Proceedings of the Sixth International Conference on Research in Intelligent and Computing in Engineering

June 3–4, 2021
Thu Dau Mot University, Vietnam

Vijender Kumar Solanki, Nguyen Ho Quang (eds.)

PTI

Annals of Computer Science and Information Systems, Volume 27

Proceedings of the Sixth International Conference on Research in Intelligent and Computing in Engineering

Vijender Kumar Solanki, Nguyen Ho Quang (eds.)

Annals of Computer Science and Information Systems, Volume 27

Proceedings of the Sixth International Conference on Research in Intelligent and Computing in Engineering

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

Cover photo: Jantar
Aleksander Denisiuk,
    Elbląg, Poland

Also in this series:

DEAR reader, it gives us immense pleasure to share the glimpses for Sixth International Conference on Research in Intelligent and Computing in Engineering. The RICe 2021 is organized by Thu Dau Mot University (TDMU), Vietnam; Jointly Co-organised by Le Quy Don Technical University, Hanoi VietNam during June 3-4, 2021. It is very tough to see that in the consecutive second year, when we have to work out in hybrid mode due to Covid19 pandemic.

This Conference provides valuable opportunities for researchers, industry practitioners as well as interdisciplinary teams to share our experiences and exchange new ideas in emerging technologies and innovative applications , emphasizing certain emerging areas like Smart manufacturing, smart cities, Smart healthcare, and digital innovations and transformation. We strongly believe that RICE2021 will not only help to reinforce our collaborations but also to create new ones—looking for integrated and innovative solutions.

We are truly thankful to the Polish Information Processing Society (PTI) , Poland for approving the proceedings of the Sixth International Conference on Research in Intelligent and Computing in Engineering (RICE 2021). It shall be published in the Annals of Computer Science and Information Systems series by the PIPS. The books of this series have been submitted to DBLP, CrossRef, BazEkon, Open Access Library. At this stage, the effort,whole hearted support and suggestions given by Editor-in-Chief Prof. Marcin and Prof. Maria Ganzha is highly applaudable and commendable.

We would like to announce that the 6th edition of RICE attracts a good virtual footfall from various country researchers. It is an honor for us to receive valuable insights and visions regarding areas of interest to the conference from eight keynote speakers from seven countries. The details are given below:

- *Assoc. Prof. Supavadee Aramvith, Department of Electrical Engineering Chulalongkorn University, Thailand*
- *Prof. Tien Tuan Dao, Centrale Lille Institut, Laboratory of Multiphysical and Multiscale Mechanics, LaMcube (UMR CNRS 9013)*
- *Assoc. Prof. Manuel Cardona, SMIEEE, Robots and Intelligence Machines Research Group, Universidad Don Bosco, El Salvador*
- *Prof. Anne Habraken, Directrice de recherches FNRS, University of Liège, Belgium*
- *Prof. Pai-Chen Lin, Department of Mechanical Engineering, National Chung Cheng University, Taiwan*
- *Assoc. Prof. Le Chi Hieu, University of Greenwich, UK*
- *Dr. Mien Van, School of Electronics, Electrical Engineering and Computer Science, Queens University, Belfast, UK*
- *Prof. Prashant Kumar Pattnaik, KIIT University, Bhubhesnhar, Odisha (IoE), India;*

As per the guidelines issued by the Ministry of Vietnam, and keeping in view of health concerns, flexible options were shared with participants to join as Online/Offline Mode.

"To become a better you, remember to be grateful to people who have contributed to making you who you are today." We are sincerely thankful to our Sponsor Vingroup Innovation Foundation (VinIF), for their sponsorship and contributions to promote further and more sustainable development of the science and technology field.

We are thankful to the leaders of TDMU, LQDTU for their sincere co- operation and whole hearted support to make RICE2021 a successful event. There are many names who play a vital role, but still we think that it is the right place to mention their support to make this congress a huge success.

Assoc. Prof. Nguyen Van Hiep, The Chairman of Thu Dau Mot University, Vietnam; RICE2021 General Chair.
Prof. Dr.Sc. Nguyen Cong Dinh, The President of Le Quy Don Technical University (LQDTU), Vietnam; RICE2021 General Chair;
Assoc. Prof. Le Chi Hieu, University of Greenwich, UK; RICE2021 Organizing Co-Chair;
Assoc. Prof. Chu Anh My, Director of Institute of Simulation Technology, Le Quy Don Technical University (LQDTU), Vietnam; RICE2021 Organizing Co-Chair;
Dr. Ngo Hong Diep; Vice President of Thu Dau Mot University; RICE2021 Organizing Co-Chair;
Dr. Le Anh Ngoc; Electric Power University , Vietnam; RICE2021 Special session chair
Dr. Nguyen Hong Quang, Thai Nguyen University of Technology, RICE2021 Special session chair
Dr. Tran Van Xuan, Thu Dau Mot University, RICE2021 Special session chair

It gives us pain to write that we have lost our beloved friend Prof. Hoang Trong Quyen, Vice-Rector, Thu Dau Mot University, due to ill health. We met during RICE 2019.For RICE, his whole hearted support and inspiration is beyond any word. May his departed soul rest in peace.

Last but not least, we would like to take this opportunity to extend our deepest gratitude to the Advisory Board, Organizing Committee, International Scientific Committee, institutions, companies, and volunteers, who have directly or indirectly supported the success of this conference.

Looking forward to welcoming you to RICE2022 at Hung Yen University of Technology and Education, Hung Yen province, Vietnam on 11-12 Nov 2022.

Proceeding's Editors – RICE2021:

Vijender Kumar Solanki, Ph.D, CMR Institute of Technology, Hyderabad, India.

Nguyen Ho Quang, Ph.D, Thu Dou Mot University, Binh Duong, Vietnam.

Annals of Computer Science and Information Systems, Volume 27

Sixth International Conference on Research in Intelligent and Computing

June 3–4, 2021. Thu Dau Mot, Vietnam

TABLE OF CONTENTS

An analysis of the effect between the heat index and Long Short-Term Memory model to electricity load forecasting

Nguyen Thi Ngoc Anh
Hanoi Uni. of Sci. and Tech
Hanoi, Vietnam
anh.nguyenthingoc@hust.edu.vn

Vu Viet Hoang
Hanoi Uni. of Sci. and Tech
Hanoi, Vietnam
hoang.vv183926@sis.hust.edu.vn

Do Thi Thanh Chau
Hanoi Uni. of Sci. and Tech
Hanoi, Vietnam
chau.dtt185434@sis.hust.edu.vn

Bui Duy Linh
National Load Dispatch Centre
Hanoi, Vietnam
linhbd@nldc.evn.vn

Nguyen Duc Huy
Hanoi Uni. of Sci. and Tech
Hanoi, Vietnam
huy.nguyenduc1@hust.edu.vn

Abstract—Accurate electricity load forecasting is essential for operating electrical systems. Most of the studies on electricity load forecasting are based on electricity load data or weather data, which is air temperature, but there are not consider the heat index. This paper proposes a short-term electricity load forecasting model using Long-Short Term Memory (LSTM) based on electricity load history and heat index data. In addition, the proposed model is applied to the data of IEVN NLDC (National Load Dispatching Center) in forecasting electricity load before 48 hours. This model is used to predict the electricity load of the Vietnamese nation and the power corporations of Vietnam. For a fair comparison, the LSTM network has fixed parameters, then compared the results when using temperature and the heat index. According to experimental results based on the Mean absolute percentage errors (MAPE) assessment, the proposed model has better accuracy than the model based on electricity load history and temperature.

Index Terms—Heat index, Long short term memory, Electricity load, Forecasting, Time series

I. Introduction

In Vietnam, the real-time load of the nation, regions separately, or power corporations are managed by the way that called top-down. This way means generating the capacity of the area's interior and the forwarding capacity of the line linking domains and regions. The electricity load forecasting is mainly based on samples from the past. In addition, affecting factors to load include days in the week, days off, holidays, or weather elements such as temperature, unusual weather phenomena like typhoons. Among all of those factors related to weather, the temperature is being considered the most important clue to determine the fluctuation of load. The other factors as humidity, amount of rain are also noticed. However, if these factors alone are considered and treated as an independent input variable, the efficiency is not high. In conclusion, analyzing and evaluating the effects of these factors on electricity load is a difficult job and have yet had many solutions and research that can solve this effectively,

especially in the specific condition of a tropical nation that has a high proportion of daily life electricity(30%-40%) like Vietnam.

Long Short-Term Memory Network (LSTM) is an improved version of the Recurrent Neural Network (RNN) model. Because Fog formation is closely related to the meteorological elements, [1] used the LSTM model to predict the short-term fog based on meteorology. The model consists of an LSTM network and a fully connected layer with time-series input data. Although there are still many challenges for the fog forecasting project, the proposed LSTM model gives better results than the current models such as AdaBoost, CNN, KNN. Because of its effectiveness in time dependency learning, [2] presented a model based on LSTM for urban sound classification. [3] Stacked bidirectional and unidirectional LSTM can be used for forecasting network-wide traffic state, while the LSTM and BiLSTM can fill missing values to yield a better prediction. This paper also shows that the 2-layer model is the most effective model and the model with 2 layers. BiLSTM is the best among LSTM and BiLSTM composite models. The proposed model is only good when the rate of missing data is small. When the rate of missing values is large, the performance of the proposed model decreases while the error of BGCP is nearly unchanged. [4] Lifan Mei et al. proposed real-time mobile bandwidth forecasting using LSTM and Bayesian fusion. In addition, to LSTM's superior demonstration work, they evaluated LSTM's sensitivity to different scenarios. A remarkable study of this paper is the construction of Bayesian fusion, which combines predictions from multiple models to adapt to the change of the moving target. [5] Some model does not give a desirable result because the information is not collected comprehensively, and the power load is not consistent. Yaquina et al. have developed probability prediction models for short-term electrical load based on Long-short term memory. The results are obtained by

regression and estimating confidence interval. Especially the independent variable is the predicted value while the dependent variable is the actual value. In this paper [6], the LSTM model was used to select features and improve classification accuracy. [7] To forecast the confidence interval, the pinball loss function was integrated with the LSTM network. The result will be the quantiles, not the random variable's expectation. That is critical for unsteady data. Indeed, the performance is pretty strong, especially when applied to data from small and medium-sized firms (SMEs) in Ireland. However, the model is unsuccessful with variable data, particularly with resident load data that includes domestic chores and working hours. Furthermore, the model does not employ optimum approaches to choose the LSTM's parameters. In general, LSTM is a popular and effective model. Especially it is important for short-term electrical load forecasting.

An important indicator used in research on the health effects of heat is the "Steadman's heat index" [8] [9]. Steadman's heat index converts current weather conditions to what humans feel like the dew point temperature of 57.2 ° F/ 14.0 ° C [8] [10]. As a result, Steadman converted the combination of air humidity and temperature and several other factors [8] [9] into an index with the same unit of measurement as the air temperature. Based on Steadman's table of original apparent temperature [8], "Heat Index"(HI) is an approximate, simple version and is based solely on air temperature and humidity. [11] Compared 21 algorithms for generating HI, NWS online 2011 algorithm was supposed to produce HI that closely resembled the apparent temperature in Steadman's table and did not show much difference from air temperature when the temperature was lower than 20 degrees C. In the paper, [12] NWS HI is used as the heat index to measure people's feelings during the day, from which to study the changes in heat in Da Nang, Vietnam in the period 2020-2049 showing the effect of heat on the health and productivity of workers in Da Nang.

In the paper [13], the authors used the forecasting model that is developed using weighted SVR models with nu-SVR and epsilon SVR. The model showed a good result: the mean absolute percentage error (MAPE) for daily energy consumption data is 5.843% and that for half-hourly energy consumption is 3.767%, respectively. Another models proposed by Yaoyao He et al. [14] solved a problem about short-term power load probability density forecasting. This model uses kernel-based support vector quantile regression (KSVQR) and Copula theory. The simulation results show that the proposed method has great potential for power load forecasting by selecting the appropriate kernel function for the KSVQR model. The best MAPE value (in all tests) is 0.81%, and the best MAE value is 47.20. In the article, [15], the short-term electrical load forecasting is solved by combining the extreme learning machine (ELM) with a new switching delayed PSO algorithm. The proposed model has better performance than other state-of-the-art ELMs. The results proved that the proposed learning algorithm gets better electrical load forecasting results in comparison with the RBFNN algorithm. The MAPE of the proposed model is 2.182%, and the MAPE

of the RBFNN method is 2.902%. Therefore, the MAPE of the proposed model achieves better performance. Similarly, the proposed model's MAE also gets better results than the RBFNN. In the paper [16], Song Li et al. propose a new synthetic method for short-term electrical load forecasting based on wavelet transform, extreme learning machine (ELM) and, partial least squares regression (PLSR). The wavelet transform is used to decompose the time series and used ELM for each sub-component from the wavelet decomposition. The results demonstrated that the proposed model could ameliorate load forecasting performance. The MAPE of the proposed model (for 1 hour ahead forecasting) is 0.2736%, which is the smallest in all results of all models used in the paper.

This paper aims to improve the accuracy of electricity load forecasting by the heat index. Concretely, the heat index is used for a feature of deep learning. Then, deep learning long short-term memory (LSTM) is used for short-term electricity load forecasting. The effective heat index for the proposed model gets outperformed.

The structure of the paper is as follows: Section II explains the methodology of feature selection using heat index, LSTM using heat index is proposed. The experiments and results are presented in Section III. Next, Section IV provides the conclusion and discussions. The final section of this paper, section V is future work.

II. METHODOLOGY

A. Formula

Let $\mathbb{T} = \{0, 1, \cdots, n, \cdots\}$ be a set of discrete-time. The electricity load time series are denoted by $y_t, t \in \mathbb{T}$.

The impact factors of electricity load are presented by $\mathbf{X}_t = (X_t^1, \cdots, X_t^n)$. In real life, the principle of the impact factor is complicated. To find a precise model to capture all factors and the perfect forecast is investigated, the electricity load y_t is the function of impact factors that influence the electricity load

$$y_t = f(\mathbf{X}_t) + \varepsilon_t. \tag{1}$$

where ε_t is uncertain part. The load forecasting model $f(.)$ is the target of this problem.

Therefore, the purpose of the forecasting model is to find the best $f(.)$ to satisfy the mean absolute percentage errors are minimum.

B. LSTM

In 1997, LSTM was an improvement of the RNN that was introduced by Hochreiterand Schmidhuber [17]. Then, Gers et al.proposed the addition of the forget gate in 2000 [18].

The forget gate f_t decides which information to get or delete from the cell state. f_t is calculated by:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{2}$$

The input gates i_t is calculated by:

$$i_t = \sigma(Wi \cdot [h_{t-1}, x_t] + b_i) \tag{3}$$
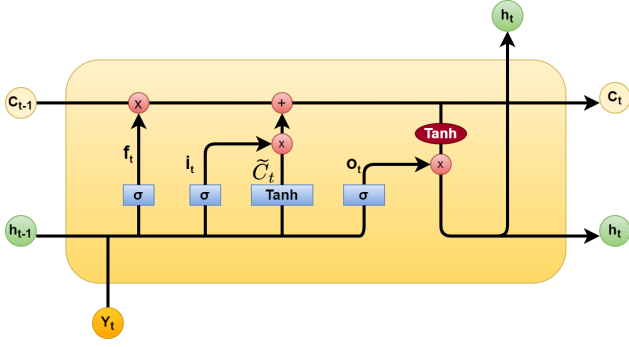
Fig. 1. LSTM network structure.

The current cell state C_t is created by all the gates. The forget gate f_t will control the long-term information, and the input gate i_t will control the short-term information:

$$\widetilde{C}_t = tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{4}$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{5}$$

The output gate is responsible for selecting and outputting the necessary information:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{6}$$

And the last step, the final output of LSTM h_t is calculated by:

$$h_t = o_t * tanh(C_t) \tag{7}$$

Where $W_f, W_i, W_o, b_f, b_i, b_o$ denote the weight matrix and bias corresponding to forget gate, input gate and output gate.

The operator $*$ stands for the element-wise multiplication.

C. Heat index

NWS online [11] creates an algorithm to calculate HI for detecting dangerous heat warnings. HI is calculated as a function of relative temperature and humidity. Figure 2 shows the heat index obtained by the HI NWS online. However, in this paper, we will make some changes in the formula to match the electricity usage forecast of Vietnamese people. The proposed algorithm shown in Figure 3 is described as follows:

Under conditions where the temperature is less than C degrees Fahrenheit, the temperature is considered exactly as HI. The parameter C is considered the cold threshold. C is changed because it affects whether people use heating devices, thanks to affecting the amount of electricity used.

HI for conditions temperature is between C and 77 ° F:

$$HI = -10.3 + 1.1 \times T + 0.04 \times RH \tag{8}$$

General HI for temperatures above 77 °F:

$$\begin{aligned}
HI = & -42.379 + 2.04901523 \times T + 10.14333127 \times RH \\
& - 0.22475541 \times T \times RH - 6.83783 \times 10^{-3} \times T^2 \\
& - 5.481717 \times 10^{-2} RH^2 + 1.22874 \times 10^{-3} T^2 \times RH \\
& + 8.5282 \times 10^{-4} T \times RH^2 - 1.99 \times 10^{-6} T^2 \times RH^2
\end{aligned} \tag{9}$$
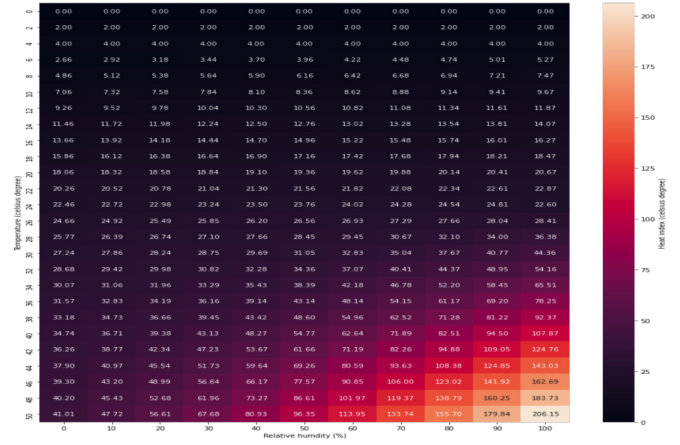
Fig. 2. The heat index is calculated using the formula NWS online [11], depending on two factors of temperature and relative humidity. The dark color shows the less heat index, and the light color show the high heat index.

If the relative humidity is less than 13% and the temperature is greater than 85 degrees F and not less than H degrees F, the above HI will subtract one adjustment below:

$$ADJ = ([13 - RH]/4) \times ([17 - |T - 95|]/17)^{0.5} \tag{10}$$

where H, which is considered the hot threshold, is also changed because it affects the use of electrical equipment for cooling. If the relative humidity is more than 85% and the temperature is between 80 and 87 degrees F, the following adjustment is added to HI:

$$ADJ = 0.02 \times (RH - 85) \times (87 - T) \tag{11}$$

For convenience, the input temperature data, C and H values in degrees Celsius will be converted to degrees Fahrenheit to match the algorithm. The calculated results are returned to degrees Celsius as temperature.

D. Proposed model

To use the heat index data for electricity load prediction with the LSTM model, we have proposed a model as follows (see Figure 4):

- Step 1: Using log transformation for the electricity load dataset.
- Step 2: Heat index algorithm: Use heat index algorithm to create heat index based on temperature data and humidity data.
- Step 3: Dataset is divided into the training set and testing set to carry out the experiments.
- Step 4: Training LSTM model to get the appropriate parameters for forecasting. That is the final model to be used for forecasting the test set.
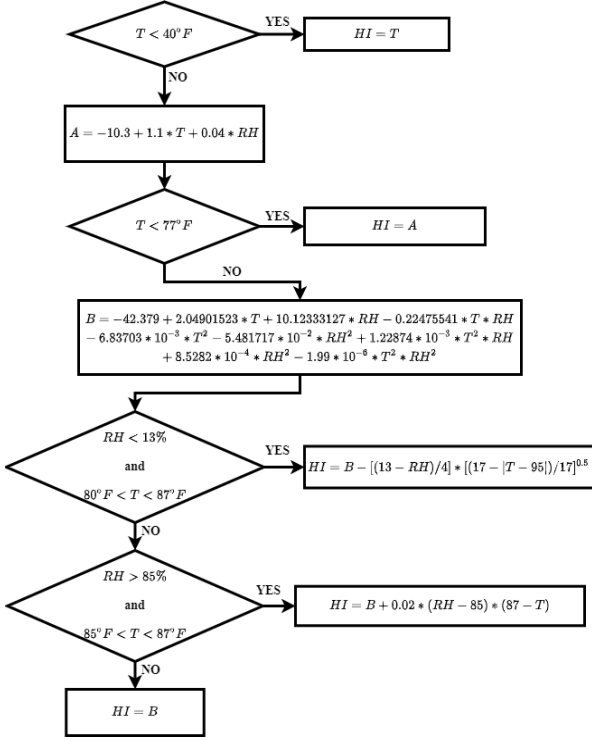
Fig. 3. The algorithm calculates the heat index (HI) feature based on air temperature in degrees Fahrenheit (T) and relative humidity in percent (RH).
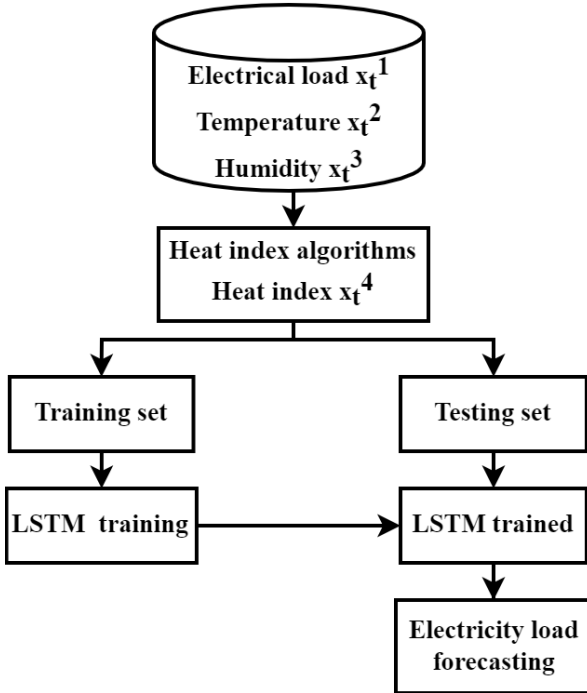
Fig. 4. Proposed model includes four steps: i) ingestion data; ii) calculating head index; iii) LSTM training; iv) LSTM forecasting.

