments discussed in this paper highlight the relevance and significance of creating open-source packages like the StyloMetrix. In the article, we have outlined the main metrics available in the tool's current version and provided some descriptive analysis with the StyloMetrix. We have also discussed the applicability of the corpus analysis tool like StyloMetrix as the baseline "cunny" model for machine learning.

Through experiments, we have traced the metrics importance in a model for classification tasks using the XAI tool – DALEX. More rigorous and detailed analysis is yet to be done in this field, and we consider it the next milestone for our study. The metrics have performed well at the validation step and can be efficient for linguistic analysis of different text genres and the cross-linguistic analysis with other languages such as Polish and English (also available in the StyloMetrix).

References

[1] Bas Aarts, Charles F Meyer, Charles J Alderson, Caroline Clapham, Dianne Wall, and Robert Beard. Livres regus. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 40:3, 1995. Doi: 10.1177/000842987300300410

Table VV.
The StyloMetrix mean values and descriptions of each metric.

Metric	Description	Mean
SY_PARATAXIS	Number of words in sentences with parataxis	0,02731253208958335
POS_ADV	Incidence of adverbs	0,04840493864411134
VF_SECOND_CONJ	Incidence of words in the second declension	0,00027041364427593664
SY_NEGATIVE	Incidence of words in the negative sentences	0,08136554265171815
L_INANIM_NOUNS	Incidence of inanimate nouns	0,013161611405387586
L_PLURAL_NOUNS	Incidence of plural nouns	0,002385407919553026
L_NOUN_MASCULINE	Incidence of masculine nouns	0,20552832194690895
L_TYPE_TOKEN_RATIO_LEMMAS	Type-token ratio for words lemmas	0,054029343395248786
L_NOM_CASE	Incidence of nouns in Nominative case	0,06036709344834804

[2] Karin Aijmer and Bengt Altenberg. English corpus linguistics. *Routledge*, 2014. Doi: 10.4324/9781315845890
[3] Hubert Baniecki,Wojciech Kretowicz, Piotr Piatyszek, JakubWisniewski, and Przemyslaw Biecek. dalex: Responsible machine learning with interactive explainability and fairness in python. *Journal of Machine Learning Research*, 22(214):1‑7, 2021. Doi: 10.48550/arXiv.2012.14406

[4] Douglas Biber. Corpus linguistics and the study of english grammar. Indonesian JELT: Indonesian Journal of English Language Teaching, 1(1):1‑22, 2005. Doi: 10.25170/ijelt.v1i1.93

[5] Dmytro Chaplynskyi. Introducing UberText 2.0: A corpus of modern Ukrainian at scale. In Proceedings of the Second Ukrainian Natural Language Processing *Workshop, Dubrovnik, Croatia,* May 2023. Association for Computational Linguistics.

[6] Rochelle Choenni and Ekaterina Shutova. What does it mean to be language-agnostic? probing multilingual sentence encoders for typological properties. arXiv preprint arXiv:2009.12862, 2020. Doi: 10.48550/arXiv.2009.12862

[7] Noam Chomsky. *Generative grammar. Studies in English linguistics and literature*, 1988.

[8] Peter Collins. It-clefts and wh-clefts: Prosody and pragmatics. *Journal of Pragmatics*, 38(10):1706‑1720, 2006. Doi: 10.1016/j.pragma.2005.03.015

[9] Maciej Eder, Jan Rybicki, and Mike Kestemont. Stylometry with r: a package for computational text analysis. *The R Journal*, 8(1), 2016. Doi: 10.32614/RJ-2016-007

[10] Svitlana Galeshchuk, Arval BNP Paribas, and France Rueil-Malmaison. Abstractive summarization for the ukrainian language: Multi-task learning with hromadske. ua news dataset. *In The Second Ukrainian Natural Language Processing Workshop (UNLP 2023),* page 49, 2023.

[11] Arthur C Graesser, Danielle S McNamara, Max M Louwerse, and Zhiqiang Cai. *Coh-metrix: Analysis of text on cohesion and language. Behavior research methods, instruments, & computers*, 36(2):193–202, 2004. Doi: 10.3758/BF03195564

[12] Sylviane Granger. Automated retrieval of passives from native and learner corpora: precision and recall. *Journal of English Linguistics*, 25(4):365–374, 1997. Doi: 10.1177/007542429702500410

[13] Sergiu Hart. Shapley value. *Springer*, 1989.

[14] Yurii Laba, Volodymyr Mudryi, Dmytro Chaplynskyi, Mariana Romanyshyn, and Oles Dobosevych. Contextual embeddings for ukrainian: A large language model approach to word sense disambiguation. *In Proceedings of the Second Ukrainian Natural Language Processing Workshop (UNLP),* pages 11–19, 2023.

[15] Piaseck Maciej, Walkowiak Tomasz, and Eder Maciej. Open stylometric system websty: Integrated language processing, analysis and visualisation. *Computational Methods in Science and Technology*, 24(1):43–58, 2018. Doi: 10.12921/cmst.2018.0000007

[16] Christian Mair. Quantitative or qualitative corpus analysis? Infinitival complement clauses in the survey of English usage corpus. *Johansson and Stenstrom (eds.)*, pages 67–80, 1991. Doi: 10.1515/9783110865967.67

[17] Danielle S McNamara, Arthur C Graesser, Philip M McCarthy, and Zhiqiang Cai. *Automated evaluation of text and discourse with Coh-Metrix.* Cambridge University Press, 2014. Doi: 10.1017/CBO9780511894664

[18] Rahul Mehta andVasudevaVarma. Llm-rmat semeval-2023 task 2: Multilingual complex ner using xlm-roberta. *arXiv preprint arXiv:2305.03300*, 2023. Doi: 10.48550/arXiv.2305.03300

[19] Benjamin Minixhofer, Fabian Paischer, and Navid Rekabsaz. WECHSEL: Effective initialization of subword embeddings for cross-lingual transfer of monolingual language models. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3992–4006, Seattle, United States, July 2022. Association for Computational Linguistics.

[20] Tempestt Neal, Kalaivani Sundararajan, Aneez Fatima, Yiming Yan, Yingfei Xiang, and Damon Woodard. Surveying stylometry techniques and applications. *ACM Comput. Surv*, 50(6), nov 2017. Doi: 10.1145/3132039

[21] Inez Okulska and Anna Zawadzka. Styles with benefits. the stylometrix vectors for stylistic and semantic text classification of small-scale datasets and different sample length.

[22] Lingwei Ouyang, Qianxi Lv, and Junying Liang. Coh-metrix model-based automatic assessment of interpreting quality. *Testing and assessment of interpreting: Recent developments in China*, pages 179–200, 2021. Doi: 10.1007/978-981-15-8554-8_9

[23] Dmytro Panchenko, Daniil Maksymenko, Olena Turuta, Mykyta Luzan, Stepan Tytarenko, and Oleksii Turuta. Ukrainian news corpus as text classification benchmark. In *ICTERI 2021 Workshops: ITER, MROL, RMSEBT, TheRMIT, UNLP 2021, Kherson, Ukraine, September 28– October 2, 2021, Proceedings*, pages 550–559. Doi: 10.1007/978-3-031-14841-5_37

[24] Andre Quispesaravia, Walter Perez, Marco Sobrevilla Cabezudo, and Fernando Alva-Manchego. Coh-metrix-esp: A complexity analysis tool for documents written in spanish. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation* (LREC16), pages 4694–4698, 2016.

[25] Carolina Scarton and Sandra Maria Aluısio. Coh-metrix-port: a readability assessment tool for texts in brazilian portuguese. In *Proceedings of the 9th International Conference on Computational Processing of the Portuguese Language, Extended Activities Proceedings, PROPOR*, volume 10. sn, 2010. Doi: 10.1007/978-3-642-16952-6_31

[26] Stefan Schweter. Ukrainian electra model, November 2020.

[27] Oleksiy Syvokon and Olena Nahorna. Ua-gec: Grammatical error correction and fluency corpus for the Ukrainian language, 2021. Doi: 10.48550/arXiv.2103.16997

[28] A Tall'on-Ballesteros and C Chen. Explainable ai: Using shapley value to explain complex anomaly detection ml-based systems. *Machine learning and artificial intelligence*, 332:152, 2020.

[29] Gunnel Tottie. Lexical diffusion in syntactic change: Frequency as a determinant of linguistic conservatism in the development of negation in English. *Historical English syntax*, pages 439–467, 1991. Doi: 10.1515/9783110863314.439

APPENDIX

TABLE V.
PARTS OF SPEECH METRICS

Metric	Description
POS_VERB	Incidence of Verbs
POS_NOUN	Incidence of Nouns
POS_ADJ	Incidence of Adjectives
POS_ADV	Incidence of Adverbs
POS_DET	Incidence of Determiners
POS_INTJ	Incidence of Interjections
POS_CONJ	Incidence of Conjunctions
POS_PART	Incidence of Particles
POS_NUM	Incidence of Numerals

TABLE VI.
LEXICAL METRICS

Metric	Description
L_PRON_RELATIVE	Incidence of relative pronoun 'що'
L_PRON_RFL	Incidence of reflexive pronoun
L_PRON_TOT	Incidence of total pronouns
L_QUALITATIVE_ADJ_SUP	Incidence of qualitative superlative adj
L_QULITATIVE_ADJ_P	Incidence of qualitative adj positive
L_RELATIVE_ADJ	Incidence of relative adj
L_SURNAMES	Incidence of surnames
L_PUNCT	Incidence of punctuation
L_PUNCT_DOT	Incidence of dots
L_PUNCT_COM	Incidence of comma
L_PUNCT_SEMC	Incidence of semicolon
L_PUNCT_COL	Incidence of colon
L_PUNCT_DASH	Incidence of dashes

TABLE VII.
GRAMMAR GROUP

Metric	Description
VF_ROOT_VERB_IMPERFECT	Root verbs and conjunctions in imperfect aspect
VF_ALL_VERB_IMPERFECT	Incidence of all verbs in imperfect aspect
VF_ROOT_VERB_PERFECT	Root verbs and conjunctions in perfect aspect
VF_ALL_VERB_PERFECT	Incidence of all verbs in perfect aspect
VF_PRESENT_IND_IMPERFECT	Incidence of verbs in the present tense, indicative mood, imperfect aspect
VF_PAST_IND_IMPERFECT	Incidence of verbs in the past tense, indicative mood, imperfect aspect
VF_PAST_IND_PERFECT	Incidence of verbs in the past tense, indicative mood, perfect aspect
VF_FUT_IND_PERFECT	Incidence of verbs in the future tense, indicative mood, perfect aspect
VF_FUT_IND_IMPERFECT_SIMPLE	Incidence of verbs in the future tense, indicative mood, imperfect aspect, simple verb form
VF_FUT_IND_COMPLEX	Incidence of verbs in the future tense, indicative mood, complex verb forms
VT_FIRST_CONJ	Incidence of verbs in the first declension
VT_SECOND_CONJ	Incidence of verbs in the second declension
VT_THIRD_CONJ	Incidence of verbs in the third declension
VT_FOURTH_CONJ	Incidence of verbs in the fourth declension
VF_TRANSITIVE	Incidence of transitive verbs
VF_PASSIVE	Incidence of verbs in the passive form
VF_PARTICIPLE_PASSIVE	Incidence of passive participles
VF_PARTICIPLE_ACTIVE	Incidence of active participles
VF_INTRANSITIVE	Incidence of intransitive verbs
VF_INFINITIVE	Incidence of verbs in infinitive
VF_IMPERSONAL_VERBS	Incidence of impersonal verbs
VF_ADV_PRF_PART	Incidence of adverbial perfect participles
VF_ADV_IMPRF_PART	Incidence of adverbial imperfect participles

\

TABLE VIII.
LEXICAL METRICS

Metric	Description
L_DIRECT_ADJ	Incidence of direct adjective
L_QUALITATIVE_ADJ_SUP	Incidence of qualitative superlative adj
L_QUALITATIVE_ADJ_CMP	Incidence of relative adj
L_RELATIVE_ADJ	Incidence of relative adj
L_QULITATIVE_ADJ_P	Incidence of qualitative adj positive
L_ANIM_NOUN	Incidence of animated nouns
L_ADV_CMP	Incidence of comparative adverbs
L_ADV_POS	Incidence of positive adverbs
L_ADV_SUP	Incidence of superlative adverbs
L_DIMINUTIVES	Incidence of diminutives
L_FEMININE_NAMES	Incidence of feminine proper nouns
L_FLAT_MULTIWORD	Incidence of flat multiwords expressions
L_INANIM_NOUN	Incidence of inanimate nouns
L_GIVEN_NAMES	Incidence of given names
L_MASCULINE_NAMES	Incidence of masculine proper nouns
L_NOUN_MASCULINE	Incidence of masculine nouns
L_NOUN_FAMININE	Incidence of feminine nouns
L_NOUN_NEUTRAL	Incidence of neutral nouns
L_NUM_CARD	Incidence of numerals cardinals
L_NUM_ORD	Incidence of numerals ordinals
L_PRON_DEM	Incidence of demonstrative pronouns
L_PRON_INT	Incidence of indexical pronouns
L_PRON_NEG	Incidence of negative pronoun
L_PRON_POS	Incidence of possessive pronoun
L_PRON_PRS	Incidence of personal pronouns
L_PRON_REL	Incidence of relative pronouns
L_TYPE_TOKEN_RATIO_LEMMAS	Type-token ratio for words lemmas
L_CONT_A	Incidence of Content words
L_FUNC_A	Incidence of Function words
L_CONT_T	Incidence of Content words types
L_FUNC_T	Incidence of Function words types
L_PLURAL_NOUNS	Incidence of nouns in plural
L_SINGULAR_NOUNS	Incidence of nouns in singular
L_PROPER_NAME	Incidence of proper names
L_PERSONAL_NAME	Incidence of personal names
L_NOM_CASE	Incidence of nouns in Nominative case
L_GEN_CASE	Incidence of nouns in Genitive case
L_DAT_CASE	Incidence of nouns in Dative case
L_ACC_CASE	Incidence of nouns in Accusative case
L_INS_CASE	Incidence of nouns in Instrumental case
L_LOC_CASE	Incidence of nouns in Locative case
L_VOC_CASE	Incidence of nouns in Vocative case
L_INDIRECT_ADJ	Incidence of indirect adjective

TABLE IX
SYNTACTIC METRICS

Metric	Description
SY_PARATAXIS	Number of words in parataxis sentences
SY_DIRECT_SPEECH	Number of words in direct speech
SY_NEGATIVE	Number of words in negative sentences
SY_NON_FINITE	Number of words in sentences without any verbs
SY_QUOTATIONS	Number of words in sentences with quotation marks
SY_EXCLAMATION	Number of words in exclamatory sentences
SY_QUESTION	Number of words in interrogative sentences
SY_ELLIPSES	Number of words in elliptic sentences
SY_POSITIONING	Number of positionings (прикладка)
SY_CONDITIONAL	Number of words in conditional sentences
SY_IMPERATIVE	Number of words in imperative sentences
SY_AMPLIFIED_SENT	Number of words in amplified sentences
SY_NOUN_PHRASES	Number of noun phrases

Communication Papers of the 18th Conference on Computer Science and Intelligence Systems pp. 313–321 DOI: 10.15439/2023F3649
ISSN 2300-5963 ACSIS, Vol. 37

The Use of AI to Determine the Condition of Corn in a Field Robot that Meets the Requirements of Precision Farming

Justyna Stypułkowska
0000-0002-8601-4483
Łukasiewicz Research Network – Institute of Aviation,
al. Krakowska 110/114, 02-256 Warsaw, Poland
Email: justyna.stypulkowska@ilot.lukasiewicz.gov.pl
and
Faculty od Electronics and Information Technology, Warsaw University of Technology
ul. Nowowiejska 15/19, 00-661 Warsaw, Poland
Email: justyna.stypulkowska.dokt@pw.edu.pl

***Abstract*—Artificial intelligence helps to solve numerous problems in modern science and technology. AI-based image recognition allows the detection of specific features. One of the fields that uses AI-based image recognition is precision agriculture. The purpose of the solutions described in this article was to create a system based on artificial intelligence methods and use it in a real project. The article describes the methodology and results of work on tasks related to detection and recognition of corn growth stages, corn hydration levels, and detection and recognition of healthy corn and pathogen-infested corn. Details of the implementation, results and their usefulness for determining selected parameters of corn condition are presented. The developed system makes it possible to monitor the condition of corn, and can be extended to other crops in the future. The presented solution meets the requirements of precision agriculture and is in line with the idea of agriculture 4.0.**

***Index Terms*—AI, image recognition, precision agriculture, determining corn condition parameters, field robot.**

I. Introduction

THE MODERN development of AI has significantly influenced the development of precision agriculture.[1] Numerous research and research centers are successfully using AI methods to achieve the primary goals facing modern agriculture [2][3], while providing an indispensable tool for efficient analysis [4], rapid prototyping and detection of selected features (e.g., object recognition in images). The achievements gained are the engine for further development of better and better solutions and implementation of the developed AI methods to more and more new applications. [5]

The present work concerns a real project implemented by the Łukasiewicz Research Network - Institute of Aviation in cooperation with the Łukasiewicz Research Network - Poznań Institute of Technology and the UNIA company titled "Polish robot - Intelligent robot that meets the requirements of precision agriculture", which fits perfectly with the theme of applying AI to modern agriculture. The main task of the aforementioned robot is to realize the assumptions that guide precision agriculture and fit into the idea of agriculture 4.0. [6] Among the numerous functions of the field robot (such as precise weeding of plants, precise fertilization of plants, navigation based on computer vision, automatic movement), of particular note from the AI point of view is the system developed for this project, the purpose of which is to automatically determine the broadly understood condition of corn (this plant was chosen as the subject of research and its cultivation is dedicated to the field robot developed in the project).

The system developed for this purpose deals with detection and recognition of developmental phases of corn, detection and recognition of corn hydration levels, and detection and recognition of healthy corn and corn infected with selected pathogens based on the RGB images acquired. Based on these values, an aggregate parameter for corn condition is determined in the next step. The system is based on the use of deep learning methods, with multiplication of calculations using a graphics card, under field conditions during field work carried out by a prototype field robot in a real time.

The system consists of four key elements: a system sensor in the form of an RGB camera mounted on a prototype field robot, a set of trained deep neural networks, a set of tagged RGB images that were used to train the deep neural networks, and an application for testing the condition of corn, through which transformations are performed using a set of trained deep neural networks.

The system that is the subject of the invention is based on a new and innovative solution that uses existing technologies, namely RGB imaging and a set of trained deep neural networks, in an area that connects agricultural producers and designers of modern agricultural machinery. This area is the support of corn cultivation, which is an important part of the domestic grain crop and an important part of the global grain crop. [7] The developed system fits perfectly into the strategy of agriculture 4.0, by using some of the latest artificial intelligence methods to automatically detect and analyze selected environmental elements.

An important element of the invention characterized by innovation is a proprietary and unique collection of labeled RGB images, containing several hundred labeled corn images in the form of RGB images, allowing neural networks to be trained to detect and recognize the developmental phases of corn and to detect and recognize corn hydration levels.

The specific challenges associated with each task are described below:

A. Recognizing the growth phases of corn

In this task, the goal was to identify the growth phases of corn, and the international BBCH scale of plant growth, or more precisely its detailing for corn, was used to accurately determine these phases. This scale encompasses several spectra of plant development, and in the conducted research fragments of the scale related to leaf development in plants were used. Thus, on this basis, the BBCH scale [8] was used, ranging from 10 to 19, where the first digit denotes the leaf scale and the second digit denotes the number of developed leaves. Fig. 1 presents an example of determining individual BBCH scale values for the indicated plants.

B. Recognizing corn hydration levels

The task was to determine in which hydration range the corn present in the photos is located. Three levels of hydration were adopted as the basic scale for which the research was conducted: low, medium and high levels of hydration, where low hydration meant exposing plants to conditions of limited access to water, medium hydration was associated with moderate access of plants to water, and high hydration was associated with regular watering of plants and maintenance of optimal levels of plant moisture. The study was conducted under soil conditions characteristic of central Poland, specifically in the village of Borowiec (Tarczyn, Mazowieckie Province).

C. Recognizing healthy corn and corn infected with selected pathogens

This task focused on determining whether the plant was fully healthy or showed some symptoms of infestation with selected pathogens.

The next chapter focuses on the datasets that served as research material for the above tasks.

II. Dataset

An important element of the described system characterized by innovation is a unique collection of tagged RGB images, which was developed strictly for the purposes of the project under implementation. The collection contains several thousand labeled images of corn in the form of RGB images, allowing to train neural networks for detection and recognition of developmental phases of corn, as well as detection and recognition of corn hydration levels.

The collection of images was acquired in the 2021 and 2022 growing seasons on a test plot established in the village of Borowiec (Tarczyn, Mazowieckie Province). Images were acquired in terms of changing developmental phases of corn (images were acquired daily throughout the growing season) and in terms of controlled access of plants to water (here, too, the relative regularity of data acquisition was maintained). In the case of the task related to the identification of healthy corn and corn infested with selected pathogens during the study, there was a problem with the controlled development of fungal infestation of plants, which, due to too low rainfall, did not develop. Therefore, the study used an external dataset [10] containing images of corn infected with selected pathogens (Helmintosporiosis, Common Rust, Gray Spot, Spodoptera Frugiperda caterpillar) and healthy corn were used during the study.

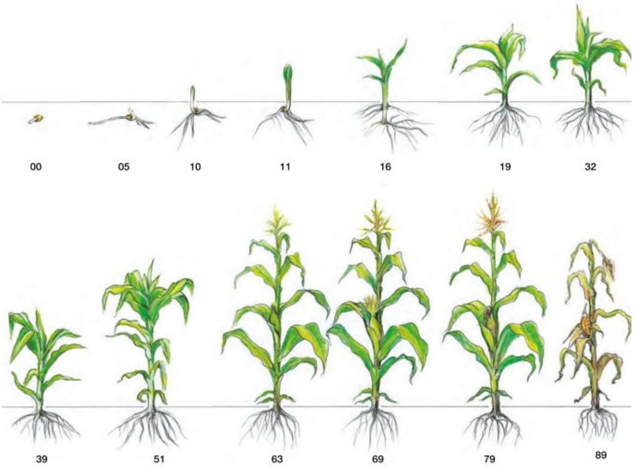


Fig. 1. Example of determining individual BBCH scale values for indicated plants (figure comes from *https://pdf.helion.pl/e_1wwu/e_1wwu.pdf*). [9]

Once the full footage was collected, the next step involved labeling selected crops. The crop concerning corn development stages was labeled using the polygon method, a very precise way, and the output of the labeling process was a .COCO file, which was then used to implement the process of training deep neural networks. The set on corn hydration levels was labeled using the ground truth method. In turn, the set on healthy corn and corn infected with selected pathogens was also labeled using the ground truth method.

Fig. 2 shows an example of the photos that went into the dataset. In some images plants are not entirely visible. It doesn’t affect the results of the algorithm for determining the growing stage, because this is only the case with higher growing stages and in the dataset there are many labeled images on which plants are not fully visible. The photos were taken from the same position of the camera in relation to the rows of plants, which allows to distinguish the structure of individual plants at different stages of development. In addition, plants may have more than 9 leaves and then all cases are still classified as BBCH stage 19, so it is not necessary to count all leaves then. Fig. 3 shows a view of several labeled images that belong to the dataset. There are plans to make the proprietary dataset publicly available in the future, but it depends only on the decision of the management of the Polish Robot project.

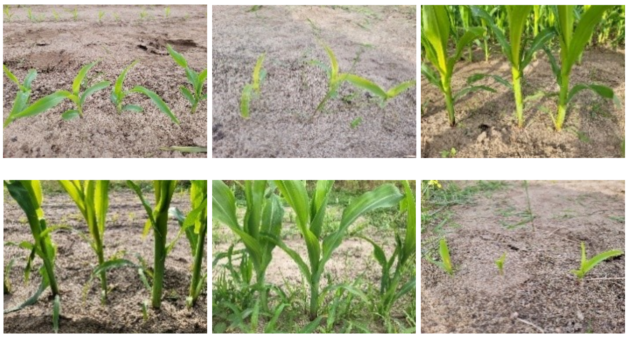


Fig. 2. Examples of photos that made it into the dataset.

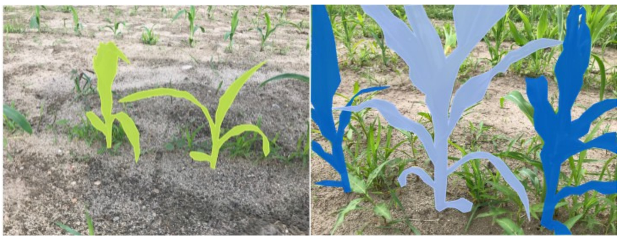


Fig. 3. Labelled photos that belong to the dataset. The colors: light green, dark blue and light purple mean respectively: phase 14, phase 17 and phase 18 of maize development on the BBCH scale.

III. Methodology

The prototype of the developed system relied on AI methods, in particular deep learning and RGB imaging, which were then used to develop a way of detecting and recognizing the developmental phases of corn, detecting and recognizing corn hydration levels, and detecting and recognizing healthy corn and corn infected with selected pathogens. One of the main goals was for the calculations to be realized in real time. In turn, the primary goal was to improve the field work carried out by the field robot and reduce the amount of crop protection products required.

The developed system is characterized by the fact that:

1) An RGB camera is mounted on the prototype field robot, which acquires images of the objects under study (corn rows). The camera acquires RGB images and is mounted in a position that allows it to acquire images of corn rows from the side, at an appropriate height, so that the places where plants grow out of the soil and whole plants or their parts are visible.

2) Then the recorded data, in the form of images of the plants under study, are sent to the computing unit.

3) In the computational unit, equipped with a corn condition application performing calculations based on a set of trained deep neural networks, analysis of the acquired images is performed. As a result of appropriate processing performed with the help of the application, network responses are obtained.

The set of trained deep neural networks includes the following artificial neural networks (characterized by the set of possible responses to their outputs):

- A network for detecting and recognizing the growth stages of corn by classifying the recognized objects: This network was previously trained on a dataset, so it can classify detected objects and assign them probabilities of belonging to the corresponding classes determined during the process of labelling images from the training set used during the training of the network.

In determining the classes and using them in the process of labelling the dataset images, the international scale of plant growth BBCH and a refinement of this scale for corn were used, resulting in a set of 10 classes as follows: "phase10", "phase11", "phase12", "phase13", "phase14", "phase15", "phase16", "phase17", "phase18", "phase19" corresponding to successive phases of corn growth. The phase numbers additionally allow to determine the moment of corn susceptibility to pathogens and the moment favorable to the use of specific plant protection products and fertilizers, which is important information from the point of view of the farmer.

- A network for detecting and recognizing corn hydration levels: this network has been pretrained on the basis of a dataset, so it can classify detected objects and assign them probabilities of belonging to the corresponding classes determined during the process of labelling images from the training dataset.

In determining the classes and using them in the process of labelling the dataset images, the parameters selected during the seeding and cultivation of corn were used, respectively: "hydration level I", "hydration level II" and "hydration level III" corresponding to low, medium and high hydration levels, respectively.

- The network used to detect and recognize healthy corn and corn infected with selected pathogens: This network was previously trained on the basis of an external dataset available on the network [10], so it can classify the detected objects and assign them a probability of belonging to the corresponding classes determined during the process of labelling images from the training dataset.

A collection of 5 classes was used in labelling the dataset images: "Spodoptera frugiperda", "Helminosporiosis", "Common rust", "Aureobasidium zeae" and "Healthy corn", corresponding to particular types of corn infestation and the case of healthy corn.

The corn fitness application is designed to continuously monitor the folder into which the images coming from the RGB camera are saved. When a new photo is detected, it is loaded by the application and then given input to a set of trained deep neural networks. The artificial neural networks then handle the processing and analysis of the received photo, and the output produces a response in the form of the name of the photo along with its location, sets of parameters calculated by the set of trained deep neural networks, a value for the length of time the photo was processed, and a timestamp value telling the time the processing was performed. These values are saved to a .csv format file with a specific location. For successive images that appear in the monitored folder and are processed by a set of trained deep neural networks, successive rows with processing results are added to the .csv file. This file can then serve as a set of data feeding the database, and being already in the database can be used, for example, to create map visualizations on the condition of corn in a given field.

The RGB camera used in the research and development of the overall system uses an array that captures the visible range, which includes the ranges:

1) blue (VIS) with a range of 400 - 500 nm
2) green (VIS) with a range of 500 - 600 nm
3) red (VIS) with a range of 600 - 700 nm

The camera during the collection of images for the dataset and during the collection of images during the operation of the field robot should be mounted at a height of about 30 cm from the ground and pointed from the side towards the row of corn at a distance of about 40 cm at an angle of 30 degrees, without the possibility of shifting.

The resulting file in .csv format contains the following columns:

1) Column 1 - "Filename"
2) Column 2 - probability of detecting level I of hydration
3) Column 3 - probability of detecting level II of hydration
4) Column 4 - probability of detecting level III of hydration
5) Column 5 - probability of detecting infestation with Spodoptera frugiperda
6) Column 6 - probability of detecting infestation with Helminosporiosis
7) Column 7 - probability of detecting infestation with Common rust
8) Column 8 - probability of detecting infestation with Aureobasidium zeae
9) Column 9 - probability of detecting healthy corn
10) Column 10 - probability of detecting growth phase 10
11) Column 11 - probability of detecting growth phase 11
12) Column 12 - probability of detecting growth phase 12
13) Column 13 - probability of detecting growth phase 13
14) Column 14 - probability of detecting growth phase 14
15) Column 15 - probability of detecting growth phase 15
16) Column 16 - probability of detecting growth phase 16
17) Column 17 - probability of detecting growth phase 17
18) Column 18 - probability of detecting growth phase 18
19) Column 19 - probability of detecting growth phase 19
20) Column 20 - value of "processing time"
21) Column 21 - value "timestamp”

The application has an image processing frequency of about 1 Hz. The application is installed on a control unit with libraries that are not expected to change during the lifetime of the system, has a permanently assigned path for downloading RGB images that does not change during the lifetime of the system, and a permanently assigned path where the output file in .csv format is saved. The adopted structure of the output file in .csv format is invariable, which guarantees that the order of columns that can feed the database at a later stage is invariable.

Below is a schematic of the data transmission in the system

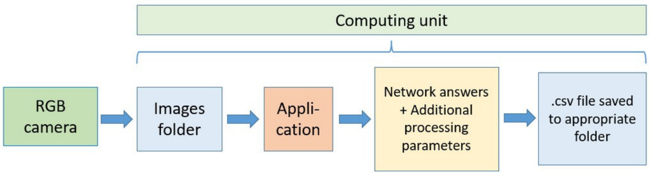

Computing unit
RGB camera
Images folder
Application
Network answers + Additional processing parameters
.csv file saved to appropriate folder

Fig. 4. Schematic of data transmission in the system.

The application was tested using the Gradio library [11], which allows visual operation of a set of trained deep neural networks, i.e. loading selected images into the inputs of individual networks, processing and analyzing the images by individual networks, and displaying the results of the networks in textual and graphical form. The trained deep neural networks for detection of corn hydration levels and corn infestation with particular pathogens produce reflections in the form of detection probability values for each class. In turn, the neural network for detection of developmental phases of corn generates responses both with regard to the value of the probability of detection of individual classes (developmental phases) and the masks visible in the image, along with the assigned value of the class and the probability of recognition of the class. The images that are processed by neural networks should be saved in typical graphic formats (e.g. jpg, png, tiff). The output of an application run through the Gradio library [11] produces numerical data and images in .jpg format with drawn masks (as opposed to an application that outputs results saved as a .csv file). Fig. 5 shows the images in the application's output, with the detected objects labeled and assigned classes and probability values.

The system can be used during the entire growing season of the tested corn crops, i.e. from April to October, depending on the region of Poland, Europe and the world. The optimal range of application of the system is to achieve plant growth from the germination period until reaching stage 19 according to the BBCH scale. The range of hydration levels that are determined by the system is from low (dry corn), through medium, to high (optimal) hydration levels.

When it comes to determining corn hydration levels, individual soil classes are key, some of which guarantee higher levels of hydration, while others don't allow as good water retention. An additional factor in determining hydration levels is the frequency and amount of precipitation and the possibility of artificially irrigating fields.