III. Experiments and results

A. Data sets

Data set description: The data used in this paper are the National and electrical load data set of Power Companies (HNPC, CPC, NPC, SPC, HCMPC), temperature (degrees C), relative humidity(%). The time series start from 1/1/2018 to 31/5/2020 with a resolution of 1 hour.

B. Evaluation criteria

Mean absolute percentage error (MAPE) and root means square error (RMSE) is used to evaluate the result of forecasting. These criteria are investigated as follows equations:

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \frac{|y_{real}(t) - y_{fore}(t)|}{y_{real}(t)} \quad (12)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (y_{real}(t) - y_{fore}(t))^2} \quad (13)$$

C. Scenarios

Three experimental scenarios are investigated corresponding with three different inputs:

- Scenario 1: The input for the forecasting model is electricity load data and temperature data.
- Scenario 2: input data is the same as scenario 1 except that the temperature data is replaced by humidity data.
- Scenario 3: Heat index is calculated carefully and is chosen as a part of input together with electricity load data.

In this paper, we will use the same LSTM model and select fixed provinces for all scenarios.

D. Results

In this paper, we chose 48 hours for the forecasting horizon because of the national power A0's operation problem. At present, we only have all the day before today's electrical load data and use it to forecast the load for the next day. Then, we make a decision based on that prediction.

In the national load forecasting model, the results obtained by the model with the input variable using heat index are very positive. In the base model using only the temperature input variable, the MAPE value obtained in the prediction national load model is 4.01%, with the corresponding RMSE being approximately 1580MW. When using heat index with many different test models of the hot-cold threshold as a part of input data, the results got an improvement. All experimental models have the error below the threshold of 3.83% that shows a decrease compared with 4.01% in the temperature model. The RMSE also improved by about 400 MW. Going into each model, we can see that the 41-4 model shows the best result with MAPE of 3.53%. The difference between the top 10 best models is not too big, at only a 0.3% difference.

In the predicted results on the PC, the heat index input model also achieves more accurate results than the model using the input variable temperature. Comparing the results of Table

I with the results of Tables IV, V, VI, VII, VIII shows that the prediction models with the input as heat index with different C-H thresholds have low MAPE and RMSE than the model predicted with the input variable temperature.

With this result, the heat index application model has a remarkable improvement in quality. Although heat index can be calculated easily from the factors of temperature and humidity, using this index as the main input parameter of the model helps the models get the error relatively low. So that heat index is applicable in the actual operation problems of the electrical system. The difficulty of data collection should also consider into account. The formula for calculating heat index is quite explicit, which is a good choice to make the model more practical because that does not have to add too many specialized meteorological factors that are difficult to collect and forecast.

TABLE I
RESULTS OF SCENARIO 1: LSTM WITH TEMPERATURE FEATURE.

	NN	n-steps	LR	MAPE	RMSE(MW)
Nation	128	12	0.001	4.01%	1580.48
HNPC	128	12	0.001	6.97%	231.69598
CPC	128	12	0.001	4.31%	133.62663
NPC	128	12	0.001	4.92%	582.87473
SPC	128	12	0.001	5.54%	658.92897
HCMPC	128	12	0.001	5.19%	220.50571

TABLE II
RESULTS OF SCENARIO 2: LSTM WITH HUMIDITY FEATURE.

	NN	n-steps	LR	MAPE	RMSE(MW)
Nation	128	12	0.001	7.92%	2634.31
HNPC	128	12	0.001	9.33%	296.12138
CPC	128	12	0.001	5.77%	179.09436
NPC	128	12	0.001	7.56%	839.57278
SPC	128	12	0.001	10.45%	1278.32204
HCMPC	128	12	0.001	10.30%	454.51833

TABLE III
RESULTS OF SCENARIO 3: LSTM WITH HEAT INDEX FEATURE FOR FORECASTING NATION LOAD.

C - H	NN	n-steps	LR	MAPE	RMSE(MW)
41 − 4	128	12	0.001	3.53%	1499.43
39 − 5	128	12	0.001	3.60%	1471.78
42 − 9	128	12	0.001	3.63%	1487.94
37 − 9	128	12	0.001	3.71%	1496.46
39 − 6	128	12	0.001	3.72%	1497.33
39 − 10	128	12	0.001	3.77%	1502.40
40 − 4	128	12	0.001	3.77%	1517.51
38 − 10	128	12	0.001	3.78%	1526.42
38 − 1	128	12	0.001	3.83%	1537.52
38 − 3	128	12	0.001	3.83%	1525.48

IV. CONCLUSION AND DISCUSSION

This paper proposes a new methodology to research heat index and electricity load forecasting. The heat index depends on the temperature and humidity that affect to behaviors of electricity load consumers. The proposed model using heat index algorithm and LSTM for short-term electricity load forecasting gets better performance.

TABLE IV
RESULTS OF SCENARIO 3: LSTM WITH HEAT INDEX FEATURE FOR FORECASTING HNPC.

C - H	NN	n-steps	LR	MAPE	RMSE(kW)
39 − 6	128	12	0.001	6.10%	215448.53
39 − 5	128	12	0.001	6.12%	213490.63
38 − 7	128	12	0.001	6.14%	219631.23
37 − 9	128	12	0.001	6.18%	217257.64
38 − 9	128	12	0.001	6.21%	214520.15
37 − 7	128	12	0.001	6.23%	228125.94
38 − 1	128	12	0.001	6.25%	223557.10
40 − 8	128	12	0.001	6.25%	214730.71
42 − 1	128	12	0.001	6.25%	214277.67
38 − 3	128	12	0.001	6.27%	219800.52

TABLE V
RESULTS OF SCENARIO 3: LSTM WITH HEAT INDEX FEATURE FOR FORECASTING CPC.

C - H	NN	n-steps	LR	MAPE	RMSE(kW)
39 − 6	128	12	0.001	3.64%	113277.69
37 − 7	128	12	0.001	3.65%	114193.14
39 − 5	128	12	0.001	3.65%	115041.49
39 − 8	128	12	0.001	3.67%	114485.71
40 − 1	128	12	0.001	3.68%	115576.97
38 − 9	128	12	0.001	3.69%	115353.33
39 − 4	128	12	0.001	3.69%	114483.58
41 − 10	128	12	0.001	3.69%	114650.49
37 − 6	128	12	0.001	3.71%	116438.16
37 − 3	128	12	0.001	3.73%	115909.76

TABLE VI
RESULTS OF SCENARIO 3: LSTM WITH HEAT INDEX FEATURE FOR FORECASTING NPC.

C - H	NN	n-steps	LR	MAPE	RMSE(kW)
39 − 8	128	12	0.001	4.34%	536113.75
39 − 5	128	12	0.001	4.38%	538198.43
37 − 6	128	12	0.001	4.40%	548028.81
37 − 10	128	12	0.001	4.40%	546065.32
37 − 1	128	12	0.001	4.41%	542417.03
41 − 5	128	12	0.001	4.41%	540391.79
41 − 8	128	12	0.001	4.42%	541764.46
38 − 9	128	12	0.001	4.44%	548004.32
38 − 5	128	12	0.001	4.46%	550851.25
40 − 5	128	12	0.001	4.46%	548939.66

TABLE VII
RESULTS OF SCENARIO 3: LSTM WITH HEAT INDEX FEATURE FOR FORECASTING SPC.

C - H	NN	n-steps	LR	MAPE	RMSE(kW)
42 − 9	128	12	0.001	4.44%	563985.90
41 − 9	128	12	0.001	4.48%	573272.89
41 − 8	128	12	0.001	4.55%	577215.30
37 − 9	128	12	0.001	4.58%	577842.66
39 − 6	128	12	0.001	4.75%	590976.58
41 − 1	128	12	0.001	4.69%	588621.45
37 − 3	128	12	0.001	4.75%	591935.00
38 − 9	128	12	0.001	4.75%	603429.64
39 − 8	128	12	0.001	4.78%	605446.99
39 − 4	128	12	0.001	4.80%	595860.43

- Developing a new approach that analyzes heat index for electricity load forecasting.
- Successfully integrating heat index and electricity load to propose a novel model ;

TABLE VIII
RESULTS OF SCENARIO 3: LSTM WITH HEAT INDEX FEATURE FOR
FORECASTING HCMPC.

C - H	NN	n-steps	LR	MAPE	RMSE(kW)
37 − 6	128	12	0.001	4.52%	206081.01
37 − 5	128	12	0.001	4.58%	210169.28
41 − 8	128	12	0.001	4.62%	208628.38
37 − 9	128	12	0.001	4.71%	210199.82
40 − 2	128	12	0.001	4.74%	212026.38
37 − 4	128	12	0.001	4.78%	205744.24
39 − 8	128	12	0.001	4.84%	212875.61
39 − 5	128	12	0.001	4.88%	218854.68
38 − 6	128	12	0.001	4.90%	218351.68
38 − 9	128	12	0.001	4.91%	218003.88

- Well applying the proposed model into data sets of Vietnam;
- Comparing with using the temperature feature, the proposed model achieves a better performance in electrical load forecasting.

V. FUTURE WORKS

In the future, we need to improve the accuracy. Firstly, the heat index should be dispensed into groups and dependently researched according to the lunar calendar or seasons. Secondly, the heat index should be clustering based on the volume to analyze the electricity load forecasting. Finally, we can combine optimization methods for feature selection and hyperparameter optimization to improve the performance of the model.

REFERENCES

[1] K. chao Miao, T. ting Han, Y. qing Yao, H. Lu, P. Chen, B. Wang, and J. Zhang, "Application of lstm for short term fog forecasting based on meteorological elements," *Neurocomputing*, vol. 408, pp. 285 – 291, 2020.
[2] I. Lezhenin, N. Bogach, and E. Pyshkin, "Urban sound classification using long short-term memory neural network," in *Proceedings of the 2019 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. Maciaszek, and M. Paprzycki, Eds., vol. 18. IEEE, 2019, pp. 57–60. [Online]. Available: http://dx.doi.org/10.15439/2019F185
[3] Z. Cui, R. Ke, Z. Pu, and Y. Wang, "Stacked bidirectional and unidirectional lstm recurrent neural network for forecasting network-wide traffic state with missing values," *Transportation Research Part C: Emerging Technologies*, vol. 118, p. 102674, 2020.
[4] L. Mei, R. Hu, H. Cao, Y. Liu, Z. Han, F. Li, and J. Li, "Realtime mobile bandwidth prediction using lstm neural network and bayesian fusion," *Computer Networks*, vol. 182, p. 107515, 2020.
[5] Y. Fan, F. Fang, and X. Wang, "Probability forecasting for short-term electricity load based on lstm," in *2019 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC)*, 2019, pp. 516–522.
[6] I. Agbehadji, R. Millham, S. Fong, and H. Yang, "Kestrel-based search algorithm (ksa) and long short term memory (lstm) network for feature selection in classification of high-dimensional bioinformatics datasets," in *FedCSIS*, 09 2018, pp. 15–20.
[7] Y. Wang, D. Gan, M. Sun, N. Zhang, Z. Lu, and C. Kang, "Probabilistic individual load forecasting using pinball loss guided lstm," *Applied Energy*, vol. 235, pp. 10 – 20, 2019.
[8] R. G. Steadman, "The Assessment of Sultriness. Part I: A Temperature-Humidity Index Based on Human Physiology and Clothing Science," *Journal of Applied Meteorology*, vol. 18, no. 7, pp. 861–873, 07 1979.
[9] ——, "The Assessment of Sultriness. Part II: Effects of Wind, Extra Radiation and Barometric Pressure on Apparent Temperature," *Journal of Applied Meteorology*, vol. 18, no. 7, pp. 874–885, 07 1979.
[10] L. P. Rothfusz and N. S. R. Headquarters, "The heat index equation (or, more than you ever wanted to know about heat index)," *Fort Worth, Texas: National Oceanic and Atmospheric Administration, National Weather Service, Office of Meteorology*, vol. 9023, 1990.
[11] G. B. Anderson, M. L. Bell, and R. D. Peng, "Methods to calculate the heat index as an exposure metric in environmental health research," *Environmental Health Perspectives*, vol. 121, no. 10, pp. 1111–1119, 2013.
[12] S. Opitz-Stapleton, L. Sabbag, K. Hawley, P. Tran, L. Hoang, and P. H. Nguyen, "Heat index trends and climate change implications for occupational heat exposure in da nang, vietnam," *Climate Services*, vol. 2-3, pp. 41 – 51, 2016.
[13] C. Cui, T. Wu, M. Hu, J. D. Weir, and X. Li, "Short-term building energy model recommendation system: A meta-learning approach," *Solar Energy*, vol. 172, pp. 251–263, 2016.
[14] Y. He, R. Liu, H. Li, S. Wang, and X. Lu, "Short-term power load probability density forecasting method using kernel-based support vector quantile regression and copula theory," *Applied Energy*, vol. 185, pp. 254–266, 2017.
[15] N. Zeng, H. Zhang, W. Liu, J. Liang, and F. E. Alsaadi, "A switching delayed pso optimized extreme learning machine for short-term load forecasting," *Neurocomputing*, 2017.
[16] L. Song, G. Lalit, and W. Peng, "An ensemble approach for short-term load forecasting by extreme learning machine," *Applied Energy*, vol. 170, pp. 22–29, 2016.
[17] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, p. 1735–1780, Nov. 1997.
[18] F. A. Gers, J. A. Schmidhuber, and F. A. Cummins, "Learning to forget: Continual prediction with lstm," *Neural Comput.*, vol. 12, no. 10, p. 2451–2471, Oct. 2000.

An IoT based Condition Monitoring System of Biogas Electrical Generator for Performance Evaluation

Hoang Duc Chinh
School of Electrical Engineering
Hanoi University of Science and Technology
Vietnam
chinh.hoangduc@hust.edu.vn

Hoang Anh
School of Electrical Engineering
Hanoi University of
Science and Technology
Vietnam

Nguyen Duy Hieu
School of Electrical Engineering
Hanoi University of
Science and Technology
Vietnam

Vu The Anh
School of Electrical Engineering Hanoi University of Science
and Technology Vietnam

Krishnanand Kaippilly Radhakrishnan
SinBerBEST Berkeley Education Alliance for Research
in Singapore (BEARS) Singapore

Abstract—**Biogas is a promising renewable energy source having great potential, especially in livestock farms. However, as biogas electric generators are usually deployed in rural areas, it would take more time and effort to repair if any fault occurs. Remote monitoring of the system condition is essential to diagnose or even predict the faults in advance and subsequently plan the maintenance schedule in time. This paper presents a monitoring system of biogas-based power generation system using Internet-of-Things (IoT) devices. Information of the generator operation is acquired by field devices and forwarded to a remote server. Data collection and management are facilitated by Lambda architecture and Apache Kafka software platform for their interoperability and strong support of big data management. The system shows that near real-time supervision of the object conditions can be obtained. Historical data analyses of a few operation scenarios are also provided to evaluate the generation system performance as well as to discuss its fault diagnosis.**

Index Terms—**Internet of Thing, Biogas generator, Lambda Architecture, Condition Monitoring.**

I. Introduction

INCREMENT in energy demand raises a serious concern worldwide, especially in developing countries. While fossil fuel resources do not provide sustainable means for sustainable development, other conventional energy sources such as hydroelectricity or nuclear power have adverse environmental impacts and constraints. Seeking alternative energy sources, agricultural countries are taking advantage of the energy generated by agricultural wastes such as biogas to replace gasoline and oil in some cases. Biogas extracted from animal manure through anaerobic digestion is a renewable energy source for livestock farms, and it can be used to produce heat or electricity. Moreover, this process also helps to reduce completely methane emissions and stabilizes the manure before its agronomic use [1]. The electrical generators using biogas are typically modified from gasoline and diesel engines or based on dual-fuel engine [2, 3]. Once installed in livestock farms, the energy produced by these generators can power the electrical loads and help to avoid disposing of contaminated substances to the surrounding area.

In order to facilitate the reliable operation of such a biogas energy system, users are required to observe and under-stand its performance while it is functioning. Machines Monitoring System (MMS), also known as SCADA (Supervisory control and data acquisition) is a system which performs real-time communication and data collection from field devices. The data is then processed, stored and analysed. The outcome would assist users to understand the operating status of machinery and equipment and take appropriate actions when an incident occurs. Businesses can be then more proactive in managing internal activities and at the same time helping to avoid loss of income due to any machine downtime.

Related studies in condition monitoring of similar systems with data analysis have been reported in the literature. One of the early works is presented in [4]. The authors monitored a natural gas-based generator; generator data was collected, calculated, and stored in a floppy disc. Recent advances in information management technologies enable data acquisition systems to handle a big amount of data and carry out sophisticated analysis, specifically in generation systems with rotary machines [5-7]. Evaluation of system conditions can be performed by processed parameters such as vibration, noise [8] or oil analysis method [9]. Variety of the data from a number of machines acquired over a long duration can provide comprehensive understanding of the machines. However, efficient data management with scalability, high security and privacy needs to be taken into account [10, 11].

In this work, an IoT based monitoring system for the biogas electrical generators has been proposed and developed to enable remote observation. Various parameters of the generator as well as other parts of the energy system are collected. The data acquisition solution is an integration of a single board computer with open source self-developed software and industrial devices like PLCs. Data acquired are pre-processed at the single board computer and then sent over the Internet to a remote for further analysis. In order to collect, store, and analyse a huge amount of information as well as to be extensible, we use a lambda architecture to build the monitoring system. It highly supports data gathering from multiple sources and processing a large amount of distributed data. The ultimate purpose of this
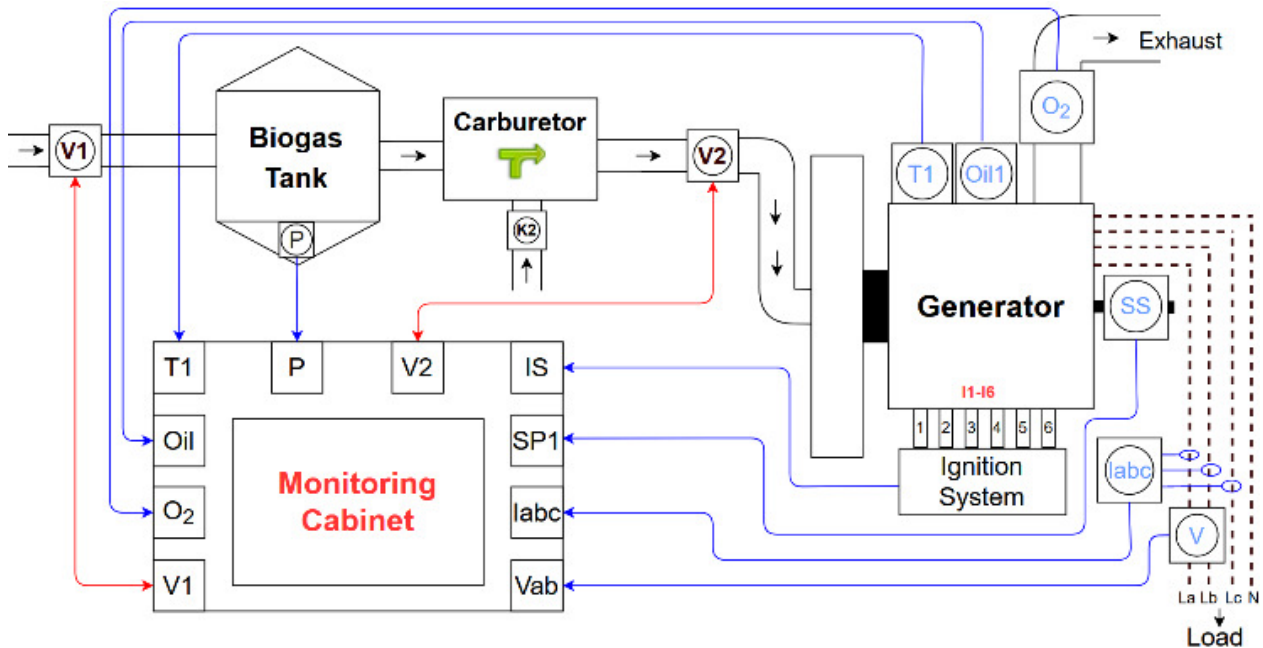
Fig. 1. Biogas based electrical generation systems.

process is to understand the biogas system comprehensively, provide the end-users recommendation of the operation and allow the manufacturers to have in-depth view of the usage patterns.

The structure of this paper is as the following: Section 1 provides the introduction of the work, Section 2 gives a description of the biogas energy system, Section 3 presents technical details of the developed system, Section 4 shows the experimental results, data analysis and discussion, and Section 5 provides the conclusion.

II. BIOGAS ENERGY SYSTEM DESCRIPTION

The overall system of this study is illustrated in Fig. 1. It is considered the biogas generator of which the engine has been modified from a diesel type to use biogas as its fuel.

The fuel. Biogas as fuel is contained in the "Biogas tank". As reported by Osorio and Torres in [12], biogas composed of various gases such as methane (60–70%), CO2 (30–40%), nitrogen (<1%), and H2S (10–2000 ppm). This various composition makes it difficult for the user in setting the engine to work stably. As shown in Fig. 1, the biogas from the main treatment tank is pushed into the biogas tank via valve V1. It then flows through the carburetor which mixes the biogas and ambient air with rich oxygen. The ratio of the mixture is pre-fixed with manual valve K2. The flow of this mixture supplied to the engine is controlled with valve V2.

The engine performance highly depends on the quality of fuel supply. The system controller has been designed in such a way that it can control the main input valve V1 and the gas supply valve V2 to obtain the proper fuel for the engine. However, changing working conditions and disturbances from an ambient environment may create some issues with the engine operation. Three parameters including the position of valves V1 and V2 as well as the main tank pressure P are acquired by the monitoring cabinet. Although these values are used by the controller to provide appropriate control

signals to the valves, they are also collected for further assessing the system performance.

The generator. The generator includes a combustion engine, a three-phase synchronous generator, and accessories such as a start-up motor, speed encoder, etc. The parameters to be monitored of this component include two groups. First, the electrical parameters such as output line voltage Vab, phase currents Iabc, and frequency calculated from the generator rotary speed SP1. Second, the environmental parameters which are the temperature of the cooling water for the engine T1, oil pressure Oil1 and O2 concentration in the exhaust.
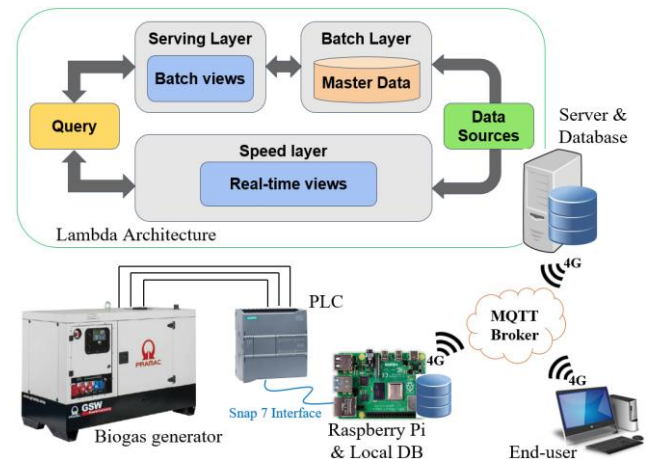
Fig. 2. The data acquisition system.

The monitoring cabinet. The cabinet consists of local data acquisition system hardware. Some of the sensors are built-in inside the system such as engine water temperature and oil pressure, valve positions, tank pressure. All of them are connected to the cabinet terminals. The others including voltage and current sensors are installed inside the cabinet. The key component in charge of data collection is a single board computer, Raspberry Pi, equipped with 3G modem. Data are stored locally in Raspberry Pi and also sent to a remote server over the Internet. The monitoring of the

engine's operating parameters can thereby give an assessment of its performance as well as predictions of its failure to some extent. In addition, it is expected that analysed data can be employed to suggest that the operator should use and adjust the controller of the machine optimally.

III. CONDITION MONITORING SYSTEM DESIGN

A. Local data acquisition scheme

The biogas generation system is equipped with a PLC Siemens S7-1200 as the main controller and perceive sensing measurements. The PLC module continuously acquires and stores sensing data in its data block. Besides the parameters mentioned in Section 2, the starting time and the stopping time of the machine are also recorded. The module allows third-party devices to connect via Ethernet port and to communicate via an open-source library called Snap7 Interface. The library supports multiple programming languages include Python. As mentioned in Section 2, a single-board computer Raspberry Pi is used as an edge computing unit for its versatility and strong support in open-source software. It is equipped with a 4G modem to communicate with the server remotely.

A python application is written and executed on Raspberry Pi to communicate with the PLC unit. It periodically accesses the PLC data block to get the data, pre-processes and stores it locally, then sends it towards the remote server. SQLite is adopted as the local database for on-site analysis and backup solution in case the Internet connection is lost. MQTT and Kafka are the communication scheme used to transfer the data over the Internet to the server.

B. Remote data collection and storage solution

The entire data collection system including the remote database server is illustrated in Fig. 2. Data Collection is built with two main components which are Message Queuing Telemetry Transport (MQTT) and Kafka. MQTT is a publishing subscription-based messaging protocol (ISO / IEC PRF 20922) widely used for IoT devices as it is light-weight and flexible. However, MQTT is not built for highly scalable, longer archival, or easily integrated into legacy systems. Apache Kafka on the other hand is a highly scalable, decentralized streaming platform. It is capable of managing and storing a massive amount of data on disks. Therefore, MQTT and Apache Kafka are the perfect combination of end-to-end IoT integration from edge to data center. MQTT and Kafka are in charge of gathering data sent by the field devices, i.e. the Raspberry Pi inside the monitoring cabinet, in JSON format packets and circulating them at server side.

Apache Kafka is part of the Lambda Architecture which also facilitates data query for providing information to users subsequently. In lambda Architecture, the data is saved into two databases: Cassandra and MongoDB which is a NoSQL database. Conventional SQL databases impose some constraints with limitations. NoSQL solves big data problems of information systems or distributed data and interact and extend easily with high performance. Fig. 3 illustrates data saved as document-oriented manner in MongoDB.
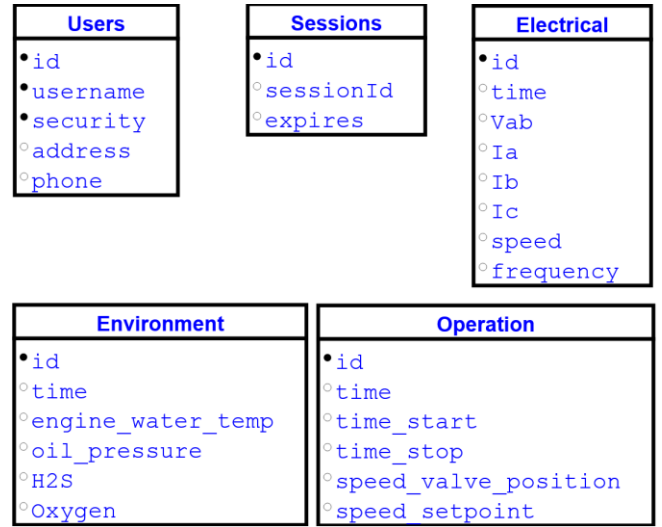
Fig. 3. Database schema in MongoDB at server side.

C. Condition monitoring with a web-based user interface

A webpage graphical user interface (GUI) is developed for end-user to monitor the system in near real-time. It displays the instantaneous values of the important parameters such as phase currents, line voltages, the generator rotary speed, frequency, etc. Besides, it also provides a brief overview of all the generators as well as the historical trend of the monitored parameters. The web application is written mainly in HTML and JavaScript with responsive design. Web-socket MQTT is used as the communication between the front-end web GUI and the server instead of having conventional backend web services. This approach enables the front-end application to be deployed in any web server as well as loaded on a browser as an independent application.

IV. DATA ANALYSIS AND DISCUSSION

Four biogas based generation systems have been built and deployed in four different hog farms distributed in the north of Vietnam. Each generator is able to supply power up to 90 kW which is sufficient for all loads on the farm. Data have been collected for a few months and analyzed to comprehend the operation of those machines. The data acquisition device read sensing measurements every 10 seconds.

Real-time monitoring. Fig.4a illustrates the web-based GUI with the four machine locations in the main dashboard page marked by 4 circles. The metadata of the machines including their locations are recorded in the database once the machines are installed. The web application also displays the measurements of machine parameters which are being sent remotely from the Raspberry Pi on site. Values shown in the parameter page in Fig.4b are of a particular machine selected to display. the fig shows the current values of few parameters at steady state such as three-phase currents, line voltage between phase A and B, the generator speed, frequency, etc. when the machine operates normally.

(b)
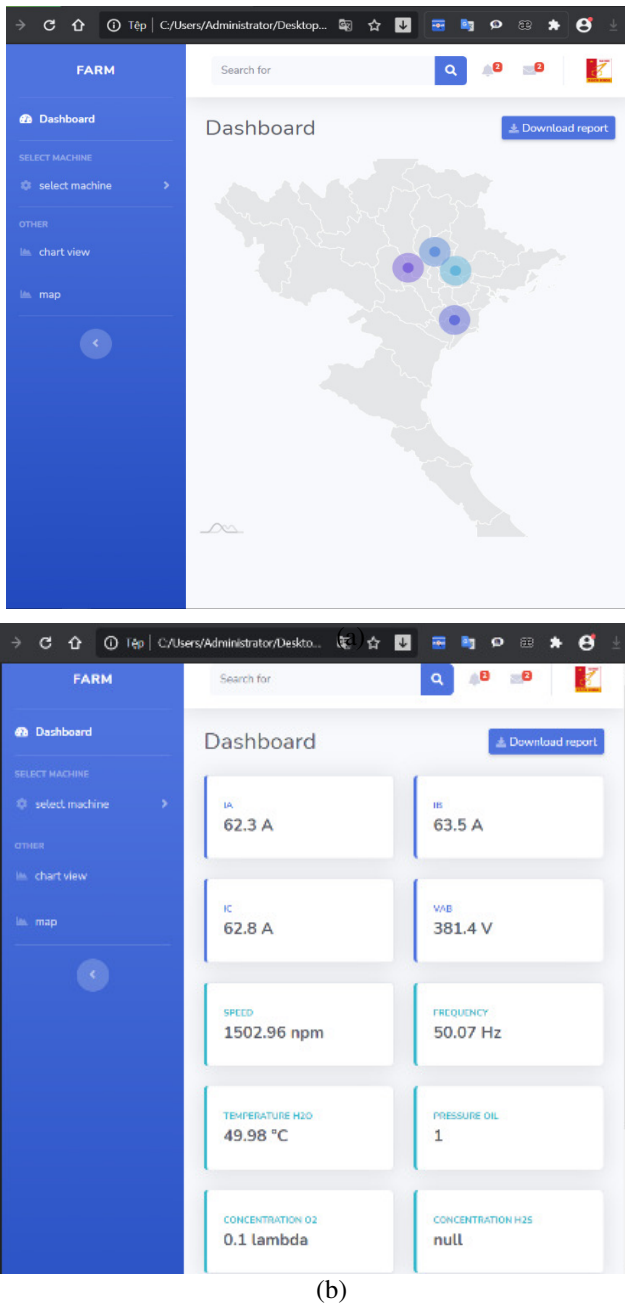
Fig. 4. (a) Main page and (b) real-time measurement page in the web-based GUI.

Data analysis. Measurements from generators are acquired every 10 seconds at the server-side with the support of Lambda architecture. Fig 5. shows the daily operating hours of a biogas generator which we deployed on a farm in Thai Nguyen province for one month. The duration varies from around 5 hours to 20 hours per day depending on the availability of the gas as well as the load demand. The generator is usually run into two sessions. Daily energy consumption of the farm when using the generator during the same period is shown in Fig. 6. The average energy consumption is around 397 kWh per day and the total energy of this month is 13126 kWh.
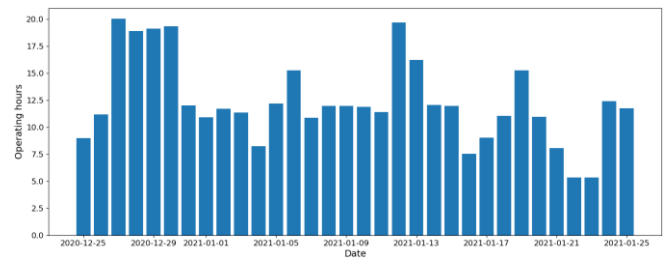
Fig. 5. Daily operating hours of a biogas generator.
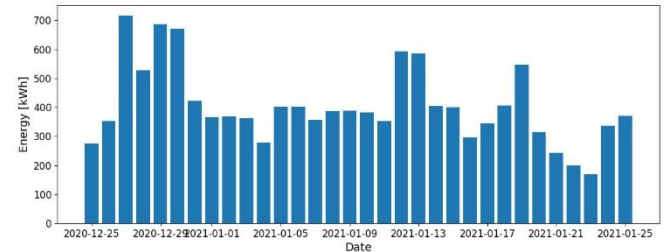
Fig. 6. Daily energy produced by the biogas generator to supply the farm.
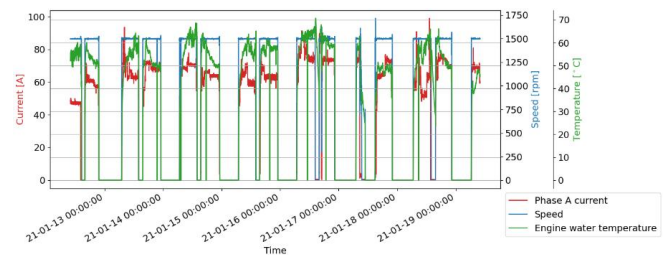
Fig. 7. Historical data of phase A current, generator speed and engine oil temperature

Fig. 7 shows the historical measurements over 7 days of one phase current, speed and temperature of the cooling water for the engine. The zero values represent that the machine is at rest. The temperature of cooling water varies around $50\ °C - 70\ °C$. Generator speed at steady state is kept at 1500 revolutions per minute (rpm) which is maintained by the speed controller in order to achieve 50Hz output frequency. The phase current is decided by the farm load requirement, average value of the current is around 60A which is usually consumed by the ventilation and lighting system and pumps. It can also be seen in this figure that at some moments the speed may increase above the desired value of 1500 rpm. No matter how the load current varies within the generator capacity, the speed should be kept constant, any fluctuation if occurs should be stabilized as soon as possible. As mentioned in Section 2, this speed is regulated by adjusting the opening position of the valve V2 to control the flow of gas mixture into the combustion engine.

The valve V2 position in terms of percentage is plotted together with the engine speed and phase A current in Fig. 8a. Normally, the position is around 20% when the current is around 70A. If the current increased or decreased the valve is also open wider or less respectively to maintain the same engine speed. However, it sometimes misbehaves as highlighted with the dashed line box in Fig. 8a. Fig. 8b shows a zoom in view of that period. At around 16.45, the load does not change but the valve open position is increased to a maximum of 100% and the speed is still the same. After 15 minutes, the speed is decreased which results

in a decrement in supply current and then the valve is almost closed subsequently. This phenomenon can happen due to the lack of fuel or biogas quality degradation. Fluctuation in engine speed would have adverse effects on the machine lifespan and should be avoided.

(a)

(b)
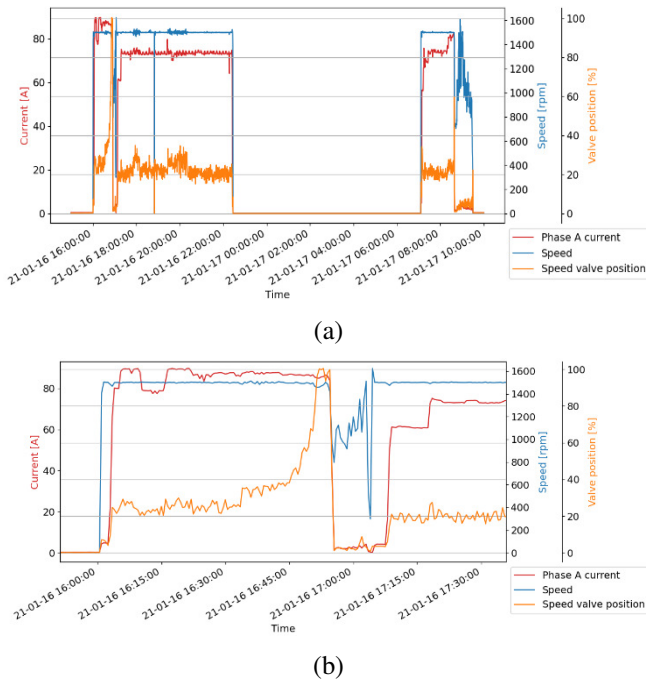
Fig. 8. Speed valve position vs. engine speed vs. phase A current over time.

V. CONCLUSIONS

In this paper, we have presented an IoT based system for monitoring performance of a biogas generator. The system architecture implemented is resilient due to edge computing and storage facilitated by IoT devices that are local to the biogas generator, thereby capable of overcoming the intermittent connectivity challenges faced in rural farms. The system is scalable for multiple edge devices and can serve many simultaneous users, owing to its adoption of lambda architecture. Through the deployment of this robust monitoring system, outages of the machines have been recognized and studied with the collected data. The common faults recorded are due to the inconsistent quality of input biogas as well as the error of the fuel feeding valve. Future work would investigate predictive techniques based on sensitivity analysis and correlation of system parameters with various fault types. Forecasting of energy generation is also worth exploring since that can improve the logistics of biogas generation planning.

REFERENCES

[1] G. Fiorese, G. Guariso, A. Polimeni: Optimizing biogas production: an application to an Italian farming district. In: Proceedings of the iEMSs fourth biennial meeting: international congress on environmental modelling and software, Barcelona, Spain (2008)

[2] I. Wayan Surata, Tjokorda Gde Tirta Nindhia, I. Ketut Adi Atmika, Dewa Ngakan Ketut Putra Negara, I. Wayan Eka Permana Putra: Simple Conversion Method from Gasoline to Biogas Fueled Small Engine to Powered Electric Generator. In: Energy Procedia, vol. 52, pp. 626-632 (2014)

[3] Chinmay Deheri, Saroj Kumar Acharya, Dhirendra Nath Thatoi, Ambica Prasad Mohanty: A review on performance of biogas and hydrogen on diesel engine in dual fuel mode. In: Fuel, vol. 260 (2020)

[4] Wall, Timothy M.P: Instrumentation and control of an engine generator set for biogas. In: Retrospective Theses and Dissertations. 8755, https://lib.dr.iastate.edu/rtd/8755, last accessed 2021/02/01 (1985)

[5] Zi Lin, Xiaolei Liu: Wind power forecasting of an offshore wind turbine based on highfrequency SCADA data and deep learning neural network. In: Energy, vol. 201 on page 117693

[6] Chenxing Sheng, Tonghai Wu, Yuelei Zhang: Non-destructive testing of marine diesel engines using integration of ferrographic analysis and spectrum analysis. Insight - Non-Destructive Testing and Condition Monitoring, vol. 54, issue 7, pp. 394-398 (2012)

[7] A. Yazidi, H. Henao, G. Capolino, F. Betin and F. Filippetti: A Web-Based Remote Laboratory for Monitoring and Diagnosis of AC Electrical Machines. In: IEEE Transactions on Industrial Electronics, vol. 58, no. 10, pp. 4950-4959 (2011)

[8] N Tandon, A Choudhury: A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. In: Tribology International, vol. 32, issue 8 pp. 469-480 (1999)

[9] X.P. Yan, C.H. Zhao, Z.Y. Lu, X.C. Zhou, H.L. Xiao: A study of information technology used in oil monitoring. In: Tribology International, vol. 38, issue 10, pp 879-886 (2005)

[10] A. Kowalska, P. Łuczak, D. Sielski, T. Kowalski, A. Romanowski and D. Sankowski, "Towards Big Data Solutions for Industrial Tomography Data Processing," 2019 Federated Conference on Computer Science and Information Systems (FedCSIS), 2019, pp. 427-431

[11] J. Jabłoński and S. Robak, "Information Systems Development and Usage with Consideration of Privacy and Cyber Security Aspects," 2019 Federated Conference on Computer Science and Information Systems (FedCSIS), 2019, pp. 1-8

[12] F. Osorio, J.C. Torres: Biogas purification from anaerobic digestion in a wastewater treatment plant for biofuel production. In: Renewable Energy, vol. 34, issue 10, pp. 2164 -2171 (2009)

Mobile robots interacting with obstacles control based on artificial intelligence

Tran Duc Chuyen
University of Economics-Technology
for Industries
Hanoi, Vietnam

Roan Van Hoa
University of Economics-Technology
for Industries
Hanoi, Vietnam

Nguyen Duc Dien
University of Economics-Technology
for Industries
Hanoi, Vietnam

Tung Lam Nguyen
Hanoi University of Science and Technology
Hanoi, Vietnam
lam.nguyentung@hust.edu.vn

Abstract—In this paper, research on the applications of artificial intelligence in implementing Deep Deterministic Policy Gradient (DDPG) on Gazebo model and the reality of mobile robot has been studied and applied. The goal of the experimental studies is to navigate the mobile robot to learn the best possible action to move in real-world environments when facing fixed and mobile obstacles. When the robot moves in an environment with obstacles, the robot will automatically control to avoid these obstacles. Then, the more time that can be maintained within a specific limit, the more rewards are accumulated and therefore better results will be achieved. The authors performed various tests with many transform parameters and proved that the DDPG algorithm is more efficient than algorithms like Q-learning, Machine learning, deep Q-network, etc. Then execute SLAM to recognize the robot positions, and virtual maps are precisely built and displayed in Rviz. The research results will be the basis for the design and construction of control algorithms for mobile robots and industrial robots applied in programming techniques and industrial factory automation control.

Index Terms—Mobile robots, artificial intelligence, DDPG algorithm, autonomous navigation, reinforcement learning.