The system in which the discussed system works is schematically illustrated in Fig. 6 (this is a side view of the field robot). The system for detection and recognition of developmental phases of corn, detection and recognition of corn hydration levels, and detection and recognition of healthy corn and corn infested with selected pathogens is composed of:

- RGB camera 1 located on the field robot and directed from the side towards the corn row,

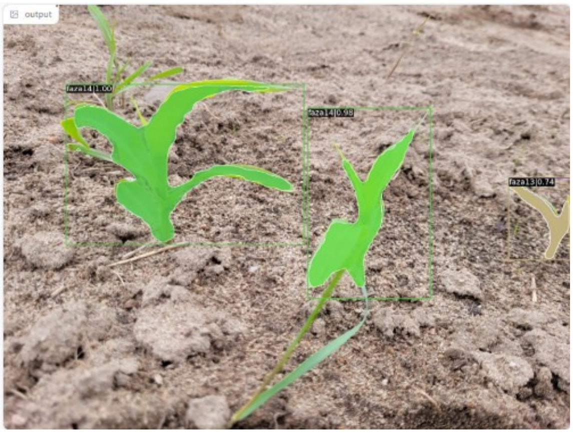


Fig. 5. View of the image on the output of the application.

- computational unit 2 connected to the RGB camera and becoming the analysis center of the system (the computational unit also takes care of placing the images coming from the RGB camera in a specific file folder),
- An application that determines the condition of corn 3 loading data from the RGB camera and performing calculations based on a set of trained deep neural networks, as well as additional calculations on the parameters and time of the performed calculations, and creating an output .csv file with the results of processing and placing it in the specified location.

The operation of the system is based on the use of an RGB camera 1 on the field robot, which acquires RGB images at a frequency of about 1 Hz. The images thus acquired are then sent directly to computing unit 2 (a computer mounted on the robot's platform), which saves the received images in an appropriate file folder. An application for determining the condition of corn is installed on the computing unit, which analyzes the acquired images. The basis of the application's operation is a set of trained deep neural networks, which have been trained using a dataset. The result of the system's operation is an output file in .csv format, which is a report of the results obtained from the output of the neural networks, which forms the basis for determining the developmental phases of corn, corn hydration levels and detection of corn infestation with selected pathogens. The generation of a .csv file makes it possible, at a further stage, to transfer the data to a suitable database and then to process them, for example, to display the results as depictions on a map. Fig. 7 shows a schematic diagram of the system, including the objects at the input of the system in the form of a row of corn 4 and the objects at the output of the system in the form of a file in .csv format 5.

IV. Results

Once the image tagging stage was completed and a complete dataset was created, it was possible to prepare and test with the dataset the selected deep neural network architectures in order to obtain the best possible results. The programming language used for this purpose was Python with appropriately selected libraries. The tests used selected architectures of artificial neural networks, the most effective of which were the following:

- Hybrid Task Cascade (HTC) with ResNet50 as a backbone
- Hybrid Task Cascade (HTC) with ResNet101 as a backbone
- Hybrid Task Cascade (HTC) with ResNeXt101 as a backbone
- Mask2Former [2021.12]

Of these, the best performance was the HTC network with ResNeXt101 as a backbone.

Fig. 8 shows a summary of the results obtained for the task of detecting and recognizing the developmental stages of corn on the BBCH scale.

Fig. 8 shows a plot of the average precision metric mAP obtained during training of deep neural networks (where the y-axis is the average precision values from all classes, and the x-axis is the successive training steps). Fig. 9 shows the results of the network with the detected objects in the form of corn, along with the identification of their developmental phases and the assignment of their class membership probability values.

The results obtained are satisfactory and allow accurate detection and recognition of developmental phases of corn. The fields drawn by the algorithm are comparable with manual marking of images, which is a great success and proves the high accuracy of the obtained solution.

A further study dealt with the identification of corn hydration levels. In this case, a collection of photos taken in a test plot was used, where hydration levels were kept under close control throughout the corn growth period. The downloaded images were divided into appropriate subsets and labeled us-

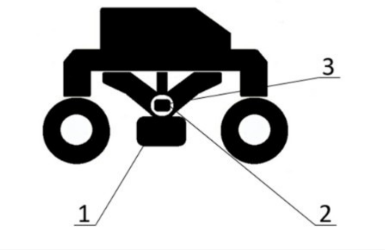


Fig. 6. Schematic view of the system mounted on a field robot.

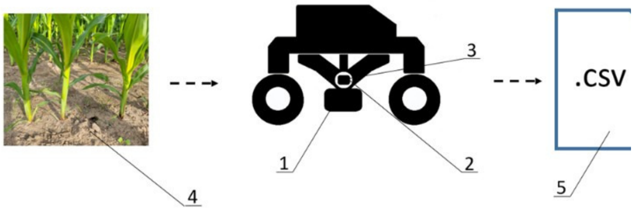


Fig. 7. Circuit schema with input and output components.

ing the ground truth label, and then the creation of the model and the training of deep neural networks were handled. A diagram of the solution concept is shown on the Fig. 10. The input shows a sample input image, then the images are fed to the input to the convolutional deep neural network, and the output gives the individual responses of the network, along with the assigned probability levels of the detected object belonging to each class. These classes are: "hydration level I", "hydration level II" and "hydration level III" corresponding to: low, medium and high levels of hydration. The best results were achieved using the ConvNeXt Small architecture network for classification.

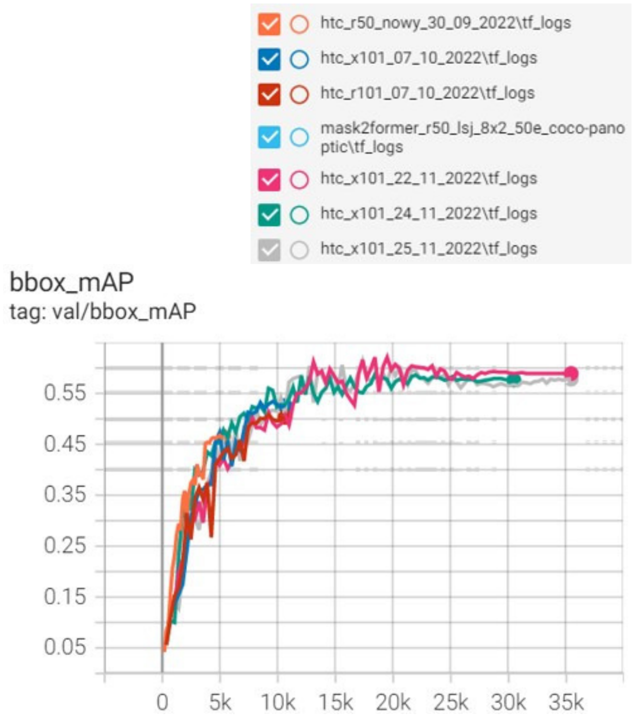


Fig. 8. Diagram of mAP average precision metric obtained during training of deep neural networks. Diagram provides a chart of mean average precision performance vs. number of training epochs.

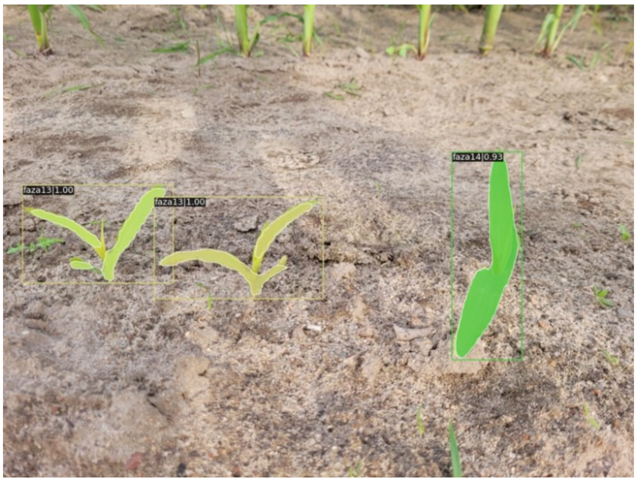


Fig. 9. Results of the network with the detected objects in the form of corn.

The next stage of the study was the recognition and detection of healthy corn and corn infested with selected pathogens. The images used to train the deep neural networks were of healthy corn and corn infected with selected pathogens (the study used a set of 5 different pathogen infestations, where in each crop the corn was infested with only one of the selected pathogens and there was no infestation with several pathogens simultaneously). The input to convolutional deep neural networks was given an input image, and the output received five responses with assigned probability levels for detection of individual pathogen infestations or a healthy plant state without pathogen infestation. A diagram of the solution concept is presented below. The best classification results were achieved using a network with the ConvNeXt Small architecture.

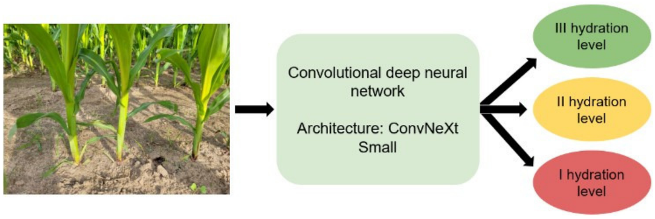

Convolutional deep neural network
Architecture: ConvNeXt Small
III hydration level
II hydration level
I hydration level

Fig. 10. Schematic of the concept for solving the problem of determining corn hydration levels.

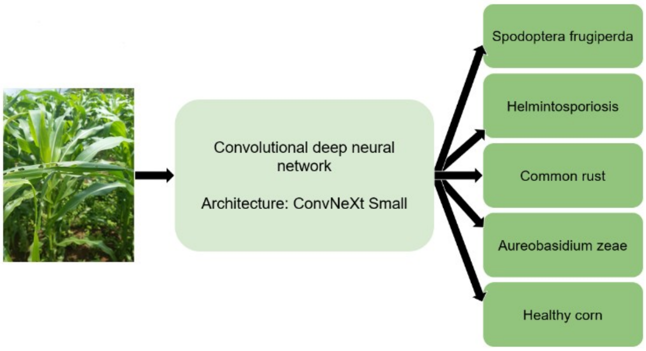


Fig. 11. Schematic of the concept for solving the problem of detection of healthy corn and pathogen-infected corn.

The approaches that were developed were then subjected to performance tests of all algorithms from the test set on the same single image. The results confirmed the validity of the approaches. The conclusion of the experiments confirms that the algorithms can successfully operate simultaneously with high detection and recognition accuracy.

The generalization ability of the trained algorithms has been tested also on a different corn fields than the one of the test site. The results of the developed system were tested on a robot prototype in a larger corn field. The conducted research confirmed the correctness of the adopted concepts and after confrontation of the obtained results with the analysis carried out using the human eye, the correctness of the obtained results was confirmed. Thanks to this, the use of the developed system in the prototype version of the robot was accepted and will be continue in the product version of the robot.

V. Discussion

The subject of corn cultivation supported by the application of modern solutions in the field of field robots and information technology (with particular emphasis on AI) is a fairly new topic, but already has a large number of implemented solutions. [12] The subject is constantly developing rapidly, and the research is a driving force for the introduction of better solutions and development of this branch of science and technology. [13] The analysis of the state of the art has been carried out focusing on the main tasks that the system in question performs, and attempts have been made to find other ways of approaching the same subject.

The first task analyzed was the detection and recognition of corn growth levels consistent with the BBCH scale. In the collection of available scientific articles, one has not come across an approach identical to the one discussed in this article. Determination of developmental stages is usually done manually, i.e., by a person who has to look at the plant to ascertain its developmental level. The BBCH plant growth scale has been developed for this purpose and detailed for the case of corn, but it is usually used for manual identification of developmental stages, and no one has yet carried out this identification in an automated manner. The prepared collection of labeled RGB images in the form of a dataset containing, among other things, a collection of images of corn with the designation of its individual growth stages according to the BBCH scale is therefore a unique thing. In addition, it was used to realize the process of automatic determination of developmental phases of corn with very high efficiency thanks to appropriately selected architecture of deep neural networks and training of these networks based on the collected dataset. The developed solution can run both on a standard computer in the office, using images taken in the field, and directly in the field as a solution applied to a computing unit mounted on a field robot. It is a fast method, which eliminates the laboriousness of the previous manual approach to this subject, and in addition it is characterized by high accuracy and allows to obtain a data set with results on many individual plants collected in a single file, which can input to the database.

Another of the tasks was the recognition and detection of corn hydration levels. A detailed analysis of the current state of the art indicates that this is also a unique solution. There are solutions that use deep learning to determine and predict soil moisture. [14][15][16][17] Machine learning is further used to predict soil properties in terms of permeability and water retention. [18][19][20] However, in no case has an approach been found that is consistent with the one developed in the present invention application. The dataset created, which also contains an image collections of corn characterized by different levels of hydration, is unique in terms of developing an approach based on training deep neural networks based on images of corn characterized by different levels of hydration. In addition, the developed solution allows efficient determination of corn hydration levels based on individual photos, and this process can be carried out both on a standard computer (e.g., in the office or at home) and directly in the field as a tool mounted on a field robot. As in the first task, here, too, the resulting data is made available in the form of a text file that can be an input to a database.

The third of the tasks, which is carried out with the help of the solution, is the recognition and detection of healthy corn and corn infected with selected pathogens. In this case, an available collection of labelled images [9] was found on the web that show pathogen-infected corn and healthy corn. AI-based solutions are also available that can begin to identify pathogen-infected plants.[21][22][23][24][25][26] However, the solution presented in this application is characterized by the use of additional deep neural network architectures, and the entire solution is composed with simultaneous determination of developmental phases of corn and recognition of corn hydration levels, which is unique. In addition, the adopted solution easily aggregates the data collected during the measurements into a .csv text file, which can conveniently be an input into the database. Thus, the solution demonstrates the several-track analysis of single images and allows the determination of parameters that are not realized by other available studies.

VI. Conclusions

The use of the system in question to support the cultivation of corn will affect the emergence of the possibility of accurate and rapid monitoring of the condition of individual plants, including: the determination of corn growth stages; corn hydration levels; and precise identification of the threat from typical corn pathogens. Such accurate identification is made possible by linking output data with data from a GPS transmitter, which can work simultaneously with the described system. Linking data from the .csv output file with data from the GPS transmitter is not complicated and can be done at the database.

The main goals that will be achieved after the implementation of the system in the actual project will be beneficial not only for the manufacturers of modern machinery used in precision agriculture, but primarily for the producers of agricultural crops, whose goal is to accurately monitorize their crops, the desire to reduce the amount of plant protection products and fertilizers used, as well as to obtain high yields of plants with minimum consumption of pesticides.

The developed approach may affect: the possibility of early detection of infestation by disease-causing pathogens attacking corn crops, the ability to accurately determine the hydration levels and development phases of corn, so that it will be possible to determine the moment of necessary application of necessary plant protection and fertilization products in a very precise way. As an additional effect, thanks to the data obtained with the system, it will be possible to visualize the field in terms of the occurrence of specific developmental phases of corn, specific levels of hydration and corn infestation with selected pathogens.

The practical use of the discussed system will translate into considerable benefits for manufacturers of agricultural machinery for precision farming: increasing the efficiency and accuracy of detection of plant pathogen infestation, speeding up the process of detecting corn development stages, and automating the determination of corn hydration levels without the need for additional soil sensors. The AI input will allow to reduce the production costs of highly specialized agricultural machinery, and will also allow to extend the developed solution to other crop species.

The system will also bring benefits to corn crop producers. These include the possibility of simultaneous and early detection of plant pathogen infestation, determination of hydration levels and precision determination of corn growth phases with the accuracy of individual plants. This will give a chance to react early to pathogen threats and overdrying of crop fields. This will directly reduce the usage of crop protection products. The process of automation of the crop field inspection will reduce the financial outlays necessary for the implementation of this process by standard means. Early response to pathogens, detection of insufficient hydration, as well as detection of differences in growth phases between plants in different parts of the fields will allow to achieve higher yields and increase income of agricultural producers.

In addition, the system will provide benefits for consuments and the environment. With early detection of pathogens, alarmingly low levels of hydration, and identification of areas of crop fields where plants are characterized by slower growth, it will be possible to decrease the use of crop protection products and fertilizers, which will directly reduce the release of harmful chemicals into the natural environment and food, thereby positively affecting human health.

The system under discussion, which will be well received by precision farming machinery producers and agricultural producers, will have tangible benefits for both these groups and for consumers themselves. This solution will be able to find interest not only in Poland, but also in the rest of Europe and around the world. In addition, the described system is easily transferable to other crop species than just corn, and can be an invaluable aid to the implementation of modern precision agriculture in many key plant food crops in the world.

REFERENCES

[1] Figiel, S. "Development of Artificial Intelligence and Potential Impact of Its Applications in Agriculture on Labor Use and Productivity/ Rozwój sztucznej inteligencji i potencjalny wpływ jej zastosowań w rolnictwie na wykorzystanie siły roboczej i produktywność. zagadnienia ekonomiki Rolnej/problems of agricultural economics, 373 (4), 5–21." (2022). https://doi.org/10.30858/zer/153583

[2] Raj, E. Fantin Erudaya, M. Appadurai, and K. Athiappan. "Precision farming in modern agriculture." Smart Agriculture Automation Using Advanced Technologies: Data Analytics and Machine Learning, Cloud Architecture, Automation and IoT. Singapore: Springer Singapore, 2022. 61-87. http://dx.doi.org/10.1007/978-981-16-6124-2_4

[3] Shaikh, Tawseef Ayoub, Tabasum Rasool, and Faisal Rasheed Lone. "Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming." Computers and Electronics in Agriculture 198 (2022): 107119. https://doi.org/10.1016/j.compag.2022.107119

[4] Dharmaraj, V., and C. Vijayanand. "Artificial intelligence (AI) in agriculture." International Journal of Current Microbiology and Applied Sciences 7.12 (2018): 2122-2128. https://doi.org/10.20546/ ijcmas.2018.712.241

[5] Bhat, Showkat Ahmad, and Nen-Fu Huang. "Big data and ai revolution in precision agriculture: Survey and challenges." IEEE Access 9 (2021): 110209-110222. https://doi.org/10.1109/ACCESS.2021.3102227

[6] Lorencowicz, Edmund. "Cyfrowe rolnictwo-cyfrowe zarządzanie." Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu 20.4 (2018). https://doi.org/10.5604/01.3001.0012.2952

[7] Popović, Aleksandar, et al. "Current status and future prospects of organic cereal production in the world." Agro-knowledge Journal 18.3 (2018): 199-207. http://dx.doi.org/10.7251/AGREN1703199P

[8] Meier, Uwe & Bleiholder, Hermann & Buhr, Liselotte & Feller, Carmen & Hack, Helmut & Heß, Martin & Lancashire, Peter & Schnock, Uta & Stauß, Reinhold & Boom, Theo & Weber, Elfriede & Zwerger, Peter. (2009). The BBCH system to coding the phenological growth stages of plants-history and publications. Journal für Kulturpflanzen. 61. 41-52. http://dx.doi.org/10.5073/JfK.2009.02.01

[9] Skala BBCH dla kukurydzy, Publikacja specjalna magazynu rolniczego Agro Profil, PS_uprawa-kukurydzy.pdf (helion.pl)

[10] Acharya, R. (October 2020) Corn Leaf Infection Dataset, Version 1. Retrieved October 2020 from https://www.kaggle.com/qramkrishna/corn-leaf-infection-dataset.

[11] https://gradio.app/

[12] Tangwannawit, Panana, and Kanita Saengkrajang. "Technology acceptance model to evaluate the adoption of the internet of things for planting maize." Life Sciences and Environment Journal 22.2 (2021): 262-273. https://doi.org/10.14456/lsej.2021.13

[13] Prabha, R., et al. "Artificial intelligence-powered expert system model for identifying fall armyworm infestation in maize (Zea mays L.)." Journal of Applied and Natural Science 13.4 (2021): 1339-1349. https://doi.org/10.31018/jans.v13i4.3040

[14] Ahmad, Sajjad, Ajay Kalra, and Haroon Stephen. "Estimating soil moisture using remote sensing data: A machine learning approach." Advances in water resources 33.1 (2010): 69-80. https://doi.org/10.1016/j.advwatres.2009.10.008

[15] Cai, Yu, et al. "Research on soil moisture prediction model based on deep learning." PloS one 14.4 (2019): e0214508. https://doi.org/10.1371/journal.pone.0214508

[16] Adab, Hamed, et al. "Machine learning to estimate surface soil moisture from remote sensing data." Water 12.11 (2020): 3223. https://doi.org/10.3390/w12113223

[17] Achieng, Kevin O. "Modelling of soil moisture retention curve using machine learning techniques: Artificial and deep neural networks vs support vector regression models." Computers & Geosciences 133 (2019): 104320.v https://ui.adsabs.harvard.edu/link_gateway/2019CG....13304320A/doi:10.1016/j.cageo.2019.104320

[18] Singh, Vijay Kumar, et al. "Modelling of soil permeability using different data driven algorithms based on physical properties of soil." Journal of Hydrology 580 (2020): 124223. https://doi.org/10.1016/j.jhydrol.2019.124223

[19] Kim, Myeong Hwan, and Chul Min Song. "Prediction of the Soil Permeability Coefficient of Reservoirs Using a Deep Neural Network Based on a Dendrite Concept." Processes 11.3 (2023): 661. https://doi.org/ 10.3390/pr11030661

[20] Tran, Van Quan. "Predicting and Investigating the Permeability Coefficient of Soil with Aided Single Machine Learning Algorithm." Complexity 2022 (2022). http://dx.doi.org/10.1155/2022/8089428

[21] Nagaraju, Mamillapally, and Priyanka Chawla. "Systematic review of deep learning techniques in plant disease detection." International journal of system assurance engineering and management 11 (2020): 547-560. http://dx.doi.org/10.1007/s13198-020-00972-1

[22] Kumar, M. Sunil, et al. "Deep Convolution Neural Network Based solution for Detecting Plant Diseases." Journal of Pharmaceutical Nega-

tive Results (2022): 464-471. http://dx.doi.org/10.47750/pnr.2022.13.S01.57

[23] Arora, Jatin, and Utkarsh Agrawal. "Classification of Maize leaf diseases from healthy leaves using Deep Forest." Journal of Artificial Intelligence and Systems 2.1 (2020): 14-26. https://doi.org/10.33969/AIS.2020.21002

[24] Paliwal, Jagrati, and Sunil Joshi. "An Overview of Deep Learning Models for Foliar Disease Detection in Maize Crop." Journal of Artifi cial Intelligence and Systems 4.1 (2022): 1-21. https://doi.org/10.17148/IJARCCE.2022.117104

[25] Proceedings of the 2022 Seventh International Conference on Research in Intelligent and Computing in Engineering, Vu Dinh Khoa, Shivani Agarwal, Gloria Jeanette Rincon Aponte, Nguyen Thi Hong Nga, Vijender Kumar Solanki, Ewa Ziemba (eds). ACSIS, Vol. 33, pages 177–182 (2022). http://dx.doi.org/10.15439/2022R34

[26] Proceedings of the 2022 Seventh International Conference on Research in Intelligent and Computing in Engineering, Vu Dinh Khoa, Shivani Agarwal, Gloria Jeanette Rincon Aponte, Nguyen Thi Hong Nga, Vijender Kumar Solanki, Ewa Ziemba (eds). ACSIS, Vol. 33, pages 227–234 (2022). http://dx.doi.org/10.15439/2022R08

DOI: 10.15439/2023F9620
ISSN 2300-5963 ACSIS, Vol. 37

Blockchain-based certification of research outputs and academic achievements: A case of scientific conference

Robert Susik
0000-0003-0653-433X
Lodz University of Technology, Institute of Applied Computer Science
Al. Politechniki 11, 90–924 Łódź, Poland
Email: rsusik@kis.p.lodz.pl

Robert Nowotniak
0000-0003-1104-4511
MetaSolid.tech, Al. Grunwaldzka 56/202, 80-241 Gdańsk, Poland
Email: rnowotniak@metasolid.tech

Emanuel Kulczycki
0000-0001-6530-3609
Adam Mickiewicz University in Poznań, Scholarly Communication Research Group
Poznań, Poland
Email: emek@amu.edu.pl

***Abstract*—We consider the problem of researcher output identification and its verification. Nowadays, researcher outputs verification is a challenging problem faced by institutions that want, for example, employ such a researcher. Usually, there is no verification, and the aforementioned institutions rely on trust that the received documents are authentic. Our solution is a based on blockchain technology, a public ledger with smart contracts, that is the root of emerging web3. We use the public wallet address as the researcher identification number and the wallet as the store of all researcher credentials. This paper presents ERC-721 standard-based solution and addresses the conference certification case. The solution proposed in this paper addresses two challenges that arise in collecting and verifying data on research output for managing, monitoring, and evaluating purposes. We show that the public wallet address can be successfully used as the researcher identification number, and the wallet can be used as a vault of all the researcher credentials.**

I. Introduction

A. Background

BLOCKCHAIN emerged as a distributed ledger used for cryptocurrency. The first successful cryptocurrency blockchain, Bitcoin, was introduced by [1]. Since that time, many alternative solutions and forks of Bitcoin have appeared. One of the recent milestones in blockchain development was the release of the Ethereum [2], the first implementation of Blockchain 2.0 concept [3]. Ethereum can be defined as a blockchain-based platform for decentralized application development. It is based on Ethereum Virtual Machine (EVM) that is Turing-complete, and which allows running any algorithms on the blockchain.

B. Areas of performance management where blockchain can be useful

The management of research institutions and research work is based on mechanisms that monitor and evaluate scientific achievements and research outputs. This includes evaluation conducted at the national level within performance-based research funding systems and evaluation of individual researchers or scholarly publication channels. One of the key elements of the whole system of monitoring are persistent identifiers. They are intended to uniquely identify a given object so that the products of scientific work can be monitored. Identifiers are commonly used to identify publications (e.g. DOI or ISBN) and recently also to identify researchers (e.g. ORCID or ScopusID) and organizations (e.g. Funder ID, Global Research Identifier Database (GRID) ID or Research Organization Registry (ROR) ID).

The solution proposed in this paper can address two challenges that arise in collecting and verifying data on research output for managing, monitoring, and evaluating purposes.

The first challenge concerns the legal aspect and relates to who owns the data or servers on which information about the output of researchers and institutions is collected. This challenge is crucial at the level of national policies, which is particularly evident in Europe with the General Data Protection Regulation. It is because it turns out that what is not problematic when identifying publications using DOIs, as it mainly contains publication metadata and possibly references, can pose significant legal problems when dealing with institutions and researchers. For example, Poland has one of the highest percentages of researchers with ORICDs [4]. It is a result of

Topical area: Information Technology for Business and Society

the announced changes in Polish science policy regarding the evaluation of research institutions. ORCID was supposed to be the primary source of information about publications and scientific activity of Polish researchers. However, it turned out that at the stage of implementation of this policy, there were doubts of legal nature as to in which country the servers containing information about the achievements of Polish researchers would be located and whether the Polish government could rely on its internal policy on such external information. Ultimately, ORCID became a recommended rather than an obligatory identifier.

The second challenge is related to the unambiguous legitimization of information by the authorized entity and verification of the validity of the presented information by the institution or researcher. When applying for a job at a scientific institution or promotion, researchers list their achievements and research outputs. In the case of scientific publications, there are no major problems with checking whether such a publication exists. However, there are publications published in the so-called hijacked journals [5], i.e., in journals pretending to be other journals. If a publication has a DOI, one can verify in which journal such a publication was actually published. However, this is no longer so easy in the case of scholarly book publications, as DOIs are rarely assigned to such publications. Moreover, many companies organize so-called questionable conferences [6] with confusingly similar names, which is done intentionally to resemble reputable events. Consequently, in the process of considering candidates for scientific positions or evaluating the output of an institution, it is not always clear whether such a conference was, in fact, organized by a reputable institution or a company organizing para-scientific events because certificates are mostly in the form of PDF files that can be produced by anyone.

C. Certification based on the blockchain

Certification based on the blockchain is known in the literature and was one of the interesting subjects during the last three years. Diverse approaches were proposed, but none of them was based on the ERC721 [7] tokenization standard, which supports multiple features (see Section II) and is compatible with the existing blockchain ecosystem (i.e. software and hardware wallets, such as MetaMask or Ledger). In [8] authors presented an outline of structure and functionality of certification system based on blockchain. They suggest a solution which is a combination of conventional database (off-chain transactions) of students and blockchain technology (on-chain transactions) where the front-end application combines information from both to present the data. This solution involves a third-party institution of Certificate Authority. [9] propose a blockchain-based solution that aims to share student results between Higher Education Institutions. [10] proposed a different approach that leverages a blockchain-based network composed of private and public blockchains to allow educational institutions, learning users, and talent markets exchange the information. An e-learning blockchain-based system was presented in [11]. Here the authors proposed an e-learning system that uses multi-chain architecture as a decentralized ledger that stores users' rewards (e-learning vouchers) and certificates. Another certification system was proposed in [12] where the platform incentives effort in grading via payments with crypto-tokens. In [13], authors proposed a blockchain-based academic certification solution for higher education institutions. They created an Ethereum-based Web3 DApp (a decentralized application [14]) with an application front-end implemented in React library using MetaMask connected to Infura node, and a back-end written in Solidity language using IPFS storage [15]. In [16] authors presented Student-Centered iLearning Blockchain framework which allows to certify, acknowledge and validate students' achievements, skills and competencies on Ethereum blockchain. Contrary to this and other previously presented solutions, in our approach an acknowledged ERC721 [7] standard is used. Thanks to this, digital certificates generated in our system are recognized and presented visually in popular cryptocurrency wallets like MetaMask or Ledger. A survey of over 30 other publications on Blockchain-based prototypes and use cases to transact digital certificates in public education was presented in [17].

D. Aim of the study

This paper aims to present an idea for certifying research outputs and academic achievements using blockchain. We show how such a certification can be designed and implemented in the example of a scientific conference. Moreover, we demonstrate how such a solution could, in the future, allow coping with specific challenges that face the collection of information about scientific activity and its verification in the research evaluation process.

Currently, mass adoption of this solution in the scientific community is difficult due to the high cost of implementation and certification of a single activity. However, the costs of using blockchain are steadily decreasing; therefore, our article may be an inspiration for discussion of whether blockchain can be used in the processes of managing scientific institutions and research evaluation, or whether it is instead a dead end.

II. Our approach

In this section, we present the design of a blockchain-based academic profile for conference attendance certification. The source codes of the solution are shared in the GitHub platform at github.com/rsusik/conference-certification. Our system consists of the following components:

1) Smart Contract (ERC721)
2) User Interface
3) Blockchain (Ethereum)
4) Certificate (Token)

The back-end of the solution is implemented as a smart contract written in Solidity language. The contract implements ERC721 interface and is deployed in the Ethereum blockchain but can also be deployed in any other compatible chain (e.g. Polygon, BSC).

We distinguish the following actors in our system: Participant, Organizer, and Others. A Participant is a person who attends the conference and gains a credential, the Organizer is an entity (university, institute, organization, etc.) that organizes the conference, and the Others are all other blockchain users who want to check or validate the user's participation in the conference. The outline of the system is presented in Fig. 1. It consists of User Interface (UI), which is a web application (front-end) and a Smart Contract (back-end).

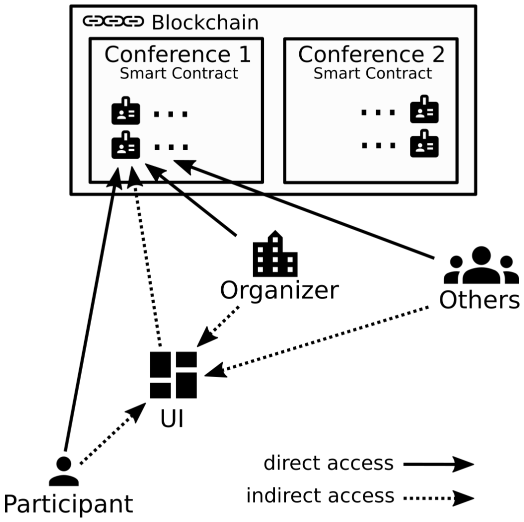


Fig. 1: System outline

The Smart Contract is created and deployed to Blockchain (Ethereum in our implementation) by the Organizer (or on his behalf) for a particular Conference Event. Once the contract is created the Organizer can mint tokens representing certifications for the Participants. The Participant can add the token to his wallet via the User Interface or any wallet manager such as MetaMask. The Others can check if the Participant attended the conference and read the metadata of his activity. The system allows to:

- issue certifications (mint the tokens),
- revoke certifications (burn the tokens),
- transfer certifications (transfer tokens),
- list Participant's certifications (list token owners),
- list the conference certifications (list tokens),
- validate the Participant's certificate.

The Organizer has permission for all of the listed activities. The Participant is the owner of the token that represents the credential. He has the same permissions as Others. The Others can list all participants and credentials of the conference.

The Participant is required to deliver his wallet public address, then the Organizer can issue a certificate for him (mint an NFT token). There are two types of participants: active (i.e. those who present their research results) and passive (listeners). According to [18] each certificate for active Participant contains such information as: Acronym, Event type (Conference, Workshop, Symposium), Year, Date (October 26-30), Location, Title (i.e. International Semantic Web Conference), Subject (what the conference is about, i.e. Semantic Web).

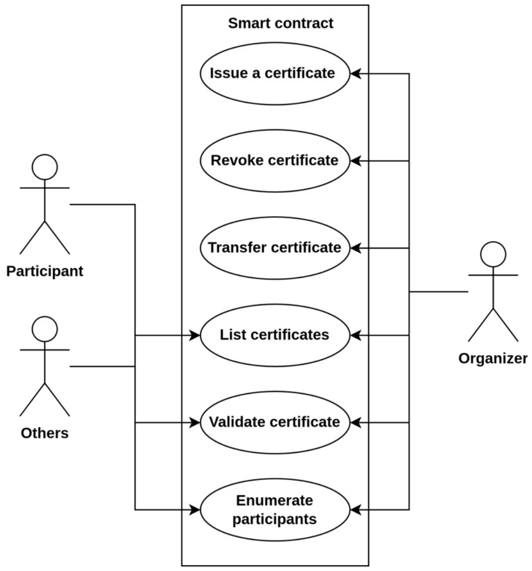


Fig. 2: Use case diagram

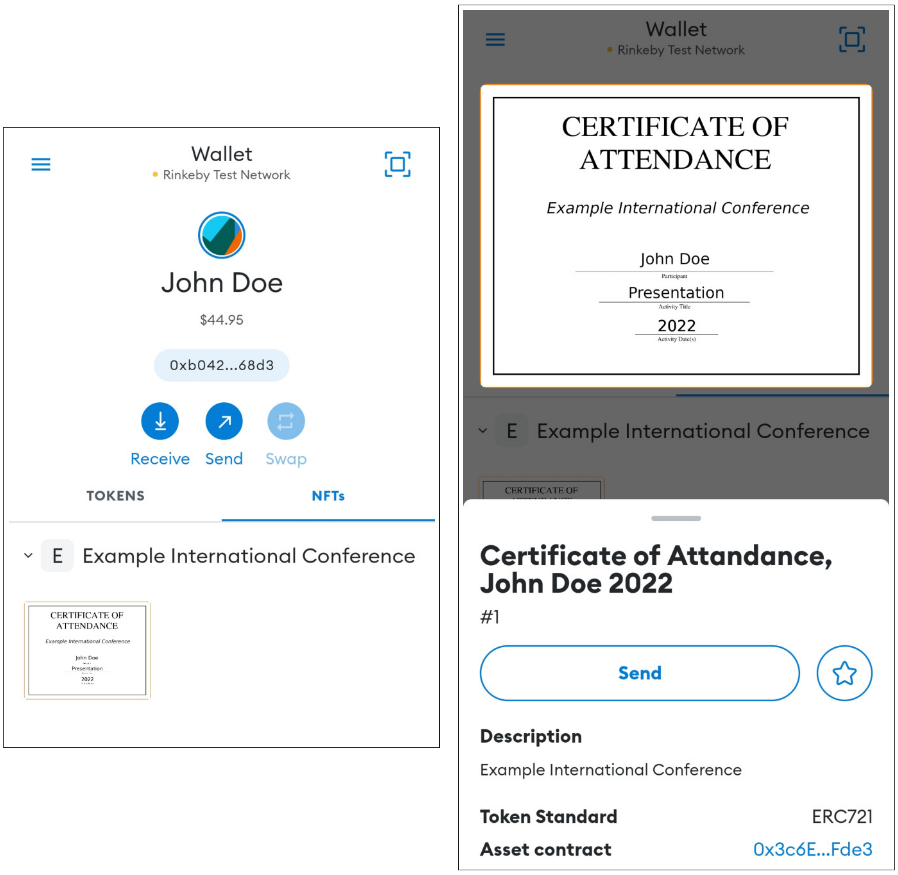
The Organizer delivers the conference Smart Contract address to Participants after the conference event. The Participants can then add the certifications to their wallets. Other users can read the conference participant list or validate credentials using either the User Interface or directly by executing a function in the Smart Contract on the blockchain.

The proposed approach is fully decentralized, there is no need to implement or use any specific API as the system is based on the public blockchain, so everyone who has access to the Internet and the blockchain nodes can read certifications and confirm their authenticity. Additionally, all the Internet users can write their applications and integrate with our solution without our permission as long as they use the same programming interfaces (the ERC721 standard and JSON Schema). Figure 3 shows a screenshot of MetaMask mobile crypto wallet containing an example conference certificate.

There are costs of smart contract deployment on blockchain and token minting. This may be perceived as a disadvantage of this system, but there are multiple options for cost optimizations (including the use of Layer2 chains or ERC115 contracts) [19], [20], [21]. On the other hand, there is no need to maintain any servers (i.e., databases, HTTP servers, DNS services, etc.) to store and share the data.

III. Conclusions and future work

In this paper, we address the problem of researcher output identification and its verification. We show that the public wallet address can be successfully used as the researcher identification number, and the wallet can be used as a vault of all the researcher credentials. Additionally, the proposed solution is based on the ERC721 standard, which makes it compatible with most existing crypto wallets and other software. Apart from the conference case, there are a number of interesting use cases for this solution. In fact, we believe that



John Doe

$44.95

0xb042...68d3

E Example International Conference

CERTIFICATE OF ATTENDANCE

Example International Conference

John Doe
Participant
Presentation
Activity Title
2022
Activity Date(s)

E Example International Conference

Certificate of Attandance, John Doe 2022

#1

Description

Example International Conference

Token Standard ERC721

Asset contract 0x3c6E...Fde3

Fig. 3: Screenshot of a certificate of attendance displayed in MetaMask mobile crypto wallet

blockchain can be used to store complete researcher profile information and replace those commonly used nowadays.

The limitation we see in a blockchain–based certification system is the need to have smart contract addresses of legitimate conferences or journals. Questionable conferences or hijacked journals may mint tokens (representing credentials) in their smart contract (pretending to represent other journals) using any address. This situation is analogous to sharing credentials on a fake HTTP conference website. However, we do not consider it a significant disadvantage because, by knowing the smart contract address of a specific legitimate conference (for instance, obtained from its official website) or having a list of such respected conferences (with their smart contract addresses), we can easily verify if a particular certificate has been issued by mentioned conference (the token is minted on their smart contracts). Moreover, verifying such fake credentials is unnecessary, as the client application wouldn't display them. Another factor that may be considered a minor inconvenience in implementing this approach is when journals or conferences use diverse blockchains for credential certifications. In such a case, we need to query multiple blockchains to perform the verification, which is not an issue but may require additional work or a higher-level API.

References

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," *Decentralized Business Review*, p. 21260, 2008.

[2] V. Buterin, "Ethereum white paper: A next generation smart contract & decentralized application platform," 2013.

[3] D. Efanov and P. Roschin, "The all-pervasiveness of the blockchain technology," *Procedia computer science*, vol. 123, pp. 116–121, 2018.

[4] S. J. Porter, "Measuring Research Information Citizenship Across ORCID Practice," *Frontiers in Research Metrics and Analytics*, vol. 7, p. 779097, Mar. 2022.

[5] M. Andoohgin Shahri, M. D. Jazi, G. Borchardt, and M. Dadkhah, "Detecting Hijacked Journals by Using Classification Algorithms," *Science and Engineering Ethics*, vol. 24, pp. 655–668, Apr. 2017.

[6] E. Kulczycki, M. Hołowiecki, Z. Taşkın, and G. Doğan, "Questionable conferences and presenters from top-ranked universities," *Journal of Information Science*, 2022.

[7] W. Entriken, D. Shirley, D. Evans, and N. Sachs, "ERC721." https://eips.ethereum.org/EIPS/eip-721, 2018. Accessed: 2022-04-20.

[8] M. Alshahrani, N. Beloff, and M. White, "Revolutionising higher education by adopting blockchain technology in the certification process," in *2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT)*, pp. 1–6, IEEE, 2020.

[9] S. Cardoso, H. São Mamede, and V. Santos, "Reference model for academic results certification in student mobility scenarios: Position paper," in *2020 15th Iberian Conference on Information Systems and Technologies (CISTI)*, pp. 1–4, IEEE, 2020.

[10] L. Gao, "Management of online education based on blockchains," in *2020 International Conference on Modern Education and Information Management (ICMEIM)*, pp. 84–89, IEEE, 2020.

[11] C. Li, J. Guo, G. Zhang, Y. Wang, Y. Sun, and R. Bie, "A blockchain system for e-learning assessment and certification," in *2019 IEEE International Conference on Smart Internet of Things (SmartIoT)*, pp. 212–219, IEEE, 2019.

[12] J. Gupta and S. Nath, "Skillcheck: An incentive-based certification system using blockchains," in *2020 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*, pp. 1–3, IEEE, 2020.

[13] M. P. Jaramillo and N. Piedra, "Use of blockchain technology for academic certification in higher education institutions," in *2020 XV Conferencia Latinoamericana de Tecnologias de Aprendizaje (LACLO)*, pp. 1–8, IEEE, 2020.

[14] W. Cai, Z. Wang, J. B. Ernst, Z. Hong, C. Feng, and V. C. Leung, "Decentralized applications: The blockchain-empowered software system," *IEEE Access*, vol. 6, pp. 53019–53033, 2018.

[15] J. Benet, "Ipfs-content addressed, versioned, p2p file system," *arXiv preprint arXiv:1407.3561*, 2014.

[16] A. S. Anwar, U. Rahardja, A. G. Prawiyogi, N. P. L. Santoso, *et al.*, "ilearning model approach in creating blockchain based higher education trust," *International Journal of Artificial Intelligence Research*, vol. 6, no. 1, 2022.

[17] A. Pfefferling and P. Kehling, "Design disclosure for blockchain-based application used in public education certificates with electronic hashes," in *Konferenzband zum Scientific Track der Blockchain Autumn School 2021*, no. 004, pp. 034–041, Hochschule Mittweida, 2021.

[18] J. Franken, A. Birukou, K. Eckert, W. Fahl, C. Hauschke, and C. Lange, "Persistent identification for conferences," *Data Science Journal*, vol. 21, no. 1, 2022.

[19] R. Susik and R. Nowotniak, "Pattern matching algorithms in blockchain for network fees reduction," *arXiv doi:10.48550/ARXIV.2207.14592*, 2022.

[20] A. Di Sorbo, S. Laudanna, A. Vacca, C. A. Visaggio, and G. Canfora, "Profiling gas consumption in solidity smart contracts," *Journal of Systems and Software*, vol. 186, p. 111193, 2022.

[21] G. A. Pierro, H. Rocha, S. Ducasse, M. Marchesi, and R. Tonelli, "A user-oriented model for oracles' gas price prediction," *Future Generation Computer Systems*, vol. 128, pp. 142–157, 2022.

DOI: 10.15439/2023F4882
ISSN 2300-5963 ACSIS, Vol. 37

An Elliptic Intuitionistic Fuzzy Portfolio Selection Problem based on Knapsack Problem

Velichka Traneva
Prof. Asen Zlatarov University
1 Prof. Yakimov Blvd, Burgas 8000, Bulgaria
Email: veleka13@gmail.com

Petar Petrov
Prof. Asen Zlatarov University
1 Prof. Yakimov Blvd, Burgas 8000, Bulgaria
Email: peshopbs2@gmail.com

Stoyan Tranev
Prof. Asen Zlatarov University
1 Prof. Yakimov Blvd, Burgas 8000, Bulgaria
Email: tranev@abv.bg

***Abstract*—This paper suggests an index-matrix approach to a knapsack-based portfolio selection model (E-IFKP) with parameters, characterized by elliptic intuitionistic fuzzy values. Elliptic Intuitionistic Fuzzy Sets are a tool to model the greater uncertainty of the environment, which is introduced in 2021. In the developed E-IFKP model, the price and the return value of the assets are determined by experts taking into account their rank. Three scenarios are proposed to the decision maker for the final choice - pessimistic, optimistic, and average. The proposed E-IFKP extends the dynamic programming approach for the Knapsack problem, which aims to select items to be placed in the knapsack to achieve the highest possible total value not exceeding its capacity. To determine the best option for an E-IFKP for certain data from the US stock exchange a software for conducting the proposed approach is developed and is used in the case study.**

I. Introduction

THE GOAL of the portfolio selection problem is to select the assets that will receive the most value from the limited resources that are available [37]. Markowitz [27], who introduced the mean-variance model and treated asset returns as random variables in the multivariate normal distribution, laid the groundwork for portfolio selection. That model defines efficient portfolios as those that maximize expected return for a given level of risk or those that minimize risk for a given level of expected return. The Markowitz portfolio selection theory and the related methods require a large amount of time sequence data. That data is required to create the statistical indices that serve as the foundation for these methods. Despite its many benefits, Markowitz's model has drawn criticism since it fails to take into account many other factors besides risk and return [50]. Over the past few decades, numerous Markowitz's model extensions have been created [58].

According to [57], historical data is either unavailable or insufficiently detailed to predict how the market will evolve in the real world. As a result, another option might be to review financial reports and the opinions of experts and/or investor preferences. Making decisions in the real financial market is frequently complicated and unclear, which is another issue. In order to take into account the actual state of the financial markets in portfolio selection models, numerous ways based on uncertain conditions have been created. These include a robust-based approach in [18], a scenario-based approach in [35], and fuzzy methods in [42], [43]. The developed theory of fuzzy logic [56] is a useful tool for working with incomplete or ambiguous information. The essential idea is to transform linguistic variables into fuzzy sets (FSs) using the appropriate membership functions [58]. It is suggested that fuzzy portfolio selection take into account the expertise of professionals, the subjective opinions of investors, or quantitative and qualitative analysis in portfolio selection challenges. Fuzzy portfolio models are another type of potential method for resolving non-probabilistic portfolio selection because the investment behaviors to new economic events cannot be precisely evaluated by the prior return rates for the selected securities due to the exclusion of many factors in the portfolio selection process. In [17], [26], [55], the objective is to maximize the fuzzy return rates while limiting the maximum investment risk by employing possibility theory, which was modeled and researched for the portfolio choice problem. Theories of possibility or believability that lead to the best selections in a fuzzy portfolio selection can be used to summarize the key studies in fuzzy portfolio models. The development of multi-objective risk measurements and fuzzy portfolio selection evaluations are presented in [30]. The entropy method is used in [55] to formulate a weighted possibility fuzzy multiobjective and higher order moment portfolio model with the efficiency and effectiveness portfolio selections. Two fuzzy-AHP approaches for portfolio selection in the Istanbul Stock Exchange are performed and compared in [43]. The capital gain tax to fuzzy portfolio selection is taken into consideration in [16] and formulated as a bi-objective mean-variance problem that is solved by a time-varying numerical integral-based particle swarm optimization algorithm. By constructing the evolutionary algorithm and fuzzy simulation approach to

Work on Sect. I and Sect. II is supported by the Asen Zlatarov University through project Ref. No. NIX-482/2023 “Modeling Management Decisions with New Analysis". Work on Sect. III and Sect. IV is supported by the Asen Zlatarov University through project Ref. No. NIX-486/2023 “Application of Intuitionistic Fuzzy Logic in Decision Making”.

demonstrate the efficient algorithm, a skewness fuzzy variable is employed in in [24] to formulate a mean-variance-skewness fuzzy portfolio selection. In order to differentiate between three different types of risk behaviors for investors, a fuzzy portfolio selection for dealing with qualitative information that is represented as hesitant fuzzy elements is suggested in [58]. A fascinating subject in the study of fuzzy portfolio selections, according to [31], is the risk behavior analysis for investors. In [49], the proposed threshold of excess investment for each security in the portfolio selection is the assured return rate. A fuzzy analytic network technique is employed for portfolio selection [37], and a great deal of other criteria other than risk and return are studied. According to [34], there are several financial applications for fuzzy logic, including portfolio optimization. Multi-objective linear programming is created in study [54] for portfolio selection in a fuzzy environment. The model, based on the investor's risk behavior in a dimension that is different from the gap between the guaranteed return rate and the return rate for each security, is suggested in [11]. According to [21], where the mean-variance model was used for portfolio selection and the risk the behavior of an investor in a different dimension distance for shortage investment and the excess investment was still not taken into account, the adjustable security proportion for excess investment and shortage investment based on the selected guaranteed return rates for profitable returns is suggested. The dimension of excess investment has been taken into consideration as the fuzzy portfolio based on guaranteed return rates has been developed for investors with various risk preferences [10]. According to Gorzaczany [19], since decision-makers aren't always able to accurately explain an element's degree of membership, formal representations of fuzzy sets are usually insufficient. There is often a degree of hesitancy between membership and non-membership in real-world issues since decision-makers frequently express their thoughts even when they are unsure of them [53]. Fuzzy set extensions are needed to address this problem. In [1], Atanassov has proposed the intuitionistic fuzzy sets (IFSs) as an extension of FSs. The difference between FS and IFS is that the elements of the latter sets have a degree of hesitancy, which complements the corresponding sum of the element membership and non-membership degrees to 1. In [52] are described as other "extensions" of the IFSs and these extensions of IFS have been compared with themselves. The authors of [52] have demonstrated that IFS can completely describe a Hesitant Fuzzy Set [44]. In [52], the authors also prove that the Picture fuzzy sets [12], the Cubic set [25], the Neutrosophic fuzzy sets [40] and the Support-intuitionistic fuzzy sets [33] are representable by interval-valued IFSs (IVIFSs) [8]. In recent years, two more generalizations of intuitionistic fuzzy sets have appeared in the form of circular [4], and elliptic IFSs [5], which also generalize interval-valued IFSs. The ultimate goal of an intuitionistic fuzzy interactive multi-objective optimization approach is to find the optimum solution that maximizes satisfaction and minimizes unhappiness, according to [36]. Interactive optimization techniques' primary objective is to actively involve the decision-maker in problem-solving. A socially responsible portfolio selection problem is solved in [20] utilizing an interactive triangular intuitionistic fuzzy approach. A portfolio selection model built on the knapsack problem with interval uncertainty is provided in the study [51]. It is suggested that the created model, which is based on the knapsack problem (KP), can be used to appropriately allocate the number of shares to various assets and may be able to determine the best asset allocation under unique circumstances involving relatively high stock values.

In this regard, our efforts are to develop an extension of the portfolio selection problem so that it can be applied to IF data and then to circular and elliptic IFSs. In our previous works [45], [46], we for the first time have suggested IF and circular IF KP (C-IFKP) for finding the optimal solution respectively of the IF and circular IF portfolio selection problem. The main parameters in the problems are IF pairs or circular IF triples, determined by experts under three different scenarios - pessimistic, optimistic, and average. E-IFSs are described as sets with an ellipse indicating the degrees of membership and non-membership for each element of the universe [5]. No developed models for optimal portfolio selection with elliptic IF data were found in the Scopus database. The index-matrix method to an elliptic knapsack-based portfolio selection model (E-IFKP), which extends the Circular IFKP and IFKP approach from the studies [46], is introduced here. Experts agree on the importance, cost, and return of each asset, and the suggested approach takes these ratings into account. Pessimistic, optimistic, and average scenarios are put out to the decision-maker for consideration before making a final decision. Software for carrying out the suggested E-IFKP is under development and utilized in the real case at a specific time to find the best alternative for an E-IFKP for specific data at a specific moment from the US stock exchange. The advantage of this model is that it can be applied to both plain and elliptic IF data. Another advantage is that it can easily be extended so that it can be applied to multidimensional IF data. Theoretical Contributions of the study are: the introduced definition of elliptic IF quads; extended comparison operations and relations on IF pairs to those on elliptic IF quads.

The Knapsack problem's (KP) goal is to maximize the total utility value of all things selected by the decision-maker within the constraints of a knapsack [28]. The phrase "Knapsack problem" first appeared in early publications by George Dantzig in the 1950s [13]. Gilmore and Gomory examined the dynamic programming method to the KP in 1966 [15]. An approximation approach for the solution of a multiple choice fuzzy KP (FKP) is provided in [22]. Ant colony optimization with environmental changes is developed in [32]. The paper presents one kind of FKP [38]. A dynamic programming strategy has been provided in [9], [38], [39] for solving FKP. In the work [14], an approach for ant colonies to optimize the Multiple-Constraint Knapsack Problem utilizing intuitionistic fuzzy (IF) pheromone updating is described. The idea of the E-IFKP and its usage for the portfolio selection problem according to three scenarios give this work its novelty.

The remaining portions of this study are structured as follows: Short remarks to the elliptic intuitionistic fuzzy quads (E-IFQs) and the index matrices (IMs) are provided in Section II. A form of 0-1 E-IFKP for portfolio selection is suggested in Section III, and with the aid of software, it is applied to a real E-IFKP for the selection of portfolio shares of the IT companies which make up the Dow Jones Industrial Average. Section IV, which summarizes the findings and offers recommendations for additional study, brings the work to a close.

II. Remarks on Elliptic Intuitionistic Fuzzy Quads and Index Matrices

We will define elliptic intuitionistic fuzzy quads (E-IFQs), index matrices (IMs), as well as some of their operations and relationships, in this section. In 2021, the IFS is extended with the E-IFS, which has a different interface. An ellipse with semi-major and semi-minor axes exists around each element of E-IFS that represents its membership degree and non-membership degree [5].

Let's consider the definition of an intuitionistic fuzzy pair (IFP) [7]: an IFP has the form of $\langle a(p), b(p)\rangle$ or $\langle \mu(p), \nu(p)\rangle$: The components of an IFP are $a(p)(\mu(p)), b(p)(\nu(p)) \in [0,1]$ and $a(p)+b(p) = \mu(p)+\nu(p) \leq 1$, respectively. These elements represent the degrees of membership and non-membership of a proposition p. Using the definition of the E-IFS [5], let us define E-IFQ as an object of the following form:

$$\langle a(p), b(p); u, v\rangle = \langle \mu(p), \nu(p); u, v\rangle,$$

where $a(p)+b(p) = \mu(p)+\nu(p) \leq 1$, which is utilized to evaluate the statement p, is regarded as the "truth degree" and "falsity degree" of the assertion p, respectively. The semi-major and semi-minor axes of the ellipse with the center $\langle a(p)(\mu(p)), b(p)(\nu(p))\rangle$ are $u, v \in [0, \sqrt{2}]$, respectively.

Two E-IFQs $x_{u_1,v_1} = \langle a, b; u_1, v_1\rangle$ and $y_{u_2,v_2} = \langle c, d; u_2, v_2\rangle$, shall be used. Let us define an operation $* \in \{\min, \max\}$. The operations over E-IFQs that follow are based on the E-IFSs operations from [5]. For the E-IFQs, the operations "subtraction" and "division" for C-IFPs [46] are modified.

$$\neg x_{u_1,v_1} = \langle b, a; u_1, u_2\rangle;$$
$$x_{u_1,v_1} \wedge_* y_{u_2,v_2} = \langle \min(a,c), \max(b,d); *(u_1,u_2), *(v_1,v_2)\rangle;$$
$$x_{u_1,v_1} \vee_* y_{u_2,v_2} = \langle \max(a,c), \min(b,d); *(u_1,u_2), *(v_1,v_2)\rangle;$$
$$x_{u_1,v_1} +_* y_{u_2,v_2} = \langle a+c-a.c, b.d; *(u_1,u_2), *(v_1,v_2)\rangle;$$
$$x_{u_1,v_1} \bullet_* y_{u_2,v_2} = \langle a.c, b+d-b.d; *(u_1,u_2), *(v_1,v_2)\rangle;$$
$$x_{u_1,v_1} @_* y_{u_2,v_2} = \langle \tfrac{a+c}{2}, \tfrac{b+d}{2}; *(u_1,u_2), *(v_1,v_2)\rangle$$
$$x_{u_1,v_1} -_* y_{u_2,v_2} = \langle \max(0, a-c), \min(1, b+d, 1-a+c); \\ *(u_1,u_2), *(v_1,v_2)\rangle$$
$$x_{u_1,v_1} :_* y_{u_2,v_2} = \begin{cases} \langle \min(1, a/c), \min(\max(0, 1-a/c), \\ \max(0, (b-d)/(1-d))); *(u_1,u_2), *(v_1,v_2)\rangle \\ \quad \text{if } c \neq 0 \ \& d \ \neq 1 \\ \langle 0, 1; *(u_1,u_2), *(v_1,v_2)\rangle \text{ otherwise} \end{cases}$$

Since the semi-major and semi-minor axes produce outputs with minimal and maximum degrees of uncertainty, respectively, the operations presented here are based on their minimum and maximum values. We propose the following relation for comparing E-IFs using a formula for the distance between C-IFSs [6], the relation for comparing two C-IFPs [46], and the distance from the element to the ideal positive alternative [41].

$$x_{u_1,v_1} \geq_{R^{elliptic}} y_{u_2,v_2} \quad \text{iff } R^{elliptic}_{x_{u_1,v_1}} \leq R^{elliptic}_{y_{u_2,v_2}} \tag{1}$$

where

$$R^{elliptic}_{x_{u_1,v_1}} = \frac{1}{6}(2-a-b)\big(|\sqrt{2}-u_1| + |\sqrt{2}-v_1| + |1-a|\big)$$

is the distance between x and the ideal positive alternative $\langle 1, 0; \sqrt{2}, \sqrt{2}\rangle$ to x. According to the Szmidt and Kacprzyk's version of the distance [6], we state that the E-IFQs x_{u_1,v_1} and y_{u_2,v_2} are in α-proximity ($\alpha \in [0;1]$): if

$$d(x_{u_1,v_1}, y_{u_2,v_2})$$
$$= \frac{1}{3}\big(|u_1-u_2| + |v_1-v_2| + 0.5(|a-c| + |b-d| + |c+d-a-b|)\big) \leq \alpha$$

In 1987, according to [2], the theory of index matrices (IMs) was developed. Over IMs, several operations, relations, and operators are defined (see [3], [48]). Assume that the set of indices $\mathscr{I}$ is fixed. Two-dimensional elliptic intuitionistic fuzzy index matrix (2-D E-IFIM) $A = [K, L, \{\langle \mu_{k_i,l_j}, \nu_{k_i,l_j}; u_{k_i,l_j}, v_{k_i,l_j}\rangle\}]$ with index sets K and L ($K, L \subset \mathscr{I}$), we denote the object analogous to circular IFIM (C-IFIM) [46]:

$$\begin{array}{c|ccc} & l_1 & \dots & l_n \\ \hline k_1 & \langle \mu_{k_1,l_1}, \nu_{k_1,l_1}; u_{k_1,l_1}, v_{k_1,l_1}\rangle & \dots & \langle \mu_{k_1,l_n}, \nu_{k_1,l_n}; u_{k_1,l_n}, v_{k_1,l_n}\rangle \\ \vdots & \vdots & \ddots & \vdots \\ k_m & \langle \mu_{k_m,l_1}, \nu_{k_m,l_1}; u_{k_m,l_1}, v_{k_m,l_1}\rangle & \dots & \langle \mu_{k_m,l_n}, \nu_{k_m,l_n}; u_{k_m,l_n}, v_{k_m,l_n}\rangle \end{array}$$

The definition of a 3-D E-IFIM extends the 2-D E-IFIM concept and is identical to those of the 3-D IM, presented in the [3]. Let us introduce some operations over the E-IFIMs.

Transposition [3]: The transposed IM of A is A'.

Let us introduce the following operations over E-IFIMs $A = [K, L, \{\langle \mu_{k_i,l_j}, \nu_{k_i,l_j}; rf_{k_i,l_j}, rs_{k_i,l_j}\rangle\}]$ and $B = [P, Q, \{\langle \rho_{p_r,q_s}, \sigma_{p_r,q_s}\rangle; \delta f_{k_i,l_j}, \delta s_{k_i,l_j}\}]$ with a similar form to that of [3], [46].

Addition-$(\circ_1, \circ_2, *)$**:**

$A \oplus_{(\circ_1,\circ_2,*)} B = [K \cup P, L \cup Q, \{\langle \phi_{t_u,v_w}, \psi_{t_u,v_w}\rangle; \eta f_{t_u,v_w}, \eta s_{t_u,v_w}\}]$, where $\langle \circ_1, \circ_2\rangle \in \{\langle \max, \min\rangle, \langle \min, \max\rangle, \langle \text{ average, average}\rangle\}$ and $* \in \{\max, \min\}$.

$\langle \phi_{t_u,v_w}, \psi_{t_u,v_w}; \eta f_{t_u,v_w}, \eta s_{t_u,v_w}\rangle = \langle \circ_1(\mu_{k_i,l_j}, \rho_{p_r,q_s}), \circ_2(\nu_{k_i,l_j}, \sigma_{p_r,q_s}); *(rf_{t_u,v_w}, \delta f_{t_u,v_w}\rangle), *(rs_{t_u,v_w}, \delta s_{t_u,v_w}\rangle)$.

Termwise subtraction-(max,min):

$$A -_{(\max,\min,*)} B = A \oplus_{(\max,\min,*)} \neg B.$$

Termwise multiplication:

$A \otimes_{(\circ_1,\circ_2,*)} B = [K \cap P, L \cap Q, \{\langle \phi_{t_u,v_w}, \psi_{t_u,v_w}; \eta f_{t_u,v_w}, \eta s_{t_u,v_w}\rangle\}]$, where $\langle \phi_{t_u,v_w}, \psi_{t_u,v_w}; \eta f_{t_u,v_w}, \eta s_{t_u,v_w}\rangle = \langle \circ_1(\mu_{k_i,l_j}, \rho_{p_r,q_s}), \circ_2(\nu_{k_i,l_j}, \sigma_{p_r,q_s}); *(rf_{t_u,v_w}, \delta f_{t_u,v_w}, *(rs_{t_u,v_w}, \delta s_{t_u,v_w}\rangle)$.

The following operations have no equivalents with these verso classical matrices. They are developed to be able to automate certain actions on IMs in order to implement various models and algorithms.

Reduction [3]: An IM A's operations-reduction $(k, \perp)$ is defined as follows:

$A_{(k,\perp)} = [K - \{k\}, L, \{c_{t_u,v_w}\}]$, where $c_{t_u,v_w} = a_{k_i,l_j}(t_u = k_i \in K - \{k\}, v_w = l_j \in L)$.

Projection [3]: Let $M \subseteq K$ and $N \subseteq L$. Then, $pr_{M,N}A = [M, N, \{b_{k_i,l_j}\}]$, where for each $k_i \in M$ and each $l_j \in N$, $b_{k_i,l_j} = a_{k_i,l_j}$.

Substitution [3]: $\left[\frac{p}{k}; \perp\right] A = \left[(K - \{k\}) \cup \{p\}, L, \{a_{k,l}\}\right]$

Internal subtraction of IMs' components [48]:

$IO_{-_{(\max,\min)}}(\langle k_i, l_j, A\rangle, \langle p_r, q_s, B\rangle) = [K, L, \{\langle \gamma_{t_u,v_w}, \delta_{t_u,v_w}\rangle\}]$.

Index type operations [48]:

$AGIndex_{(\max_R^{elliptic}),(\not\perp)}(A) = \langle k_i, l_j\rangle$, where $\langle k_i, l_j\rangle$ (for $1 \le i \le m, 1 \le j \le n$) is the index of the maximum E-IFQ of A in the sense of the relation (1) that has no empty value.

$Index_{(\max_R^{elliptic}),k_i}(A) = \{\langle k_i, l_{v_1}\rangle, \ldots, \langle k_i, l_{v_x}\rangle, \ldots, \langle k_i, l_{v_V}\rangle\}$, where $\langle k_i, l_{v_x}\rangle$ (for $1 \le i \le m, 1 \le x \le V$) are the indices of the largest element in A's k_i-th row.

Aggregation operations Let us extend the operations $\#_q, (q \le i \le 3)$ from [47] such that they can be applied over E-IFQs $x = \langle a, b; rf_1, rs_1\rangle$ and $y = \langle c, d; rf_2, rs_2\rangle$:

$x\#_1, *y = \langle min(a,c), max(b,d); *(rf_1, rf_2), *(rs_1, rs_2)\rangle$;

$x\#_2, *y = \langle average(a,c), average(b,d); *(rf_1, rf_2), *(rs_1, rs_2)\rangle$;

$x\#_3, *y = \langle max(a,c), min(b,d); *(rf_1, rf_2), *(rs_1, rs_2)\rangle$.