I. Introduction

NOWADAYS Artificial Intelligence (AI), Internet of Things (IoT), and robot controls are receiving a lot of attention. Robot technology has changed since the first introduction of robots in 1917. Today, machines are present in our lives, supporting us in everyday life, [1] - [5]. One of these new technologies is artificial intelligence that has come to life as well as robotics and machine tools technology, so robots can now properly process and manage information, and automatically perform certain tasks without human assistance, replacing humans in industrial factories. However, the ability to perceive the environment (feel) and make decisions (to take action) is a very difficult task for the computerized machines. Therefore, the field of Artificial Intelligence (AI) is needed for mobile robots to solve such problems, [3, 4, 5, 6].

In this paper, the authors present a robot control problem based on an intelligent and modern Deep Deterministic Policy Gradient (DDPG). The designed help the robot to navigate in a complex environment with obstacles. Experimental studies are per-formed on automated navigation for the mobile robot. More specifically, the author introduces the neural network structure to generalize and approximate the values of all states based on the DDPG artificial intelligence algorithm. This is a policy-based, online learning intensive learning algorithm, and is backboned by the Actor - Critic intelligent network architecture. The tests are conducted on the Gazebo emulator using a high-profile computer with a mobile robot, with its open-source extension to perform automated navigation tasks for mobile robots, and then carried out to experiments. The research results will be the basis for the research and application of mobile robots in practical production, in industrial plants, [6, 8, 10, 12].

II. Mobile robot model

Consider a mobile robot as in Fig. 1. Two coordinate frames are used to describe the motion of the mobile robot, the global coordinate frame (X, Y) which is earth-fixed and the other is the local coordinate frame (X_l, Y_l) which is attached to the mobile robot. The angle between the two coordinate frames is denoted as θ. The robot's motion will be defined for the navigation stack. As the global coordinate chosen in Fig. 1, it is clear that the robot's velocity contains three components: the linear velocity along OX and OY axes, and the angular speed, [6].
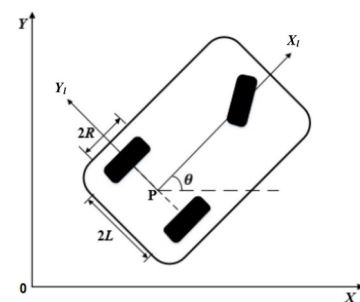
Fig. 1. The model mobile robot

To identify the position of the multi-directional mobile robot, the P-coordinate is selected on the robot's frame as its control center position reference point. Reference point P is positioned by the coordinates (x, y) in the global frame of the entire control environment for the robot. In order to formulate the motion of the mobile robot as component movements, it is necessary to define the motion mapping along the axes of the spherical frame with the motion along the coordinate of the local frame. This mapping is represented as the following expression:

$$\begin{bmatrix} \dot{x}_l \\ \dot{y}_l \\ \dot{\theta}_l \end{bmatrix} = \begin{bmatrix} cos\theta & sin\theta & 0 \\ -sin\theta & cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \end{bmatrix} \quad (1)$$

By considering the limit of wheel slip $\dot{y}_l = 0$ implies that the wheel cannot slide orthogonal to the wheel plane, then it is straightforward to have:

$$-\dot{x}sin\theta + \dot{y}cos\theta = 0 \quad (2)$$

Defining forward velocity $\dot{x}_l$ of the multi-directional mobile robot as v and rotation speed θ as w, the kinematics of the multi-directional mobile robot becomes:

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} cos\theta & 0 \\ sin\theta & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} \quad (3)$$

The forward velocity and rotation speed of the robot have the following relation:

$$v = \frac{1}{2}(v_l + v_r), \qquad w = \frac{v_l - v_r}{2L} \quad (4)$$

If we consider the linear velocities at the wheels and the angular velocities of the two qualified wheels, then we can get: $v_l = w_l R$ and $v_r = w_r R$. The angular velocities of the two standard wheels can be represented according to the forward velocity and rotation speed of the multi-directional mobile robot as:

$$w_l = \frac{v + wL}{R}, \quad w_r = \frac{v - wL}{R} \quad (5)$$

Considering the acceleration of the multi-directional mobile robot in its local working coordinate limit, then the robot dynamic can be written as [12]:

$$\begin{bmatrix} \dot{v} \\ \dot{w} \end{bmatrix} = \begin{bmatrix} \dfrac{1}{Rm} & \dfrac{1}{Rm} \\ \dfrac{L}{RI} & \dfrac{L}{RI} \end{bmatrix} \begin{bmatrix} \tau_1 \\ \tau_2 \end{bmatrix} \quad (6)$$

where m and I are the robot mass and inertia, respectively; R is the radius of the two fixed wheels; L is the half of the distance between the two fixed standard wheels; $\tau = [\tau_1 \ \tau_2]^T$ is the input torque vector exerted to the two fixed standard wheels, [1, 4, 8, 29]. The goal is to teach multi-directional mobile robots to track and follow certain trajectories within the right spaces and work environments with $e_{x,k} \approx 0$, $e_{y,k} \approx 0$, $e_{\theta,k} \approx 0$.

III. APPLICATION OF ARTIFICIAL INTELLIGENCE TO MOBILE ROBOT CONTROL

A. DDPG algorithm learning method

DDPG algorithm is built similar to the idea of Double Deep Q-Network. It is a model with reinforcement learning, online learning and an off-policy algorithm group with an Actor - critic network structure, [6, 7].

$$\max_a Q_{\theta_Q}(s,a) = Q_{\theta_Q}\left(s, arg \max_a Q_{\theta_Q}(s,a)\right) \quad (7)$$

If we build a neural network to choose the optimal action for a particular state, $\mu_{\theta_\mu}(s) arg \max_a Q_{\theta_Q}(s,a)$, then optimize component Q_{θ_Q} according to the network parameters θ_μ just created, then we have:

$$\theta_\mu \leftarrow arg \max_{\theta_\mu} Q_{\theta_Q}\left(s, \mu_{\theta_\mu}(s)\right) \quad (8)$$

This optimization considers the change of Q_{θ_Q} according to the variable θ_μ, or in other words, the evaluation of the action. This can be calculated using a string rule like the following expression:

$$\frac{dQ_{\theta_Q}}{d\theta_\mu} = \frac{dQ_{\theta_Q}}{d\mu} \cdot \frac{d\mu}{d\theta_\mu} \quad (9)$$

So building a function that approximates the value of action by a $\mu_{\theta_\mu}(s)$ neural network here is the main difference of DDPG. Thus, in each DDPG algorithm structure, there are always two components, one is Actor $\mu_{\theta_\mu}(s)$ and the other is Critic $Q_{\theta_Q}(s,a)$ lattice. The relationship between the two aforementioned components and their connection to the enhanced learning environment of this algorithm when controlling the multi-directional mobile robot in a random open environment with many fixed and obstacle obstacles mobile. At that time, the DDPG algorithm always meets the requirements of the control quality as well as the working quality of the robot.

In fact, we can understand more clearly that the DDPG algorithm is improved from other algorithms to have the ability to compute continuous action space problems for multi-directional mobile robots, when Then we go to update the target function as follows: For an input sample set of $(s_i, a_i, R_i, s_i^{'})$, then the formula updates the target function value as follows:

$$y_i = R_i + \gamma Q'(s_i^{'}, \mu'(s_i^{'})) \quad (10)$$

Then we compute the loss function for the sample M value $(s_i, a_i, R_i, s_i^{'})$ to train the written network:

$$J = \frac{1}{M} \sum_{i=1}^{M} (y_i - Q(s_i, a_i))^2 \quad (11)$$

According to the string rule in expression (9), the gradient is calculated as follows:

$$\nabla_{\theta_\mu} J = \frac{1}{M} \sum_{i=1}^{M} G_{ai} G_{\mu i} \quad (12)$$

In which, $G_{ai} = \nabla_a Q(s_i, a)$ is the output gradient of the Critic network according to variable a, estimated by the Actor

network $a = \mu(s_i)$. And $G_{\mu i} = \nabla_{\theta\mu}\mu(s_i)$ is the gradient of the Actor network input according to the model parameter θ_μ.

The DDPG algorithm with neural network training and training always ensures the requirements for accurate and reliable control, DDPG Agent updates the network parameter θ of the Q rating (*S, a*) at each step of the process. network trainer, to do action a, receive new algorithm R, then it will significantly improve the performance of the control model for multi-directional mobile robot when using DDPG algorithm control programming. Moreover, DDPG always explores the space for action constantly and this is also a great challenge for scientists, [6, 13, 14].

B. The robot navigation using DDPG algorithm

The DDPG is a member of the actor-critic algorithm, which contains four neural networks: Current critic network $Q(s,a|\theta^Q)$, current actor network $\mu(s|\theta^\mu)$, target critic network $Q'(s,a|\theta^{Q'})$, and target actor network $\mu'(s|\theta^{\mu'})$, where $\theta^Q, \theta^\mu, \theta^{Q'}, \theta^{\mu'}$ are the network weights. The ingredient Q' and μ' are copy of Q and μ respectively in the structures. Both $\theta^{Q'}$ and $\theta^{\mu'}$ are partially updated from the current networks at each timestep. The current critic network is updated by minimizing the loss function. Then, the gradient function is continuous, ensuring that the robot's agent action when controlled in an obstacle environment and now the algorithm is updated in a continuous space. The specific process of the DDPG algorithm for navigating multi-directional mobile robots is described in detail as follows, [6, 7].

The performance of the DDPG algorithm deployed is very positive on the multi-directional mobile robot control model. One of the reasons authors chose to study this algorithm for the primary control of multi-directional robots was to develop something industrially controllable. Comparing the DDPG algorithms with other algorithms has also been successful for the goal of mobile navigation for robots. Some tests for each form have been given and it is clear that the DDPG algorithm works better than other algorithms like Q-learning, RL, etc.

Algorithm: Deep Deterministic Policy Gradient algorithm for the mobile robot

1: Randomly initialize critic network $Q(s,a|\theta^Q)$, actor network $\mu(s|\theta^\mu)$ with weights θ^Q and θ^μ.
2: Initialize target network Q' and μ' with weights $\theta^{Q'} \leftarrow \theta^Q, \theta^{\mu'} \leftarrow \theta^\mu$
3: Initialize replay buffer R
4:   **for** $episode = 1$ to M **do**
5:       Initialize a random process N
6:       Receive initial state s_1 from environment E
7:       **for** $t = 1$ to T **do**
8:           Select action $a_t = \mu(s_t|\theta^\mu) + N_t$ according to the current actor network
9:           Execute action a_t in the environment E, and receive reward r_t and new state s_{t+1}
10:          Store transition (s_t, a_t, r_t, s_{t+1}) in buffer R
11:          Sample a random minibatch of N transitions (s_i, a_i, r_i, s_{i+1}) from R
12:          Set $y_i = r_i + \gamma Q'(s_{i+1}, \mu'(s_{i+1}|\theta^{\mu'})|\theta^{Q'})$
13:          Update the critic by minimizing the loss: $L = \frac{1}{N}\sum_i [y_i - Q(s_i,a_i|\theta^Q)]^2$
14:          Update the actor policy using the sampled gradient:
$$\nabla_{\theta^\mu}\mu|_{s_i} \approx \frac{1}{N}\sum_i \nabla_a Q(s,a|\theta^Q)|_{s=s_i,a=\mu(s_i)} \nabla_{\theta^\mu}\mu(s|\theta^\mu)|_{s=s_i}$$
15:          Update the target networks:
$$\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}, \theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'}$$
16:      **end for**
17: **end for**

Therefore, to build a complete DDPG algorithm, it is always necessary to meet the needs of selecting a control action to a robot, executing the action, receiving rewards, storing and sampling to train the algorithm, calculating of the target function. Subsequently updating the model parameters by minimizing the loss function on all selected samples, followed by selecting the method to update the target neural network parameters, and finally updating the environmental discovery coefficient during the control process [3, 5, 6, 15, 16].

IV. EXPERIMENTAL RESULTS

A. The research TurtleBot mobile robot

Here the authors go to study the model of mobile robot: with the actual hardware architecture of this robot is shown as shown in figure 3, in which each hardware module will perform a number of tasks, in the sequence of activities of this mobile robot: such as finding a path, crossing obstacles, etc.

In the tests, the authors perform several tasks in the sequence of the mobile robot's operations, such as the Ubuntu-powered Raspberry Pi 3 Model B +. The Raspberry Pi embedded computer directly processes information from a range of sensors including the Astra smart camera and the smart sensor then transmits commands to a smart microcontroller. To record images from the environment as well as measure the distance between mobile robots and unknown obstacles, mobile robots are equipped with cameras and smart sensors, in which the smart camera can do 360 degrees laser scanning and ranges within 12m generate map data to be used for the mapping process. The smart microprocessor control circuit receives control signals from Raspberry Pi 3 Model B+.

360⁰ LiDAR for SLAM and Navigtion
Scalable structure
Single Board Computer (Raspberry Pi)
OpenCR (ARM Cortex-M7)
DYNAMIXEL x 2 for Wheels
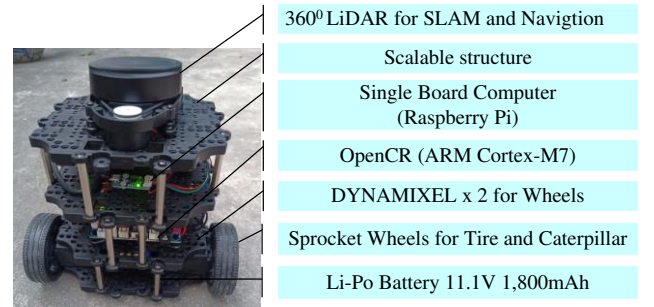Sprocket Wheels for Tire and Caterpillar
Li-Po Battery 11.1V 1,800mAh

Fig. 3. Pictures of the actual TurtleBot robot

B. The experimental results

In this section, the actual mobile robot TurtleBot is tested in a real environment, which is the environment used for real world testing consisting of flowerpots as obstacles.
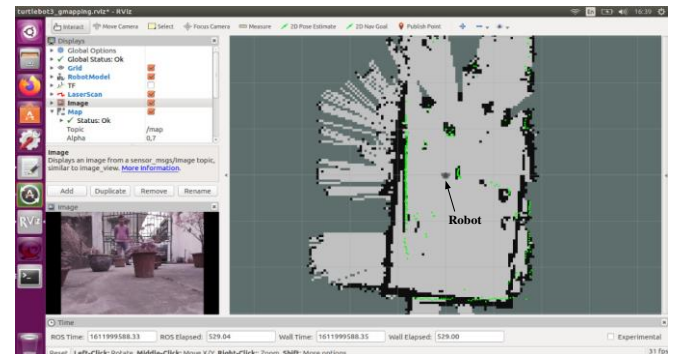

Fig. 5. The TurtleBot robot is in the process of creating concurrent mapping

In order to set up an operating environment for the purpose of controlling and navigating the robot, the authors have created a map with the goal of creating an obstacle environment of flowerpots with many different colors, then programmed for the controller, for the mobile robot as depicted in Fig. 5 and 6. The environment here includes obstacles created by different flowerpots, the robot's path will be taught in advance via computer, Wifi network, and the actual TurtleBot Robot. based on DDPG artificial intelligence algorithm and SLAM algorithm. This is primarily a visualization tool that can provide live updates of maps generated from the SLAM and DDPG algorithms. Furthermore, the vehicle's trajectory in the map can also be displayed in the real-world environment where the training and teaching and identification process so that the robot knows during the obstacle avoidance process intelligently and perfectly.
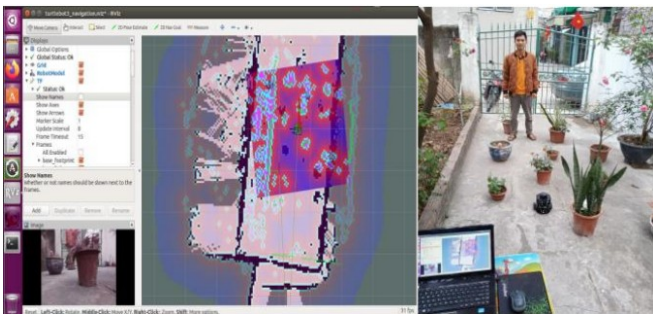
Fig. 6. Results of navigating Robot TurtleBot done on Rviz

From the simulation results and experimental results, compared with some other algorithms: Q-learning, DDPG algorithm is better than DQN, Q-learning in terms of value accuracy and control strategy are presented in [13, 16], this is also consistent with the DDPG algorithm that the authors have proposed in the paper. Accelerated learning technology and rapid action processing in large environments, can be used to achieve action status maps and meet the mobility needs of mobile robots. The data also demonstrated that the robot's path planning method based on DDPG method is better than previous studies, as shown in [13]. The above results prove the strength of the algorithm, the optimal problem of the proposed method in planning the path of the mobile robot that allows interaction with objects based on artificial intelligence.

V. Conclusions

The paper has presented the construction and control formulation of the mobile robot TurtleBot control structure. It is shown that the good effect of Slam when using Gmapping stack to construct 2D map. This article also performs a successful route planning for mobile robot when interacting with objects that are static and unknown obstacles due to constant update of local path in the environment. Tasks are based on data generated from smart

cameras and smart sensors. Furthermore, robot activity can be monitored through visual tools.

References

[1] M. N. Cirstea, A. Dinu, J.G. Khor, M. McCormick, "*Neural and Fuzzy Logic Control of Drives and Power Systems*", Linacre House, Jordan Hill, Oxford OX2 8DP , First published 2002.

[2] Charu C. Aggarwal, "*Neural Networks and Deep Learning*", Springer International Publishing AG, part of Springer Nature, 2018.

[3] Nils J. Nilsson, "*The quest for artificial interlligence a history of ideas and achievements*", Web Version Print version published by Cambridge University Press, Publishing September 13, 2010, http://www.cambridge.org/us/0521122937.

[4] Mohit Sewak, "*Deep Reinforcement Learning*", Springer Nature Singapore Pte Ltd. 2019.

[5] Latombe, J.C. "*Robot Motion Planning*"; Kluwer Academic Publishers: Norwell, MA, USA, 1992.

[6] Vu Thi Thuy Nga, Ong Xuan Loc, Trinh Hai Nam, "*Enhanced learning in automatic control with Matlab simulink*", Hanoi Polytechnic Publishing House, 2020.

[7] Nguyen Thanh Tuan, "*Base Deep learning*", The Legrand Orange Book. Version 2, last update, August 2020.

[8] Tran Hoai Linh, "*Neural network and its application in signal processing*", Hanoi Polytechnic Publishing House, 2015.

[9] Do Quang Hiep, Ngo Manh Tien, Nguyen Manh Cuong, Pham Tien Dung, Tran Van Manh, Nguyen Tien Kiem, Nguyen Duc Duy, "*An Approach to Design Navigation System for Omnidirectional Mobile Robot Based on ROS*", (IJMERR); pp: 1502-1508, Volume 11; Issue 9; 2020.

[10] Roan Van Hoa, Tran Duc Chuyen, Nguyen Tung Lam, Nguyen Duc Dien, Tran Ngoc Son, Vu Thi To Linh, "*Reinforcement Learning based Method for Autonomous Navigation of Mobile Robots in Unknown Environments*", Proceedings of the 2020 International Conference on Advanced Mechatronic Systems, Hanoi, Vietnam, December 10 - 13, 2020.

[11] Evan Prianto, MyeongSeop Kim, Jae-Han Park, Ji-Hun Bae, and Jung-Su Kim, "*Path Planning for Multi-Arm Manipulators Using Deep Reinforcement Learning: Soft Actor–Critic with Hindsight Experience Replay*", Sensors, Published: 19 October 2020.

[12] Deepak Ramachandran, Rakesh Gupta, "*Smoothed Sarsa: Reinforcement Learning for Robot Delivery Tasks*", 2009 IEEE International Conference on Robotics and Automation, Kobe, Japan, May 12-17, (2009).

[13] M. U. KHAN, Mobile Robot Navigation Using Reinforcement Learning in Unknown Environments, BALKAN JOURNAL OF ELECTRICAL & COMPUTER ENGINEERING, Vol. 7, No. 3, July 2019 (2019).

[14] G. A. Cardona, C. Bravo, W. Quesada, D. Ruiz, M. Obeng, X. Wu, and J. M. Calderon, *Autonomous Navigation for Exploration of Unknown Environments and Collision Avoidance in Mobile Robots Using Reinforcement Learning*, Conference Paper, April 2019, DOI: 10.1109/ SoutheastCon42311.2019.9020521, (2020).

[15] Luis V. Calderita, Araceli Vega, Sergio Barroso-Ramírez, Pablo Bustos and Pedro Núñez, *Designing a Cyber-Physical System for Ambient Assisted Living: A Use-Case Analysis for Social Robot Navigation in Caregiving Centers*, pp 2-24, Sensor. (2020).

[16] A. D. Pambudi, T. Agustinah and R. Effendi, "*Reinforcement Point and Fuzzy Input Design of Fuzzy Q-Learning for Mobile Robot Navigation System*," 2019 International Conference of Artificial Intelligence and Information Technology, 2019.

Control design of an UAV-Q based on feedback linearization and optimum modulus methods

Tran Duc Chuyen
University of Economics-
Technology for Industries
Hanoi, Vietnam

Hoang Van Huy
Hanoi University of Industry
Hanoi, Vietnam

Tung Lam Nguyen
Hanoi University of Science and Technology
Hanoi, Vietnam
lam.nguyentung@hust.edu.vn

Abstract—In the paper, we present the formulation of quadrotor control loops that are based on a decomposition into a cascade structure and the use of feedback linearization and optimum modulus methods to determine controller parameters. The dynamic model used in this paper considers the dynamics of the propeller rotor drive systems. The propeller rotor drive systems are considered as a linear actuated system. After the synthesizing of the controllers is completed, the system is simulated in MATLAB/Simulink. The results from this work can be useful for the development of autonomous algorithms for UAV - Q (Unmanned Aerial Vehicle - Quadrotor). The research results serve as the basis for control algorithms development for other similar systems.