Let the fixed index be $k_0 \notin K$. The expanded definition of the aggregation operation $\alpha_{K,\#_q,*}(A, k_0)$ by the dimension K over 3-D E-IFIM A utilizing that of ([46], [47]) is as follows:

$$\begin{array}{c|cc}
h_g \in H & l_1 & \ldots \\
\hline
k_0 & \mathop{\#_{q,*}}\limits_{i=1}^{m} \langle \mu_{k_i,l_1,h_g}, \nu_{k_i,l_1,h_g}; rf_{k_i,l_1,h_g}, rf_{k_i,l_1,h_g}\rangle & \ldots
\end{array},$$

$$\begin{array}{cc}
\ldots & l_n \\
\hline
\ldots & \mathop{\#_{q,*}}\limits_{i=1}^{m} \langle \mu_{k_i,l_n,h_g}, \nu_{k_i,l_n,h_g}; rf_{k_i,l_1,h_g}, rf_{k_i,l_1,h_g}\rangle
\end{array}$$

Aggregate global internal operation [48]: $AGIO_{\oplus_{(\#_q,*)}}(A)$. If $q = 1, q = 2$ or $q = 3$, we get a pessimistic, averaged, or optimistic scenario.

Operation “Purge" of IM A The following is how we define the new operation “Purge” by the dimension K as follows: $Purge_K(A)$ decreases each row k_x of A, if $a_{k_x,l_j} \le a_{k_y,l_j}$, but $a_{k_x,l_e} \ge a_{k_y,l_e}$ for $1 \le x \le m$, $1 \le y \le m$, $1 \le j \le n$ and $1 \le e \le n$.

III. An Elliptic Intuitionistic Fuzzy Portfolio Selection Problem based on Knapsack Problem

In this part, we extend the dynamic programming strategy for C-IFKP [46] to develop an IM interpretation for a set method for 0-1 E-IFKP for the portfolio selection problem. The problem is: An investor has a budget of $Bu = \langle \rho, \sigma; rf_{Bu}, rs_{Bu}\rangle$ to spend on assets. A set of m assets from $\{k_1, \ldots, k_i, \ldots, k_m\}$ must be evaluated by experts $\{d_1, \ldots, d_s, \ldots, d_D\}$ (for $s = 1, ..., D$) with a given IFP rating $re_s = \langle \delta_s, \varepsilon_s\rangle$ $(1 \le s \le D)$ using the criteria c_1 and c_2. Let us use the symbols a_{k_i,c_1} (for $i = 1, ..., m$) and a_{k_i,c_2} (for $i = 1, ..., m$) to represent the return and the price, respectively, of the k_i-th asset. The objective of the problem is to choose the investor's portfolio's assets as optimally as possible while staying within his financial constraints. The parameters of this optimization problem are highly unknown because of market dynamics. Through the use of the E-IF logic toolkit, helped by IMs, it is required to look for the best answer for the investment portfolio with three decision-making scenarios (optimistic, averaged, and pessimistic).

A. An Elliptic IF Portfolio Selection Problem with a Dynamic Programming Approach Using a Type E-IFKP

The following are the stages in the IM interpretation for the Elliptic Intuitionistic Fuzzy Portfolio Selection Problem, based on the Circular IFKP technique [46]:

Step 1. 3-D evaluation IFIM $EV[K, C, E, \{ev_{k_i,c_j,d_s}\}]$ is created in compliance with the aforementioned problem, where $K = \{k_1, k_2, \ldots, k_m\}$, $C = \{c_1, c_2\}$, $E = \{d_1, \ldots, d_s, \ldots, d_D\}$ and the element $\{ev_{k_i,c_j,d_s}\} = \langle \mu_{k_i,c_j,d_s}, \nu_{k_i,c_j,d_s}\rangle$ (for $1 \le i \le m, 1 \le j \le n, 1 \le s \le D$) is the estimate of the d_s-th expert for the k_i-th asset by the c_j-th criterion $(j = 1, 2)$. Due to changes in some uncontrolled elements, the expert is unsure about the evaluation according to a particular criterion, and his evaluations take the shape of IFPs. Let the s-th expert's score (rating) $re_s, s \in E$ be specified by an IFP $\langle \delta_s, \varepsilon_s\rangle$, where δ_s and ε_s are considered as the expert's level of competence and incompetence, respectively.

Next, we calculate

$$EV^*[K, C, E, \{ev^*{}_{k_i,c_g,d_s}\}] = re_1 pr_{K,C,d_1} EV \oplus_{(\circ_1,\circ_2)} \cdots$$
$$\cdots \oplus_{(\circ_1,\circ_2)} re_D pr_{K,C,d_D} EV.$$
$$EV := EV^*(ev_{k_i,l_j,d_s} = ev^*_{k_i,l_j,d_s}, \ \forall k_i \in K, \forall l_j \in L, \forall d_s \in E).$$

The degrees of membership and non-membership of the E-IFQs are determined using the three aggregating operations $\alpha_{K,\#_1,*}, \alpha_{K,\#_3,*}$ and $\alpha_{K,\#_2,*}$, which provide the evaluations of the k_i-th asset against the c_j-th criterion in a present moment $h_f \notin E$:

$PI_{min}[K, C, h_m, \{pi_{min_{k_i,c_g,h_f}}\}] = \alpha_{E,\#_1}(EV^*, h_m)$

$$= \begin{array}{c|cc}
h_m & c_1 & c_2 \\
\hline
k_1 & \mathop{\#_1}\limits_{s=1}^{D} \langle \mu_{k_1,c_1,d_s}, \nu_{k_1,c_1,d_s}\rangle & \mathop{\#_1}\limits_{s=1}^{D} \langle \mu_{k_1,c_2,d_s}, \nu_{k_1,c_2,d_s}\rangle \\
\vdots & \vdots & \vdots \\
k_m & \mathop{\#_1}\limits_{s=1}^{D} \langle \mu_{k_m,c_1,d_s}, \nu_{k_m,c_1,d_s}\rangle & \mathop{\#_1}\limits_{s=1}^{D} \langle \mu_{k_m,c_2,d_s}, \nu_{k_m,c_2,d_s}\rangle
\end{array}.$$

$PI_{max}[K, C, h_m, \{pi_{max_{k_i,c_g,h_f}}\}] = \alpha_{E,\#_3}(EV^*, h_m)$=

$$\begin{array}{c|cc}
h_m & c_1 & c_2 \\
\hline
k_1 & \mathop{\#_3}\limits_{s=1}^{D} \langle \mu_{k_1,c_1,d_s}, \nu_{k_1,c_1,d_s}\rangle & \mathop{\#_3}\limits_{s=1}^{D} \langle \mu_{k_1,c_2,d_s}, \nu_{k_1,c_2,d_s}\rangle \\
\vdots & \vdots & \vdots \\
k_m & \mathop{\#_3}\limits_{s=1}^{D} \langle \mu_{k_m,c_1,d_s}, \nu_{k_m,c_1,d_s}\rangle & \mathop{\#_3}\limits_{s=1}^{D} \langle \mu_{k_m,c_2,d_s}, \nu_{k_m,c_2,d_s}\rangle
\end{array}.$$

Then construct $PI* = PI_{min} \oplus_{(\circ_1,\circ_2,*)} PI_{max}$ and $PI[K, C, h_f, \{pi_{k_i,c_g,h_f}\}] = \alpha_{E,\#_2}(PI^*, h_f)$, whose elements are the coordinates of the centers of the E-IFQs evaluating the shares.

We now determine E-IFIM $A[K, C\{a_{k_i,c_g}\}]$, which represents current evaluations of the assets utilizing the approach from [5] by criteria for return and price:

$$\begin{array}{c|cc} & c_1 & c_2 \\ \hline k_1 & \langle \mu^a_{k_1,c_1}, \nu^a_{k_1,c_1}; rf^a_{k_1,c_1}, rs^a_{k_1,c_1}\rangle & \langle \mu^a_{k_1,c_2}, \nu^a_{k_1,c_2}; rf^a_{k_1,c_2}, rs^a_{k_1,c_2}\rangle \\ \vdots & \vdots & \vdots \\ k_m & \langle \mu^a_{k_m,c_1}, \nu^a_{k_m,c_1}; rf^a_{k_m,c_1}, rs^a_{k_m,c_1}\rangle & \langle \mu^a_{k_m,c_2}, \nu^a_{k_m,c_2}; rf^a_{k_m,c_2}, rs^a_{k_m,c_2}\rangle \end{array},$$

where $K = \{k_1, \dots, k_i, \dots, k_m\}, i = 1, \dots, m; C = \{c_1, c_2\}$, a_{k_i,c_g} (for $i = 1, \dots, m; g = 1, 2$) are created as E-IFQs by converting the IFPs pi_{k_i,c_j,d_s} using the following steps

for $g = 1$ to 2, $i = 1$ to m $\Big\{ \mu^a_{k_i,c_g} = \mu^{pi}_{k_i,c_g,h_f}; \nu^a_{k_i,c_g} = \nu^{pi}_{k_i,c_g,h_f},$
$rf^a_{k_i,c_g}$

$$= \sqrt{\min_{1\le s\le D} \mu^{ev}_{k_i,c_g,d_s}{}^2 + \left\{ \frac{\max_{1\le s\le D} \mu^{ev}_{k_i,c_g,d_s}{}^2 - \min_{1\le s\le D} \mu^{ev}_{k_i,c_g,d_s}{}^2}{\max_{1\le s\le D} \nu^{ev}_{k_i,c_g,d_s}{}^2 - \min_{1\le s\le D} \nu^{ev}_{k_i,c_g,d_s}{}^2} \right\}^2 \cdot \min_{1\le s\le D} \nu^{ev}_{k_i,c_g,d_s}{}^2}$$

and $rs^a_{k_i,c_g}$

$$= \sqrt{\min_{1\le s\le D} \mu^{ev}_{k_i,c_g,d_s}{}^2 \cdot \left\{ \frac{\max_{1\le s\le D} \nu^{ev}_{k_i,c_g,d_s}{}^2 - \min_{1\le s\le D} \nu^{ev}_{k_i,c_g,d_s}{}^2}{\max_{1\le s\le D} \mu^{ev}_{k_i,c_g,d_s}{}^2 - \min_{1\le s\le D} \mu^{ev}_{k_i,c_g,d_s}{}^2} \right\}^2 + \min_{1\le s\le D} \nu^{ev}_{k_i,c_g,d_s}{}^2}$$

$\Big\}$

The input data for the portfolio's budget is then checked to ensure that it does not exceed the investor's specified budget, Bu. If the price of a given asset k_i exceeds the budget Bu, then the corresponding row of IM A is reduced by it.

for $i = 1$ to m {If $a_{k_i,c_2} > Bu$ then $A_{(k_i,\perp)}$ }

Let us denote by $|K| = m$ the number of the elements of the set K, then $|C| = 2$. As well, we define E-IFIM $X[K, C]$ containing the elements x_{k_i,c_g} (for $1 \le i \le m$, $\le g \le 2$) and:

$$\{x_{k_i,c_g}\} \in \begin{cases} \langle 1, 0; 0, 0\rangle, & \text{if the request } k_i \text{ is selected} \\ \langle 0, 1; 0, 0\rangle & \text{otherwise} \end{cases}$$

Let us assume that at the start of the algorithm, all components of IM X are identical to $\langle 0, 1; 0, 0\rangle$.

Create IM

$$S^0[u_0, L] = \begin{array}{c|cc} & c_1 & c_2 \\ \hline u_0 & s^0_{u_0,c_1} & s^0_{u_o,c_2} \end{array} = \begin{array}{c|cc} & c_1 & c_2 \\ \hline u_0 & \langle 0, 1; 0, 0\rangle & \langle 0, 1; 0, 0\rangle \end{array}$$

Step 2. for $i = 1$ to m do $\big\{$Create IMs

$$R_i[k_i, C] = pr_{k_i,C}A; SH_1^{i-1} = \left[\frac{u_i}{u_{i-1}}; \perp\right] S^{i-1}$$

for $h = 1$ to $i + 1$ do {

$$SH_1^{i-1} = SH_1^{i-1} \oplus_{(\circ_1,\circ_2,*)} \left[\frac{u_h}{k_i}; \perp\right] R_i$$

}

$$S^i[U^i, L] = S^{i-1} \oplus_{((\circ_1,\circ_2,*))} SH_1^{i-1};$$

for $h = 1$ to $i + 1$ do $\big\{$ *Checks the conditions for the capacity of the knapsack*

If $s^i_{h,w} > Bu$ then $S^i_{(h,\perp)}$ $\big\}$

The "Purge" procedure is currently underway by $S^i = Purge_{U^i} S^i$ } Go to *Step 3.*

Step 3. This step finds the index of the highest stock return by

$$Index_{(\max_{R elliptic}),c_1}(A) = \langle u_g, c_1\rangle$$

for $i = m$ to 1 do

{Find the α-nearest elements of $s^i_{u_g,c_1}$ (or $s^i_{u_g,c_2}$) ($\alpha = 0.5$) of S^i and choose the closest element from them - $s^i_{u_{g*},c_1}$ (or $s^i_{u_{g*},c_2}$).

If $\{s^i_{u_g*,c_1}$ (or $s^i_{u_{g*},c_2}$)$\} \in S^i$ and $\{s^i_{u_g*,c_1}$ (or $s^i_{u_{g*},c_2}$)$\} \notin S^{i-1}$ then

$$\{x_{k_i,p} = \langle 1, 0; 0, 0\rangle \text{ and } x_{k_i,w} = \langle 1, 0; 0, 0\rangle;$$
$$s^i_{u_g,c_1} = s^i_{u_g*,c_1} -_* a_{k_i,c_1}; s^i_{u_g,c_2} = s^i_{u_g*,c_2} -_* a_{k_i,c_2}\}$$

Go to *Step 4.*

Step 4. The optimal return and price of the investment portfolio are:

$$AGIO_{\oplus_{(\#_q,*)}}\left(pr_{K,c_1}A \otimes_{(\circ_1,\circ_2,*)} pr_{K,c_1}X\right);$$
$$AGIO_{\oplus_{(\#_q,*)}}\left(pr_{K,c_2}A \otimes_{(\circ_1,\circ_2,*)} pr_{K,c_2}X\right).$$

If $q = 1, q = 2$ or $q = 3$ then we determine the optimal benefit's pessimistic, averaging, or optimistic value. The optimistic scenario has been accepted if $\langle \circ_1, \circ_2\rangle = \langle \max, \min\rangle$ is used in all operations of the algorithm. On the other hand, if $\langle circ_1, circ_2\rangle = \langle \min, \max\rangle$ is employed, the pessimistic scenario is used. The operation "$* = \max$" is used when there is more ambiguity, else "$* = \min$". Therefore several optimal solutions could be generated according to the investor's opinion. Thus, an investor may have greater confidence in the obtained solution.

The complexity of the normal dynamic programming algorithm [28], [59] and the E-IFKP approach are both comparable - $O(m.C)$.

To study how the algorithm affects a range of input data, we are now developing software that uses the E-IFKP approach. After reading a file with a matrix of the return predictions and stock price, it completes the aforementioned procedures. It was developed in C++. This objective was accomplished by building an IM structure using the STL's std::tuple types. This structure is then used to create the fundamental IM protocols [29]. The app requires the share E-IFQs and the knapsack budget Bu as input. When the program is finished, a suggested solution is displayed on the computer screen along with a thorough output of each algorithm iteration.

In the scientific literature, no portfolio optimization model on elliptic IF fuzzy data was found, a suitable tool for representing vague or incomplete data in conditions of large fluctuations in market parameters. In this model, there are three scenarios according to the attitudes of the decision maker. Evaluations of the returns and the price of financial assets for the purpose of optimal selection of the portfolio are carried out by experts and their rating is taken into account in the evaluation process. Therefore, the developed E-IF optimal portfolio selection task is socially oriented and reflects the preferences of both the experts and the decision maker. The Markowitz [27] portfolio cannot be applied under conditions of fuzziness and large parameter fluctuations and his model cannot reflect the investor's attitude towards the market environment – whether it is pessimistic, optimistic or average.

B. An E-IFKP Real Case Study for Portfolio Selection

Here, a real case study for the best asset selection for the investor's portfolio within his budget $B = \langle 0.99, 0.0; \sqrt{2}, \sqrt{2}\rangle$ clarifies the proposed E-IFKP in this part. A group of the experts $\{d_1, d_2, d_3\}$ with the specified IFP rating $re_s = \langle \delta_s, \varepsilon_s\rangle$ $(1 \le s \le D)$ is required to evaluate a set of assets $\{k_1, k_2, k_3, k_4\}$ from the IT firms Microsoft Corp., Apple Inc, Cisco Systems Inc., and Intel Corporation, which make up the Dow Jones Industrial Average for the last 5 years, by the criteria c_1 and c_2: the return a_{k_i,c_1} (for $i = 1,...,m$) of the k_i-th asset and its price as a_{k_i,c_2} (for $i = 1,...,m$). The objective of the problem is to choose the investor's portfolio's assets as efficiently as possible while staying within his financial constraints using three different decision-making scenarios.

The solution to the problem:

Step 1. The initial form of the 3-D evaluation IFIM $EV[K, C, E, \{ev_{k_i,c_g,d_s}\}]$ is the following:

$$EV = \left\{ \begin{array}{c|cc} d_1 & c_1 & c_2 \\ \hline k_1 & \langle 0.54, 0.29\rangle & \langle 0.35, 0.48\rangle \\ k_2 & \langle 0.552, 0.31\rangle & \langle 0.38, 0.514\rangle \\ k_3 & \langle 0.546, 0.265\rangle & \langle 0.399, 0.461\rangle \\ k_4 & \langle 0.486, 0.316\rangle & \langle 0.144, 0.694\rangle \end{array} \right. ,$$

$$\begin{array}{c|cc} d_2 & c_1 & c_2 \\ \hline k_1 & \langle 0.504, 0.365\rangle & \langle 0.2, 0.752\rangle \\ k_2 & \langle 0.504, 0.32\rangle & \langle 0.238, 0.74\rangle \\ k_3 & \langle 0.675, 0.154\rangle & \langle 0.0099, 0.865\rangle \\ k_4 & \langle 0.672, 0.157\rangle & \langle 0.024, 0.96\rangle \end{array} ,$$

$$\left. \begin{array}{c|cc} k_1 & \langle 0.65, 0.13\rangle & \langle 0.035, 0.944\rangle \\ k_2 & \langle 0.081, 0.91\rangle & \langle 0.0019, 0.982\rangle \\ k_3 & \langle 0.072, 0.91\rangle & \langle 0.0024, 0.99\rangle \\ k_4 & \langle 0.126, 0.823\rangle & \langle 0.007, 0.98\rangle \end{array} \right\} ,$$

$\{ev_{k_i,c_j,d_s}\}$ (for $1 \le i \le 4, 1 \le g \le 2, 1 \le s \le 3$) is the expert's assessment in accordance with the c_g-th criterion for the k_i-th stock. Assign the experts the corresponding rating coefficients shown below:

$$\{r_1, r_2, r_3\} = \{\langle 0.9, 0.1\rangle, \langle 0.8, 0.08\rangle, \langle 0.7, 0.07\rangle\}.$$

We create

$$EV^*[K, C, E, \{ev^*{}_{k_i,c_g,d_s}\}]$$
$$= re_1 pr_{K,C,d_1} EV \oplus_{(\circ_1,\circ_2)} \dots \oplus_{(\circ_1,\circ_2)} re_D pr_{K,C,d_D} EV;$$
$$EV := EV^*.$$

Then, we determine E-IFIM $A[K, C]$, which consists of the assessments of the shares made at a moment h_f by 2 criteria:

$$\begin{array}{c|cc} & c_1 & c_2 \\ \hline k_1 & \langle 0.55, 0.29; 0.56, 0.98\rangle & \langle 0.38, 0.49; 0.47, 0.68\rangle \\ k_2 & \langle 0.5, 0.34; 0.51, 0.97\rangle & \langle 0.19, 0.72; 0.33, 0.77\rangle \\ k_3 & \langle 0.66, 0.14; 0.8, 0.22\rangle & \langle 0.02, 0.91; 0.011, 1.76\rangle \\ k_4 & \langle 0.099, 0.87; 0.09, 1.3\rangle & \langle 0.004, 0.99; 0.003, 1.21\rangle \end{array} ,$$

where $K = \{k_1, k_2, k_3, k_4\}$, $C = \{c_1, c_2\}$ and $\{a_{k_i,c_1}, a_{k_i,c_2}\}$ are, respectively, the k_i-th stock's price and E-IF return. $X[K, C]$ is formed with the elements $\langle 0, 1; 0, 0\rangle$.

TABLE I
THE RESULTS FOR THE OPTIMISTIC, PESSIMISTIC, AND AVERAGE SCENARIOS.

Scenario	c_1	c_2
Optimistic	$\langle 0.099, 0.867; 0.56, 1.31\rangle$	$\langle 0.19, 0.72; 0.47, 0.77\rangle$
Pessimistic	$\langle 0.099, 0.867; 0.093, 0.966\rangle$	$\langle 0.19, 0.72; 0.33, 0.685\rangle$
Average	$\langle 0.38, 0.498; 0.389, 1.085\rangle$	$\langle 0.19, 0.73; 0.269, 0.89\rangle$

Step 2. The following IMs are calculated consecutively by the algorithm: $S_1^0[u_1, C], S^1[U_1, C], S_1^1[U{*}_1, C], S^2[U_2, C]$. The operation "Purge" has reduced the u_2 and u_3 row of $S^2[U_2, C]$. Then IMs are created: $S_1^2[U{*}_2, C], S^3[U_3, C]$. The operation "Purge" has reduced the u_4 row of $S^3[U_3, C]$. Then IMs are created: $S_1^3[U{*}_3, C], S^4[U_4, C]$. The following is the final IM $S^4[U{*}_4, C]$ after the "Purge" operation:

$$\begin{array}{c|cc} & c_1 & c_2 \\ \hline u_1 & \langle 0.546, 0.285; 0.56, 0.99\rangle & \langle 0.375, 0.487; 0.47, 0.69\rangle \\ u_2 & \langle 0.495, 0.341; 0.51, 0.97\rangle & \langle 0.191, 0.723; 0.33, 0.69\rangle \\ u_3 & \langle 0.546, 0.285; 0.56, 0.22\rangle & \langle 0.022, 0.913; 0.01, 0.69\rangle \\ u_4 & \langle 0.099, 0.867; 0.09, 0.99\rangle & \langle 0.004, 0.986; 0.003, 0.69\rangle \end{array}$$

Step 3. In this step, using the results from Step 2. we determine that the fourth, second, and first IT businesses' stocks are included in the investor's ideal portfolio in this problem.

Step 4. The outcomes for the optimistic, pessimistic, and average scenarios for the greatest benefit are shown in the following table (cf. table I):

In conditions of high inflation and great uncertainty, the decision-maker will choose the pessimistic scenario, in case of small fluctuations in the market parameters, the decision-maker will prefer the averaged scenario, and in the case of stability of the market parameters, the optimistic scenario will be preferred.

A comparative analysis between the proposed E-IFKP method for portfolio optimization could not be performed because we could not find methods for similar type of problems under conditions of high uncertainty modeled by E-IF logic.

After the results are obtained in E-IFKP portfolio method, the question arises whether small deviations in the values of the input parameters used change the results of the model. Checking the robustness of the results in the developed model and analyzing the sensitivity to the changes in the input variables of the obtained results is a critical step for E-IFKP portfolio problem.

The weights of the experts are of great importance on the results of the E-IFKP portfolio problem. A sensitivity analysis consisting of 8 different scenarios has been conducted to analyze the effect of the change in weight of each expert on the ranking results. In the analysis, a total of 8 different changes have been applied in the weights of the three experts included in the study, and the final results are different in the cases indicated. Based on the software final results in these cases, we can conclude that the optimal portfolio selections in the described cases differ. The results of the software show that there is sensitivity in the output results when including 1, 2 and 4 assets; or 1 and 4 assets; 2 and 4 assets. In some of these cases, the optimization problem is invalid from the point of view of IF logic.

A sensitivity analysis was performed by changing the input data by $\pm10\%, \pm25\%, \pm50\%$ and $\pm75\%$ respectively. In all these cases, the input data was invalid from the IF point of view.

IV. Conclusion

A 0-1 E-IFKP approach for portfolio selection was established in this study expanding 0-1 C-IFKP from [46] and the classical dynamic optimization algorithm for this problem [59]. The software being developed for the performance of the E-IFKP approach is applied to a real case for the selection of portfolio shares of the IT companies which make up the Dow Jones Industrial Average. Three scenarios are proposed to the decision maker for the final choice - pessimistic, optimistic, and average. Future research will include expanding this E-IFKP technique to three-dimensional intuitionistic fuzzy data [3] as well as developing software for its implementation.

References

[1] K. Atanassov, "Intuitionistic Fuzzy Sets," VII ITKR Session, Sofia, 20-23 June 1983 (Deposed in Centr. Sci.-Techn. Library of the Bulg. Acad. of Sci., 1697/84) (in Bulgarian). Reprinted: *Int. J. Bioautomation,* vol. 20(S1), 2016, pp. S1-S6.

[2] K. Atanassov, "Generalized index matrices," *Comptes rendus de lÁcademie Bulgare des Sciences,* vol. 40(11), 1987, pp. 15-18.

[3] K. Atanassov, "Index Matrices: Towards an Augmented Matrix Calculus," *Studies in Computational Intelligence,* Springer, Cham, vol. 573, 2014, DOI: 10.1007/978-3-319-10945-9.

[4] K. Atanassov, "Circular Intuitionistic Fuzzy Sets," *Journal of Intelligent & Fuzzy Systems,* vol. 39 (5), 2020, pp. 5981-5986.

[5] K. Atanassov, "Elliptic Intuitionistic fuzzy sets," *Comptes rendus de l'Academie bulgare des Sciences,* vol. 74 (6), 2021, pp. 812-819.

[6] K. Atanassov K, E. Marinov, "Four Distances for Circular Intuitionistic Fuzzy Sets," *Mathematics,* vol. 9 (10), 2021, pp. 11-21, DOI: 10.3390/math9101121.

[7] K. Atanassov, E. Szmidt, J. Kacprzyk, "On intuitionistic fuzzy pairs," *Notes on Intuitionistic Fuzzy Sets,* vol. 19 (3), 2013, pp. 1-13.

[8] K. Atanassov, G. Gargov, "Interval valued intuitionistic fuzzy sets," *Fuzzy sets and systems,* vol. 31 (3), 1989, 343-349.

[9] D. Chakraborty, V. Singh, "On solving fuzzy knapsack problem by multistage decision making using dynamic programming," *AMO,* vol. 16(3), 2014, pp. 575-585.

[10] K-S Chen, Y-Y Huang, R-C Tsaur, N-Y Lin, "Fuzzy Portfolio Selection in the Risk Attitudes of Dimension Analysis under the Adjustable Security Proportions," *Mathematics*, vol. 11 (5), 2023, 1143.

[11] K.-S. Chen, R.-C. Tsaur, N.-C. Lin," Dimensions analysis to excess investment in fuzzy portfolio model from the threshold of guaranteed return rates," *Mathematics,* vol. 11, 2023, 44.

[12] C. B. Cuong, V. Kreinovich, "Picture fuzzy sets-a new concept for computational intelligence problems," In: *Proceedings of the Third World Congress on Information and Communication Technologies WICT'2013*, Hanoi, Vietnam, 2013, pp. 1-6.

[13] G. Dantzig, *Linear programming and extensions,* Princeton University Press Oxford; 1963.

[14] S. Fidanova, K. Atanassov, "ACO with Intuitionistic Fuzzy Pheromone Updating Applied on Multiple-Constraint Knapsack Problem," *Mathematics,* vol. 9 (13), 2021, pp. 1456.

[15] P. Gilmore, R. Gomory, "The theory and computation of knapsack functions," *Operations research,* vol. 14, 1966, pp. 1045-1074.

[16] S. Guo, W.-K. Ching, W.-K. Li, T.-K. Siu, Z. Zhang, "Fuzzy hidden Markov-switching portfolio selection with capital gain tax," *Expert Syst. Appl.*, vol. 149, 2020, 113304.

[17] P. Gupta, M. K. Mehlawat, S. Yadav, A. Kumar, "A polynomial goal programming approach for intuitionistic fuzzy portfolio optimization using entropy and higher moments," *Appl. Soft Comput.,* vol. 85, 2019, 105781.

[18] D. Goldfarb, G. Iyengar, "Robust portfolio selection problems," *Mathematics of operations research,* vol. 28 (1), 2003, pp. 1–38.

[19] M. B. Gorzałczany, "A Method of Inference in Approximate Reasoning Based on Interval-Valued Fuzzy Sets," *Fuzzy Sets Syst.,* vol. 21, 1987, pp. 1–17.

[20] Y. Hanine, Y. Lamrani Alaoui, M. Tkiouat, "Lahrichi, Y. Socially Responsible Portfolio Selection: An Interactive Intuitionistic Fuzzy Approach," *Mathematics,* vol. 9 (23), 2021, pp. 1-13.

[21] Y.-Y. Huang, I.-F. Chen, C.-L. Chiu, R.-C. Tsaur, "Adjustable security proportions in the fuzzy portfolio selection under guaranteed return rates," *Mathematics,* vol. 9, 2021, 3026.

[22] D. Kuchta, "A generalization of an algorithm solving the fuzzy multiple choice knapsack problem," *Fuzzy sets and systems,* vol. 127 (2), 2002, pp. 131-140.

[23] J. Li, "Multi-Objective Portfolio Selection Model with Fuzzy Random Returns and a Compromise Approach-Based Genetic Algorithm," *Inf. Sci.,*, vol. 220, 2013, pp. 507–521.

[24] X. Li, Z. Qin, S. Kar, "Mean-variance-skewness model for portfolio selection with fuzzy returns," *Eur. J. Oper. Res.,* vol. 202, 2010, 239–247.

[25] T. Mahmood, S. Abdullah, S. -ur-Rashid, M. Bilal, "Multicriteria decision making based on a cubic set," *Journal of New Theory*, vol. 16, 2017, pp. 1-9.
Man

[26] N. Mansour, M. S. Cherif, W. Abdelfattah, "Multi-objective imprecise programming for financial portfolio selection with fuzzy returns," *Expert Syst. Appl.,* vol. 138, 2019, 112810.

[27] HM. Markowitz, Portfolio selection: Efficient diversification of investment, *John Wiley & Sons,* New York, USA; 1959.

[28] S. Martello, P. Toth, *Knapsack problems,* Algorithms and computer implementations, John Wiley & Sons; 1990.

[29] D. Mavrov, "An Application for Performing Operations on Two-Dimensional Index Matrices," *Annual of "Informatics" Section, Union of Scientists in Bulgaria,* vol. 10, 2019 / 2020, pp. 66-80.

[30] R. Mehralizade, M. Amini, B. S. Gildeh, H. Ahmadzade, "Uncertain random portfolio selection based on risk curve," *Soft Comput.,* vol. 24, 2020, 13331–13345.

[31] G. Michalski, "Portfolio Management Approach in Trade Credit Decision Making," *Romanian J. Econ. Forecast.* vol. 3, 2007, 42–53.

[32] A. Mucherino, S. Fidanova, M. Ganzha, "Ant colony optimization with environment changes: An application to GPS surveying," *Proceedings of the 2015 FedCSIS*, 2015, pp. 495 - 500.

[33] X. T. Nguyen, V. D. Nguyen, "Support-Intuitionistic Fuzzy Set: A New Concept for Soft Computing," *I.J. Intelligent Systems and Applications,* 2015, 04, 2015, pp. 11-16.

[34] M. Pandey, V. Singh, N. K. Verma, "Fuzzy Based Investment Portfolio Management," *Fuzzy Manag. Methods,* 2019, pp. 73–95.

[35] MC. Pınar, "Robust scenario optimization based on downside-risk measure for multi-period portfolio selection," *OR Spectrum,* vol. 29(2), 2007, 295–309.

[36] J. Razmi, E. Jafarian, S. H. Amin, "An intuitionistic fuzzy goal programming approach for finding Pareto-optimal solutions to multi-objective programming problems," *Expert Syst. Appl.,* vol. 65, 2016, pp. 181–193.

[37] M. Rahiminezhad Galankashi, F. Mokhatab Rafiei, M. Ghezelbash, "Portfolio Selection: A Fuzzy-ANP Approach," *Financ. Innov.,* vol. 6 (17), 2020, pp. 1-34.

[38] V. Singh, "An Approach to Solve Fuzzy Knapsack Problem in Investment and Business Model," *in: Nogalski, B., Szpitter, A., Jaboski, A., Jaboski, M. (eds.),* Networked Business Models in the Circular Economy, 2020. DOI: 10.4018/978-1-5225-7850-5.ch007

[39] V. P. Singh, D. Chakraborty, "A Dynamic Programming Algorithm for Solving Bi-Objective Fuzzy Knapsack Problem," *in: Mohapatra, R., Chowdhury, D., Giri, D. (eds.),* Mathematics and Computing. Proceedings in Mathematics & Statistics, Springer, New Delhi, vol. 139, 2015, pp. 289-306.

[40] F. Smarandache, *Neutrosophy. Neutrosophic Probability,* Set, and Logic, Amer. Res. Press, Rehoboth, USA; 1998.

[41] E. Szmidt, J. Kacprzyk, "Amount of information and its reliability in the ranking of Atanassov's intuitionistic fuzzy alternatives," *in: Rakus-Andersson, E., Yager, R., Ichalkaranje, N., Jain, L.C. (eds.),* Recent Advances in Decision Making, SCI, Springer, vol. 222, 2009, pp. 7–19.

[42] H. Tanaka, P. Guo, IB Türksen, "Portfolio selection based on fuzzy probabilities and possibility distributions," *Fuzzy sets and systems,* vol. 111(3), 2000, 387–397.

[43] F. Tiryaki, B. Ahlatcioglu, "Fuzzy portfolio selection using fuzzy analytic hierarchy process," *Information Sciences,* vol. 179 (1–2), 2009, 53–69.

[44] V. Torra, “Hesitant fuzzy sets,” *International Journal of Intelligent Systems*, vol. 25 (6), 2010, pp. 529-539.

[45] V. Traneva, P. Petrov, S. Tranev, “Intuitionistic Fuzzy Knapsack Problem through the Index Matrices Prism,” *in: I. Georgiev, M. Datcheva, Kr. Georgiev, G. Nikolov (eds.),* Proceedings of 10th International Conference NMA 2022, Borovets, Bulgaria, Lecture Notes in Computer Science, Springer, Cham, vol. 13858, 2023, pp. 314-326.

[46] V. Traneva, P. Petrov, S. Tranev, “Circular IF Knapsack problem,” *Lecture Notes in Computer Science,* Springer, Cham, vol. 758, 2023. (in press)

[47] V. Traneva, S. Tranev, M. Stoenchev, K. Atanassov, “ Scaled aggregation operations over two- and three-dimensional index matrices,” *Soft computing,* vol. 22, 2019, pp. 5115-5120.

[48] V. Traneva, S. Tranev, *Index Matrices as a Tool for Managerial Decision Making,* Publ. House of the USB; 2017 (in Bulgarian).

[49] R. C. Tsaur, C.-L. Chiu, Y.-Y Huang, “Guaranteed rate of return for excess investment in a fuzzy portfolio analysis,” *Int. J. Fuzzy Syst.,*vol. 23, 2021, 94–106.

[50] S. Utz, M. Wimmer, M. Hirschberger, R. E. Steuer, “Tri-Criterion Inverse Portfolio Optimization with Application to Socially Responsible Mutual Funds,” *Eur. J. Oper. Res.,*, vol. 234, 2014, pp. 491–498.

[51] F. Vaezi, S. Sadjadi, A. Makui, “A portfolio selection model based on the knapsack problem under uncertainty,” *PLOS ONE,* vol. 14 (5), 2019, pp. 1-19, DOI: 10.1371/journal.pone.0213652

[52] P. Vassilev, K. Atanassov, *Modifications and extensions of Intuitionistic Fuzzy Sets,* “Prof. Marin Drinov” Academic Publishing House, Sofia, 2019.

[53] X. Xu, Y. Lei, W. Dai, “Intuitionistic Fuzzy Integer Programming Based on Improved Particle Swarm Optimization,” *J. Comput. Appl.,*, vol. 9, 2008, pp. 062.

[54] G.-F. Yu, D.-F. Li, D.-C. Liang, G.-X. Li, “An Intuitionistic Fuzzy Multi-Objective Goal Programming Approach to Portfolio Selection,” *Int. J. Inf. Technol. Decis. Mak.,* vol. 20, 2021, pp. 1477–1497.

[55] W. Yue, Y. Wang, H. Xuan, “Fuzzy multi-objective portfolio model based on semi-variance–semi-absolute deviation risk measures," *Soft Computing,* vol. 23, 2019, pp. 8159–8179.

[56] L. Zadeh, “Fuzzy Sets,” *Information and Control,* vol. 8 (3), 1965, pp. 338-353.

[57] Y. Zhang, X. Li, S. Guo, “Portfolio Selection Problems with Markowitz’s Mean–Variance Framework: A Review of Literature,” *Fuzzy Optim. Decis. Mak.*, vol. 17, 2018, pp. 125–158.

[58] W. Zhou, Z. Xu, “Score-Hesitation Trade-off and Portfolio Selection under Intuitionistic Fuzzy Environment,” *Int. J. Intell. Syst.,* vol. 34, 2019, pp. 325–341.

[59] Knapsack problem using dynamic programming, https://codecrucks.com/knapsack-problem-using-dynamic-programming/. Last accessed 18 May 2023

DOI: 10.15439/2023F9747
ISSN 2300-5963 ACSIS, Vol. 37

An Elliptic Intuitionistic Fuzzy Model for Franchisor Selection

Velichka Traneva
Prof. Asen Zlatarov University
1 Prof. Yakimov Blvd, Burgas 8000, Bulgaria
Email: veleka13@gmail.com

Stoyan Tranev
Prof. Asen Zlatarov University
1 Prof. Yakimov Blvd, Burgas 8000, Bulgaria
Email: tranev@abv.bg

***Abstract*—Choosing a successful franchise company in the ever-changing business environment is a challenge for any investor. The work suggests the creation of an optimal algorithm (E-IFFr) for selecting a franchise company using the concepts of index matrices and elliptic intuitionistic fuzzy sets for modeling this variability in the business environment to optimally solve this optimal problem with elliptic intuitionistic fuzzy parameters. The E-IFFr approach involves experts with dynamic ranks performing evaluations by the selection criteria while also taking into consideration the relative importance of the criteria for each investor. The efficacy of the suggested strategy is demonstrated by a numerical example of the best franchisor selection for the courier business. In an optimistic, average, and pessimistic scenario, the investor has three options to choose from.**

I. Introduction

A PROFITABLE company strategy for entering new markets is franchising. An entrepreneur looking for a franchisor must make the best decision possible for the franchise firm. The developed theory of fuzzy logic [34] is a useful tool for working with incomplete or ambiguous information. The concept of fuzzy logic has successfully been utilized in multi-criteria decision-making problems because human judgments are usually not precise when choosing an alternative concerning multiple criteria with different levels of significance. An approach that is suitable for solving multi-criteria decision-making problems characterized by fuzzy criteria is introduced in [18], based on linguistic criteria values. An Analytic Hierarchical Process (AHP) and neural networks are used in the studies [15], [16] to develop fuzzy franchisee selection models.

Real-world situations typically involve some degree of hesitation between membership and non-membership since decision-makers frequently voice their opinions even when they are undecided about them [33]. One of the first generalizations of fuzzy sets, intuitionistic fuzzy sets (IFSs), exhibit some hesitancy. They are a more potent tool for illustrating environmental uncertainty. We have proposed a software application for the resolution of an optimal interval-valued intuitionistic fuzzy multicriteria outsourced decision-making problem in the paper [28]. Additionally, utilizing the concept of index matrices (IMs, [2]), we have developed an intuitionistic fuzzy approach (IFIMFr) and software to choose the most qualified franchise candidates (see [25], [26]). The study presents an integrated approach [19], based on stepwise weight assessment ratio analysis (SWARA) and complex proportional assessment (COPRAS) approaches, for the selection of optimal bioenergy production technology alternatives. The parameters of the contemporary economic environment are rife with uncertainty. For modeling optimal algorithms, the apparatus of intuitionistic fuzzy sets is insufficient. “Extensions” of the IFSs are detailed in the study [32] and contrasted with one another. The authors of [32] have shown that a Hesitant Fuzzy Set can be completely described by IFS [24]. In [32], the authors further demonstrate that interval-valued IFSs (IVIFSs) [6] can represent the Picture fuzzy sets [14], the Cubic set [17], the Neutrosophic fuzzy sets [22] and the Support-intuitionistic fuzzy sets [20]. Two more generalizations of intuitionistic fuzzy sets, known as circular [4], and elliptic IFSs [5], which also generalize interval-valued IFSs, have emerged in recent years.

Our work in this area is focused on creating an extension of the franchisor selection problem that can be used with circular and elliptic IFSs.

Circular and elliptic IFSs, two IFS extensions that are currently increasing in popularity, can reduce accuracy and ambiguity by enclosing the degrees of membership and non-membership in a circle or an ellipse [4], [5]. The paper [12] develops a novel current worth analysis based on interval-valued IF and C-IF sets. An integrated MCDA technique that combines the C-IF AHP and VIKOR is suggested in the work [21]. Circular IFSs are applied in Multi-Criteria Decision Making in [13]. The development of Circular Intuitionistic Fuzzy Multicriteria Analysis (C-IFFr) for Petrol Station Franchisor Selection is presented in [29].

E-IFSs are described as sets with an ellipse indicating the degrees of membership and non-membership for each element of the universe [5]. The Scopus database does not contain any established models for elliptic intuitionistic fuzzy models for franchisor selection with elliptic IF data. Using the toolset of index matrices (IMs) theories and elliptic intuitionistic fuzzy sets (E-IFSs), we enhance C-IFFr [29] in our work and create an elliptic intuitionistic fuzzy algorithm (E-IFFr) for the best

Work on Sect. I and Sect. II is supported by the Asen Zlatarov University through project Ref. No. NIX-486/2023 “Modeling Management Decisions with New Analysis". Work on Sect. III and Sect. V is supported by the Asen Zlatarov University through project Ref. No. NIX-482/2023 “Application of Intuitionistic Fuzzy Logic in Decision Making”.

selection of a franchise organization. The criteria values in this model are determined by experts with dynamic ranks and are expressed as E-IF numbers. The investor decides if each criterion is significant or not. The main contributions of the article are: definition of elliptic IF quads; extending comparison operations and relations on IF pairs to those on elliptic IF quads; extending the definition of three-dimensional IF index matrix (3-D IFIM) and some operations with them to those of 3-D elliptic IFIM (3-D E-IFIM); developed a model for ranking franchisors on elliptical IF data, describing to a greater extent the uncertainty in the economic environment; in this model, the evaluation of franchisors against criteria with weights set by the investor is carried out by dynamic rating experts; an application of E-IFFr in selecting a chain franchisor for the courier business in Bulgaria. The pessimistic, optimistic, and intermediate scenarios are presented to the decision maker for consideration before making a final decision. The advantage of this model is that it can be applied to both regular and elliptical IF data. Another advantage is that it can be easily extended so that it can be applied to multidimensional IF data. A numerical example of the best franchisor choice for the courier industry in Bulgaria serves as an illustration of the effectiveness of the suggested approach. The remainder of our investigation is organized as follows: Section II contains preliminary information for IM concepts and E-IF numbers. In Section III, an optimal E-IF problem with an IM creative solution for selecting a franchisor is provided, and also the actual C-IFFr problem of selecting a franchisor for the courier firm is solved by the software being developed. Future-oriented suggestions are provided in Section V.

II. IMs and Elliptic Intuitionistic Fuzzy Pairs Preliminary

In this section, the preliminaries of E-IF pairs and IMs are introduced. One of the most modern extensions of IFS is the elliptic IFSs, proposed by Atanasov in 2021. They are a powerful tool for representing data fuzziness.

A. Elliptic Intuitionistic Fuzzy Quads (E-IFQs)

An intuitionistic fuzzy pair (IFP) is defined as having the form $\langle a(p), b(p)\rangle$ or $\langle \mu(p), \nu(p)\rangle$: The components of an IFP are $a(p)(\mu(p)), b(p)(\nu(p)) \in [0,1]$ and $a(p)+b(p) = \mu(p)+\nu(p) \leq 1$, respectively. These components are used as an evaluation of some object or process and are interpreted as degrees of membership and non-membership, degrees of validity and non-validity, or degrees of correctness and non-correctness, etc. of a proposition p. Let us define here the elliptic IFQ (E-IFQ) as an object with the following form based on the definition of the E-IFS [5]:

$$\langle a(p), b(p); u, v\rangle = \langle \mu(p), \nu(p); u, v\rangle,$$

where $a(p)+b(p) = \mu(p)+\nu(p) \leq 1$. The "truth degree" and "falsity degree" of the statement p are considered to be $a(p)(\mu(p))$ and $b(p)(\nu(p))$ and $a(p)+b(p) \leq 1$. The ellipse's semi-major and semi-minor axes are $u, v \in [0, \sqrt{2}]$.

Let two E-IFQs be given: $x_{u_1,v_1} = \langle a, b; u_1, v_1\rangle$ and $y_{u_2,v_2} = \langle c, d; u_2, v_2\rangle$. Let us define an operation called $* \in \{\min, \max\}$. The operations over E-IFQs that come after are based on the operations for E-IFSs [5].

$$x_{u_1,v_1} \wedge_* y_{u_2,v_2} = \langle \min(a,c), \max(b,d); *(u_1,u_2), *(v_1,v_2)\rangle;$$
$$x_{u_1,v_1} \vee_* y_{u_2,v_2} = \langle \max(a,c), \min(b,d); *(u_1,u_2), *(v_1,v_2)\rangle;$$
$$x_{u_1,v_1} +_* y_{u_2,v_2} = \langle a+c-a.c, b.d; *(u_1,u_2), *(v_1,v_2)\rangle;$$
$$x_{u_1,v_1} \bullet_* y_{u_2,v_2} = \langle a.c, b+d-b.d; *(u_1,u_2), *(v_1,v_2)\rangle;$$

We suggest the following relation for comparing E-IFQs using a formula for the distance between C-IFSs [8], the relation for comparing two C-IFQs [27], and the distance from the element to the ideal positive alternative [23]:

$$x_{u_1,v_1} \geq_{R^{elliptic}} y_{u_2,v_2} \quad \text{iff } R^{elliptic}_{x_{u_1,v_1}} \leq R^{elliptic}_{y_{u_2,v_2}} \tag{1}$$

where

$$R^{elliptic}_{x_{u_1,v_1}} = \frac{1}{6}(2-a-b)\left(|\sqrt{2}-u_1| + |\sqrt{2}-v_1| + |1-a|\right)$$

is the distance between x and the ideal positive alternative $\langle 1, 0; \sqrt{2}, \sqrt{2}\rangle$ to x.

B. Three-Dimensional Elliptic Intuitionistic Fuzzy Index Matrices (3-D E-IFIM)

In 1987, according to [1], the theory of index matrices (IMs) appeared. Over IMs, several operations, relations, and operators are defined (see [2], [31]). Assume that the set of indices $\mathscr{I}$ is fixed. Using the definition of 3-D IFIM from [2], [31], let us we define a 3-D E-IFIM $A = [K, L, H, \{\langle \mu_{k_i,l_j,h_g}, \nu_{k_i,l_j,h_g}; rf_{k_i,l_j,h_g}, rs_{k_i,l_j,h_g}\rangle\}]$ as follows:

$$\begin{array}{c|ccc}
h_g \in H & l_1 & \ldots & l_n \\
\hline
k_1 & \langle \mu_{k_1,l_1,h_g}, \nu_{k_1,l_1,h_g}; rf_{k_1,l_1,h_g}, rs_{k_1,l_1,h_g}\rangle & \ldots & \langle \mu_{k_1,l_n,h_g}, \nu_{k_1,l_n,h_g}; rf_{k_1,l_n,h_g}, rs_{k_1,l_n,h_g}\rangle \\
\vdots & \vdots & \ldots & \vdots \\
k_m & \langle \mu_{k_m,l_1,h_g}, \nu_{k_m,l_1,h_g}; rf_{k_m,l_1,h_g}, rs_{k_m,l_1,h_g}\rangle & \ldots & \langle \mu_{k_m,l_n,h_g}, \nu_{k_m,l_n,h_g}; rf_{k_m,l_n,h_g}, rs_{k_m,l_n,h_g}\rangle
\end{array},$$

where $(K, L, H \subset \mathscr{I})$ and its elements are E-IFQs.

There are many defined operations over the IMs [2]. Let E-IFIMs $A = [K, L, H, \{\langle \mu_{k_i,l_j,h_g}, \nu_{k_i,l_j,h_g}; rf_{k_i,l_j,h_g}, rs_{k_i,l_j,h_g}\rangle\}]$ and $B = [P, Q, R\{\langle \rho_{p_r,q_s,t_e}, \sigma_{p_r,q_s,t_e}; \delta f_{p_r,q_s,t_e}, \delta s_{p_r,q_s,t_e}\}]$ be given.

We for the first time introduce some operations performed over E-IFIM that are comparable to those performed over IFIMs [2].

Addition-$(\circ_1, \circ_2, *)$:

$$A \oplus_{(\circ_1,\circ_2,*)} B$$
$$= [K \cup P, L \cup Q, H \cup R, \{\langle \phi_{t_u,v_w,x_y}, \psi_{t_u,v_w,x_y}; \eta_{t_u,v_w,x_y}\rangle\}],$$

where $\langle \circ_1, \circ_2\rangle \in \{\langle \max, \min\rangle, \langle \min, \max\rangle, \langle average, average\rangle\}$ and $* \in \{\max, \min\}$.

$$\langle \phi_{t_u,v_w,x_y}, \psi_{t_u,v_w,x_y}; \eta_{t_u,v_w,x_y}\rangle$$
$$= \langle \circ_1(\mu_{k_i,l_j,x_y}, \rho_{p_r,q_s,x_y}), \circ_2(\nu_{k_i,l_j,x_y}, \sigma_{p_r,q_s,x_y});$$
$$*(rf_{t_u,v_w,x_y}, \delta f_{t_u,v_w,x_y}, *(rs_{t_u,v_w,x_y}, \delta s_{t_u,v_w,x_y}\rangle).$$

Multiplication:

$$A \odot_{(\circ_1,\circ_2,*)} B$$
$$= [K \cup (P-L), Q \cup (L-P), H \cup R, \{\langle \phi_{t_u,v_w,x_y}, \psi_{t_u,v_w,x_y};$$
$$\eta f_{t_u,v_w,x_y}, \eta s_{t_u,v_w,x_y}\rangle\}],$$

where

$$\langle \phi_{t_u,v_w,x_y}, \psi_{t_u,v_w,x_y}\rangle$$

is defined in [2],

$$\eta f_{t_u,v_w,x_y} = *(rf_{t_u,v_w,x_y}, \delta f_{t_u,v_w,x_y})$$
$$\text{and } \eta s_{t_u,v_w,x_y} = *(rs_{t_u,v_w,x_y}, \delta s_{t_u,v_w,x_y}).$$

The following operations cannot be performed on these conventional verso matrices. They are designed with the ability to automate specific IM operations to implement different models and algorithms.

Aggregation operation by one dimension:

Let us extend the operations $\#_q, (q \leq i \leq 3)$ from [30] such that they can be applied over E-IFQs $x = \langle a, b; rf_1, rs_1 \rangle$ and $y = \langle c, d; rf_2, rs_2 \rangle$:

$$x\#_1, *y = \langle min(a,c), max(b,d); *(rf_1, rf_2), *(rs_1, rs_2) \rangle;$$
$$x\#_2, *y = \langle average(a,c), average(b,d); *(rf_1, rf_2), *(rs_1, rs_2) \rangle;$$
$$x\#_3, *y = \langle max(a,c), min(b,d); *(rf_1, rf_2), *(rs_1, rs_2) \rangle.$$

Let the fixed index be $k_0 \notin K$. The expanded definition of the aggregation operation $\alpha_{K,\#_q,*}(A, k_0)$ by the dimension K over 3-D E-IFIM A utilizing that of [27], [30] is as follows:

$$\begin{array}{c|ccc}
h_g \in H & l_1 & \dots \\ \hline
k_0 & \underset{i=1}{\overset{m}{\#_q,*}} \langle \mu_{k_i,l_1,h_g}, \nu_{k_i,l_1,h_g}; rf_{k_i,l_1,h_g}, rs_{k_i,l_1,h_g} \rangle & \dots
\end{array}$$
$$\begin{array}{cc}
\dots & l_n \\ \hline
\dots & \underset{i=1}{\overset{m}{\#_q,*}} \langle \mu_{k_i,l_n,h_g}, \nu_{k_i,l_n,h_g}; rf_{k_i,l_1,h_g}, rs_{k_i,l_1,h_g} \rangle
\end{array}.$$

We may perform a super pessimistic aggregation operation in conditions of high inflation using $\#_1^*$, an average aggregation operation in anticipation of slight fluctuations in the market situation using $\#_2^*$, and a super optimistic aggregation operation in conditions of stability of the market parameters using $\#_3^*$.

Projection [2]: Let $W \subseteq K$, $V \subseteq L$ and $U \subseteq H$. Then,
$$pr_{W,V,U}A = [W, V, U, \{\langle R_{p_r,q_s,e_d}, S_{p_r,q_s,e_d} \rangle\}],$$
where for each $k_i \in W, l_j \in V$ and $t_g \in U$,
$$\langle R_{p_r,q_s,e_d}, S_{p_r,q_s,e_d} \rangle = \langle \mu_{k_i,l_j,h_g}, \nu_{k_i,l_j,h_g} \rangle.$$

Reduction [2]: An IM A's operations-reduction $(k, \perp, \perp)$ is defined as follows:
$$A_{(k,\perp,\perp)} = [K - \{k\}, L, H, \{c_{t_u,v_w,e_d}\}], \text{where}$$
$$c_{t_u,v_w,e_d} = a_{k_i,l_j,h_g}(t_u = k_i \in K - \{k\}, v_w = l_j \in L, e_d = h_g \in H).$$

Substitution [2]:
$$\left[\frac{p}{k_i}; \perp, \perp\right] A = \left[(K - \{k_i\}) \cup \{p\}, L, H, \{a_{k_i,l_j,h_g}\}\right]$$

A Level Operator for Decreasing the Number of Elements of E-IFIM: Let $\langle \alpha, \beta; r_1, r_2 \rangle$ is an E-IFQ and $A = [K, L, H, \{a_{k_i,l_j,h_g}\}] = [K, L, H, \{\langle \mu_{k_i,l_j,h_g}, \nu_{k_i,l_j,h_g}; rf_{k_i,l_j,h_g}, rs_{k_i,l_j,h_g} \rangle\}$ is a 3-D E-IFIM, then according to [10] let us define the operator $N^{>_{R^{elliptic}}}_{\langle \alpha,\beta,r_1,r_2 \rangle}(A) = [K, L, H, \{\langle \rho_{k_i,l_j,h_g}, \sigma_{k_i,l_j,h_g}; rf_{k_i,l_j,h_g}, rs_{k_i,l_j,h_g} \rangle\}],$ where

$$\langle \rho_{k_i,l_j,h_g}, \sigma_{k_i,l_j,h_g}; rf^n_{k_i,l_j,h_g}, rs^n_{k_i,l_j,h_g} \rangle$$

$$= \begin{cases} a_{k_i,l_j,h_g} & \text{if } a_{k_i,l_j,h_g} >_{R^{elliptic}} \langle \alpha, \beta; r_1, r_2 \rangle \\ \langle 0, 1; 0, 0 \rangle & \text{otherwise} \end{cases} \tag{2}$$

III. An Elliptic Intuitionistic Fuzzy Method for Selecting the Most Beneficial Franchisor (E-IFFr)

This section will extend the intuitionistic fuzzy (IF) algorithm for the best franchisee selection [26] to suggest an algorithm for a specific type of E-IF franchisor selection problem (E-IFFr). The Elliptical IFS is a better tool for characterizing this fuzziness than the IFS in situations of galloping inflation and quick changes in the economic environment because in these situations, the degrees of truth and falsity of a given element shift in the shape of an ellipse.

The optimal E-IF franchisor selection problem is posed: An entrepreneurial company has created an evaluation system with criteria $\{c_1, \dots, c_j, \dots, c_n\}$ (for $j = 1, ..., n$) for franchise companies from a certain business. The business wants to choose a successful business franchisor. It is necessary to do a professional evaluation by experts $\{d_1, \dots, d_s, \dots, d_D\}$ of franchise businesses $\{k_1, \dots, k_i, \dots, k_m\}$ in the pertinent business sector. The ranking coefficients of the experts $\{r_1, \dots, r_s, \dots, r_D\}$ are calculated based on their qualitative involvement in the evaluation of franchise procedures and are given to the experts in the form of IFPs $\langle \delta_s, \varepsilon_s \rangle (1 \leq s \leq D)$. The interpretation of the elements δ_s and ε_s is the level of competence and incompetence of the s-th expert, respectively. The expert evaluations of the franchise chains are made and presented as IF data ev_{k_i,c_j,d_s} (for $1 \leq i \leq m, 1 \leq j \leq n, 1 \leq s \leq D$). The final estimates of the franchisors are calculated in the form of E-IFQs fi_{k_i,v_e,h_f} (for $1 \leq i \leq m$), taking into consideration the E-IF priorities pk_{c_j,v_e} of the criteria c_j (for $j = 1, ..., n$) from the view of the entrepreneur v_e in a given moment h_f. The optimal aim is to determine which franchise chain is the most suitable for the entrepreneurial company.

A. Index-matrix Interpretation of the Optimal Elliptic Intuitionistic Fuzzy Franchisor Selection Problem

The following operations are part of the index-matrix approach to the optimal elliptic intuitionistic fuzzy franchisor selection problem (E-IFFr), defined above:

Step 1. An IF index matrix $EV[K, C, E, \{ev_{k_i,c_j,d_s}\}]$, $K = \{k_1, k_2, \dots, k_m\}$, $C = \{c_1, c_2, \dots, c_n\}$ and $E = \{d_1, d_2, \dots, d_D\}$ is constructed. Due to the uncertainty of the economic environment, the elements $\{ev_{k_i,c_j,d_s}\} = \langle \mu_{k_i,c_j,d_s}, \nu_{k_i,c_j,d_s} \rangle$ (for $1 \leq i \leq m, 1 \leq j \leq n, 1 \leq s \leq D$) of the matrix EV are the IF valuations of the d_s-th expert for the k_i-th franchisor by the c_j-th criterion. Next, we go on to *Step 2*.

Step 2. An IFP $r_s = \langle \delta_s, \varepsilon_s \rangle, (s \in E)$, whose components might be interpreted as showing how competent or incompetent experts are, should be used to specify each expert's score coefficient.

The IM has been built by:
$$EV^*[K, C, E, \{ev^*{}_{k_i,c_j,d_s}\}]$$
$$= r_1 pr_{K,C,d_1} EV \oplus_{(\circ_1,\circ_2)} r_2 pr_{K,C,d_2} EV \dots \oplus_{(\circ_1,\circ_2)} r_D pr_{K,C,d_D} EV.$$
$$EV := EV^*(ev_{k_i,l_j,d_s} = ev^*_{k_i,l_j,d_s}, \ \forall k_i \in K, \forall l_j \in L, \forall d_s \in E).$$

The degrees of membership and non-membership of the E-IFQs are determined by the elements of the matrix EV using the three aggregating operations $\alpha_{K,\#_1,*}, \alpha_{K,\#_3,*}$ and $\alpha_{K,\#_2,*}$, which provide the evaluations of the k_i-th franchisor against the c_j-th criterion in a present moment $h_f \notin E$:

$$PI_{min}[K, h_f, C, \{pi_{min k_i,h_f,c_g}\}] = \alpha_{E,\#_1}(EV^*, h_f)$$

$$= \left\{ \begin{array}{c|c} c_j & h_f \\ \hline k_1 & \overset{D}{\underset{s=1}{\#_1}} \langle \mu_{k_1,c_j,d_s}, \nu_{k_1,c_j,d_s} \rangle \\ \vdots & \vdots \\ k_m & \overset{D}{\underset{s=1}{\#_1}} \langle \mu_{k_m,c_j,d_s}, \nu_{k_m,c_j,d_s} \rangle \end{array} \middle| \; c_j \in C \right\};$$

$$PI_{max}[K,h_f,C,\{pi_{max k_i,h_f,c_g}\}] = \alpha_{E,\#_3}(EV^*,h_f)$$

$$= \left\{ \begin{array}{c|c} c_j & h_f \\ \hline k_1 & \overset{D}{\underset{s=1}{\#_3}} \langle \mu_{k_1,c_j,d_s}, \nu_{k_1,c_j,d_s} \rangle \\ \vdots & \vdots \\ k_m & \overset{D}{\underset{s=3}{\#_1}} \langle \mu_{k_m,c_j,d_s}, \nu_{k_m,c_j,d_s} \rangle \end{array} \middle| \; c_j \in C \right\}$$

$$PI* = PI_{min} \oplus_{(\circ_1,\circ_2,*)} PI_{max}$$

Then the centers of the E-IFQs used to evaluate the franchise companies are represented as elements in a matrix as follows: $PI[K,h_f,C,\{pi_{k_i,h_f,c_g}\}] = \alpha_{E,\#_2}(PI^*,h_f),(h_f \notin E)$. Next, we continue to *Step 3.*

Step 3. At this point, the evaluation system for the franchise business candidate will be optimized. We recommend removing slower or more expensive criteria to measure that has been found to closely connect with other criteria under intuitionistic fuzzy settings from the franchisee evaluation system utilizing inter-criteria analysis (ICrA, [7], [9]). Let $\langle \alpha,\beta \rangle$ be an IFP. The criteria C_k and C_l are in

(α,β)-positive consonance, if $\mu_{C_k,C_l} > \alpha$ and $\nu_{C_k,C_l} < \beta$;
(α,β)-negative consonance, if $\mu_{C_k,C_l} < \beta$ and $\nu_{C_k,C_l} > \alpha$;
(α,β)-dissonance, otherwise.

The transposed IM $PI^T = [K,C,h_f,\{pi^T{}_{k_i,c_g,h_f}\}]$ is searched for consonant criteria using the ICrA algorithm. More expensive, slower, or more complicated criteria are eliminated from the evaluation franchise system using the IM reduction operation over matrix PI^T. The following step is *Step 4.*

Step 4. Now we can calculate E-IFIM $A[K,C,h_f\{a_{k_i,c_g,h_f}\}]$, which represents current assessments of the franchisors using the methodology from [5] according to the system of criteria:

$$\begin{array}{c|ccc} h_f & c_1 & \dots & c_n \\ \hline k_1 & \langle \mu^a_{k_1,c_1}, \nu^a_{k_1,c_1}; rf^a_{k_1,c_1}, rs^a_{k_1,c_1} \rangle & \dots & \langle \mu^a_{k_1,c_n}, \nu^a_{k_1,c_n}; rf^a_{k_1,c_n}, rs^a_{k_1,c_n} \rangle \\ \vdots & \vdots & \dots & \vdots \\ k_m & \langle \mu^a_{k_m,c_1}, \nu^a_{k_m,c_1}; rf^a_{k_m,c_1}, rs^a_{k_m,c_1} \rangle & \dots & \langle \mu^a_{k_m,c_n}, \nu^a_{k_m,c_n}; rf^a_{k_m,c_n}, rs^a_{k_m,c_n} \rangle \end{array},$$

where $K = \{k_1,\dots,k_i,\dots,k_m\}, i = 1,\dots,m$; $C = \{c_1,\dots,c_j,\dots,c_n\}, j = 1,\dots,n$; its elements a_{k_i,c_g,h_f} (for $i = 1,\dots,m; g = 1,\dots,n$) are created as E-IFQs by transforming the IFPs $pi^T{}_{k_i,c_j,h_f}$ using the following steps

for $g = 1$ to n, $i = 1$ to m

$$\Big\{ \mu^a_{k_i,c_g,h_f} = \mu^{pi^T}_{k_i,c_g,h_f}; \nu^a_{k_i,c_g,h_f} = \nu^{pi^T}_{k_i,c_g,h_f},$$

$$rf^a_{k_i,c_g,h_f} =$$

$$\sqrt{ \min_{1\le s\le D} {\mu^{ev}_{k_i,c_g,d_s}}^2 + \left\{ \frac{\max\limits_{1\le s\le D} {\mu^{ev}_{k_i,c_g,d_s}}^2 - \min\limits_{1\le s\le D} {\mu^{ev}_{k_i,c_g,d_s}}^2}{\max\limits_{1\le s\le D} {\nu^{ev}_{k_i,c_g,d_s}}^2 - \min\limits_{1\le s\le D} {\nu^{ev}_{k_i,c_g,d_s}}^2} \right\}^2 . \min_{1\le s\le D} {\nu^{ev}_{k_i,c_g,d_s}}^2 }$$

$$\text{and } rs^a_{k_i,c_g} =$$

$$\sqrt{ \min_{1\le s\le D} {\mu^{ev}_{k_i,c_g,d_s}}^2 . \left\{ \frac{\max\limits_{1\le s\le D} {\nu^{ev}_{k_i,c_g,d_s}}^2 - \min\limits_{1\le s\le D} {\nu^{ev}_{k_i,c_g,d_s}}^2}{\max\limits_{1\le s\le D} {\mu^{ev}_{k_i,c_g,d_s}}^2 - \min\limits_{1\le s\le D} {\mu^{ev}_{k_i,c_g,d_s}}^2} \right\}^2 + \min_{1\le s\le D} {\nu^{ev}_{k_i,c_g,d_s}}^2 }$$

$$\Big\}$$

Next, we go on to *Step 5*.