Index Terms—GameUAV-Q, propeller rotor drive system, feedback linearization, modulus optimum.

I. Introduction

IN THE recent years, UAV-Q has been an active research topic because of their broad applications, especially in the field of military and media services. Among many types of UAV-Q have been widely used because of their advantages such as they have simple structure, compact size, etc. Despite of having simple and symmetric structures, the dynamic models of quadrotor are nonlinear ones. Therefore, the control of these types of UAV requires advanced techniques in order to get good control quality. There have been a lot of studies on the control system design of such UAVs. Some of them can be listed such as the use of PID controller based on linearized models of quadrotor [1, 5, 6], or a number of other approaches that use the sliding mode control, backstepping [11] or robust control $H_\infty 1$ [5]. Additionally, most of the previous works have mainly considered the control inputs to quadrotor as the forces or moments, neglecting the dynamics of the rotors that driving the propellers. In fact, the dynamics of the propeller rotor drive systems should be taken into account. This, of course, will increase the complexity of the UAV-Q system, [6, 13, 14].

In this paper, we present the synthesis of position and attitude controllers by breaking down the system into a cascade structure and the use of the feedback linearization and optimum modulus methods to determine the controllers for the quadrotor control loops. The dynamic models of quadrotor take into account the dynamics of the propeller rotor drive system.

II. The dynamic Model of UAV-Q

The dynamics of a quadrotor is presented in [5]. Earth inertial frame (E frame) and body frame (B frame) whose origin is chosen the as quadrotor center of mass are shown in Fig. 2.
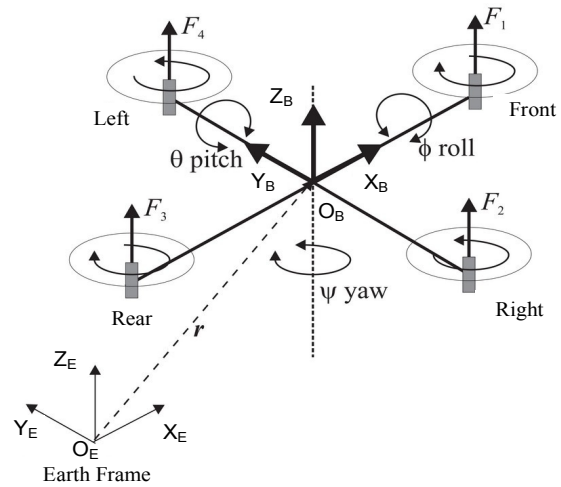
Fig. 1. The system coordinate of unmanned aerial vehicle quadrotor

The dynamics of a quadrotor [2, 5, 7, 8, 11] is described as follows:

$$\begin{bmatrix} m I_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & I \end{bmatrix} \begin{bmatrix} \dot{V}^B \\ \dot{\omega}^B \end{bmatrix} + \begin{bmatrix} \omega^B \times (m V^B) \\ \omega^B \times (I\ \omega^B) \end{bmatrix} = \begin{bmatrix} F^B \\ \tau^B \end{bmatrix} \quad (1)$$

where, $I_{3\times3}$ is the [3x3] identity matrix, $V^B_{[m.s^{-1}]}$, $\dot{V}^B_{[m.s^{-2}]}$ are the quadrotor linear velocity and acceleration vector expressed in the B-frame, while $\omega^B_{[rad.s^{-1}]}$, $\dot{\omega}^B_{[rad.s^{-2}]}$ is the quadrotor angular velocity and acceleration expressed in the B-frame, $F^B_{[N]}$ is the quadrotor forces vector with respect to B-frame and $\tau^B_{[N\,m]}$ is the quadrotor moment vector expressed in the B-frame. A generalized force vector Λ is defined as:

$$\Lambda = \begin{bmatrix} F^B & \tau^B \end{bmatrix}^T = \begin{bmatrix} F_x & F_y & F_z & \tau_x & \tau_y & \tau_z \end{bmatrix}^T \quad (2)$$

Equation (1) is rewritten in a matrix form as:

$$M\,\dot{\upsilon} + C(\upsilon)\upsilon = \Lambda \quad (3)$$

where $\dot{\upsilon}$ is generalized acceleration vector, M_B is inertia matrix, and $C_B(\upsilon)$ is Coriolis-centripetal matrix. The dynamic equations of quadrotor [3, 5] are given as follows:

$$\begin{cases} \ddot{X} = (\sin\psi\sin\phi + \cos\psi\sin\theta\cos\phi)U\,/m \\ \ddot{Y} = (\cos\phi\sin\theta\sin\psi - \sin\phi\cos\psi)U\,/m \\ \ddot{Z} = -g + \cos\phi\cos\theta)U\,/m \\ \ddot{\phi} = \dfrac{(I-I)}{I}\dot{\psi}\dot{\theta} - \dfrac{J}{I}\dot{\theta}\Omega + \dfrac{lb}{I}(-\Omega+\Omega) \\ \ddot{\theta} = \dfrac{(I-I)}{I}\dot{\psi}\dot{\phi} + \dfrac{J}{I}\dot{\phi}\Omega + \dfrac{lb}{I}(-\Omega+\Omega) \\ \ddot{\psi} = \dfrac{(I-I)}{I}\dot{\phi}\dot{\theta} + \dfrac{d}{I}(-\Omega+\Omega-\Omega+\Omega) \end{cases} \quad (4)$$

where the propellers' speed inputs U_1, U_2, U_3 and U_4 with respect to B-frame are given as:

$$\begin{cases} U_1 = b(\Omega_1^2 + \Omega_2^2 + \Omega_3^2 + \Omega_4^2) \\ U_2 = lb(-\Omega_2^2 + \Omega_4^2) \\ U_3 = lb(-\Omega_1^2 + \Omega_3^2) \\ U_4 = d(-\Omega_1^2 + \Omega_2^2 - \Omega_3^2 + \Omega_4^2) \\ \Omega = -\Omega_1 + \Omega_2 - \Omega_3 + \Omega_4 \end{cases} \quad (5)$$

where, U_1 is responsible for the X, Y, Z coordinates of the quadrotor and their rates of change. U_2, U_3 and U_4 are responsible for the roll (ϕ), pitch (θ), and yaw (ψ) rotations and their rates of change. This model can be written in state space form $\dot{\mathbf{X}} = f(\mathbf{X}, \mathbf{U})$, where $\mathbf{U}^T = [U_1, U_2, U_3, U_4]$ is the input variables and $\mathbf{X}^T = (\dot{X}, \dot{Y}, \dot{Z}, \varphi, \theta, \psi, \dot{\varphi}, \dot{\theta}, \dot{\psi}, X, Y, Z)$ with:

$X_1 = \dot{X}$	$X_4 = \varphi$	$X_7 = \dot{\varphi}$	$X_{10} = X$
$X_2 = \dot{Y}$	$X_5 = \theta$	$X_8 = \dot{\theta}$	$X_{11} = Y$
$X_3 = \dot{Z}$	$X_6 = \psi$	$X_9 = \dot{\psi}$	$X_{12} = Z$

Therefore, equations (4) are rewritten into (6). In the state space model (6), one can see the common model in the works on quadrotor control with control input U_1, U_2, U_3 and U_4. Considering the dynamics of the propeller rotor drive systems, an additional system with nonlinear equations must be included which has the following general form:

$$\dot{\mathbf{X}}_{dc} = f(\mathbf{X}_{dc}, \mathbf{U}_{dc}) \quad (6)$$

$$\dot{\mathbf{X}} = \begin{bmatrix} (\sin X_6 \sin X_4 + \cos X_6 \sin X_5 \cos X_4)U_1/m \\ (\cos X_4 \sin X_5 \sin X_6 - \sin X_4 \cos X_6)U_1/m \\ -g + (\cos X_4 \cos X_5)U_1/m \\ X_7 \\ X_8 \\ X_9 \\ \dfrac{(I_{YY} - I_{ZZ})}{I_{XX}} X_9 X_8 - \dfrac{J_{TP}}{I_{XX}} X_8 \Omega + \dfrac{U_2}{I_{XX}} \\ \dfrac{(I_{ZZ} - I_{XX})}{I_{YY}} X_9 X_7 + \dfrac{J_{TP}}{I_{YY}} X_7 \Omega + \dfrac{U_3}{I_{YY}} \\ \dfrac{(I_{XX} - I_{YY})}{I_{ZZ}} X_7 X_8 + \dfrac{U_4}{I_{ZZ}} \\ X_{10} \\ X_{11} \\ X_{12} \end{bmatrix} \quad (7)$$

Where $\mathbf{X}_{dc} \in R^n$ is the vector of the state variables of rotors; $\mathbf{U}_{dc} \in R^m$ is the vector of the input variables to control rotors. The output variables of (6) are the rotor velocities, that are the components $\Omega_1, \Omega_2, \Omega_3, \Omega_4$ in expression (5). The decomposition technique is used to transform the state space equations (5), (7) and (6) into the below subsystems: The first subsystem S_1 includes the different equations, which describes the quadrotor's angular rates.

$$\begin{bmatrix} \dot{X}_7 \\ \dot{X}_8 \\ \dot{X}_9 \end{bmatrix} = \begin{bmatrix} \dfrac{(I_{YY} - I_{ZZ})}{I_{XX}} X_9 X_8 - \dfrac{J_{TP}}{I_{XX}} X_8 \Omega + \dfrac{U_2}{I_{XX}} \\ \dfrac{(I_{ZZ} - I_{XX})}{I_{YY}} X_9 X_7 + \dfrac{J_{TP}}{I_{YY}} X_7 \Omega + \dfrac{U_3}{I_{YY}} \\ \dfrac{(I_{XX} - I_{YY})}{I_{ZZ}} X_7 X_8 + \dfrac{U_4}{I_{ZZ}} \end{bmatrix} \quad (8)$$

The UAV - Quadrotor's Euler angles can be calculated by simply integrating $\dot{X}_7, \dot{X}_8$ and $\dot{X}_9$. The second subsystem includes the different equations, which describe the velocities of UAV-Q inputs of the subsystem are Euler angles and variable U_1.

$$\begin{bmatrix} \dot{X}_1 \\ \dot{X}_2 \\ \dot{X}_3 \end{bmatrix} = \begin{bmatrix} (\sin X_6 \sin X_4 + \cos X_6 \sin X_5 \cos X_4)\dfrac{U_1}{m} \\ (\cos X_4 \sin X_5 \sin X_6 - \sin X_4 \cos X_6)\dfrac{U_1}{m} \\ -g + (\cos X_4 \cos X_5)\dfrac{U_1}{m} \end{bmatrix} \quad (9)$$

The third subsystem is described by system of nonlinear different equations (7), that describes the propeller rotor drive systems of quadrotor.

III. THE CONTROLLER DESIGN

According to the paper's approach, structure of quadrotor control system includes the propeller rotor speed control loop, the attitude loop and the position control loop. In this control architecture, the inner attitude control loop faster dynamics compared to the outer loop responses. In this paper a linear model of the propeller rotor drive systems depicted as:

$$W_m = \frac{\Omega_m}{U_m} = \frac{K_m}{s(T_1 s + 1)(T_2 s + 1)} \qquad (10)$$

where Ω_m is the rotor speed, U_m is the voltage input. The controller synthesis is based on the optimum modulus method [8, 15, 16] and select factor $a = 2$, a desired transfer function of open loop is defined by below expression:

$$W_{ol} = \frac{1}{2T_1 s(T_1 s + 1)} \text{ then}$$

$$W_{cm} = \frac{W_{ol}}{W_m} = \frac{1}{2T_1 s(T_1 s + 1)} \cdot \frac{s(T_1 s + 1)(T_2 s + 1)}{K_m} = \frac{T_2 s + 1}{2K_m T_1} \qquad (11)$$

According to [3], [10], the transfer function of the controller (C0) in this case is determined by equation (11). This is a Proportional-Derivative (PD) controller.

A. Attitude control design

It is assumed that the gyroscopic terms are ignored. The attitude loop stabilizes the Euler angles following a desired vector $(\phi_d, \theta_d, \psi_d)$ or (X_{4d}, X_{5d}, X_{6d}). The subsystem is described in (7) is of nonlinear system. Applying feedback linearization [1, 2] and [11] the following linear system can be obtained

$$\begin{cases} U_2 = f_2(X_7, X_8, X_9) + U_2^* \\ U_3 = f_3(X_7, X_8, X_9) + U_3^* \\ U_4 = f_4(X_7, X_8, X_9) + U_4^* \end{cases} \qquad (12)$$

Where U_2^*, U_3^*, U_4^* are new control variables. Substituting (12) into the equation (8) and neglecting the gyroscopic terms, we are received the equation (13).

$$\begin{cases} \dot{X}_7 = I_1 X_9 X_8 + \left(f_2(X_7, X_8, X_9) + U_2^* \right) / I_{XX} \\ \dot{X}_8 = I_2 X_9 X_7 + \left(f_3(X_7, X_8, X_9) + U_3^* \right) / I_{YY} \\ \dot{X}_9 = I_3 X_7 X_8 + \left(f_4(X_7, X_8, X_9) + U_4^* \right) / I_{ZZ} \end{cases} \qquad (13)$$

where $I_1 = (I_{YY} - I_{ZZ})/I_{XX}$, $I_2 = (I_{ZZ} - I_{XX})/I_{YY}$ and $I_3 = (I_{XX} - I_{YY})/I_{ZZ}$. In order to obtain a linear system, the

new control variables U_2^*, U_3^*, U_4^* are selected in the right side of the equation system (13), which becomes a linear system. Toward this end, the following conditions must be fulfilled:

$$\begin{cases} I_1 X_9 X_8 + \left(f_2(X_7, X_8, X_9) + U_2^* \right) / I_{XX} = K_2 X_7 \\ I_2 X_9 X_7 + \left(f_3(X_7, X_8, X_9) + U_3^* \right) / I_{YY} = K_3 X_8 \\ I_3 X_7 X_8 + \left(f_4(X_7, X_8, X_9) + U_4^* \right) / I_{ZZ} = K_4 X_9 \end{cases} \qquad (14)$$

with unknown parameters K_2, K_3, K_4. From above expression, it is shown that

$$\begin{cases} f_2(X_7, X_8, X_9) = I_{XX} \left(K_2 X_7 - I_1 X_9 X_8 \right) \\ f_3(X_7, X_8, X_9) = I_{YY} \left(K_3 X_8 - I_2 X_9 X_7 \right) \\ f_4(X_7, X_8, X_9) = I_{ZZ} \left(K_4 X_9 - I_3 X_7 X_8 \right) \end{cases} \qquad (15)$$

From equation (13) and (15) one can derive a linear equation system (16) as the following:

$$\begin{cases} \dot{X}_7 = K_2 X_7 + U_2^* / I_{XX} \\ \dot{X}_8 = K_3 X_8 + U_3^* / I_{YY} \\ \dot{X}_9 = K_4 X_9 + U_4^* / I_{ZZ} \end{cases} \qquad (16)$$

and

$$\begin{cases} U_2 = K_2 X_7 - I_1 X_9 X_8 + U_2^* \\ U_3 = K_3 X_8 - I_2 X_9 X_7 + U_3^* \\ U_4 = K_4 X_9 - I_3 X_7 X_8 + U_4^* \end{cases} \qquad (16)$$

To determine the coefficients K_2, K_3, K_4, we use the following Lyapunov candidate function

$$V(X_7, X_8, X_9) = (X_7^2 + X_8^2 + X_9^2) / 2 \qquad (17)$$

Taking the first time derivative of V results in:

$$\dot{V} = X_7 \dot{X}_7 + X_8 \dot{X}_8 + X_9 \dot{X}_9 V \qquad (18)$$

From equation (16), equation (18) can be rewritten as:

$$\dot{V} = X_7 \dot{X}_7 + X_8 \dot{X}_8 + X_9 \dot{X}_9 = K_2 X_7^2 + K_3 X_8^2 + K_4 X_9^2 \qquad (20)$$

where $K_2, K_3, K_4 < 0$. After performing the linearized transformation of attitude control systems of quadrotor we obtain the following equations:

$$\begin{aligned} \ddot{X}_4 &= K_2 \dot{X}_4 + U_2^* / I_{XX} \\ \ddot{X}_5 &= K_3 \dot{X}_5 + U_3^* / I_{YY} \\ \ddot{X}_6 &= K_4 \dot{X}_6 + U_4^* / I_{ZZ} \end{aligned} \qquad (21)$$

From (22) we can get the transfer function of each channel. In this case we implement for roll motion:

$$W_\phi = \frac{\phi}{U_{c\phi}} = \frac{1/(-I_{XX}K_2)}{s((-1/K_2)s+1)} \qquad (22)$$

The control design for this channel uses optimum modulus method [7, 8, 15, 16]. The the roll angle controller is determined as:

$$W_{c\phi} = K_\phi = 0.5I_{XX}K_2^2 \qquad (23)$$

The other controllers of pitch and yaw channel can be done in the same way, and they are presented in the following expressions: $W_{c\theta} = K_\theta = 0.5I_{YY}K_3^2$; $W_{c\psi} = K_\psi = 0.5I_{ZZ}K_4^2$.

B. Velocity control design

If the attitude control loop is sufficiently fast, it is assumed that the desired roll, pitch, and yaw angles are achieved very fast. Therefore, without loss of generality, the attitude loop can be regarded as a unity gain. According to (8), the position subsystem can be depicted follows

$$\begin{cases} \dot{X}_1 = (\sin X_{6d} \sin X_{4d} + \cos X_{6d} \sin X_{5d} \cos X_{4d})\dfrac{U_1}{m} \\ \dot{X}_2 = (\sin X_{4d} \cos X_{6d} - \cos X_{4d} \sin X_{5d} \sin X_{6d})\dfrac{U_1}{m} \\ \dot{X}_3 = -g + (\cos X_{4d} \cos X_{5d})\dfrac{U_1}{m} \end{cases} \qquad (24)$$

where X_{4d}, X_{5d}, X_{6d} and U_1 are the input variables. These are the nonlinear equations and can be rewritten as follows:

$$\begin{aligned} \dot{X}_1 &= \tilde{U}_1 = f_1(X_{4d}, X_{5d}, X_{6d}, U_1) \\ \dot{X}_2 &= \tilde{U}_2 = f_2(X_{4d}, X_{5d}, X_{6d}, U_1) \\ \dot{X}_3 &= \tilde{U}_3 = f_3(X_{4d}, X_{5d}, X_{6d}, U_1) \end{aligned} \qquad (26)$$

$$\begin{aligned} & & \tilde{U}_1 = n_1(X_{1d} - X_1) \\ \text{and } & & \tilde{U}_2 = n_2(X_{2d} - X_2) \\ & & \tilde{U}_3 = n_3(X_{3d} - X_3) \end{aligned} \qquad (25)$$

with newly defined input variables $\tilde{U}_1, \tilde{U}_2, \tilde{U}_3$ are selected for a proportional controller in the following form (27). Here, the parameters of the controllers n_1, n_2, n_3 could be selected such that the outer loop is sufficiently fast compared to attitude control loop. These transformed input variables is used to calculate the real input variables X_{4d}, X_{5d}, X_{6d} and U_1 by evaluating (25). It is noted that any desired velocity vector can be reached regardless any yaw rotation, so the system of equations (25) can be simplified as follows:

$$\tilde{U}_1 = \sin X_{5d} \cos X_{4d} \frac{U_1}{m}$$

$$\tilde{U}_2 = \sin X_{4d} \frac{U_1}{m} \qquad (26)$$

$$\tilde{U}_3 = -g + \cos X_{4d} \cos X_{5d} \frac{U_1}{m}$$

These above equations can be solved analytically by applying the considering relation:

$$\begin{cases} \lambda = \sin X_{4d} & \Rightarrow & \cos X_{4d} = \pm\sqrt{1-\lambda^2} \\ \gamma = \sin X_{5d} & \Rightarrow & \cos X_{5d} = \pm\sqrt{1-\gamma^2} \end{cases} \qquad (27)$$

Substitutin λ, γ into equation (27), we obtain the following relations:

$$\begin{cases} \tilde{U}_1 = \gamma \dfrac{U_1}{m}.\pm\sqrt{1-\lambda^2} \\ \tilde{U}_2 = \lambda \dfrac{U_1}{m} \\ \tilde{U}_3 = -g + \left(\pm\sqrt{1-\lambda^2}.\pm\sqrt{1-\gamma^2}.\dfrac{U_1}{m}\right) \end{cases} \qquad (30)$$

$$\Rightarrow \begin{cases} \gamma = \pm \dfrac{1}{\sqrt{\left[(g+\tilde{U}_3)/\tilde{U}_1\right]^2 + 1}} \\ U_1 = \pm m\sqrt{\tilde{U}_1^2/\gamma^2 + \tilde{U}_2^2} \\ \lambda = \tilde{U}_2 \dfrac{m}{U_1} \end{cases} \qquad (28)$$

If $\tilde{U}_1 \neq 0$, we solve the equation system (26) and obtain the following solution λ, γ and U_1. U_1 is always positive, so from (27) for U_1 we obtain the unique solution:

$$U_1 = m\sqrt{(\tilde{U}_1^2/\lambda^2) + \tilde{U}_2^2} \qquad (29)$$

λ is unique, hence $X_{4d} = \arcsin \lambda$ is uniquely obtained in $[\pm\pi/2]$. In the similar fashion, it can be show that $X_{5d} = \arcsin \gamma$, but γ and X_{5d} could be positive or negative. This is explained in the following:

$$\tilde{U}_1 = -\cos X_{4d} \sin X_{5d}.U_1/m \qquad (30)$$

In (30), the first term is positive in the interval $[\pm\pi/2]$ and the last term (U_1/m) is also positive. Therefore, X_{5d} is negative if $\tilde{U}_1$ is positive and vice versa.