Step 5. At this stage, a 3-D E-IFIM PK is created, and the coefficients used in the following operation determine the weighting of each evaluation criterion from the view of the entrepreneur v_e for the franchise business:

$$PK[C,v_e,h_f,\{pk_{c_j,v_e,h_f}\}] = \begin{array}{c|c} h_f & v_e \\ \hline c_1 & pk_{c_1,v_e,h_f} \\ \vdots & \vdots \\ c_j & pk_{c_j,v_e,h_f} \\ \vdots & \vdots \\ c_n & pk_{c_n,v_e,h_f} \end{array},$$

where $C = \{c_1,c_2,\dots,c_n\}$. The evaluation E-IFIM $FI[K,v_e,h_f,\{fi_{k_i,v_e,h_f}\}] = A \odot_{(\circ_1,\circ_2,*)} PK$ (for $1 \le i \le m$) for the entrepreneur v_e includes all of the E-IF estimates for k_i-th franchisor. Go to *Step 6.*

Step 6. At this stage, based on the aggregation operation $\alpha_{K,\#_q,*}(FI,k_0)$, the business owner v_e chooses the franchisor that is the most advantageous. Depending on the value of q, utilizing pessimistic, average, or optimistic scenarios:

$$\alpha_{K,\#_q,*}(FI,k_0)$$

$$= \begin{array}{c|c} h_f & v_e \\ \hline k_0 & \overset{m}{\underset{i=1}{\#_{q},*}} \langle \mu_{k_i,v_e,h_f}, \nu_{k_i,v_e,h_f}, rf_{k_i,v_e,h_f}, rs_{k_i,v_e,h_f} \rangle \end{array}, \qquad (3)$$

where $k_0 \notin K, 1 \le q \le 3$. Go to *Step 7.*

Step 7. The updated rating coefficients for the experts who participated in the evaluation process are obtained in this stage. The expert's new score will be altered by the method used in the work after he participates in the present procedure. Let's assume the expert d_s $(s = 1,...,D)$ has participated in γ_s evaluation procedures for the selection of a franchisee, based on which his score $r_s = \langle \delta_s, \varepsilon_s, \phi^1_s, \phi^2_s \rangle$ is determined, then after his participation in the current procedure, his new score will be changed by ideas from [3]:

$$\langle \delta_s^{'}, \varepsilon_s^{'}; {\phi^1}^{'}_s, {\phi^2}^{'}_s \rangle$$

$$= \begin{cases} \langle \frac{\delta\gamma+1}{\gamma+1}, \frac{\varepsilon\gamma}{\gamma+1}; *({\phi^1}^{'}_s, \phi_s^1), *({\phi^2}^{'}_s, \phi_s^2) \rangle, \text{ if the expert's assessment was accurate} \\ \langle \frac{\delta\gamma}{\gamma+1}, \frac{\varepsilon\gamma}{\gamma+1}; *({\phi^1}^{'}_s, \phi_s^1), *({\phi^2}^{'}_s, \phi_s^2) \rangle, \text{ if the expert has not provided any estimates} \\ \langle \frac{\delta\gamma}{\gamma+1}, \frac{\varepsilon\gamma+1}{\gamma+1}; *({\phi^1}^{'}_s, \phi_s^1), *({\phi^2}^{'}_s, \phi_s^2) \rangle, \text{if the expert has made an inaccurate assessment} \end{cases}$$

The algorithm is complete.

If the operations $\langle \circ_1, \circ_2 \rangle = \langle \min, \max \rangle$ are used in the E-IFFr, the pessimistic scenario has been utilized.

If $\langle \circ_1, \circ_2 \rangle = \langle \max, \min \rangle$ are applied, the optimistic scenario has been obtained, and if the operations $\langle \circ_1, \circ_2 \rangle = \langle average, average \rangle$ are used, the averaged scenario has been obtained. The operation "$* = \max$" is used in cases of more ambiguity, otherwise "$* = \min$." As a result, different ideal solutions could be produced based on the investor's viewpoint. As a result, an investor may have more faith in the discovered solution.

Based on the complexity of ICrA [11]), the suggested E-IFFr method has a complexity $O(Dm^2n^2)$. Crisp and E-IF data can be used using the suggested E-IFFr approach. It can be used without limitations and is easy to adapt to the different kinds of data that are present in a fuzzy environment.

There is no model for the best franchise chain selection based on elliptic IF fuzzy data in the scientific literature that can represent ambiguous or incomplete data under situations of significant market parameter volatility. There are three scenarios in this model, each based on the decision-makers attitudes. Experts evaluate franchisors to make the best choice, and their ratings as well as the importance of the criteria are taken into consideration during the review process. Because of this, the proposed E-IF optimum franchisor selection task is socially oriented and takes into account the preferences of the decision-maker as well as the experts.

IV. Using E-IFFr Approach to Overcome the Difficulty of Selecting a Courier Franchise Company

Finding the ideal franchisor in the courier services industry is possible with the help of the E-IFFr technique of Sect. III. Let us formulate the following problem for this purpose:

A business investor v_e wants to make an optimal choice of a courier brand offering a franchise such as Econt, Leo Express, Fasto Courier, and T-Post. For this purpose, he creates a system of criteria for evaluating potential franchisors k_i (for $1 \le i \le 4$) using the expertise of experts d_1, d_2, and d_3. The four groups of criteria that make the system for franchisee selection using the requirements of the four courier franchise companies are as follows:

- C_1 - the choice of a well-known, respected brand with a significant market share with reputation, market share, and corporate capabilities;
- C_2 - expected profit in the form of commission, which is a dynamic % of the realized turnover of the office depending on the quality and volume of the courier services provided.
- C_3 - operational and initial costs for starting the business model: to calculate initial and ongoing operational costs, such as initial and monthly franchise fees; royalties and marketing expenses; costs for satisfying the brand's requirements for the look of the offices and their equipment, for the look of the cars and their number; costs of providing office security measures; costs of providing equipment for servicing large-volume shipments
- C_4 - evaluation of the franchisor's level of training and support, including ongoing marketing and technical support.

Ranking coefficients of the experts $\{r_1, r_2, r_3\}$ are given in the form of IFPs $\langle \delta_s, \varepsilon_s \rangle (1 \le s \le 3)$. The four courier franchise chains' expert assessments were made by the criteria and presented as IF data ev_{k_i,c_j,d_s} (for $1 \le i \le 4, 1 \le j \le 4, 1 \le s \le 3$). The final E-IF evaluations fi_{k_i,v_e,h_f} (for $1 \le i \le 4$) of the courier brands are based on the priorities pk_{c_j,v_e} of criteria c_j (for $j = 1, ..., 4$) from the point of view of entrepreneur v_e at time h_f. The optimal purpose is to determine which franchise structure for courier services is the best for the startup business.

Solution of the problem:

Step 1. At this stage, the 3-D expert assessment IFIM $EV[K, C, E, \{es_{k_i,c_j,d_s}\}]$ is created with the expert's estimations in the c_j-th criterion for the k_i-th franchisor (for $1 \le i \le 4, 1 \le j \le 4, 1 \le s \le 3$), and its form is:

$$\left\{ \begin{array}{c|cccc} d_1 & c_1 & c_2 & c_3 & c_4 \\ \hline k_1 & \langle 0.30, 0.30 \rangle & \langle 0.20, 0.50 \rangle & \langle 0.60, 0.20 \rangle & \langle 0.20, 0.50 \rangle \\ k_2 & \langle 0.10, 0.60 \rangle & \langle 0.40, 0.40 \rangle & \langle 0.40, 0.50 \rangle & \langle 0.40, 0.40 \rangle \\ k_3 & \langle 0.40, 0.20 \rangle & \langle 0.10, 0.70 \rangle & \langle 0.20, 0.40 \rangle & \langle 0.60, 0.20 \rangle \\ k_4 & \langle 0.10, 0.75 \rangle & \langle 0.20, 0.70 \rangle & \langle 0.205, 0.70 \rangle & \langle 0.40, 0.50 \rangle \end{array} \right. ,$$

$$\begin{array}{c|cccc} d_2 & c_1 & c_2 & c_3 & c_4 \\ \hline k_1 & \langle 0.40, 0.40 \rangle & \langle 0.10, 0.70 \rangle & \langle 0.70, 0.10 \rangle & \langle 0.30, 0.50 \rangle \\ k_2 & \langle 0.20, 0.80 \rangle & \langle 0.30, 0.50 \rangle & \langle 0.60, 0.20 \rangle & \langle 0.60, 0.10 \rangle \\ k_3 & \langle 0.30, 0.40 \rangle & \langle 0.30, 0.60 \rangle & \langle 0.10, 0.70 \rangle & \langle 0.40, 0.40 \rangle \\ k_4 & \langle 0.15, 0.60 \rangle & 0.25, 0.30 \rangle & \langle 0.20, 0.60 \rangle & \langle 0.30, 0.30 \rangle \end{array} ,$$

$$\left. \begin{array}{c|cccc} d_3 & c_1 & c_2 & c_3 & c_4 \\ \hline k_1 & \langle 0.10, 0.70 \rangle & \langle 0.20, 0.70 \rangle & \langle 0.40, 0.40 \rangle & \langle 0.40, 0.40 \rangle \\ k_2 & \langle 0.10, 0.80 \rangle & \langle 0.30, 0.60 \rangle & \langle 0.20, 0.60 \rangle & \langle 0.50, 0.20 \rangle \\ k_3 & \langle 0.30, 0.50 \rangle & \langle 0.20, 0.70 \rangle & \langle 0.30, 0.60 \rangle & \langle 0.40, 0.50 \rangle \\ k_4 & \langle 0.10, 0.80 \rangle & \langle 0.30, 0.50 \rangle & \langle 0.10, 0.70 \rangle & \langle 0.30, 0.60 \rangle \end{array} \right\}$$

Step 2. These are the ranking experts' rank coefficients: $\{r_1, r_2, r_3\} = \{\langle 0.80, 0.10 \rangle, \langle 0.70, 0.10 \rangle, \langle 0.90, 0.10 \rangle\}$.

The evaluation IM $EV^*[K, C, E, \{ev^*\}]$ is made using the subsequent procedures:

$$EV^* = r_1 pr_{K,C,d_1} EV \oplus_{(\circ_1,\circ_2)} r_2 pr_{K,C,d_2} EV \oplus_{(\circ_1,\circ_2)} r_3 pr_{K,C,d_3} EV;$$

$$EV := EV^*$$

Then the IMs are created:

$$PI* = PI_{min} \oplus_{(\circ_1,\circ_2,*)} PI_{max}$$

and

$$PI[K, h_f, C, \{pi_{k_i,h_f,c_g}\}] = \alpha_{E,\#_2}(PI^*, h_f), (h_f \notin E)$$

whose elements are the coordinates of the centers of the E-IFQs evaluating the courier brands.

Step 3. At this step, we ran the ICrA over the matrix PI^T with $\alpha = 0.80$ and $\beta = 0.10$. Following ICrA, it is concluded that

TABLE I
THE IFPS PROVIDE THE INTERCRITERIA CORRELATIONS

	C_1	C_2	C_3	C_4
C_1	—	$\langle 0.76;0.20\rangle$	$\langle 0.79;0.17\rangle$	$\langle 0.71;0.20\rangle$
C_2	$\langle 0.76;0.20\rangle$	—	$\langle 0.73;0.23\rangle$	$\langle 0.76;0.12\rangle$
C_3	$\langle 0.79;0.17\rangle$	$\langle 0.73;0.23\rangle$	—	$\langle 0.65;0.26\rangle$
C_4	$\langle 0.71;0.20\rangle$	$\langle 0.76;0.12\rangle$	$\langle 0.65;0.26\rangle$	—

no consonant-dependent criteria exist. The results are shown as an IM in μ - ν view result matrix (cf. table I):

Step 4. Now we can calculate E-IFIM $A[K,C,h_f\{a_{k_i,c_g,h_f}\}]$, which represents recent assessments of the courier brands using the following criteria:

$$\begin{array}{c|cccc}
h_f & c_1 & c_2 & c_3 & c_4 \\
\hline
k_1 & \langle 0.21,0.54;0.06,0.04\rangle & \langle 0.14,0.68;0.04,0.02\rangle & \langle 0.44,0.34;0.05,0.03\rangle & \langle 0.24,0.55;0.05,0.03\rangle \\
k_2 & \langle 0.11,0.77;0.04,0.02\rangle & \langle 0.26,0.57;0.04,0.02\rangle & \langle 0.31,0.51;0.04,0.02\rangle & \langle 0.39,0.34;0.04,0.02\rangle \\
k_3 & \langle 0.26,0.45;0.03,0.01\rangle & \langle 0.16,0.62;0.04,0.02\rangle & \langle 0.16,0.62;0.03,0.01\rangle & \langle 0.36,0.45;0.03,0.01\rangle \\
k_4 & \langle 0.10,0.74;0.03,0.01\rangle & \langle 0.20,0.57;0.03,0.01\rangle & \langle 0.14,0.74;0.03,0.01\rangle & \langle 0.26,0.54;0.04,0.02\rangle
\end{array}$$

Step 5. At this stage, the priority of each evaluation criterion from the viewpoint of the franchisor v_e is determined by the coefficients employed in the subsequent process on a 3-D E-IFIM PK:

$$PK[C,v_e,h_f,\{pk_{c_j,v_e,h_f}\}] = \begin{array}{c|c}
h_f & v_e \\
\hline
c_1 & \langle 0.90,0.10;0.02,0.01\rangle \\
c_2 & \langle 0.80,0.10;0.02,0.01\rangle \\
c_3 & \langle 0.60,0.20;0.02,0.01\rangle \\
c_4 & \langle 0.80,0.10;0.02,0.01\rangle
\end{array} \quad (4)$$

The evaluation E-IFIM

$$FI[K,v_e,h_f,\{fi_{k_i,v_e,h_f}\}] = A \odot_{(\circ_1,\circ_2,min} PK$$

(for $1 \le i \le m$) includes the full estimates of the k_i-th franchise chain for the business owner v_e based on the optimistic case:

$$\text{and } FI = \begin{array}{c|c}
h_f & v_e \\
\hline
k_1 & \langle 0.674,0.046;0.02,0.01\rangle \\
k_2 & \langle 0.688,0.040;0.02,0.01\rangle \\
k_3 & \langle 0.694,0.047;0.02,0.01\rangle \\
k_4 & \langle 0.539,0.170;0.01,0.01\rangle
\end{array} \quad (5)$$

Step 6. According to the optimistic aggregation operation $\alpha_{K,\#_3,min}(FI,k_0)$, k_3 is the best courier franchise brand in Bulgaria from the point of view of the entrepreneur, with a maximum acceptance degree of 0.694 and a minimum rejection degree of 0.047. The decision-makers will select the candidate k_4 with the minimum degree of membership 0.539 and the greatest degree of non-membership 0.17 in a pessimistic scenario if the future is unclear and the decision-making environment is unpredictable.

Step 7. At the last step, we assume that the correctness of the experts' evaluations was evaluated by senior experts and they are correct from the point of view of intuitionistic fuzzy logic [3] and their new rating coefficients are equal to $\{\langle 0.82,0.09;0.02,0.01\rangle,\langle 0.73,0.09;0.02,0.01\rangle,$ $\langle 0.91,0.09;0.02,0.01\rangle\}$.

The decision-maker will favor the pessimistic scenario when there is high inflation and significant uncertainty, the averaged scenario when there are only minor variations in the market parameters, and the optimistic scenario when the market parameters are stable. We were unable to locate techniques for issues of a comparable type under circumstances of high uncertainty characterized by E-IF logic, preventing us from performing a comparative analysis between the suggested E-IFKP approach for franchisor optimization.

The question of whether minor variations in the values of the input parameters utilized impact the outcomes of the model emerges following the results of the E-IFKP franchisor approach. A crucial step in solving the E-IFKP franchisor problem is to evaluate the robustness of the findings in the constructed model and their sensitivity to changes in the input variables.

The weights of the criteria are of great importance to the results of the E-IFKP franchisor problem. A sensitivity analysis consisting of 8 different scenarios have been conducted to analyze the effect of the change in weight of each criterion on the ranking results by ±10%, ±25%, ±50% and ± 75% respectively. In the analysis, a total of 8 different changes have been applied to the weights of the criteria included in the study, and the final results are different in the cases indicated.

The sensitivity of franchisor ranking to criteria weights is explored, and four distinct weighting methodologies have been taken into consideration to reveal different answers.

The results of the sensitivity analysis show that there is sensitivity in the output results when criteria are assigned different weights assets. The best option is candidate k_3 if the first criterion is given top priority; the best option is candidate k_1 if the second criterion is given top priority; candidate k_3 if the third criterion is given top priority; and the best option is candidate k_2 if the fourth criterion is given top priority.

V. CONCLUSION

The instruments of E-IF logic and the theory of index matrices were used in the study to construct an optimal algorithm (E-IFFr) for the most effective selection of franchise firms in conditions of significant parameter uncertainty. Additionally, it considered the opinions of the experts as well as the order in which entrepreneurs should rank the evaluation criteria. A case study of the franchisor selection for a courier brand is used to illustrate the proposed strategy based on the criteria of 4 leading courier franchise companies in Bulgaria. The created E-IFFr method can be used with both crisp and elliptic values. There are no restrictions on its use, and it is simple to modify to various types of data that are present in a fuzzy environment. In the future, research will continue with the development of a franchisor selection software program to automate the proposed approach and with its extension so that it can be applied to three-dimensional data and with its extension so that it can be applied to three-dimensional elliptic intuitionistic fuzzy data.

REFERENCES

[1] K. Atanassov, "Generalized index matrices," *Comptes rendus de l'Academie Bulgare des Sciences,* vol. 40(11), 1987, pp. 15-18.

[2] K. Atanassov, "Index Matrices: Towards an Augmented Matrix Calculus," *Studies in Computational Intelligence,* Springer, vol. 573, 2014.

[3] K. Atanassov, "On Intuitionistic Fuzzy Sets Theory," *STUDFUZZ,* vol. 283, Springer, Heidelberg, 2012. DOI: 10.1007/978-3-642-29127-2.

[4] K. Atanassov, "Circular Intuitionistic Fuzzy Sets," *Journal of Intelligent & Fuzzy Systems*, vol. 39 (5), 2020, pp. 5981-5986.

[5] K. Atanassov, "Elliptic Intuitionistic Fuzzy Sets," *Comptes rendus de l'Académie bulgare des Sciences,* vol. 74 (65), 2021, pp. 812-819 (2021)

[6] K. Atanassov, G. Gargov, "Interval valued intuitionistic fuzzy sets," *Fuzzy sets and systems,* vol. 31 (3), 1989, 343-349.

[7] K. Atanassov, D. Mavrov, V. Atanassova, "Intercriteria decision making: a new approach for multicriteria decision making, based on index matrices and intuitionistic fuzzy sets," *Issues in IFSs and Generalized Nets,* vol. 11, 2014, pp. 1-8.

[8] K. Atanassov, E. Marinov, "Four Distances for Circular Intuitionistic Fuzzy Sets," *Mathematics,* vol. 9 (10), 2021, pp. 11-21.

[9] K. Atanassov, E. Szmidt, J. Kacprzyk, V. Atanassova, "An approach to a constructive simplification of multiagent multicriteria decision making problems via ICrA," *Comptes rendus de lAcademie bulgare des Sciences,* vol. 70 (8), 2017, pp. 1147-1156.

[10] K. Atanassov, P. Vassilev, O. Roeva, "Level Operators over Intuitionistic Fuzzy Index Matrices," *Mathematics,* vol. 9, 2021, pp. 366.

[11] V. Atanassova, O. Roeva, "Computational complexity and influence of numerical precision on the results of ICrA in the decision-making process," *Notes on IFSs,* vol. 24 (3), 2018, pp. 53-63.

[12] E. Boltürk, C. Kahraman, "Interval-valued and circular intuitionistic fuzzy present worth analyses," *Informatica,* vol. 33 (4), 2022, pp. 693-711.

[13] E. Çakır, M. A. Taş, Z. Ulukan, "Circular Intuitionistic Fuzzy Sets in Multi-Criteria Decision Making, *in Aliev, R.A., Kacprzyk, J., Pedrycz, W., Jamshidi, M., Babanli, M., Sadikoglu, F.M. (eds) 11th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions and Artificial Intelligence. ICSCCW 2021. Lecture Notes in Networks and Systems,* Springer, Cham, vol. 362, 2022, pp. 34-42.

[14] C. B. Cuong, V. Kreinovich, "Picture fuzzy sets-a new concept for computational intelligence problems," In: *Proceedings of the Third World Congress on Information and Communication Technologies WICT'2013*, Hanoi, Vietnam, 2013, pp. 1-6.

[15] P. Hsu, B. Chen, "Developing and Implementing a Selection Model for Bedding Chain Retail Store Franchisee Using Delphi and Fuzzy AHP," *Quality & Quantity,* vol. 41, 2007, pp. 275-290.

[16] R. J. Kuo, S. C. Chi, S. S. Kao, "A decision support system for selecting convenience store location through the integration of FAHP and artificial neural network," *Computers in Industry,* vol. 47 (2), 2002, pp. 199-214.

[17] T. Mahmood, S. Abdullah, S. -ur-Rashid, M. Bilal, "Multicriteria decision making based on a cubic set," *Journal of New Theory*, vol. 16, 2017, pp. 1-9.

[18] A. Mieszkowicz-Rolka, L. Rolka, "Multi-Criteria Decision-Making with Linguistic Labels," *in: M. Ganzha, L. Maciaszek, M. Paprzycki, D. Ślęzak (eds),* Proceedings of the 17th Conference FedCSIS, ACSIS, vol. 30, (2022), pp. 263–267.

[19] A. R. Mishra, P. Rani, K. Pandey, A. Mardani, J. Streimikis, D. Streimikiene, M. Alrasheedi, "Novel Multi-Criteria Intuitionistic Fuzzy SWARA–COPRAS Approach for Sustainability Evaluation of the Bioenergy Production Process," *Sustainability,* vol. 12, 2020, pp. 4155. DOI: 10.3390/su12104155

[20] X. T. Nguyen, V. D. Nguyen, "Support-Intuitionistic Fuzzy Set: A New Concept for Soft Computing," *I.J. Intelligent Systems and Applications,* 2015, 04, 2015, pp. 11-16.

[21] İ. Otay, C. Kahraman, "A novel circular intuitionistic fuzzy AHP & VIKOR methodology: An application to a multi-expert supplier evaluation problem," *Pamukkale Univ. J. Eng. Sci.*, vol. 28 (1), 2021, pp. 194–207.

[22] F. Smarandache, *Neutrosophy. Neutrosophic Probability,* Set, and Logic, Amer. Res. Press, Rehoboth, USA; 1998.

[23] E. Szmidt, J. Kacprzyk, "Amount of information and its reliability in the ranking of Atanassov's intuitionistic fuzzy alternatives," *in: Rakus-Andersson, E., Yager, R., Ichalkaranje, N., etc. (eds.),* Recent Advances in Decision Making, SCI, Springer, Heidelberg, vol. 222, 2009, pp. 7–19.

[24] V. Torra, "Hesitant fuzzy sets," *International Journal of Intelligent Systems*, vol. 25 (6), 2010, pp. 529-539.

[25] V. Traneva, D. Mavrov, S. Tranev, "Software Implementation of the Optimal Temporal Intuitionistic Fuzzy Algorithm for Franchisee Selection," *In: Proceedings of the 17th Conference on Computer Science and Intelligence Systems (FedCSIS),* 2022, pp. 387-390.

[26] V. Traneva, S. Tranev, "Intuitionistic Fuzzy Model for Franchisee Selection," *in: Kahraman, C., Tolga, A.C., Cevik Onar, S., Cebi, S., etc. (eds) Intelligent and Fuzzy Systems, INFUS 2022, Lecture Notes in Networks and Systems,* Springer, Cham, vol. 504, 2022, pp. 632-640.

[27] V. Traneva, P. Petrov, S. Tranev, "Circular IF Knapsack problem," *Lecture Notes in Networks and Systems,* (2023) (in press)

[28] V. Traneva, S. Tranev, D. Mavrov, "Interval-Valued Intuitionistic Fuzzy Decision-Making Method using Index Matrices and Application in Outsourcing," *in: Proceedings of the 16th Conference on Computer Science and Information Systems (FedCSIS),* Sofia, Bulgaria, vol. 25, 2021, pp. 251–255.

[29] V. Traneva, P. Petrov, S. Tranev, "Petrol Station Franchisor Selection through Circular Intuitionistic Fuzzy Multicriteria Analysis," *Lecture Notes in Networks and Systems,* (2023) (in press)

[30] V. Traneva, S. Tranev, M. Stoenchev, K. Atanassov, " Scaled aggregation operations over two- and three-dimensional index matrices," *Soft computing,* vol. 22, 2019, pp. 5115-5120.

[31] V. Traneva, S. Tranev, *Index Matrices as a Tool for Managerial Decision Making,* Publ. House of the USB; 2017 (in Bulgarian).

[32] P. Vassilev, K. Atanassov, *Modifications and extensions of Intuitionistic Fuzzy Sets,* "Prof. Marin Drinov" Academic Publishing House, Sofia, 2019.

[33] X. Xu, Y. Lei, W. Dai, "Intuitionistic Fuzzy Integer Programming Based on Improved Particle Swarm Optimization," *J. Comput. Appl.,*, vol. 9, 2008, pp. 062.

[34] L. Zadeh, "Fuzzy Sets," *Information and Control,* vol. 8 (3), 1965, pp. 338-353.

Communication Papers of the 18th Conference on Computer Science and Intelligence Systems pp. 345–351
DOI: 10.15439/2023F6208
ISSN 2300-5963 ACSIS, Vol. 37

A modular and verifiable software architecture for interconnected medical systems in intensive care

Marc Wiartalla, Frederik Berg, Florian Ottersbach, Jan Kühn, Mateusz Buglowski,
Stefan Kowalewski, André Stollenwerk
Informatik 11 - Embedded Software, RWTH Aachen University
Ahornstraße 55, Aachen, Germany

***Abstract*—Medical device interoperability enables new therapy methods and the automation of existing ones. Due to different medical device manufacturers and protocols, we need auxiliary hardware and software for the interconnection. In this paper we propose a service-oriented software architecture built on a real-time operating system in order to create a modular medical cyber-physical system consisting of networked embedded nodes. In particular we highlight the need for the application of formal methods to ensure the functional safety of the system.**

I. Introduction

IN MEDICAL intensive care many different medical devices from various manufacturers are used for therapies. More and more future therapy methods are based on interconnected devices, called medical cyber-physical systems or cyber-medical systems [1], [2]. One particular case is the class of physiological closed loop control systems that aim to control one or several physiological parameter based on sensor measurements. These systems enable new therapies and the automation of existing ones and thus relieve the clinic personnel. In such systems flexibility and modularity is essential. Firstly, because clinics use different medical devices by manufacturers with different interfaces and protocols. Secondly, because it is always necessary to react to new insights or changes in the patient's condition, e.g. extend the therapy with more devices.

However due to the current legislation, a medical device which consists of an interconnection of different medical devices that are already authorized for the market, needs to pass the complete authorization process again. Hence, today many of these devices are interconnected manually by clinic personnel, e.g. reading measurements from one device and feeding these values into another device. Of course, for the future this legislation issue is supposed to be solved. Once this interconnection of authorized medical devices is integrated in the legislation, we need a software architecture to allow for the interconnection of these devices in a safe manner. As it is necessary to react to new insights in an agile way during therapy, the software architecture needs to be reliable and modular.

Without a doubt, medical systems are safety-critical as any fault can lead to a patient's harm or even death. In our opinion, it is therefore essential to apply existing state-of-the-art formal methods for the verification and validation of the medical software. Even though the use of formal methods is recommended by authorities, the application is still not enforced today [3]. Our thesis is that a future software architecture must support verification as best as possible.

In the use-case of physiological closed loop control systems the reaction time to external events like a change in physiological parameters is crucial. It might also be necessary to fulfill real-time constraints. This is why the vast majority of these algorithms are executed on embedded devices. Instead of using one centralized complex system, this allows us to distribute the system among multiple smaller nodes. Thus, resulting in significantly reduced software complexity on each node without reducing the overall processing capabilities. In this paper we present a modular extension of an existing real-time operating system in order to create a medical cyber-physical system consisting of networked embedded nodes.

A. Worked example

As a worked example for the next chapters we use the automation of extracorporeal membrane oxygenation (ECMO), where a patient with severe lung failure is supported by blood-based gas exchange outside the patient's body [4]. Figure 1 shows a sketch of the setup.

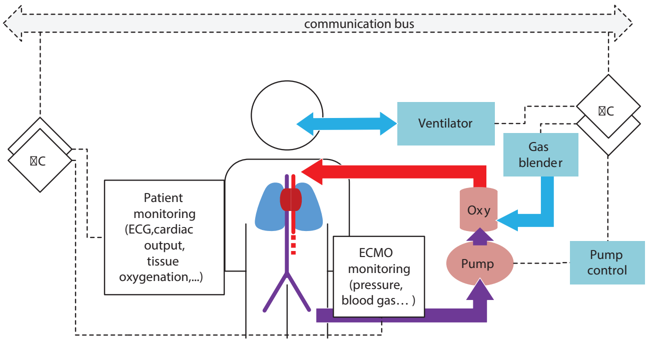

communication bus
⊠C
Ventilator
⊠C
Gas blender
Patient monitoring (ECG,cardiac output, tissue oxygenation,...)
Oxy
ECMO monitoring (pressure, blood gas…)
Pump
Pump control

Fig. 1. Worked example: Automated ECMO therapy setup [5]

In the setup, a blood pump creates a blood flow through an oxygenator. Within the oxygenator, there is a blood and gas flow separated by a special membrane. This membrane allows for gas exchange, in particular the enrichment of blood with oxygen and the elimination of carbone-dioxide out of the blood. As a second actuator besides the blood pump, the gas-blender is in charge of controlling the gas composition and flow through the oxygenator. The sensors in the worked example are an online blood gas monitoring system as well

Thematic track: Software Engineering for Cyber-Physical Systems

as different pressure and flow sensors. In this setup we also implemented an interface for the patient's ventilator to efficiently cover the interaction between both therapy devices.

II. State of the Art

For the related work, we will analyze three aspects. First, we will present existing real-time operating systems (RTOS), on which our software architecture is built. Secondly, we will give an overview over the existing messaging protocols for the communication between nodes. Finally, we will present several projects working in the field of medical device interoperability. With these three aspects in mind we can highlight the key differences to our proposed architecture. In addition, we will present related work from non-medical domains.

The basis for our software architecture is the operating system underneath. For safety-critical embedded systems, the preferred option are real-time operating systems for microcontrollers due to the resulting determinism. The market-leading real-time operating system is FreeRTOS [6], which includes a kernel and Internet of Things (IoT) libraries. An alternative operating system is ChibiOS [7], which includes an operating system as well as a Hardware Abstraction Layer (HAL), peripheral drivers and a complete development environment. Furthermore, there exist several real-time variants of the Linux operating system, which are designed for the use in safety-critical software but are far more complex than the previously listed operating systems for microcontrollers [8]. These Linux operating systems might also contain pre-compiled libraries without source-code access.