C. The Synthesis of position control system

The design of position controller is implemented after the inner-loop controllers are synthesized. The way to

design for controller X, Y, Z is the same. We assume that the velocity loop is of second order. In this section we synthesize the controller for altitude channel Z. Therefore, the transfer function of Z channel has a form:

$$W_{pz} = \frac{Z}{U_{cz}} = \frac{K_z}{s(T_{z1}s+1)(T_{z2}s+1)} \tag{31}$$

According to the optimum modulus method [3], [10], we can obtain the transfer function of Z channel controller in the following form:

$$W_{cz} = \frac{T_{z2}s+1}{2K_z T_{z1}} \tag{32}$$

IV. NUMERICAL SIMULATIONS

From equations describe the quadrotor dynamics (4) and controllers for control loop are synthesized above with quadrotor parameters simulation, we implement the numerical simulation via Matlab. The parameters of controller in loops are chosen as: $K_P=3$, $K_D=0.5$, $T_1 = 0.01$, $T_2 = 0.1$, $K_1=K_2=K_3=-80$; Coefficients of velocity controller: $n_1=n_2=n_3=1$, $K_P=0.25$, $K_D=0.1$. The thrust factor b = $2,92.10^{-6}$ kg.m, total rotational moment of inertia around the propeller axis $J_{TP} = 3,36.10^{-5}$ kg.m^2, the aerodynamic drag factor d = $1,1.10^{-7}$ kg.m^2, mass of quadrotor m = 0.5 kg, the body moment of inertia $I_{XX} = I_{YY} = 4,85.10^{-3}$ kg.m^2 and $I_{ZZ} = 8,81.10^{-3}$ kg.m^2, the center-to-center distance between the quadrotor and the propeller l = 0,24 m.

Case 1: The desired inputs $\psi_d = 0$ rad, $X_d = 0$ [m], $Y_d = 0$ m, $Z_d = 15$ m: as shown in the picture Fig 5 and 6 presents the results of X, Y, Z and Euler angles responses.
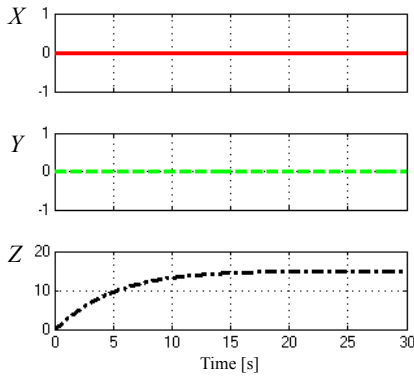
Fig. 2.
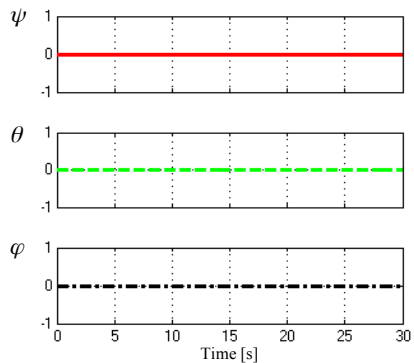
Fig. 3.

Case 2: The desired inputs $\psi_d = 0$ [rad], $X_d = 10$ [m], $Y_d = 0$ [m], $Z_d = 15$ [m]: Fig 7 and 8 present the results of value X, Y, Z and Euler angles responses.
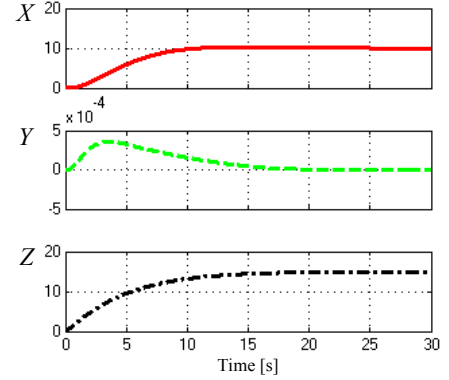
Fig. 4.
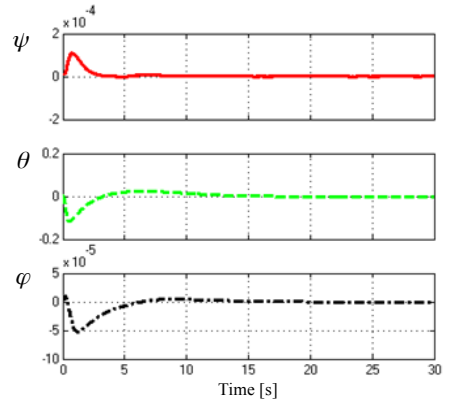
Fig. 5.

From the simulation results, we see that all the desired state reached along channel Z, X, Y with fast response. These results prove that the controllers work well with good tracking performance. Compared with previous studies, as shown in [9, 13], the research results of the paper are much better than previous studies.

CONCLUSIONS

The paper has presented the synthesis results for quadrotor control loops (position control loop, attitude control loop, propeller rotor control loop), which used the feedback linearization and optimum modulus methods. The dynamic model of the quadrotor is derived and numerically implemented. Through the simulation, the nonlinear vehicle control system is verified and its efficiency is demonstrated. The results, which obtained from this paper, contribute to the development of algorithms for autonomous UAV-Q.

REFERENCES

[1] A. J. Fossard and D. Normand-Cyrot (Eds.). "*Nonlinear Systems*", Vol. 3: Control, Springer, (1996).
[2] A. Isidori, "*Nonlinear Control Systems*", 3rd Edition, Springer, (1995).
[3] S.-K. Kim, K.-G. Lee, and K.-B. Lee, "*Singularity-free adaptive speed tracking control for uncertain permanent magnet synchronous motor*", IEEE Transactions on Power Electronics, vol. 31, no. 2, pp. 1692-1701, (2016).
[4] J. J. Slotine and W. Li, "*Applied Nonlinear Control*". Englewood Cliffs, NJ: Prentice-Hall, (1991).

[5] J. Ghandour, S. Aberkane, J-C. Ponsart, "*Feedback Linearization approach for Standard and Fault Tolerant control: Application to a Quadrotor UAV Testbed*", Journal of Physics: Conference Series 570, (2014).

[6] Utkin V., Guldner J., Shi J., *Sliding Mode Control in Electromechanical Systems*, CRC Press LLC, (1999).

[7] Pedro Castillo, Rogelio Lozano and Alejandro E.Dzul, *"Modelling and Control of Mini-Flying Machines"*, Springer, Compiègne, France, (2004).

[8] Ali Emadi, "*Advanced Electric Drive Vehicles"*, CRC Press is an imprint of Taylor & Francis Group, an Informa business, Springer International Publishing; USA, (2016).

[9] S. Bouabdallah, P. Murrieri, and R. Siegwart, "*Design and control of an indoor micro quadrotor*". In Robotics and Automation, Proceedings. ICRA'04. IEEE (2004).

[10] Hyeonbeom Lee and H. Jin Kim, "*Trajectory Tracking Control of Multirotors from Modelling to Experi- ments: A Survey*", International Journal of Control, Automation and Systems, pp. 1-12, (2017).

[11] Tommaso Bresciani, "*Modelling, Identification and Control of a Quadrotor Helicopter*", master thesis, October (2008).

[12] László Keviczky, Ruth Bars, Jenő Hetthéssy, "*Csilla Bányász, Control Engineering: MATLAB Exercises*", Springer Nature Singapore Pte Ltd, USA ISSN 1439-2232, (2019).

[13] M. Navabi, H. Mirazei, "*Robust optimal adaptive trajectory tracking control of quadroto helicopter*", Lat. Am. J. solids struct. vol 14 No 6 Rio de Janeiro June (2017),

[14] Luo, Shaohua, Gao, Ruizhen, *Chaos control of the permanent magnet synchronous motor with time-varying delay by using adaptive sliding mode control based on DSC*, Volume 355, Issue 10 Pages 4147-4163, July (2018).

[15] Ключев В. И, Теория электропривода Учебник для вузов Изд. Энергоатомиздат, 3-е (2001)

[16] Б. К. Чемоданов - Следящие приводы Т1, 2.- М.: Изд. МГТУ им Баумана (1999).

Design of the mobile-robot-based surveillance system on university campuses to reduce the effects of COVID-19 pandemic

Quy Xuan Dao
Eastern International University
quy.dao@eiu.edu.vn

Viet Thanh Cao
Eastern International University
viet.cao@eiu.edu.vn

Duc Ngoc Trinh
Eastern International University
duc.trinh@eiu.edu.vn

Linh Thi Kim Linh
Eastern International University
linh.ngo@eiu.edu.vn

Abstract—**This paper introduces a new surveillance system to detect wearing-mask and monitor social distancing and body temperature to reduce the effects of the COVID-19 pandemic on university campuses. This surveillance system was designed and implemented to SunBot, an autonomous mobile, based on hardware including Jetson Nano, camera, and thermal camera, and open-source software including OpenCV, YOLOv3, MobilNetv2, TensorFlow, Keras. Both hardware and software are basic, simple to deploy, and affordable cost. Experimental results showed that the surveillance system deployed on university campuses to reduce the effects of the COVID-19 pandemic worked as expected.**

Index Terms—**mask detection, social distancing monitoring, body temperature monitoring, covid-19**

I. Introduction

WE CURRENTLY face the COVID-19 pandemic due to the agent SARS-COV-2 virus which is recognized as a global pandemic by the World Health Organization [1] because of its effects and damages. The COVID-19 virus can be fatal because it damages the respiratory human [2]. As of December 15, 2020, the COVID-19 pandemic has been crossing over 200 countries in the world, there were 73,143,329 cases, 1,627,046 deaths, 51,263,541 cures [3]. In which, the United States is the most affected country with 16,909,295 cases, 307,853 deaths, 9,828,401cures. Everyone is affected and has a high risk of death [2]. In the UNESCO report [4], the COVID-19 pandemic has affected billions of students, and thousands of schools. According to UNESCO monitoring, 1,576,021,818 learners are currently affected due to school closures in response to the pandemic, impacting about 91.3% of the total number of learners in 188 countries. In order to reduce the spread of COVID-19, the government decided to temporarily close educational institutions. But this solution is only according to a short-term solution. Teaching and learning must continue when the vaccine is currently developed and tested.

This paper aims to highlight the benefits of the surveillance system in fighting COVID-19 on university campuses. The World Health Organization recommended people should wear a face mask in public areas because it is one of the effective protection methods since the COVID-19 pandemic becomes a global health crisis. In order to determine wearing a mask, the authors [5] proposed a face mask detection model based on a hybrid model using deep and classical machine learning including two components: (1) Resnet50; (2) Decision trees, Support Vector Machine (SVM). This method achieved more than 99% testing accuracy. Similarly to [5], the authors [6] proposed mask detection based on YOLOv2 network with ResNet-50. This result showed the highest average precision percentage of 81% by using the adam optimizer. In [7], the author presented a simplified approach to performing mask detection using the basic Machine Learning (ML) packages such as TensorFlow, Keras, OpenCV, and Scikit-Learn.

Another author [8] deployed an IoT-based solution that used Arduino Uno to perform body temperature checks based on an infrared sensor or thermal camera, and Raspberry Pi to perform mask detection and social distancing check. In [9], the authors proposed a B5G framework to develop a mass surveillance system to monitor social distancing, mask-wearing, and body temperature based on three deep learning models including ResNet50, Deep tree, and Inceptionv3. In another research, the author [10] presented a method to determine the social distance between two people by using a camera placed on a mobile robot to estimate the distance. The robot was also equipped with a thermal camera to monitor body temperature.

In [11], the author proposed use of the technologies such as IoT, UAVs, Blockchain, AI, and 5G in order to reduce the impact of the COVID-19 pandemic. More detail of style and technology robots in the COVID-19 pandemic, Murply et al. [12] presented the application of robots for COVID-19 response by summarizing 262 reports appearing between March 27 and July 4, 2020. These reports described 203 instances of the actual use of 104 different robot models for the COVID-19 response. In 203 instances, public safety was the largest number of the reported instance (74), clinical care (46),

Fig. 1: Robots in the COVID-19 time ([14]).

quality of life (27), continuity of work and education (22/203), laboratory and supply chain automation (21), and non-clinical care (13). To control robots, teleoperation, automation, and autonomy were reported as 105, 74, and 24, respectively. In other research, the author [13] summarized some exciting and important new research on mobile robots which is deployed in the COVID-19 pandemic time. As shown in Fig. 1, the author [14] presented some different applications of mobile robots in order to fight the COVID-19 pandemic such as population awareness and control, delivery services, disinfection of facilities, transport of material and supplies, etc. In another research, the author [15] concerned about how robotic and autonomous systems and smart wearable complement and support healthcare delivery and the healthcare staff during the COVID-19 pandemic.

Leveraging artificial intelligence, especially deep learning, the implementation of the surveillance system is becoming simpler and more effective. It is possible to implement this system using the available hardware such as Arduino, Raspberry Pi, and Jetson Nano and open-source software including machine learning libraries such as OpenCV, YOLOv3, MobileNetv2, TensorFlow, Keras, etc. To continue learning as well as ensure social distancing, a surveillance system to detect mask-wearing and monitor social distancing and body temperature are expected to deploy on university campuses. In this paper, we present a surveillance system implemented into SunBot (multiple functional mobile robots based on Jetson Nano, depth camera, and thermal camera) that serve as a healthcare assistant to detect mask-wearing and monitor social distancing and body temperature.

This paper is organized as follows: Section 2 describes the SunBot surveillance system that we consider healthcare functions including mask detection, social distancing monitoring, and body temperature monitoring, Section 3 shows the experimental results when we deployed SunBot surveillance system on the Eastern International University (EIU) campus during the social distancing time, and Section 4 are some conclusion as well as future work.

II. SunBot Surveillance System

This section presents SunBot surveillance system to detect wearing-mask as well as monitor social distancing and body temperature. SunBot was designed and illustrated in Fig. 2. SunBot was designed to perform five functions such as (1) delivery, (2) telepresence, (3) guide, (4) healthcare, (5) security, (6) mask detection and (7) social distancing monitoring. The proposed mobile robot contains two main parts: platform

Fig. 2: Full optional SunBot:(1) thermal camera, (2) – (3) cameras, (4) tablet/screen, (5) speaker/mic, (6) hand clean bottle, (7) lidar, (8) depth camera, (9) ultra-sonic senor and (10) Omni wheels.

and tool. Jetson Nano and Microcontroller STM32F4 are used to program and operate the mobile robot based on the ROS framework. The sensors, cameras, and lidar are chosen in order to provide information that helps the mobile robot to move in autonomy. Jet-son Nano, Raspberry Pi, camera and thermal camera perform mask detection, social distancing, and body temperature monitoring based on open-source such as OpenCV, MobileNetv2 [16], YOLOv3 [17], TensorFlow, Keras, etc. In this paper, we only focus on mask detection and social distancing monitoring of SunBot.

A. Autonomous Mobile Module

This module allows SunBot to move on university campuses in autonomy. This module was built based on self-driving-cars technology including computer vision and sensor technologies. We built SunBot platform based on open-source hardware and software: TurtleBot2 [18] and self-driving-cars technology using a convolution neural network (CNN) [19]. We also built Information and Communications Technology (ICT) infrastructure on the EIU campus to support SunBot. Fig. 3 shows hardware and software requirements to build this module. Jetson Nano, Lidar, and RGB-D camera are important hardware while ROS, OpenCV, and CNN are required software. Information from Lidar and RGB-D camera are processed by Jetson Nano based on CNN [19] to control mobile robots. Two sub-modules are navigation (ROS) and avoidance (OpenCV+CNN) that help SunBot to move around university campuses.

B. Virtual Assistant Module

This module allows SunBot to communicate with users based on natural language processing. SunBot can communicate with users by speech or text. In our previous work [20], [21] we developed a virtual assistant for online learning. This system was implemented into SunBot to communicate with users based on healthcare functions. Our virtual
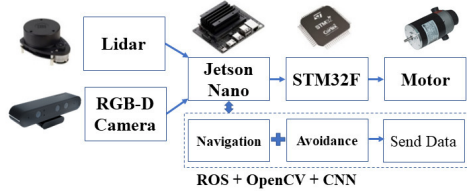
Fig. 3: Hardware and Software for Autonomous Mobile Module.

assistant is a hybrid method that is a combination of rule-based chatbots and AI-based chatbots. Our virtual assistant was developed based on Rasa's "Open-source machine learning framework to automate text- and voice-based conversations" [22]. Fig. 4 shows hardware and software requirements to build the virtual assistant module. We use Jetson Nano, Camera, Speaker, and Monitor as hardware while OpenCV and Google Text-to-Speech (T2S)/Speech-to-Text (S2T) API as software. The face/speech detection module allows SunBot to recognize users while natural language processing helps SunBot to communicate with users via speech and/or text.
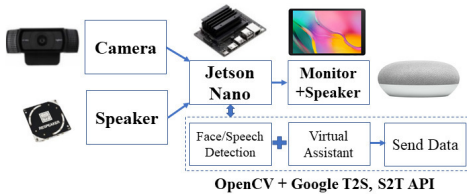
Fig. 4: Hardware and Software for Virtual Assistant Module.

C. Mask Detection Module

During the COVID-19 time, wearing-mask becomes an important way to fight the COVID-19 pandemic. SunBot surveillance system was designed to detect wearing-mask by using OpenCV, and MobileNetv2 [16]. This is one of the simplest ways to detect wearing-mask. There are many ways to detect wearing-mask but the one that we used because of its simplicity and efficiency and accepted price. Fig. 5 shows hardware and software requirements. Hardware includes Jetson Nano, camera, monitor, and speaker while software includes OpenCV, TensorFlow, Keras, and MobileNetv2. The Mask detection module allows us to distinguish whether to wear a mask or not. This function is very important in deploying SunBot surveillance system on university campuses. Furthermore, SunBot surveillance system reminds students to wear a mask on university campuses via virtual assistant module by speech.

D. Social Distancing Monitoring Module

To practice social distancing or estimate the number of the people [23], we need to stay at least 6 feet from other people. The social distancing monitoring module was built based on hardware such as Jetson Nano, camera, monitor, speaker, and software such as OpenCV and YOLOv3 [17]
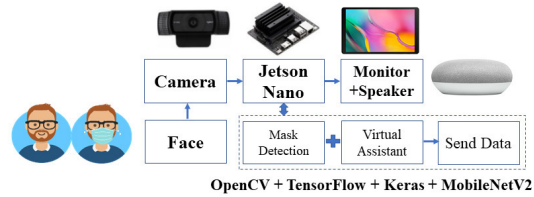
Fig. 5: Hardware and Software for Mask Detection Module.

illustrated in Fig. 6. Universities need to keep social distancing during the COVID-19 time. We estimate a distance between two people by using OpenCV and YOLOv3. We introduce one of the simplest methods to determine social distancing. People are recognized and label by rectangle boxes. The Euclidean distance of the center of these rectangle boxes is considered social distancing. When the social distancing is less than a setup threshold, SunBot surveillance system reminds students as well as requires students to keep a social distancing through three levels of distance as (1) high risk if the distance between two peoples is less than the set-up threshold (ex. within about 6 feet), (2) safe if the distance between two peoples is greater than the setup threshold (ex. within about 6-12 feet), and (3) low risk if the distance between two peoples is large enough (ex. from 12 feet).
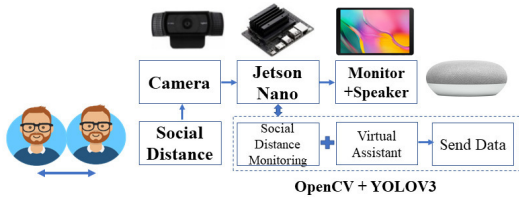
Fig. 6: Hardware and Software for Social Distancing Monitoring Module.

E. Body Temperature Monitoring Module

One of the most common and recognizable symptoms of covid-19 is a fever. Therefore, we need to check body temperature to perform the COVID-19 screening. The COVID-19 pandemic is thought to be spread mainly through close person-to-person contact. Body temperature monitoring by using a thermal camera [24] is the right selection in this situation. Fig. 7 shows hardware and software requirements. We use an MLX90640 thermal camera and Raspberry Pi as hardware and OpenCV as software to monitor body temperature. When the body temperature is greater than a setup threshold, SunBot surveillance system recognizes people at risk of disease.

III. EXPERIMENTAL RESULTS

SunBot surveillance system was deployed on the EIU campus to detect wearing-mask and monitor social distancing and body temperature. Sunbot moves around the EIU campus to perform surveillance functions at gates, libraries, laboratories, and cafeterias.
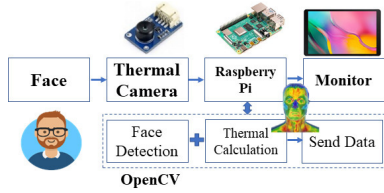
Fig. 7: Hardware and Software for Body Temperature Monitoring Module.

A. Mask Detection

First of all, we choose a dataset in [25] to train a mask detection model. Fig. 8 shows how to wear a mask in the COVID-19 pandemic. When we use this dataset (first case), we could obtain a wrong result illustrated in Fig. 9. In the first case, the training dataset is labeled "mask" containing images of wearing a mask under the nose. This dataset needs to be modified to train the mask detection model. In the second case, the result becomes better when we replace images wearing a mask under the nose with the "without a mask" dataset. Therefore, we use the modified dataset (second case) to train the mask detection model.
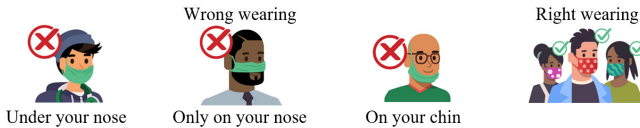
Fig. 8: How to wear a mask.

Fig. 9: Dataset for Mask Detection.

We implemented the mask detection model above to Jetson Nano and camera to detect wearing-mask on the EIU campus. The scenario of mask detection is given as (1) SunBot moves around the EIU campus to detect wearing-mask and capture video; (2) SunBot surveillance system reminds "without mask" cases; (3) SunBot sends information to cloud-management center to store data. Since a high proportion of people comply with wearing-mask that have a substantial impact on COVID-19 transmission on campus, SunBot statistics the number of people who do not wear masks to notify the administrative center. Fig. 10 shows experimental results for mask detection task on the EIU campus. Although people wear different masks

from medical masks to cloth masks, SunBot surveillance system can distinguish between people who wear masks or not.
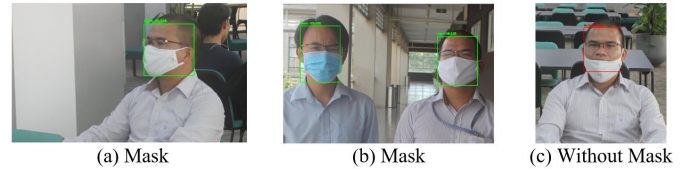
Fig. 10: Mask Detection on the EIU campus.

B. Social Distancing Monitoring

Universities need to keep social distancing between individuals in order to reduce transmission of the COVID-19 spread. SunBot moves around university campuses to ensure everyone that must consistently follow social distancing practices. We implemented the social distancing monitoring model based on YOLOv3 to Jetson Nano and camera to determine social distancing measurement on the EIU campus. Fig. 11 and 12 shows experimental results on the EIU campus. SunBot surveillance system establishes policy and reminders to keep at least 6 feet of space between individuals including three warning levels as high risk, safe, and low risk.
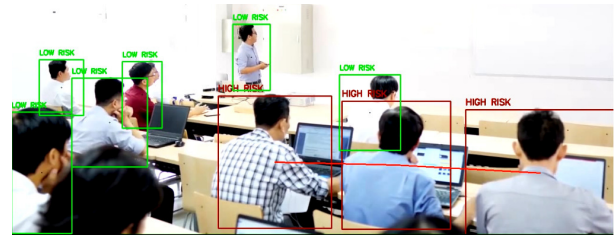
Fig. 11: Social Distancing Monitoring in Classroom.
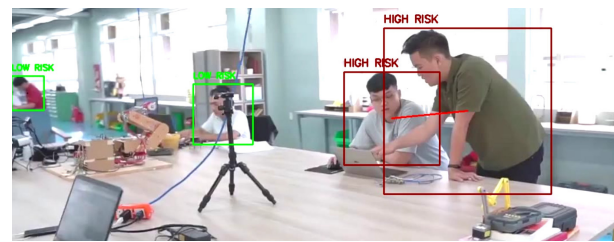
Fig. 12: Social Distancing Monitoring in Laboratory.

C. Body Temperature Monitoring

Fig. 13 illustrates the experimental result of body temperature monitoring based on the MLX90640 thermal camera which captures temperature from -40°C to 300°C. We easily configure a body temperature monitoring alert at 37°C. The scenario of body temperature monitoring is given as follows (1) users move before the thermal camera, (2) the thermal camera checks the users' body temperature, shows information,

and sends data to the cloud management system, (3) SunBot communicates with users about their status and provides health advice and guidance. Furthermore, the virtual assistant also provides Questions and Answers about the COVID-19 pandemic.
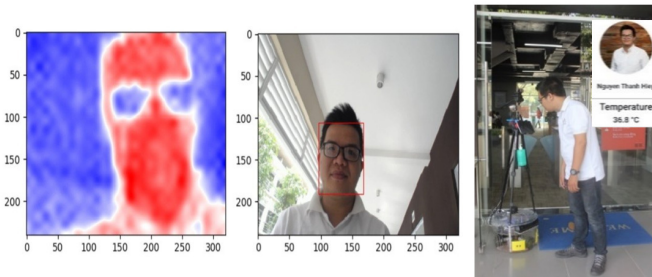
Fig. 13: Body Temperature Monitoring at 36.8^{o}C.

D. Discussion

To achieve the goal of reopening universities as safely and quickly as possible, we need to reduce the spread of COVID-19 in universities and out in the community. In the previous sections, we presented how to detect mask-wearing and monitor social distancing and body temperature on university campuses. Universities can deploy these systems and put together corresponding policies to guarantee continuing learning as (1) encourage behaviors that help reduce the spread of COVID-19, (2) maintain a healthy environment, (3) maintain healthy activities, and (4) be prepared when a coronavirus case is recognized. Students, teachers, and staff must obey fully with steps to protect themselves and others such as the correct use of masks, social distancing, and hand hygiene.

IV. CONCLUSION

In this paper, a surveillance system including face mask detection and social distancing monitoring, and body temperature monitoring was presented. The proposed surveillance system was designed and implemented to the SunBot autonomous mobile robot. This surveillance system was developed based on open-software including OpenCV, MobileNetv2, YOLOv3, TensorFlow, Keras, and affordable hardware including Jetson Nano, camera, and thermal camera. Experimental results demonstrated that the proposed surveillance system is one of the performance ways to keep physical distancing as well as ensure continuous learning in the COVID-19 time. In the future, we will develop a complete surveillance system in small urban cities.

ACKNOWLEDGEMENT

REFERENCES

[1] WHO. Who director-general's opening remarks at the media briefing on covid-19 - 11 march 2020. [Online]. Available: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020

[2] L. Meng, F. Hua, and Z. Bian, "Coronavirus disease 2019 (covid-19): emerging and future challenges for dental and oral medicine," *Journal of dental research*, vol. 99, no. 5, pp. 481–487, 2020.

[3] Worldometer. Coronavirus update (live): 87,976,426 cases and 1,898,009 deaths from covid-19 virus pandemic - worldometer. [Online]. Available: https://www.worldometers.info/coronavirus/

[4] UNESCO. Coronavirus update (live): 87,976,426 cases and 1,898,009 deaths from covid-19 virus pandemic - worldometer. [Online]. Available: School closures caused by Coronavirus (Covid-19)

[5] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, "A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the covid-19 pandemic," *Measurement*, vol. 167, p. 108288, 2021.

[6] ——, "Fighting against covid-19: A novel deep learning model based on yolo-v2 with resnet-50 for medical face mask detection," *Sustainable Cities and Society*, vol. 65, p. 102600, 2021.

[7] A. Das, M. W. Ansari, and R. Basak, "Covid-19 face mask detection using tensorflow, keras and opencv," in *2020 IEEE 17th India Council International Conference (INDICON)*. IEEE, 2020, pp. 1–5.

[8] N. Petrovic and j. v. p. y. Kocic, DJ, "Iot-based system for covid-19 indoor safety monitoring."

[9] M. S. Hossain, G. Muhammad, and N. Guizani, "Explainable ai and mass surveillance system-based healthcare framework to combat covid-i9 like pandemics," *IEEE Network*, vol. 34, no. 4, pp. 126–132, 2020.

[10] A. J. Sathyamoorthy, U. Patel, Y. A. Savle, M. Paul, and D. Manocha, "Covid-robot: Monitoring social distancing constraints in crowded scenarios," *arXiv preprint arXiv:2008.06585*, 2020.

[11] V. Chamola, V. Hassija, V. Gupta, and M. Guizani, "A comprehensive review of the covid-19 pandemic and the role of iot, drones, ai, blockchain, and 5g in managing its impact," *Ieee access*, vol. 8, pp. 90 225–90 265, 2020.

[12] R. R. Murphy, V. B. M. Gandudi, and J. Adams, "Applications of robots for covid-19 response," *arXiv preprint arXiv:2008.06976*, 2020.

[13] D. Feil-Seifer, K. S. Haring, S. Rossi, A. R. Wagner, and T. Williams, "Where to next? the impact of covid-19 on human-robot interaction research," 2020.

[14] M. Cardona, F. Cortez, A. Palacios, and K. Cerros, "Mobile robots application against covid-19 pandemic," in *2020 IEEE ANDESCON*. IEEE, 2020, pp. 1–5.

[15] M. Tavakoli, J. Carriere, and A. Torabi, "Robotics, smart wearable technologies, and autonomous intelligent systems for healthcare during the covid-19 pandemic: An analysis of the state of the art and future vision," *Advanced Intelligent Systems*, vol. 2, no. 7, p. 2000071, 2020.

[16] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.

[17] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[18] ROS-Fundation. Turtlebot2. [Online]. Available: https://www.turtlebot.com/turtlebot2/

[19] M. Bojarski, D. Del Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang *et al.*, "End to end learning for self-driving cars," *arXiv preprint arXiv:1604.07316*, 2016.

[20] T.-M.-T. Nguyen, T.-H. Diep, B.-N. Bac, N.-B. Le, and X.-Q. Dao, "Design of online learning platform with vietnamese virtual assistant," in *2021 6th International Conference on Intelligent Information Technology*, 2021, pp. 51–57.

[21] X.-Q. Dao, N.-B. Le, and T.-M.-T. Nguyen, "Ai-powered moocs: Video lecture generation," in *2021 3rd International Conference on Image, Video and Signal Processing*, 2021, pp. 95–102.

[22] RasaHQ. Open source machine learning framework to automate text- and voice-based conversations: Nlu, dialogue management, connect to slack, facebook, and more - create chatbots and voice assistants. [Online]. Available: https://github.com/RasaHQ/rasa

[23] K. Przybylek and I. Shkroba, "Crowd counting´ a la bourdieu: Automated estimation of the number of people," *Computer Science and Information Systems*, no. 00, pp. 29–29, 2020.

[24] S. Ansari and S. Salankar, "An overview on thermal image processing." in *RICE*, 2017, pp. 117–120.

[25] Chandrikadeb. Face-mask-detection: Face mask detection system based on computer vision and deep learning using opencv and tensorflow/keras. [Online]. Available: https://github.com/chandrikadeb7/Face-Mask-Detection

Time Series Forecasting with Data Transform and Its Application in Sport

Hao Do*†, Duc Chau*†, Thuc Cai‡, Han Lam‡, Dat Nguyen‡, Tuong Lam‡, Son Tran*†

* University of Science, HCM City, Vietnam
† Vietnam National University, HCM City, Vietnam
‡ OLLI Technology JSC, HCM City, Vietnam
Corresponding author: hao@olli-ai.com

Abstract—Forecasting time series data is an exciting challenge. Although being complex, this is a high potential for industrial use. One of the most significant gaps in the forecasting process is the quality of data representation, especially with the time-series data. This paper proposes an effective method using an integral transform that can show hidden information of the time series data. The integral transform exploits data as a composition of many basic functions and then use this set to present the data. Mathematically, this transform converts the data into another space with another feature, showing many properties hidden in the original form. The experimental result demonstrates our suggestion can learn the transformation rules and then can be applied for many applications.

Index Terms—time series forecasting, integral transform, periodic data

I. Introduction

MANY applications such as predicting gold price or predicting stock price can be formed as time series forecasting problems [1]. These applications' data are usually sequences of numbers or vectors corresponding with time. The fluctuated range of data can be vary depending on the kind and properties of data. In a time series forecasting problem, the primary mission is to predict the next value in the series based on the values in the past.

Time series forecasting [2] [3] is very useful in many real-world applications including economic, management, environment, healthcare, stock prediction [4] [5], etc. This problem means constructing a model from available data to predict the next values in the near future. If the data can be analyzed with rules, the predicted values will be approximated with the truth or the value in the real world. These results should be applied to the world to adjust human behaviours and then gain a better state. The results of the forecasting model provide a quantitative method to measure the impact of past and present on the future.

Forecasting is not an easy problem [6]. There are at least two agents that make the predicting process is so complicated. The first one is whether the predicted value can be predicted. It cannot predict that whether an Adidas store will sell off tomorrow. That event mostly depends on the store owner's decision or the store policy. Many forecasting methods focus on finding the rules or the trends in the data to make a guest, so these methods could not work well with the random data or emotion-based data. The second aspect is the ability to

collect enough data to build the predicting model. Data should be collected enough in terms of amount and related context. While the amount can be satisfied, the data context is much more challenging to collect and organize. How many things affect the gold price is a tricky question. On the other hand, the way we structure a database with its own contexts is complicated. These two reasons, especially the second one, mainly cause the difficulty of time series forecasting.

We propose a simple method to predict the following values for time series data in this work. Our approach focuses on the comparative periodic data, which is a common data type in real-world applications. With this type, data is followed by some rules or trends. Remarkably, these data are repeated during the time. We exploit these rules with an integral transform [7]. The transform, in nature, is a mathematical way to separate an occasional series into many periodic simple series called basic. The set of basic functions can present all the rules and trends in the data, and then apply these factors to compute the values in the future. With a cheap computing cost, this method can forecast the next value for the nearly repeated data very efficiently.

The remaining of this paper includes four sections. Section two will present the general approach for the problem of time series forecasting. The next section shows our proposed method. This section presents our transform and its application to calculate the next value in the series. Section four is the application of our data transform to sports data. Particularly, we apply the transform method to sports data to exploit its trend. The final section will summarize our work and then make a conclusion.

II. A general framework for time series forecasting

To deal with a forecasting problem, there are three main stages that we should complete:

- Defining the problem and collecting data: Identifying what the model will predict, data volume, data properties, and how to collect data.
- Building model: Identifying the input and output of the model, what the main processing model should be.
- Predicting the next values: Specifying the way to apply the constructed model to predict the values in the near future.

In the three stages above, the second stage, building model, can be the most important step. The accuracy and performance of the final stage mostly depend on the quality of the built model in this stage. Constructing a right and efficient model will learn and present the data rules very well. This is fundamental for a good prediction.

III. DATA TRANSFORM AND ITS APPLICATION TO DATA FORECASTING

In many applications, the data's current form is difficult to analyse or process. We need to transform the data into another form for more informative data. There are many techniques for data processing in general, and integral transform is one of the most popular choices.

Integral transform [7] [8] is a mathematics tool to transform data from this form to another form via an integral operator. Let $f(t)$ and $K(t, u)$ denote the original data, and the kernel, the integral with the kernel $K(\cdot)$ can be defined as:

- Forward transform:

$$Tf(u) = \int_{t_1}^{t_2} f(t)K(t, u)dt \qquad (1)$$

- Inverse transform:

$$f(t) = \int_{u_1}^{u_2} Tf(u)K^{-1}(t, u)du \qquad (2)$$

Forward transform is used to explore the hidden properties of the data in the other space. The inverse transform is applied to reconstruct the data after being processed in the other space. Generally, we use all these transforms in reality for many purposes.

IV. TRANSFORM AND FORECASTING IN SPORT DATA

We use the integral transform for analyzing the sport data. Particularly, we consider some football teams including Barcelona, Bayern Munich, and Manchester City, which Pep Guardiola coaches. Using the proposed method, we try to predict the result of Pep's team via predicting the average conceded goal per match.

A. Analyzing data

Pep Guardiola is a genius in football, especially with the role of a coach. His teams always have a clear brand identity due to their playing styles. Pep inherits the style from the traditional Netherland style and develops it to reach a new level. Pep and his teams won and lost due to his strategy for every match.

Pep usually spends from three to five years with each team. In the final years in the teams, his achievements are so bad, and then he would be fired. One or two years before, his results were much better and win many champions. On the other hand, the results in the first years are not good compared to the second year.

One of the most aspects, which affects directly to the performance of a team, is the ratio of conceded goals per match. When the teams perform well and win some champions, this ratio is low. If this ratio is high, the teams play too badly and

TABLE I
THE CONCEDED GOAL/MATCH RATE OF THE TEAMS COACHED BY PEP
GUARDIOLA

Team	Year	Match	Conceded goal	Conceded goal/match
Barcelona	2008-2009	38	35	0.92
Barcelona	2009-2010	38	24	0.63
Barcelona	2010-2011	38	21	0.55
Barcelona	2011-2012	38	29	0.76
Bayern Munich	2013-2014	34	23	0.68
Bayern Munich	2014-2015	34	18	0.53
Bayern Munich	2015-2016	34	17	0.50
Man. City	2016-2017	38	39	1.03
Man. City	2017-2018	38	27	0.71
Man. City	2018-2019	38	23	0.61
Man. City	2019-2020	38	35	0.92
Man. City	2020-2021	38	?	?
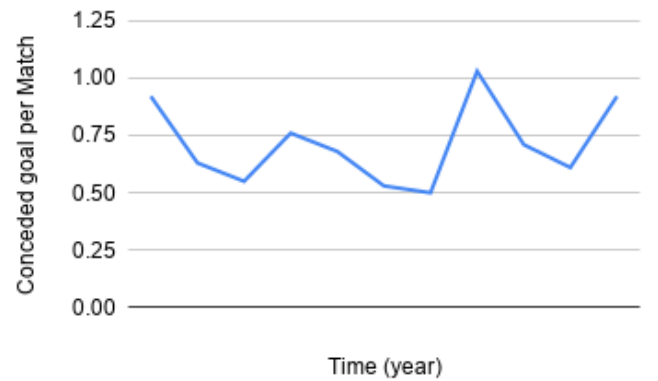
Fig. 1. The average conceded goal per match of Pep's team.

lose too much. The achievements of Pep is present in the table I with the rightmost column. The number for the years from 2008 to 2019 are given due to the statistics while the number for this year (2020-2021) is still a question. The forecasting problem means predicting the value for this year. For more visibly, these values are illustrated in the figure 1.

It is easy to observe in the figure 1, the distribution of conceded goal is near a periodic waveform. Pep's achievements can be separated into three-stage corresponding with three teams and three periods in the figure 1. The local bottom values are with the third, seventh, and tenth years. They locate at the two to three years in each team and are the most successful year of these teams. This fact can be interpreted by analyzing Pep's strategy as follow:

- Mostly depending on unique techniques, not strong or power.
- Needing the time for the player to adapt with the strategy.
- First year: Not all player plays well because they need more time.
- The second to the fourth year: The team is formed, they play well and perform full strengths of the strategy.
- The final year: The opponents are familiar with the strategy and counter-work to defeat Pep's team.
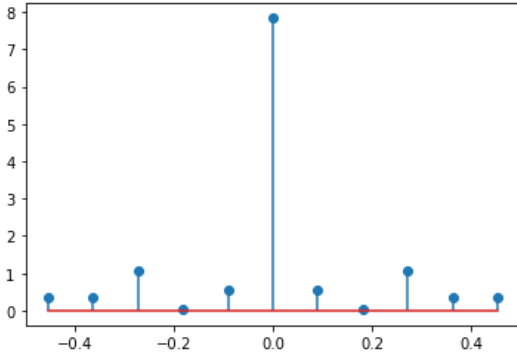
Fig. 2. Frequency representation for Pep's achievement
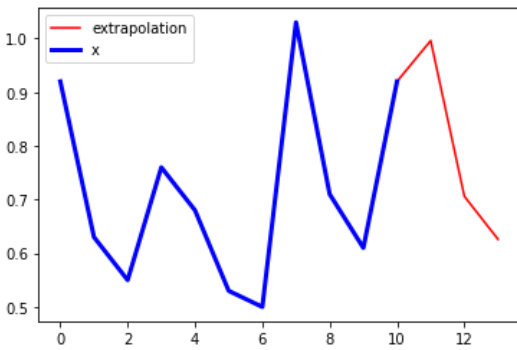
Fig. 3. Predicted value for Pep's team in the next three years

B. Framework for forecasting

As presented in the previous subsection, Pep's teams are usually followed by a fixed strategy and form a fixed achievement curve with three to five years in terms of period. This aspect is very fitted to the strength of Fourier transform. Thus, in this research, we apply this mathematical tool to exploit the data. We apply the below framework for this process:

- Modeling phase: In this phase, we transform the data into the frequency domain via Discrete Fourier Transform [9] [10] with data $f(t)$ and eleven values for ω_k:

$$F(\omega_k) = \frac{1}{2\pi} \sum_{t=1}^{11} f(t)e^{-i\omega_k t} \tag{3}$$

- Forecasting phase: Applying Inverse Discrete Fourier Transform to compute the next value in $f(t)$:

$$f(t) = \frac{2\pi}{11} \sum_{k=1}^{11} F(\omega_k)e^{i\omega_k t} \tag{4}$$

C. Experimental result

In the modelling phase, the frequency coefficients are presented in the figure 2. These elements correspond with the eleven basic waveforms of the data. As can be seen, these values distribute very differently from the original data during the time.

Figure 3 shows the Forecasting phase. After computing the next values in the data series, the results are:

$$f(12) = 0.996 \tag{5}$$

$$f(13) = 0.706 \tag{6}$$

$$f(14) = 0.626 \tag{7}$$

This means that the model predicts Pep's team will not win a lot in 2021 and can lose so much. Let us check the results in reality and then compare with the result showed by the model.