Next, we will discuss which messaging protocols exist for the data exchange between nodes. Here we will focus on Ethernet for the communication between nodes. Many current research projects use the data distribution service (DDS) standard for interconnectivity, standardized by the object management group (OMG) [9]. DDS allows participants to communicate by publishing and subscribing to topics. The data can be shared with Quality of Service (QoS) specifications to ensure certain properties like reliability or periodicity. While DDS is an Application Programming Interface (API) specification, there exist several different DDS implementations, for example the C++ implementation eprosima FastDDS [10] and embeddedRTPS [11], a portable and open-source implementation of the Real-Time Publish-Subscribe Protocol for embedded systems. There also exist several alternative messaging transport protocols such as the MQTT protocol [12], a lightweight publish/subscribe messaging protocol which is often used in the IoT.

After discussing the state of the art of the underlying operating system and messaging protocol, we will present related work in medical device interoperability. Together with the Center for Integration of Medicine & Innovative Technology (CIMIT), the Medical Device 'Plug-and-Play' Interoperability Program (MD PnP) proposed the open standards for the Patient-Centric Integrated Clinical Environment (ICE) [13]. The standard defines the conceptual model, general requirements and different clinical use cases. Additional standards like the data logger are planned and in work. In the MD PnP program, an open source implementation of an interconnection environment called OpenICE was developed [14]. In an OpenICE system distributed network nodes are connected via Ethernet. OpenICE Device Adapters act as bridges to connect medical devices with the network. In addition, a central OpenICE supervisor unit runs clinical applications and logs data. All nodes in an OpenICE system have to be Java-capable devices like Linux computers. OpenICE uses DDS as the messaging protocol, in the current version the DDS implementation by Real-Time Innovations (RTI) [15]. For the OpenICE architecture several supervisor application examples were implemented, e.g. the synchronization between a ventilator and the shutter of an x-ray as a closed loop control use-case.

In the project OR.NET [16] concepts for open medical device interoperability in the operating room and clinic were developed. The concept of a service-oriented medical device architecture (SOMDA) including the technical specification was standardized in the IEEE 11073 SDC family [17], [18], [19]. In the OR.NET project, the service-oriented device architecture (SODA) was refined to the SOMDA paradigm. Besides standardized interface descriptions, a standardized way to describe provided and exchanged data was developed. As a base technology, the Devices Profile for Web Services (DPWS) is used for communication. Several open source frameworks implement the IEEE 11073 SDC standards in different programming languages, e.g. openSDC (Java) and SDCLib/C (C++).

Another research project called Smart Cyber Operating Theater (SCOT) works on an integrated operating room [20]. The SCOT project is based on the ORiN network interface for robot systems [21].

From the previously mentioned projects, the ICE project proposes concepts for the complete clinical IT infrastructure and follows a more centralized approach, with a central ICE Supervisor unit, that runs all applications. This differs from our approach, where the system is distributed among multiple smaller nodes. The OR.NET and SCOT projects mostly focus on an integrated operating room. The focus on the operating room results in specific requirements, e.g. a centralized console for visualization, and many scenarios have no real-time requirements [16]. In addition, a major part of the OR.NET project is the standardization of interfaces, to specify which data are exchanged between devices.

In contrast to the related work, our proposed software architecture is aimed at medical systems in the intensive care unit. In this paper we also exclude the direct communication with the clinical network. Notably, physiological closed loop control systems are a special use-case as these are highly automated. The main difference is that the patient in an operating room is under constant human monitoring. This is not the case for the intensive care unit, where patients are in critical state but not always under direct monitoring by the staff. In the future, it is even conceivable that such closed loop control systems are used outside of the hospital without constant monitoring by medical professionals. In these use cases, the safety of these systems is essential, as medical professionals can not immediately react to faults in the system. Our architecture therefore includes a dedicated safety layer for various safety measures.

Additionally, we emphasize that it is necessary to use formal methods in the development process and increase safety through various design decisions. However, the presented OR.NET SDC and OpenICE projects are built on conventional Java-capable operating systems like Windows and Linux. Due to the high complexity, the formal verification and testing of the whole Linux operating system is still an open research topic. With Windows the verification and testing are even harder, as the operating system is not open-source. In addition, the usage of the Java runtime environment further increases the complexity and abstraction. There exists a real-time specification for Java to be used in real-time software development, which changes the semantics of the scheduling and memory management [22].

In contrast, our proposed software architecture builds on a real-time operating system for embedded systems with a small code scope and open-source libraries. This operating system can be verified through formal methods as it is far less extensive and complex. In addition, we aid the usage of formal methods through certain design decisions presented in the later chapters.

The previously mentioned interoperability projects all focus on the medical environment. However, the interconnection of nodes and data exchange is also relevant in different non-medical domains. The robot operating system 2 (ROS 2) [23] is a middleware for building robot applications including software libraries and tools. The second version improved many limitations of ROS 1. Similar to our approach a ROS system consists of a network of nodes which communicate over DDS. However, the provided tools are specialized for robot applications, while our proposed software is aimed at medical applications. With ROS-Health an extension of the Robot Operating System was developed for the use in neurorobotics [24].

III. Embedded Software Architecture

The software architecture proposed for interconnected intensive care applications is based on a real-time operating system (RTOS) for embedded systems. In our architecture, we use different features of the operating system like multitasking, timers and synchronization. In particular, the operating system should include a hardware abstraction layer. This abstraction allows us to freely change the used hardware platform. As there exist several operating systems with these features, the specific operating system can be abstracted. Figure 2 shows the resulting software architecture.

The following chapter describes the proposed software architecture from the top to the bottom. First, the application layer is presented with different development methods for data processing algorithms. Next, the data provisioning layer with the used communication protocol is introduced. Finally, we will discuss how our software architecture supports the application of formal methods to ensure safety.

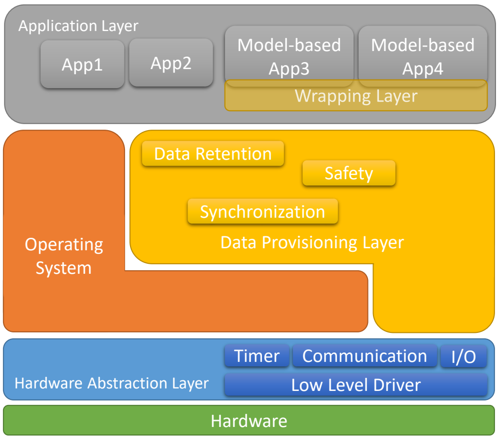


Fig. 2. Reference software architecture for medical applications in intensive care

A. Application Layer

Topmost in the architecture is the application layer. The applications are the actual data processing algorithms. In our architecture one node can run multiple applications depending on the requirements, e.g. the required computational power. Applications can also be moved between nodes without the need to adapt the application.

Applications can either be implemented directly or be generated from specific models. For model-based development we create models, e.g. diagram-based, using suitable software tools and expertise by medical personnel. From these models we can automatically generate code which is supported by several modern modeling tools. The generated code has to be connected with the data retention layer. Therefore we introduce a wrapping layer. If suitable physiological models of the patient or models of the physiological processes are available, we can also simulate the system and test our applications against these models.

As the software architecture is based on an embedded real-time operating system, each application has to be registered as a task by the developer with a declaration of the required stack size for this application. In addition, we do not use dynamic memory allocation or management offered by the operating system, in order to prevent memory allocation failures during runtime and aid the application of formal methods in such safety-critical systems.

B. Data Provisioning Layer

The main task of the data provisioning layer (DPL) is to store the data needed by the applications on the individual node. In addition, the data generated by this node needs to be communicated throughout the network. For this the concrete communication medium can be abstracted as it depends on the application requirements.

We implemented the data provisioning as a separate layer to enable modularity. This allows us free movement of applications between the different nodes in the system without the need to adapt the application. For this we create a global listing of all possible measurement and internally generated values in the system. This global listing is called communication matrix as we adopted the name from the Controller Area Network (CAN). The communication matrix contains unique identifiers for each message as well as additional information like label, description, scaling factor and unit. The wrapping layer and the DPL are automatically generated from this matrix, so the communication matrix will be referenced in the following architectural parts. In our architecture each application has to announce which data it uses, so that only the required data is stored on each node. This is later referenced as the selective part of the data provisioning layer.

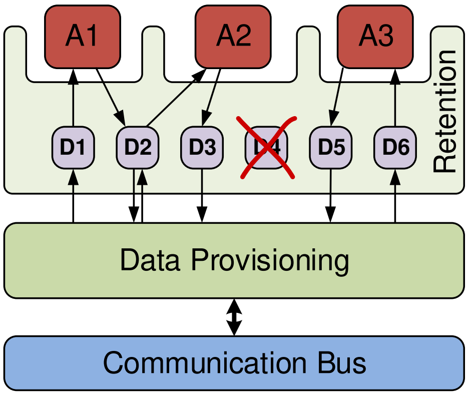


Fig. 3. Realization of the Data Retention [25]

Figure 3 shows the resulting data retention in our architecture. In this example the communication matrix consists of six different pieces of information (data) D1 to D6. Three different applications A1 to A3 run on the depicted node and read and write different data. As the data D4 is not used by any application it is locally eliminated. This methodology facilitates the movement of applications between nodes.

1) Time Synchronization: The simplest implementation for the data retention would be to just store the last broadcasted measurement value. However, in addition it is also necessary to store the timestamps of measurements and generated values. As an example, we describe two scenarios from the ECMO. First, we might want to keep track of a patient's body temperature during the treatment. To put the temperature measurements in a chronological order, we need timestamps for each measurement. However, in this scenario an accuracy of one second is sufficient. Second, it might be necessary to compute the patient's dynamic lung compliance during mechanical ventilation, which can be calculated from the tidal volume and pressure. For the calculation of the compliance we therefore need to correlate pressure measurements and flow measurements. Thus, we need precise measurement timestamps with an accuracy in the range of few milliseconds.

To support such scenarios, we need to keep track of time locally on each specific node, but also globally for the whole network of nodes. This global time allows us to check if a value is outdated but also correlate values from different nodes. The required accuracy of the time synchronization depends on the specific application requirements and used synchronization protocol.

For this we define a master-node and synchronize all other nodes to this node. There are several existing network synchronization protocols depending on the choice of the communication medium, for example OCS-CAN [26] for CAN and the Precision Time Protocol (PTP) [27] for Ethernet communication.

Overall the synchronization has to be defined in a way, such that the local clock-error on a specific node does not exceed an upper bound. In the example of PTP the accuracy depends on whether we use a hardware or software implementation. In our worked example we use a software implementation and can achieve an accuracy of 1 millisecond. Next, we present the data retention layer.

2) Data Retention: In medical applications it is not only necessary to work with single measurement values but also time series. In the calculation of the lung compliance during mechanical ventilation, we need to store series of flow and pressure measurements to compute the change in volume and pressure.

In order to support such medical scenarios, the data retention is able to store a time series of data. We can configure the length of the series as well as the sample rate. This can be configured for each individual app with reference to the used measurements. Since it is not always necessary to work with the complete series of data, we realized general and specific operations on these time series. These resulting values are constantly calculated in the background as new generated data is received. As general operators we implemented the minimum, maximum, median and mean operations. These should already address many needs during the processing of medical data. In addition, it is also possible to define specialized data operations tailored for specific applications. This is realized by allowing the user to register a callback function, which is executed on the time series.

3) Safety: As mentioned before, medical applications in particular are safety critical. Therefore, we include an additional safety layer in the software architecture [28]. In this layer, the control values for actuators can be safeguarded but also measurement values of sensors can be annotated with a metric regarding aspects like plausibility, data quality or age.

The basic approach to safeguard transmitted values are general boundaries defined in the communication matrix. Medical devices often have maximum ratings for their operation. In our worked example, the used oxygenator has a maximum rating for the set-values of gas and blood flow. Moreover, to maintain a basic patient support in the context of the treatment we might want to define a minimum gas and blood flow depending on the patient's parameters. In addition to this, even more intelligent safeguarding is possible. We can derive mathematical equations from physical or physiological models and integrate them in the safety layer. Also the usage of artificial intelligence is possible. In a third step, it is also important to consider the age of a measurement value. If a measurement in our system is several minutes old, we have to react accordingly and for example, use a fallback value. Furthermore, we might need information about the quality of measured data, which can be transmitted over the network with additional messages if this information is available.

C. Formal Methods

Since software errors can lead to harm of the patient and even death, medical software is highly safety-critical. We therefore emphasize that the application of formal methods in the software development process is essential.

In our embedded software architecture, we made several design decisions, which aid the application of formal method. These are mainly the fully static architecture and knowledge of the underlying operating system. In the application layer we have to declare the stack size for each registered task. With this information we can then perform a stack size analysis to prevent stack overflows. Many compilers already offer stack size analysis capabilities [29] and from our point of view, this is one of the most important measures to safeguard a medical cyber-physical system.

Furthermore, static analysis enables us to prove the absence of critical run-time errors without having to execute the code. Common errors that can be found with static analysis techniques are dead or unreachable code, uninitialized variables, null pointer dereference and invalid arithmetic operations. It is advisable to use static analysis early in the development process as bugs are cheaper to fix compared to later development stages. [30]

The adherence and documentation of compliance with a suitable coding guideline, like MISRA C [31], MITRE Common Weakness Enumeration (CWE) [32] or the SEI CERT C coding standard [33] is highly recommended when dealing with safety-critical systems. The MISRA C coding guidelines define a subset of the C language without parts with e.g. undefined behavior. This is aimed at error prevention and is supposed to increase code readability and explainability [34]. The MISRA C rules can be checked by algorithms, which will be used in the worked example implementation.

IV. Worked Example Implementation

In our worked example, we implemented the automation of an extracorporeal membrane oxygenation therapy for patients who suffer from Acute Respiratory Distress Syndrome (ARDS) [35], [36]. In the setup, microcontroller nodes were connected to medical devices and communicate over a network or communication bus. Therefore, we implemented the software architecture as presented in the previous chapter. The implemented architecture with the used RTOS and hardware is shown in figure 4.

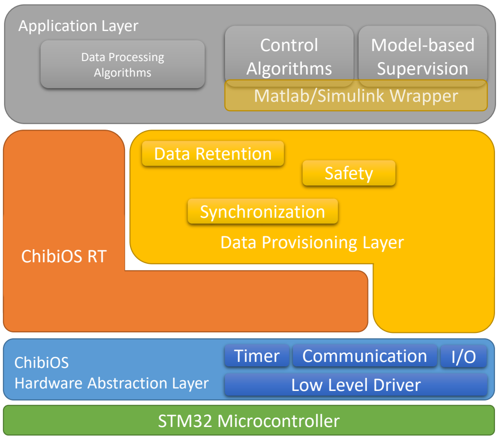


Fig. 4. Implementation of software architecture

The implementation of the software architecture with an example project for an OxiMax N-560 pulse oximeter is available in [37]. As the real-time operating system we use ChibiOS as it offers a wide range of features including a fully static architecture and hardware support for the used STM32 microcontroller. Because of the existing hardware abstraction layer within ChibiOS, we were able to switch from our previous hardware platform with an Atmel AT91 microcontroller to a STM32 microcontroller with minimal changes in the software. The communication matrix is defined in multiple TOML (Tom's Obvious Minimal Language) files. Based on these TOML files we generate parts of the data provisioning and safety layer. The selective part of the data provisioning layer is based on the C compiler's pre-processor. Thus, we still only use static memory allocation.

In the first version of the architecture, the nodes were connected via CAN. In the latest version, we added Ethernet for the communication in order to allow better integration of mobile devices for monitoring.

We use an existing end-to-end middleware, because we have less code to maintain as the specification and different implementations already exist. We decided to use DDS, as it offers us great flexibility to control the communication behavior through the quality of service policies and has an integrated discovery system. With these QoS policies we can enforce additional safety requirements like the periodicity of measurements or the usage of redundant sensors. If the periodicity of data is known, we can notify the user about missing measurements. In addition, DDS is already established in several safety-critical scenarios, like aviation or military applications [38]. Another advantage is that DDS is suited for the use in embedded systems and offers a wide language and platform support, which eases the integration of other devices. For the specific DDS implementation, we use embeddedRTPS [11] as it is suited for embedded systems.

In our worked example implementation, it is possible to prioritize certain messages like alarms, as these should be treated with higher priorities. These alarm message have to be prioritized in the software as well as in the network. Using Ethernet, these messages are prioritized with the differentiated services code point (DSCP) in the IP header. However, we also need infrastructure, like switches, that support this prioritization. For CAN communication the CAN message identifier determines the priority with a low message identifier representing a high priority.

The actual applications, e.g. control algorithms, are either implemented directly in the language C or generated C-Code from Matlab Simulink (The Mathworks, Natick, MA) [39] models is used. We automatically generate Simulink blocks from the TOML files to allow the developer to use any data out of the communication matrix without the need to take care of its retention.

In the described implementation, we solely use static memory structures. Additionally, we apply different formal methods. First of all, we carry out a stack size analysis and compare the results with the declared stack size registered to the operating system. In addition, we carried out static analysis using Polyspace (The Mathworks, Natick, MA) [40] to prove the absence of critical run-time errors. Furthermore, we used Polyspace to check the compliance of our software with the MISRA C rules.

Finally, we also carried out a worst-case execution time analysis of the used algorithms to get an over-approximation of the total CPU consumption on a specific node in the medical cyber-physical system. The analysis was conducted using aiT from AbsInt [41]. One analyzed algorithm is the model-based blood pump supervision, which can detect and predict events like the suction of the withdrawing cannula to the wall of the surrounding vessel or the presence of gas bubbles in the blood tubing. This analysis leads to a worst-case CPU time of 0.138 milliseconds on the Atmel AT91SAM7 hardware, which makes the algorithm suited for the use in an embedded environment. [36]

V. Results

In section III, we presented a software architecture for interconnected medical systems. The architecture enables modularity, as we can move applications between nodes depending on the required memory and CPU time. However, the network delay has to be considered. We might prefer to process sensor data on the same node as the control algorithm, which uses this data as it's input. Through a loopback mechanism this data does not need to be sent over the network, which decreases the delay.

Additionally, the data provisioning layer provides basic safety mechanism which can be easily extended with more complex methods like physiological models. This goes hand in hand with the need for precise global timestamps of measurements or information on the time passed between measurements. This enrichment of the bare data with additional information also helps with the safeguarding of measurements. The safety layer can be extended depending on the requirements, however, often the difficulty is to find physiological models in the appropriate abstraction to derive mathematical formulas. In addition, we highlighted the need for formal methods in the development process and aided the application by various design decisions.

One major advantage of the presented software architecture is the free choice for the used hardware and operating system. Because of the abstraction layer, we do not need to modify our existing software when enhancing an already existing setup.

VI. Conclusion

In this paper, we presented a modular and verifiable software architecture based on a real-time operating system suited for interconnected medical systems in intensive care environments. We presented the general concept of our proposed entities, of which such a medical cyber-physical system can consist of, and presented a worked example implementation for the automation of an extracorporeal membrane oxygenation therapy. The main component we introduced was the generated data provisioning layer and its interactions to the other parts of the software architecture. The DPL takes care of the data storage, needed by the different applications and algorithms. Additionally, the DPL offers time synchronization as well as communicating the data between nodes. The included safety layer allows us to safeguard transmitted values and measurements. Finally, we enable the integration of model-based generated code into this software architecture by a wrapping layer.

We presented a strictly static software architecture, which allows for the efficient use of formal methods. This allows us to prove the absence of possibly safety related issues like the violation of real-time constraints or memory/stack overflows, which need to be avoided in medical systems. In conclusion, the combination of these measures allows for the safe operation of medical cyber-physical systems.

However, the applied formal methods presented in this work are just the standard methods and the verification of more properties is possible. A next step would be the validation and verification of the control algorithms to ensure the patient's safety. Though, the verification in a physiological closed-loop system requires an accurate physiological model of the system, which is challenging due to the complexity and variability of the human physiology.

As an further outlook, we plan to improve the safety layer by deriving safeguarding algorithms out of publicly available physiological models. Through the use of publicly available databases, there is also a high potential to automatically learn the needed relations between different physiological values by means of artificial intelligence. In addition, we plan to add different methods for the detection and diagnosis of hardware faults and medical complications in the system.

VII. Acknowledgments

The authors gratefully acknowledge the funding of the German Research Foundation (PAK 138/1 and 2) and the Federal Ministry of Education and Research (031L0253B), which allowed for these results.

References

[1] I. Lee and O. Sokolsky, "Medical cyber physical systems," in *Design Automation Conference*, pp. 743–748, ACM, 2010.

[2] G. De Micheli, "Cyber-medical systems: Requirements, components and design examples," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 64, no. 9, pp. 2226–2236, 2017.

[3] S. Bonfanti, A. Gargantini, and A. Mashkoor, "A systematic literature review of the use of formal methods in medical software systems," *Journal of Software: Evolution and Process*, vol. 30, no. 5, p. e1943, 2018.

[4] Rüdger Kopp, Ralf Bensberg, Marian Walter, Jutta Arens, Rolf Rossaint, and André Stollenwerk, "Automation of extracorporeal membrane oxygenation using a combined safety and control concept.," *Intensive Care Medicine*, vol. 37, no. S1, 2011.

[5] J. Kühn, C. Brendle, A. Stollenwerk, M. Schweigler, S. Kowalewski, T. Janisch, R. Rossaint, S. Leonhardt, M. Walter, and R. Kopp, "Decentralized safety concept for closed-loop controlled intensive care," *Biomedical Engineering/Biomedizinische Technik*, vol. 62, no. 2, pp. 213–223, 2017.

[6] Richard Barry, "FreeRTOS," 2023. https://www.freertos.org/.

[7] Giovanni Di Sirio., "ChibiOS," 2023. https://www.chibios.org.

[8] F. Reghenzani, G. Massari, and W. Fornaciari, "The real-time linux kernel: A survey on preempt_rt," *ACM Computing Surveys*, vol. 52, no. 1, pp. 1–36, 2020.

[9] Object Management Group, "Data distribution service specification, version 1.4," 10.04.2015.

[10] eProsima, "Fast DDS," 2023. https://www.eprosima.com/index.php/products-all/eprosima-fast-dds.

[11] A. Kampmann, A. Wustenberg, B. Alrifaee, and S. Kowalewski, "A portable implementation of the real-time publish-subscribe protocol for microcontrollers in distributed robotic applications," in *The 2019 IEEE Intelligent Transportation Systems Conference - ITSC*, (Piscataway, NJ), pp. 443–448, IEEE, 2019.

[12] OASIS MQTT Technical Committee, "MQTT, Version 5.0," 07.03.2019.

[13] ASTM, "Medical devices and medical systems - essential safety requirements for equipment comprising the patient-centric integrated clinical environment (ice) - part 1: General requirements and conceptual model," 2013.

[14] J. Plourde, D. Arney, and J. M. Goldman, "OpenICE: An open, interoperable platform for medical cyber-physical systems," in *2014 ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS 2014)*, (Piscataway, NJ), p. 221, IEEE, 2014.

[15] Real-Time Innovations, "Connext DDS," 2023. https://www.rti.com/.

[16] M. Kasparick, M. Schmitz, B. Andersen, M. Rockstroh, S. Franke, S. Schlichting, F. Golatowski, and D. Timmermann, "OR.NET: a service-oriented architecture for safe and dynamic medical device interoperability," *Biomedizinische Technik. Biomedical engineering*, vol. 63, no. 1, pp. 11–30, 2018.

[17] IEEE Engineering in Medicine and Biology Society, "IEEE Standard - Health informatics – Point-of-care medical device communication - Part 10207: Domain Information and Service Model for Service-Oriented Point-of-Care Medical Device Communication," 2017.

[18] IEEE Engineering in Medicine and Biology Society, "IEEE Standard - Health informatics – Point-of-care medical device communication - Part 20702: Medical devices communication profile for web services," 2016.

[19] IEEE Engineering in Medicine and Biology Society, "IEEE Standard - Health informatics – Point-of-care medical device communication - Part 20701: Service-Oriented Medical Device Exchange Architecture and Protocol Binding," 2019.

[20] J. Okamoto, K. Masamune, H. Iseki, and Y. Muragaki, "Development concepts of a smart cyber operating theater (scot) using orin technology," *Biomedizinische Technik. Biomedical engineering*, vol. 63, no. 1, pp. 31–37, 2018.

[21] M. Mizukawa, H. Matsuka, T. Koyama, T. Inukai, A. Noda, H. Tezuka, Y. Noguchi, and N. Otera, "Orin: open robot interface for the network - the standard and unified network interface for industrial robot applications," in *SICE 2002*, (Tōkyō), pp. 925–928, SICE, 2002.

[22] G. Bollella and J. Gosling, "The real-time specification for java," *Computer*, vol. 33, no. 6, pp. 47–54, 2000.

[23] S. Macenski, T. Foote, B. Gerkey, C. Lalancette, and W. Woodall, "Robot operating system 2: Design, architecture, and uses in the wild," *Science robotics*, vol. 7, no. 66, p. eabm6074, 2022.

[24] G. Beraldo, N. Castaman, R. Bortoletto, E. Pagello, J. del R. Millan, L. Tonin, and E. Menegatti, "Ros-health: An open-source framework for neurorobotics," in *2018 IEEE International Conference on Simulation, Modeling, and Programming for Autonomous Robots (SIMPAR)* (H. Kurniawati, ed.), (Piscataway, NJ), pp. 174–179, IEEE, 2018.

[25] A. Stollenwerk, F. Göbe, M. Walter, J. Arens, R. Kopp, and S. Kowalewski, "Smart Data Provisioning for Model-Based Generated Code in an Intensive Care Application," in *3rd Joint Workshop On High Confidence Medical Devices, Software, and Systems & Medical Device Plug-and-Play Interoperability : HCMDSS/MDPnP 2011 ; in conjunction with CPSweek 2011*, (Chicago), HCMDSS/MDPnP 2011, Apr 2011.

[26] G. Rodriguez-Navas, S. Roca, and J. Proenza, "Orthogonal, fault-tolerant, and high-precision clock synchronization for the controller area network," *IEEE transactions on industrial informatics*, vol. 4, no. 2, pp. 92–101, 2008.

[27] IEEE Instrumentation and Measurement Society, "IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems," 2019.

[28] J. Kühn, A. Stollenwerk, C. Brendle, T. Janisch, M. Walter, R. Rossaint, S. Leonhardt, S. Kowalewski, and R. Kopp, "Sensor supervision and control value limitations in networked intensive care," in *Gemeinsamer Tagungsband der Workshops der Tagung Software Engineering 2016 (SE 2016), Wien, 23.-26. Februar 2016* (W. Zimmermann, L. Alperowitz, B. Brügge, J. Fahsel, A. Herrmann, A. Hoffmann, A. Krall, D. Landes, H. Lichter, D. Riehle, I. Schaefer, C. Scheuermann, A. Schlaefer, S. Schupp, A. Seitz, A. Steffens, A. Stollenwerk, and R. Weißbach, eds.), vol. 1559 of *CEUR Workshop Proceedings*, pp. 187–194, CEUR-WS.org, 2016.

[29] E. Botcazou, C. Comar, and O. Hainque, "Compile-time stack requirements analysis with gcc: Motivation, development, and experiments results," in *Proc. GCC Developers Summit*, pp. 93–105, 2005.

[30] A. Gosain and G. Sharma, "Static analysis: A survey of techniques and tools," in *Intelligent Computing and Applications*, pp. 581–591, Springer, 2015.

[31] A. Burnard, P. Burden, L. Whiting, C. Tapp, G. McCall, M. Hennell, C. Hills, and S. Montgomery, "MISRA C:2012," 2013.

[32] P. Anderson, B. Curtis, P. Braione, A. Summers, C. Eng, J. Fung, J. Gazlay, A. Hoole, J. Jarzombek, J. Lam, C. Levendis, J. Oberg, K. Seifried, C. Turner, and A. van der Stock, "Common weakness enumeration," *Mitre Corporation*, 2007.

[33] Software Engineering Insitute CERT, "C coding standard: Rules for developing safe, reliable, and secure systems," *Reliable, and Secure Systems*, 2016.

[34] R. Bagnara, A. Bagnara, and P. M. Hill, "The MISRA C Coding Standard and its Role in the Development and Analysis of Safety- and Security-Critical Embedded Software," in *Static Analysis* (A. Podelski, ed.), vol. 11002 of *Lecture Notes in Computer Science*, pp. 5–23, Cham: Springer International Publishing, 2018.

[35] R. Kopp, R. Bensberg, A. Stollenwerk, J. Arens, O. Grottke, M. Walter, and R. Rossaint, "Automatic control of veno-venous extracorporeal lung assist," *Artificial organs*, vol. 40, no. 10, pp. 992–998, 2016.

[36] A. Stollenwerk, J. Kühn, C. Brendle, M. Walter, J. Arens, M. N. Wardeh, S. Kowalewski, and R. Kopp, "Model-based supervision of a blood pump," *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 6593–6598, 2014.

[37] M. Wiartalla, F. Berg, F. Ottersbach, J. Kühn, M. Buglowski, S. Kowalewski, and A. Stollenwerk, "A modular and verifiable software architecture for interconnected medical systems in intensive care," 2023. https://doi.org/10.18154/RWTH-2023-07342.

[38] Object Management Group, "Who's Using DDS?," accessed 21.12.2022. https://www.dds-foundation.org/who-is-using-dds-2/.

[39] The Mathworks, Inc., "MATLAB Simulink (R2022b)," 2022.

[40] The Mathworks, Inc., "Polyspace (R2022b)," 2022.

[41] C. Ferdinand, "Worst case execution time prediction by static program analysis," in *Proceedings / 18th International Parallel and Distributed Processing Symposium*, (Los Alamitos, Calif.), pp. 125–127, IEEE Computer Society, 2004.

Communication Papers of the 18th Conference on Computer Science and Intelligence Systems pp. 353–358

DOI: 10.15439/2023F9991
ISSN 2300-5963 ACSIS, Vol. 37

Experimental Assessment of MPTCP Congestion Control Algorithms for Streaming Services in Open Internet

Łukasz Piotr Łuczak
0000−0003−0892−7276
Institute of Information Technology
Lodz University of Technology
Łódź, Poland
lukasz.luczak@dokt.p.lodz.pl

Przemysław Ignaciuk, *IEEE Senior Member*
0000−0003−4420−9941
Institute of Information Technology
Lodz University of Technology
Łódź, Poland
przemyslaw.ignaciuk@p.lodz.pl

Michał Morawski
0000−0002−8902−1259
Institute of Information Technology
Lodz University of Technology
Łódź, Poland
michal.morawski@p.lodz.pl

***Abstract*—Efficient data transfer is required in various fields such as entertainment, business communications, image processing systems, and industrial applications. A fast speed, low latency, and stable transmission parameters are required to ensure high-quality streaming, which is difficult to achieve with a single data channel. Using multiple communication paths is a promising solution for elevating performance. The Multipath TCP (MPTCP) protocol allows for splitting the application stream among a few connections. A key element determining the overall transmission quality is the MPTCP congestion control algorithm. In this paper, the most common MPTCP congestion control algorithms are evaluated in the open Internet in the context of streaming applications. The results obtained indicate that for a streaming service that utilizes multiple paths the most effective pair of CC algorithms are BALIA at the MPTCP level and BBR at the path level. These algorithms provide the smallest path delay and Head-of-Line blocking degree under consistent throughput. Delay-based wVegas shows the weakest performance in terms of multipath streaming.**

***Index Terms*—MPTCP, congestion control, streaming applications, tactile Internet**

I. Introduction

IP-based systems are gradually replacing other network solutions in traditional telecommunications, medicine, and industrial automation, as well as in new areas like entertainment, Internet of Things (IoT), and tactile Internet [13]. It occurs despite the fact that other solutions can provide better Quality of Service (QoS) measures, e.g., guaranteed minimum bandwidth, fault tolerance, or maximum latency. An unquestioned advantage of IP networks is their universality and ease of expansion, which results in economic benefits. Unlike other network solutions, IP networks require a generic connection to the network, only, utilizing its dynamic routing capabilities as a transport basis. Despite continuous efforts to improve QoS [17], a disadvantage of IP networks is the lack of control over the transmission quality. For time-sensitive transmissions, UDP/RTP protocols may be used. However, due to security requirements or application restrictions, the preferred form of data transmission is the TCP protocol.

The widespread use of mobile appliances has created new possibilities and challenges. The link parameters vary with time, rapidly. In addition, the movement of devices makes it necessary to smoothly switch to another network. Although the logical IP address of the device may remain unchanged, the link parameters may be radically different. In the considered class, the terminals are often equipped with more than one network interface, e.g., a cellular (LTE, 5G) and a Wi-Fi one. Already, these two interfaces have completely different characteristics. In addition, changing the location entails a change of the access point. Various phenomena, outside the control of communicating agent, such as interference from other users and appliances, aggravate the system uncertainty and limit the available range of services. In order to address these problems, it has been proposed to simultaneously engage multiple transmission channels using different physical interfaces [17, 18, 19], thus mitigating the impact of uncertainty. However, early attempts to materialize this idea failed [20], until a new version of the TCP protocol tailored for multi-interface traffic – Multipath TCP (MPTCP) was proposed [21, 22]. Conveniently, the reference implementation of MPTCP [23] addresses the general aspects of the protocol's behavior, only, which allows for potential innovations in its implementation [24], in particular the choice of congestion control (CC) algorithm.

Streaming applications, such as on-demand entertainment or video systems, often utilize adaptive data compression methods and do not require high bandwidth. Rather, they call for short latency, low error rate, and low jitter, while extensive buffering is to be avoided. However, ensuring these parameters in the case of multipath transmissions can be challenging, as the constraints on delay and its variation are difficult to impose [25]. It should also be noted that the principal objective in the design of MPTCP CC algorithms so far was to boost efficiency without compromising fairness [14], rather than cope with delay constraints [15], which are critical for streaming and tactile applications. Although some research on streaming transmission in the multipath framework has been carried out from the perspective of schedulers [16], the literature lacks works investigating the role of CC algorithms in this context. The objective of this paper is to examine whether the popular CC algorithms designed for MPTCP are suitable for streaming applications.

Frequently, the research on network protocols relies on simulations or tests conducted in a closed environment. However, the conclusions drawn from such findings may not be reliable. In this regard, this article tests the parameters of various CC algorithms for their application in a real-world setting using a public network. It follows from the conducted study that the MPTCP CC algorithm BALIA is found capable of achieving the lowest values of path delay and

This work has been performed in the framework of a project "Robustcontrol solutions for multi-channel networked flows" no. 2021/41/B/ST7/00108 financed by the National Science Centre, Poland.

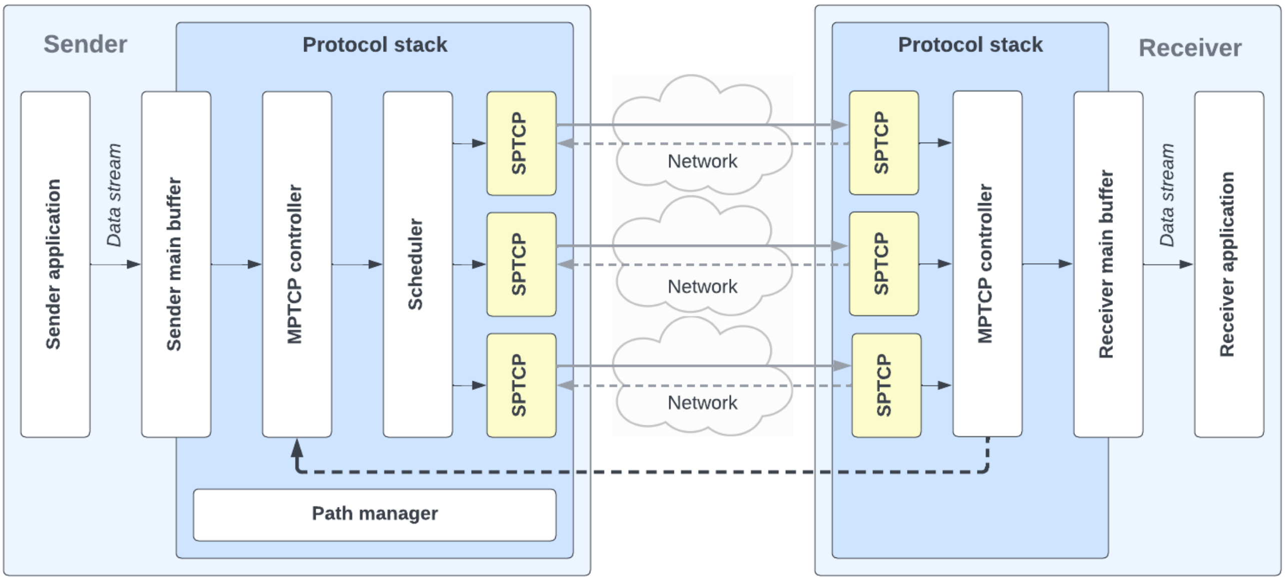


Fig. 1. Data flow in MPTCP architecture: solid lines – data, dashed lines – acknowledgements. Protocol stack and user application are sharing MPTCP buffer, whereas the SPTCP controllers have their own buffers.

head-of-line degree, as well as the highest throughput among four tested algorithms, whereas wVegas performed the worst in the context of streaming applications.

II. Multipath Transmission

The data flow in an MPTCP-capable network is illustrated in Fig. 1. When a request for data transfer is initiated, the standard TCP procedure is used to establish a connection. During the negotiation process, it is determined whether both parties support an appropriate version of the MPTCP protocol. When the multipath extension is available, an attempt is made to open as many transmission channels as possible. The path controller then decides on the number of channels and which paths to use. With multiple transmission channels, communication can be maintained even in the event of a channel failure, which would otherwise lead to a path break.

The stream of data generated by the application first passes through Master Controller, which shapes the data transfer characteristics, and then to Scheduler. Scheduler distributes the data among the active paths that have been established by Path Manager. The data in each path is then transmitted using the single path (SPTCP) controller module. The Scheduler is typically adjusted to achieve a desired strategy, such as reducing power consumption or delay [14], [15]. However, it does not directly influence the intensity of generated traffic. Instead, it responds to signals originating from the logic of the ordinary single-path TCP, which manages the data stream separately for each path. The Master Controller, SPTCP controllers, and Scheduler interact with one another in a complex manner to meet the communication objectives.

It should be noted that the TCP architecture was primarily designed to maximize throughput rather than minimize delay. However, one can shift this priority by an appropriate selection of SPTCP and MPTCP CC algorithms, which exert a significant impact on data transfer dynamics and the resulting quality of data streaming.

III. Congestion Control Algorithms

Currently, the TCP protocol serves as the foundation for the majority of Internet services. However, TCP versions operate differently in various conditions, corresponding to the method of adjusting the transfer speed. The default SPTCP CC algorithm in both Windows and Linux systems is Cubic. In addition to Cubic, the Linux kernel also includes 15 other versions of SPTCP (in alphabetical order): BBR, BIC, CDG, DCTCP, HSTCP, Hybla, Illinois, LP, NV, Reno, Scalable, Vegas, Westwood+, and YeAH. With the deployment of MPTCP, additional algorithms become available [16]: BaLIA, DMCTCP, LIA, OLIA, and wVegas, which are the primary emphasis of this work

None of these algorithms are specifically designed for streaming traffic. It presents a challenge for the transmission of data-intensive multimedia content, where factors such as low latency, low error rates, and low jitter are critical for the intended user experience [24]. There is a need to examine the performance of MPTCP CC algorithms in the context of multipath streaming traffic and to determine which algorithm is best suited for such type of content.

The following CC algorithms were selected for analysis:

On the SPTCP level: Reno and BBR. The legacy Reno algorithm [17] linearly increases the transfer speed and reduces it multiplicatively when packet loss is detected. Although it is no longer used in practical settings, it remains a common reference algorithm for TCP CC evaluation. Furthermore, both LIA and OLIA algorithms can be considered as multipath versions of Reno, since they also adjust the transmission speed in response to drops. In contrast, the BBR [18] algorithm operates by observing the speed at which the network is already delivering the traffic, along with the changes in the smoothed round-trip time (SRTT). Currently, BBR is promoted by Google and gains in popularity as an alternative to the traditional TCP CC algorithms.

TABLE 1.
IMPACT OF CONGESTION CONTROL ALGORITHMS ON STREAMING TRANSMISSION

		LIA		OLIA		BALIA		wVegas	
		Reno	**BBR**	**Reno**	**BBR**	**Reno**	**BBR**	**Reno**	**BBR**
Protocol delay	υ_{av}	139	115	122	118	109	112	266	120
[ms]	υ_{max}	726	565	556	526	491	533	860	556
HoL Degree	ζ_{av}	61	39	43	39	35	35	193	44
[ms]	ζ_{max}	83	425	423	390	353	370	726	426
SRTT [ms]	$\tau_{1,av}$	71	69	71	71	67	68	66	69
path 1	$\tau_{1,max}$	202	223	225	209	226	219	206	223
SRTT [ms]	$\tau_{2,av}$	64	63	65	65	62	64	60	63
path 2	$\tau_{2,max}$	221	229	242	217	229	219	218	218
Mean drop rate	d_1	9.6	10.2	10.3	8.2	7.9	7.8	9.6	12.3
[seg/s]	d_2	8.0	12.2	12.7	13.0	10.5	11.4	8.1	9.2
Throughput	*av*	4.09	4.57	4.47	4.72	5.49	5.47	3.69	3.75
[Mbps]	*max*	10.74	12.03	11.01	11.69	13.66	13.52	8.76	9.59

On the MPTCP level: LIA [19], OLIA [20], BALIA [21], and wVegas [22]. Their design premise is protocol fairness. LIA increases the transmission speed faster than the slowest path, whereas OLIA analyzes the underlying SPTCP control variables and responds to channel disparities and fluctuations. Therefore, OLIA is better adapted for heterogeneous environments. BALIA is a hybrid algorithm that combines the strengths of LIA and OLIA, which allows it to perform well in both homogeneous and heterogeneous environments. The main advantage of BALIA is its ability to dynamically adjust the aggressiveness of CC based on the network conditions. Finally, wVegas is a window-based algorithm that modifies the congestion window size based on the estimated round-trip time and packet loss rate. It has been designed to perform well in high-speed and long-distance connectivity. However, it is less effective in congested or lossy networks.

IV. QUALITY MEASURES

Transmission parameters are affected by numerous factors, such as congestion or buffering, whose impact cannot be predicted *a priori*. Therefore, to fairly evaluate the performance of different CC algorithms, the following quality metrics have been used.

A. Path Delay

The path delay refers to the time it takes a packet to traverse the path from the sender to the receiver. The length of this delay depends on the distance between the sender and receiver, the number of routers and switches along the path, and the degree of congestion on the path.

In the model developed in this paper, the total delay on path *i*, denoted by T_i, comprises the SRTT of this path τ_i and the waiting time for processing the data stream θ_i:

$$T_i = \tau_i + \theta_i\,. \tag{1}$$

The waiting time θ_i is influenced by the scheduler algorithm. The value of τ_i can be reduced by limiting the buffer bloat via a prudent selection of a CC algorithm, as studied in this work.

As the transmission progresses, the delays on individual paths change, and another path may become the "slowest". The path delay is a metric for assessing the quality of service, particularly for streaming traffic. High path delay can result in prolonged buffering and poor user experience.

B. Protocol Delay

The protocol delay is defined as the time that a given piece of data, e.g., a packet, waits in the buffer for stream reassembly. It is measured from the instant when Master Controller receives the user data from the transmit buffer and ends when the corresponding data acknowledgment is received. The protocol delay is equivalent to the delay on the slowest path

$$T_{over}(k) = \max_i T_i(k)\,. \tag{2}$$

The average protocol delay is calculated as

$$\upsilon_{av}^r = \frac{1}{K}\sum_{k=1}^{K} T_{over}(k) \tag{3}$$

and maximum protocol delay as

$$\upsilon_{\max}^r = \max_{k\in[1,K]} T_{over}(k)\,. \tag{4}$$

The average protocol delay for all the experiment runs is determined as

$$\upsilon_{av} = \frac{1}{R}\sum_{r=1}^{R} \upsilon_{av}^r \tag{5}$$

and the average maximum protocol delay as

$$\upsilon_{\max} = \frac{1}{R}\sum_{r=1}^{R} \upsilon_{\max}^r\,. \tag{6}$$

K represents the number of samples collected in a single experiment run indexed by *r,* while *R* refers to the number of experiment runs. Streaming performance improves with lowering the values of υ_{av} and υ_{max}.

C. HoL Degree

In MPTCP, the Head-of-Line (HoL) blocking degree refers to the number of packets queuing up and waiting to be transmitted in a certain path before the head-of-line packet being delivered.

Real-time applications such as video streaming, online gaming, and video conferencing are highly sensitive to latency and packet loss, as these factors can significantly degrade the user experience. In particular, the HoL blocking [23] can have a substantial impact on the quality of service provided. From the application perspective, the actual visible value is T_{over}, i.e., the delay on the slowest path. Based on that value, the waiting time is defined as

$$T_{over}(k) - \max_{i\in[1,m]} \tau_i(k) \,. \qquad (7)$$

Then, the average waiting time

$$\zeta_{av}^{r} = \frac{1}{K}\sum_{k=1}^{K}\left(\max\left(T_{over}(k) - \max_{i\in[1,m]} \tau_i(k), 0\right)\right) \qquad (8)$$

and maximum waiting time for each experiment

$$\zeta_{\max}^{r} = \max_{k\in[1,K]}\left(T_{over}(k) - \max_{i\in[1,m]} \tau_i(k)\right). \qquad (9)$$

The average waiting time across all the experiments

$$\zeta_{av} = \frac{1}{R}\sum_{r=1}^{R}\zeta_{av}^{r} \qquad (10)$$

and the average maximum protocol delay across all the experiments

$$\zeta_{\max} = \frac{1}{R}\sum_{r=1}^{R}\max(\zeta_{\max}^{r}, 0) \,. \qquad (11)$$

D. Mean Drop Rate

The mean drop rate is defined as the proportion of packets lost in the course of transmission. Packet drops can happen for a number of causes, such as route failure, network congestion, or packet reordering.

V. Experimental Setup

The test setup depicted in Fig. 2 was employed to assess the effect of CC algorithm interoperability on the quality of streaming content delivery within the MPTCP framework. The created test setup represents a typical data transmission scenario in which a client device connects to a high-end server device to retrieve the content. The server, accessed through a public IP address, is located in a remote data center. A specialized program is utilized to generate the streaming content. Both the client and server devices run under the Linux operating system version 4.19, which had been patched to support MPTCP version 0.95. The client device has two communication interfaces – one connected to an LTE router through an Ethernet cable and the other linked through Wi-Fi 802.11bgn to the same LTE router. Two different LTE networks from different operators were used, with good signal quality ensured. The packets transmitted through one interface arrive at their destination after 10 hops, while packets transmitted through the other arrive after 12 hops. A single scenario lasts 10 seconds and each is repeated 30 times.

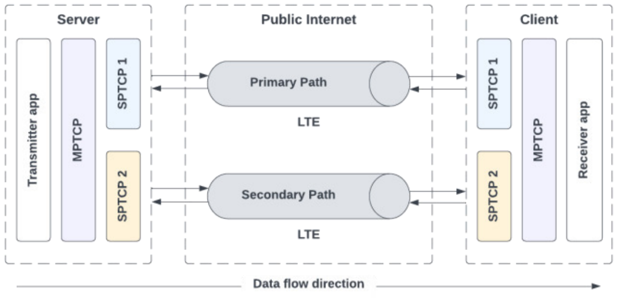


Fig. 2. Experimental setup

VI. Tests and Results

Two CC algorithms: Reno and BBR, were examined for SPTCP, and four algorithms: LIA, OLIA, BALIA, and wVegas, for MPTCP. The tests were performed for each combination, resulting in eight different scenarios. Table 1 summarizes the obtained measurements, whereas Fig. 3 depicts graphically a chosen test run.

The gathered data show that using BALIA at the MPTCP level leads to the best overall performance. This was particularly visible in the case of path delays, where the algorithm achieved the lowest average and maximum delay values. The graphs reveal that the most significant differences occur at the beginning of transmission, where the path delay is nearly three times longer, and the protocol delay and HoL degree are almost five times higher than those observed after stabilization which occurred approximately three seconds after that period. Although peaks resulting from the fluctuation of network parameters were observed, they are negligible after averaging all test runs.

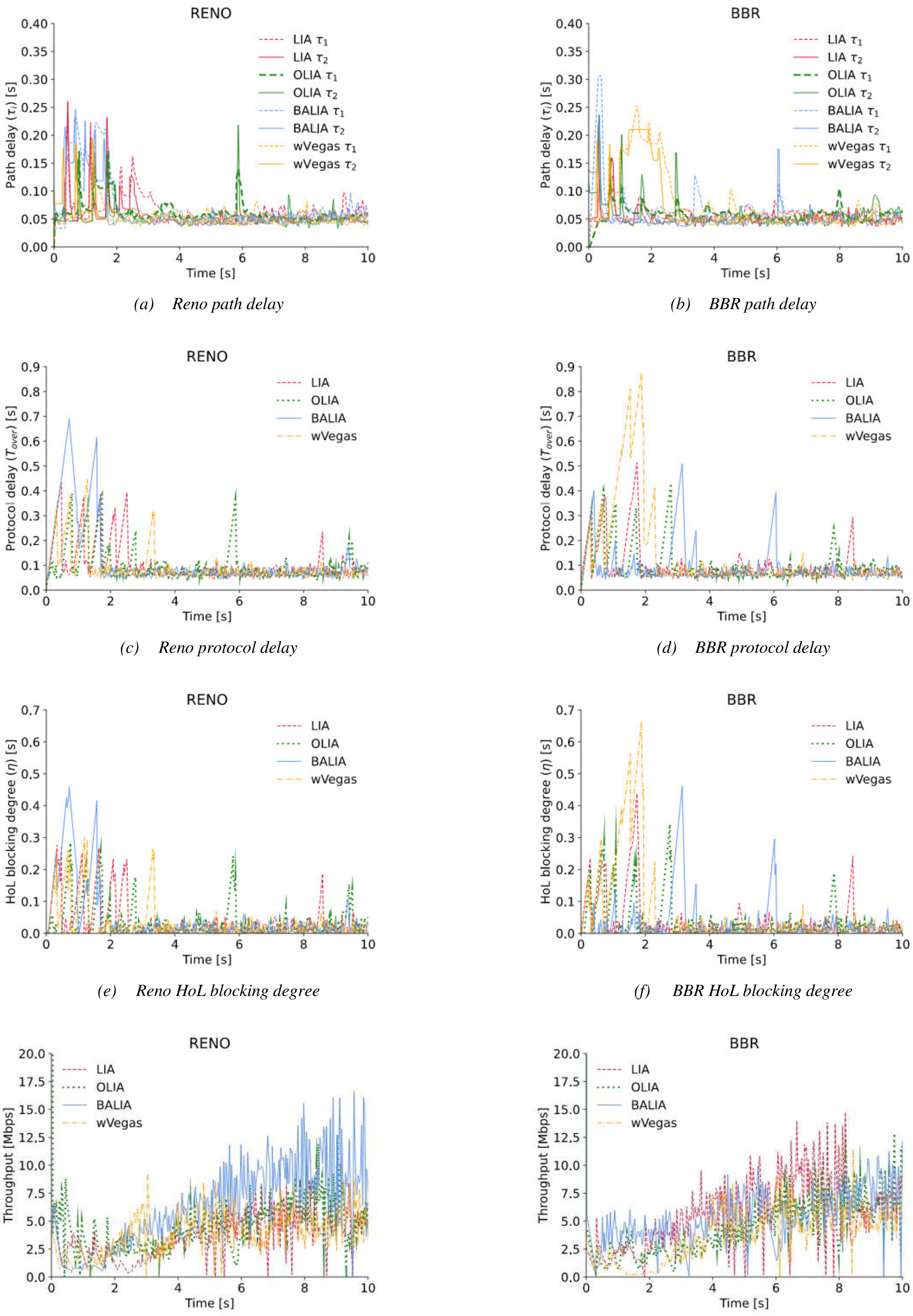
Similar observations apply to the protocol delays, where BALIA also exhibits the lowest average and maximum delays. Moreover, BALIA happened to achieve the lowest HoL degree, implying the smallest proclivity to the multipath queue build-up.

It is worth noting that wVegas underperformed in all the scenarios. Although the path delay was consistent between the different scenarios, the protocol delay was slightly worse for wVegas with BBR, and over two times worse in the scenario with RENO. The HoL degree was over five times larger for wVegas with RENO, and almost one and a half times larger for wVegas with BBR. The throughput was worse when Vegas was used, scoring 33% lower than the other scenarios. Consequently, the wVegas protocol is not recommended for use in MPTCP streaming transmissions.

Finally, the throughput data show that the BALIA algorithm was more efficient in utilizing the available resources, resulting in a maximum value that was about 15% (2 Mbps) higher and an average that was about 20% (1 Mbps) higher than the other scenarios.

VII. Conclusions

The paper's focus was to investigate how the main MPTCP CC algorithms handle streaming transmission over heterogeneous public networks. The use of multiple communication paths can be an answer for achieving high data speed, low latency, and stable transmission parameters, which are essential for quality streaming. Indeed, it was also



(a) Reno path delay

(b) BBR path delay

(c) Reno protocol delay

(d) BBR protocol delay

(e) Reno HoL blocking degree

(f) BBR HoL blocking degree

(g) Reno throughput

(h) BBR throughput

Fig. 3 Measured transmission properties – left column Reno, right column BBR acting on the paths

found that the choice of CC protocol is a decisive factor affecting the transmission quality.

The results show that data transfer over multiple paths does increase latency, but this is acceptable and should not negatively affect streaming applications. The BALIA algorithm was found to be the most effective CC algorithm on the MPTCP level and BBR at the path level. Using these two algorithms together provides the lowest latency, lowest error, and highest throughput, making them the best choice out of the off-the-shelf algorithms for efficient streaming over multiple communication channels. In turn, mVegas achieved the lowest transmission parameters and its use in multimedia transmission is ill-advised.

Acknowledgement

This work has been completed while the first author was the Doctoral Candidate in the Interdisciplinary Doctoral School at the Lodz University of Technology, Poland.

References

[1] Cisco, "Cisco Annual Internet Report (2018-2023) White Paper, " 2020. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

[2] C. Raiciu, M. J. Handley, and D. Wischik, "Coupled Congestion Control for Multipath Transport Protocols, " Request for Comments, no. 6356, RFC Editor, Oct. 2011. [Online]. Available: https://www.rfc-editor.org/info/rfc6356, doi: 10.17487/RFC6356.

[3] M. Morawski and P. Ignaciuk, "Choosing a Proper Control Strategy for Multipath Transmission in Industry 4.0 Applications," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 6, pp. 3609-3619, 2022, doi: 10.1109/TII.2021.3105499.

[4] C. Paasch, S. Ferlin, O. Alay, and O. Bonaventure, "Experimental Evaluation of Multipath TCP Schedulers," *Proceedings of the 2014 ACM SIGCOMM Workshop on Capacity Sharing Workshop*, CSWS '14, Chicago, Illinois, USA, 2014, pp. 27-32, doi: 10.1145/2630088.2631977.

[5] M. Barreiros and P. Lundqvist, "QoS-Enabled Networks: Tools and Foundations," Wiley & Sons, 2016.

[6] M. Morawski and P. Ignaciuk, "Network nodes play a game – a routing alternative in multihop ad-hoc environments," *Comp. Netw.*, vol. 122, pp. 96-104, 2017, doi: 10.1016/j.comnet.2017.04.031.

[7] J. Qadir, A. Ali, K. A. Yau, A. Sathiaseelan, and J. Crowcroft, "Exploiting the power of multiplicity: a holistic survey of network-layer multipath," *IEEE Commun. Surv. Tut.*, vol. 17, no. 4, pp. 2176-2213, "4Q", 2015, doi: 10.1109/COMST.2015.2453941.

[8] M. Li et al., "Multipath Transmission for the Internet: A Survey, " *IEEE Commun. Surv. Tut.*, vol. 18, no. 4, pp. 2887-2925, Q4, 2016 , doi: 10.1109/COMST.2016.2586112.

[9] S. Barré, C. Paasch, and O. Bonaventure, "MultiPath TCP: From theory to practice, " *Tech. Report*, Université Catholique de Louvain, 2011, doi: 10.1007/978-3-642-20757-0_35.

[10] A. Ford, C. Raiciu, M. Handley and O. Bonaventure, "TCP extensions for multipath operation with multiple addresses," RFC 6824, 2013, doi: 10.17487/RFC6824.

[11] S. Barré and C. Paasch, "MultiPath TCP – Linux kernel implementation," http://www.multipath-tcp.org.

[12] M. Morawski and P.Ignaciuk, "A green multipath TCP framework for Industrial Internet of Things applications," *Comp. Netw.*, vol. 187, 107831, 2021, doi: 10.1016/j.comnet.2021.107831.

[13] K. Yedugundla, S. Ferlin, T. Dreibholz, Ö. Alay, N. Kuhn, P. Hurtig, and A. Brunstrom, "Is multipath transport suitable for latency sensitive traffic?, " *Comp. Netw.*, vol. 105, pp. 1-21, 2016, doi: 10.1016/j.comnet.2016.05.008.

[14] C. Paasch, S., Ferlin, O., Alay and O., Bonaventure, "Experimental evaluation of multipath TCP schedulers," *Proc. on ACM SIGCOMM CSWS*, pp. 27–32, 2014, Chicago, USA, doi: 10.1145/2630088.2631977.

[15] M. Morawski and. P. Ignaciuk, "Energy-efficient scheduler for MPTCP data transfer with independent and coupled channels," *Comp. Commun.*, vol. 132, pp. 56-64, 2018, doi: 10.1016/j.comcom.2018.09.008.

[16] C. Xu, J. Zhao, G. Muntean, "Congestion control design for multipath transport protocols: A survey," *IEEE Commun. Surv. Tut.*, vol. 18, no. 4, pp. 2948–2969, 2016, , doi: 10.1109/COMST.2016.2558818.

[17] T. Henderson, S. Floyd, A. Gurtov, and Y. Nishida, "The NewReno modification to TCP's fast recovery algorithm," RFC 6582, Apr. 2012, doi: 10.17487/RFC3782.

[18] N. Cardwell, Y. Cheng, C. S. Gunn, S. H. Yeganeh and V. Jacobson, "BBR: Congestion-Based Congestion Control: Measuring Bottleneck Bandwidth and Round-Trip Propagation Time," Queue, vol. 14, no. 5, pp. 20-53, Sep.-Oct. 2016, doi: 10.1145/3012426.3022184.

[19] C. Raiciu, M. J. Handley and D. Wischik, "Coupled Congestion Control for Multipath Transport Protocols," RFC 6356, RFC Editor, Oct. 2011, pp. 1-12, doi: 10.17487/RFC6356.

[20] N. Gast, R. Khalili, J.-Y. Le Boudec and M. Popovic, "Opportunistic Linked-Increases Congestion Control Algorithm for MPTCP," *Proceedings of the 13th International IFIP Networking Conference (Networking)*, Trondheim, Norway, 2014, pp. 1-9.

[21] Q. Peng, A. Walid, J. Hwang, and S. H. Low, "Multipath TCP: Analysis, Design, and Implementation, " IEEE/ACM Transactions on Networking, vol. 24, no. 1, pp. 596-609, 2016, doi: 10.1109/TNET.2014.2379698.

[22] Y. Cao, M. Xu, and X. Fu, "Delay-based congestion control for multipath TCP, ", 2012 20th *IEEE International Conference on Network Protocols (ICNP)*, pp. 1-10, 2012, doi: 10.1109/ICNP.2012.6459943.

[23] R. Khalili, N. Gast, M. Popovic, and J.-Y. Le Boudec, "MPTCP is not Pareto-optimal: performance issues and a possible solution," *IEEE/ACM Trans. Netw.*, vol. 21, no 5, pp. 1651-1665, 2013, doi: 10.1109/TNET.2013.2274462.

[24] S. Grzyb and P. Orłowski, "Congestion feedback control for computer networks with bandwidth estimation, ", *2015 20th International Conference on Methods and Models in Automation and Robotics (MMAR)*, pp. 1151-1156, 2015, doi: 10.1109/MMAR.2015.7284041.

Author Index

Abrahams, Nathier 1
Alqahtani, Yahya 9
Antkiewicz, Michał 21

Balon, Tomasz 29
Barski, Aleksy 289
Belle, Jean-Paul Van 1
Beloff, Natalia 9
Berg, Frederik 345
Bielecki, Wlodzimierz 37
Bindu, Farsha 223
Blaszynski, Piotr 37
Bluemke, Ilona 103
Bochenek, Adam 43
Borst, Janos 51
Buglowski, Mateusz 345
Burghardt, Manuel 51

Chaber, Wojciech 273
Chadam, Szymon 61
Cyganek, Bogusław 29
Czerniak, Jacek M. 91

Daszczuk, Wiktor 65
Ding, Shitao 141
Dobrosielski, Wojciech 91
Dörpinghaus, Jens 81
Dutta, Soma 273
Dzido, Tomasz 75

Ewald, Dawid 91

Flotyński, Jakub 265
Fomin, Oleksandr 97
Fontes, Raphael 281

García-Vico, Ángel M. 131
Ghosh, Anonnya 223
Gitlin, Erik Sebastian Skibinsky 131
Grabski, Adam 103
Groningen, Tomas van 111

Hassmann, Tobias 121
Hernandez, Manuel Luis González 131
Huang, Qiang 141

Ignaciuk, Przemysław 353

Jabłoński, Piotr 147
Jałowiec, Jakub 65
Jędrzejowicz, Piotr 157, 237
Jełowicki, Franciszek 153
Júnior, Methanias Colaço 281

Kasprowicz, Paulina 273
Knapik, Mateusz 29
Kniveton, Nicholas 195
Kowalewski, Stefan 345
Krishna, Aneesh 181
Krüger, Kai 165
Krykun, Valentyn 97
Kubica, Bartłomiej Jacek 173
Kühn, Jan 345
Kulczycki, Emanuel 323
Kumar, Lov 181
Kuzmych, Lyudmyla 189
Kuzmych, Stepan 189

Lima, Admilson de Ribamar 251
Lins, Gabriel 281
Lipka, Richard 215
Liu, Yongbo 141
Lozada, Randy 131
Łuczak, Łukasz 353

Maik, Mikołaj 265
Marinello, Francesco 141
Mażbic-Kulma, Barbara 289
Mishra, Pavan Kumar 229
Misra, Sanjay 181
Morawski, Michał 353
Murthy, Lalita Bhanu 181
Myszkowski, Pawel 21

Nakayama, Minoru 201
Niekler, Andreas 51
Nowak, Wioletta 201
Nowotniak, Robert 323

Okulska, Inez 303
Orlov, Andrii 97
Ottersbach, Florian 345
Overbeek, Sietse 111
Owsiński, Jan 289

Paiva, Jailton 281
Palkowski, Marek 37
Petcusin, Felix 209
Petrov, Petar 329
Poliwoda, Maciej 37
Polozhaenko, Sergii 97
Potuzak, Tomas 215
Prokofiev, Andrii 97

Rab, Raqeebir 223
Rajak, Prince 229
Ratajczak-Ropel, Ewa 237
Rudnik, Katarzyna 245
Ruiz, Jorge 131
Ruszczak, Bogdan 245

Sanjida, Sheikh 223
Santos, Josimar dos 251
Sedano, Javier 131
Sęp, Krzysztof 289
Shaghaghi, Navid 195
Silva, Caldeira 281
Silva, Gilton José Ferreira da 251
Silva, Rodrigo 281
Singh, Vikram 181
Skubalska-Rafajłowicz, Ewa 259
Śliwicki, Michał 265
Sobociński, Paweł 265
Sommer, Martin W. 81
Sosnowski, Łukasz 273
Souza, Luan 281
Spahiu, Cosmin Stoica 209
Stańczak, Jarosław 289
Stanescu, Liana 209
Steinbrink, Nicholas 295
Stetsenko, Daria 303
Stollenwerk, André 345
Stypułkowska, Justyna 313
Susik, Robert 323
Szymusik, Iwona 273

Tabassoum, Nafiza 223
Tao, Youzhi 141
Topa, Paweł 61
Traneva, Velichka 329, 337
Tranev, Stoyan 329, 337
Tsibolane, Pitso 1

Valetim, Ricardo 281
Villar, José Ramón 131
Voropai, Halyna 189

Wahab, Tamima Binte 223
Walczak, Krzysztof 265
Wehrheim, Lino 51
Weil, Vera 81
Westner, Markus 121
White, Martin 9
Wiartalla, Marc 345
Wierzbowska, Izabela 157

Zarowska, Anna 201
Zarzycki, Hubert 91