At the end of 2020-2021 season, Pep's team plays 38 matches, and concedes 32 goals. So we have:

$$\hat{f}(12) = 32/38 = 0.842 \tag{8}$$

The season 2021-2022 are continuing (at Dec, 2021) so we can use the current result as a representative. The team has played 20 matches and received 12 conceded goals. So we can estimate:

$$\hat{f}(13) = 12/20 = 0.600 \tag{9}$$

As can be observed, the numbers predicted by our method and their real values are different, but there are two things meaningful in the results. The first is that the differences not too significant. The results can be accepted in a prediction task if the distance from an estimated value to its real value is small enough. The second, and more important, is the trend of sequence value. From $f(12)$ to $f(13)$, in the values estimated by our model, the decreasing rate is:

$$d_r = 1 - \frac{0.996}{0.706} = 0.411 \tag{10}$$

On the other hand, this rate in reality is computed as:

$$\hat{d}_r = 1 - \frac{0.842}{0.600} = 0.403 \tag{11}$$

The trends are decreasing, and the decreasing rates are too similar. This is evidence to prove that our method can exploit many hidden behaviours in the data and predict the values and the trend in the time series data.

V. CONCLUSION

We have presented an application of integral transform, a simple and effective approach, for time series forecasting problems. This method can exploit many hidden behaviours of the data in a latent space including the rules for change, distribution, and trend. When applied to real data such as average conceded goal, this method performs well with an interesting result. The experiments shows that our proposed method can be a potential solution for time series prediction in real-world applications.

REFERENCES

[1] C. R. Madhuri, M. Chinta and V. V. N. V. P. Kumar, "Stock Market Prediction for Time-series Forecasting using Prophet upon ARIMA." In: 7th International Conference on Smart Structures and Systems (ICSSS), Chennai, India, doi: 10.1109/ICSSS49621.2020.9202042, pp. 1-5, 2020

[2] G. Liu, F. Xiao, C. -T. Lin and Z. Cao, "A Fuzzy Interval Time-Series Energy and Financial Forecasting Model Using Network-Based Multiple Time-Frequency Spaces and the Induced-Ordered Weighted Averaging Aggregation Operation." In: IEEE Transactions on Fuzzy Systems. 28(11), doi: 10.1109/TFUZZ.2020.2972823, pp. 2677-2690, 2020

[3] S. G. N and G. S. Sheshadri, "Electrical Load Forecasting Using Time Series Analysis." In: 2020 IEEE Bangalore Humanitarian Technology Conference (B-HTC), Vijiyapur, India, doi: 10.1109/B-HTC50970.2020.9297986, pp. 1-6, 2020

[4] Ming-Che Lee, Jia-Wei Chang, Jason C. Hung, and Bae-Ling Chen, "Exploring the Effectiveness of Deep Neural Networks with Technical Analysis Applied to Stock Market Prediction." In: Computer Science and Information Systems, Vol. 18, No. 2, 401–418, 2021, https://doi.org/10.2298/CSIS200301002L

[5] Radojičić, D., Radojičić, N., Kredatus, S., "A multicriteria optimization approach for the stock market feature selection." In: Computer Science and Information Systems, Vol. 18, No. 3, 749–769, 2021, https://doi.org/doi.org/10.2298/CSIS200326044R

[6] N. D. Hieu, N. Cat Ho and V. N. Lan, "An efficient fuzzy time series forecasting model based on quantifying semantics of words." In: 2020 RIVF International Conference on Computing and Communication Technologies (RIVF), Ho Chi Minh, Vietnam, Doi: 10.1109/RIVF48685.2020.9140755, pp. 1-6, 2020

[7] Brian Davies, "Integral Transforms and Their Applications." Springer-Verlag, New York, ISBN 978-1-441-92950-1, 2002

[8] Lokenath Debnath and Dambaru Bhatta, "Integral Transforms and Their Applications." Chapman and Hall/CRC, ISBN ISBN 978-1-584-88575-7, 2016

[9] Smith Steven, "Chapter 8: The Discrete Fourier Transform." The Scientist and Engineer's Guide to Digital Signal Processing (Second ed.), California Technical Publishing, ISBN 978-0-9660176-3-2, 1999

[10] Thomas H. Cormen, Leiserson Charles, Ronald L. Rivest, Clifford Stein, "Chapter 30: Polynomials and the FFT." Introduction to Algorithms (Second ed.), MIT Press and McGraw-Hill, pp. 822–848, ISBN 978-0-262-03293-3, 2001

Ankle joint rehabilitation system—the preliminary results

Tran Vi Do
Electrical and Electronic Faculty
Ho Chi Minh City University of
Technology and Education
Ho Chi Minh City, Vietnam
dotv@hcmute.edu.vn

Nguyen Tran Luu Phuong
Electrical and Electronic Faculty
Ho Chi Minh City University of
Technology and Education
Ho Chi Minh City, Vietnam
17129054@student.hcmute.edu.vn

Nguyen Thi Thuy Trang
Electrical and Electronic Faculty
Ho Chi Minh City University of
Technology and Education
Ho Chi Minh City, Vietnam
17129037@student.hcmute.edu.vn

Le Thi Hong Lam
Electrical and Electronic Faculty
Ho Chi Minh City University of
Technology and Education
Ho Chi Minh City, Vietnam
lamhth@hcmute.edu.vn

Phung Son Thanh
Electrical and Electronic Faculty
Ho Chi Minh City University of
Technology and Education
Ho Chi Minh City, Vietnam
thanhps@hcmute.edu.vn

Nguyen Tu Duc
Electrical and Electronic Faculty
Ho Chi Minh City University of
Technology and Education
Ho Chi Minh City, Vietnam
ducnt@hcmute.edu.vn

Abstract—Ankle joint injury is often caused due to sport activities or accidents. This makes it difficult for the daily life. Therefore, the rehabilitation of ankle joint plays an essential role, which helps patients to enhance their life qualities. Practicing physical therapy is an effective approach to aid the patients recover from injury. Currently, not only are the conventional physiotherapy methods used, but the application of hi-tech advancements, especially robots in physiotherapy, is increasingly ubiquitous and creates possible and optimistic results. This paper concentrates on the building of a supportive system for the low-cost physiotherapy and manipulating other devices by physical buttons. A mobile application programmed for smartphone displays the pieces of system information, which brings convenience to users. On the other hand, application is also connected to Bluetooth as showing the angle of device and controlling it with an intentional angle. The preliminary results of system consist of a model and a mobile smartphone application presented in this article. The proposed system operates exactly as design, this is promising for application with clinical trials on patients in the following studies.

Index Terms—Ankle-joint, rehabilitation, mobile-app, physical training.

I. Introduction

ANKLES are the important component in human movements, they act as the essential part in daily-living activities such as walking, standing, and running. Due to joint osteoarthritis, sports activities or accidents, the damage in ankle joint occurred frequently. For instance, a survey was carried out with 70 different sports from 38 countries and 201,600 patients in total. The final statistics has indicated that ankle has been the most injured site. It accounts for 34.3% that is equivalent to 24 sports and 32,509 ankle injuries [1]. When the ankle is injured, it will cause difficulties in daily life activities such as falling, limping and this depends on the severity of the injury.

It is considered to be two common treatments that would be applied for the ankle joint treatment. The first one is the total ankle replacement (TAR) surgery. This is an invasive solution which is appropriate to older patients low demands, and multiple joint osteoarthritis or inflammatory arthropathy [2]. For the minor traumas, surgery is not the most optimal and effective solution. At this moment, physiotherapy will be an alternative plan. It is mentioned to be a non-invasive treatment and safe for both the young and the old, even patients are in recovering period after TAR surgery. With the ongoing progression of the modern medical field, there are also combinations of medical devices, in addition to traditional exercises. Some traditional physical exercise methods such as flexion-extension feet movement. This treatment uses an elastic band to create resistance around the soles of the feet and pull hands to hold the ends. In opposite the backs of legs, there is a band restores resistance around the ankles and instep of the foot with both ends of the strip securely attached to the floor or wall. In addition, the training could be stepping on a platform, stepping, heel walking, ankle swivel, folding and stretching of the foot [3, 4].

Nowadays, the application of robotics or supportive systems in physiotherapy offers several advantages. There are two types of robots used in ankle therapy are Plat-form based ankle rehabilitation robots and Wearable ankle rehabilitation robots [5]. Platform based ankle rehabilitation devices have a fixed platform and thus cannot be used during gait training [6]. Parallel mechanisms are typically used for multiple degrees of freedom (DOF) systems to diminish the size of robots. Wearable robots known as exoskeleton robots or as powered orthoses are being developed in contrast to Platform based rehabilitation robots [7, 8], robots mainly referred to wearable anklebot and AFOs. The AFO (Ankle-Foot Orthosis) is a single-joint orthosis designed to assist and support movements of the ankle joint. It plays an important role during human walking [5]. The application of robots in treatment helps patients to monitor their health, propose a specific exercise plan. Moreover, exercising along with robots also contributes to control the correct movements of the patient, which minimizes the subjective errors and improves the quality of recovery treatment.

For the low-income country like in Vietnam, these systems are not affordable since the investigate cost is still high. Owning an individual device becomes hard to be possible. Moreover, going to training officials takes much time and is inflexible. The need of a low-cost supportive system for ankle rehabilitation is high. The research of a team in Pham Van Dong University has proposed a device for rehabilitating ankle joints using a linear actuator [9]. However, the model did not included the software or a mobile-app to make it easy for user.

From these cornerstones have been mentioned above, our team has studied, de-signed and built an application on smartphone what is intended to the ankle joint physiotherapy equipment. With this topic, the proposed system could be used to assist the injured people to have a flexible and easily controlled training device. The system was designed with the mobile application, which can reduce the existence of therapist to control the device.

This paper comprises of the following section: the brief information on ankle joint is described in the Section 2; Section 3 presents the design of the system; the preliminary results of the system operation are shown in the Section 4.

II. METHOD

A. Ankle joint structure

The anatomy of ankle joint is a fair complication due to its connection among bones, muscles, ligaments, tendons, nerves, blood vessels and with foot, limb. Fundamentally, It is a complex of the tibia, the fibula and the Talus at the tibiotalar joint [10, 11]. There are about 20 muscles in the foot and they are classified as intrinsic or extrinsic. The gastrocnemius or calf muscle is the largest one in all of them and aids with movement of the foot [10, 11]. Ligaments are the soft tissues that attach bones to bones. Ligaments are very similar to tendons. The difference is that tendons attach muscles to bones [10, 11]. There are also several nerves such as the superficial and deep peroneal, sural, saphenous, tibial, medial and lateral plantar, me-dial and inferior calcaneal, common digital, and medial proper plantar digital nerves [12]. They stretch and intertwine with bones, muscles, tendons, and ligaments.

Generally, the ankle joint allows the foot to move in six different ways: dorsiflex-ion (DF), plantarflexion (FP), inversion, eversion, and medial and lateral rotation [10]. Within the research scope of this topic merely focuses on two movements that are dorsiflexion and plantarflexion. Two motions are greatly common when people walk or run.
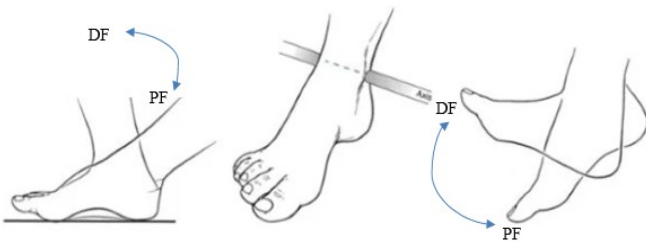
Fig. 1. Dorsiflexion (DF) and Plantarflexion (FP).

B. The ankle joint range of motion

The ankle range of motion is proved to be diverse dramatically amid individuals. Some reasons are considered like geographical region, cultural activities of daily living, or even ancestral origins. Motion of the ankle is primarily examined in the sagittal plane, with plantarflexion and dorsiflexion at the tibiotalar joint [13]. The overall ROM in the ankle joint can be shown in the Table I. Normally, dorsiflexion and plantarflexion are examined at the sagittal plane. Inversion and eversion are at frontal plane.

TABLE I. THE TYPICAL RANGE OF ANKLE JOINT MOTION

	Range of ankle joint motion	
	Sagittal plane	Frontal plane
Dorsiflexion	0 to 30°	None
Plantarflexion	0 to 55°	None
Iversion	None	0 to 23°
Eversion	None	0 to 12°

III. SYSTEM DESCRIPTION

A. The block diagram

The block diagram of system is illustrated throughout the Fig. 2.
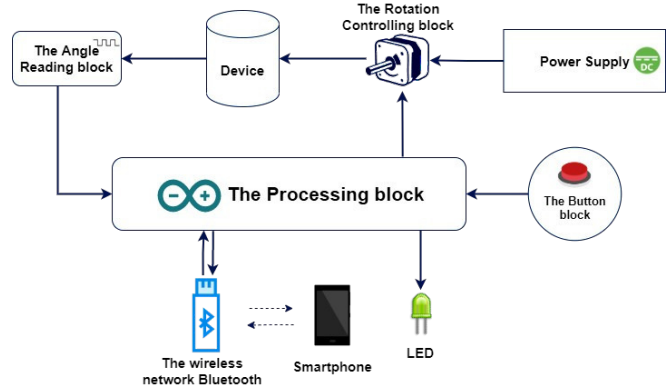
Fig. 2. The block diagram of the system.

As shown in the figure, the dash line stands for the wireless connections and the solid line represent the wired connections. Diagram includes as the following blocks and their functions respectively:

- The Processing block: KIT Arduino Mega 2560 is used to program and control all of actions in system. Arduino receives signals from the Button block, the Angle Reading block and module Bluetooth. After processing these signals, microcontroller outputs signals for controlling action of the Rotation Control-ling block and lightening LED. Furthermore, Arduino also supplies power to the Angle Reading block and module Bluetooth HC-05.
- Power Supply: A power source has 24V/5A output supplying energy for the Rotation Controlling block.
- The Rotation Controlling block: the main part of block is a 2-phase stepper motor whose rate is 57/11 (~5/1). This motor is wired directly with Micro-step driver TB6600 to run the rotational direction as well as speed of device. TB6600 is powered by the Power Supply.
- The Button block and LED: There are two pushbuttons. One is used to activate ON or OFF state of stepper motor. Each state is alarmed by LED. The other button drives device rotate as Sine wave.
- The Angle Reading block: A Rotary Encoder with 600 pulse per revolution comprises two outputs A and B producing pulses. Output signals are passed to the Processing block to calculate the angle of device.

- The wireless network Bluetooth: Module BLE HC-05 sends data from the Processing block to smartphone and vice versa. Module interacts with the Processing block via UART protocol.

- A mobile app that is programmed by MIT app inventor will be paired with module HC-05 by its address before showing data on screen.

B. Control diagram

Devices and peripherals interact with Arduino are shown in the Fig. 3.
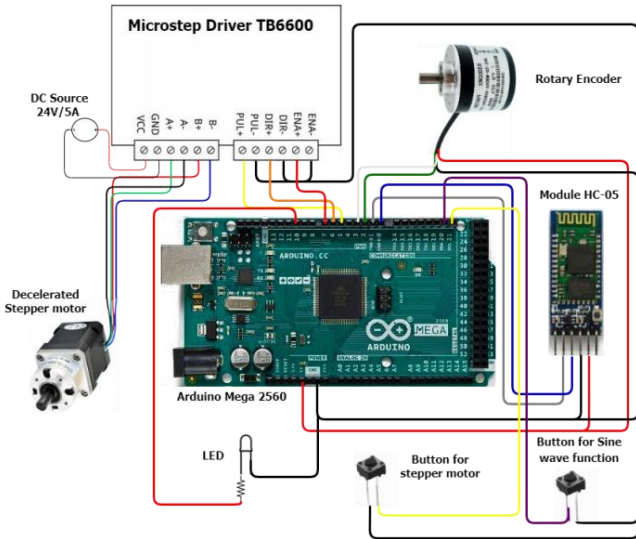
Fig. 3. The hardware connection diagram beween KITArduino Mega 2560 and other devices, peripherals.

Arduino charges 5V power to module HC-05 and Rotary Encoder. TX and RX pin of module are wired respectively with RX pin (pin 0 on Arduino) and TX pin (pin 1). Pin 2 on MCU is connected to phase A while pin 3 is linked with phase B of Encoder. Port VCC of TB6600 receives energy source 24V/5A. Four other ports A +, A-, B + and B- will be connected to the stepper motor pins respectively. PUL-, DIR- and ENA- are generally grounded to the Arduino. Pin 5, 6, 7 are connected to PUL +, DIR +, ENA + of the TB6600 driver respectively. The signal LED is connected to pin 10 through a 300Ohm resistor. The two buttons respectively control are connected to pin 20 and 21 of Arduino.

C. Algorithm flowchart

Main program of the proposed system is presented as in an algorithm flowchart in Fig. 4. On condition that Arduino detects data are send from module Bluetooth, it will process them and then go back to check data. On the other hand, Arduino does not detect any data from module Bluetooth, it will be in the following order: calculating angle, controlling motor direction, controlling motor rotation, controlling automatic motor rotation, checking motor button state and checking sine wave button state. After all of these steps, Arduino will go back to check data.
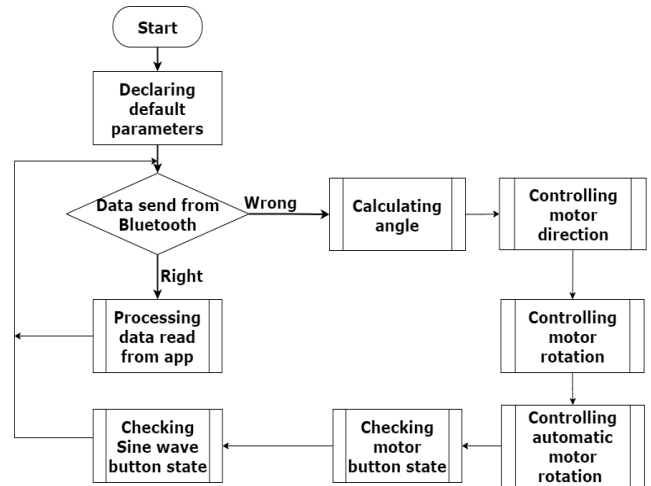
Fig. 4. Algorithm flowchart of the main programme.

D. System working principle

Two main activities of this device are lifting and lowering movement as in Fig. 6a and Fig. 6b. These movements could be caused from foot or the stepper motor.

At beginning, keep device at an equal position (Fig. 5c). Phase A and B produce no pulses so the default angle is 0 degree. There are two solutions to control this sys-tem.

The first solution is the mechanical impact from foot. At any time device is lifted or lowered, axis of Encoder will rotate and produce pulses at phase A and B output. They are transmitted to Arduino for figuring the current angle of device. Because revolution of Encoder is 600 BM, this means that there are 600 pulses created after Encoder axis has completed a 360 degree rotation.

The second solution is the interaction between stepper motor and device. First of all, it is necessary to push button (physical button or mobile app button). Arduino gets signals, activates the "ON" state of motor and makes LED bright. Signals will be transmitted to MCU by TX pin of module on condition that button is pressed on app.

Next, it is compulsory to fill any angle and press "ENTER" button on the application. This data is sent to the microcontroller via Bluetooth network. Arduino will control the device to rotate with the deliberate angle via the stepper motor.

Rotating automatically as Sine wave is an additional function. Modifying device at the balancing position is requisite. Arduino will control device to rotate automatically after button has pressed (physical button or mobile app button).

Concurrently, Encoder detects the modifications of angle while device rotates. Phase A and B transmit pulses at the output to Arduino for conducting the calculation. Final result which is the current angle is updated on mobile app. During the rotating process, if the button (mechanical button or button on the app is pressed), signal is sent back to Arduino to stop the stepper motor immediately.

IV. RESULTS

A. Hardware

Below pictures illustrate for the completed model. It is made up and laser cut from iron materials. Rotary Encoder and Stepper motor are attached as in Fig. 5. Pedal is located at the top center position for up and down moving. The weight of the system is about 5 kilograms.

Detail expenditures of components in the system are listed in the TABLE II. In general, the total estimated cost of the system is about 130 USD.

TABLE II. ESTIMATED COST OF THE SYSTEM COMPONENTS

Component	Quantity	Price (USD)
Rotary Encoder	1	11
Stepper motor	1	34
Microstep Driver	1	5
Module Bluetooth HC-05	1	5
Arduino Mega	1	10
Laser cutting		65
Total price		**130**

a) Device on the side   b) Device from above

c) Device at balancing position

Fig. 5. Some pictures of the device

a) Pedal is lifted.   b) Pedal is lowered.

Fig. 6. Device illustrates up and down moving

B. Mobile app

Outcomes of the mobile application are shown below. The very first GUI of app is exact as Fig. 7.
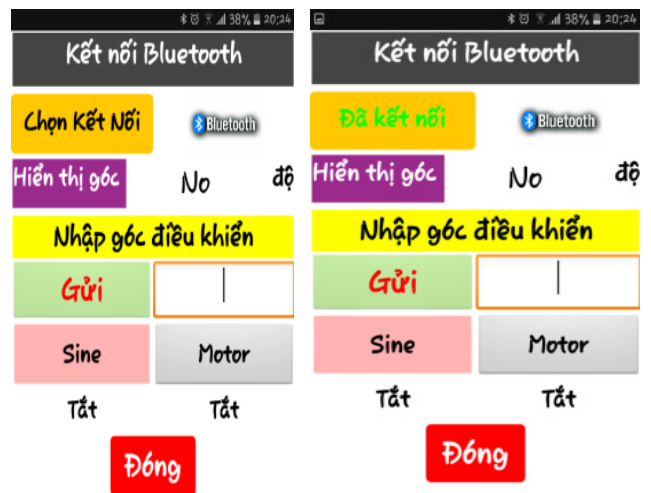
a) App is open.   b) App is connected to Blue-tooth.

Fig. 7. The original interface of mobile application.
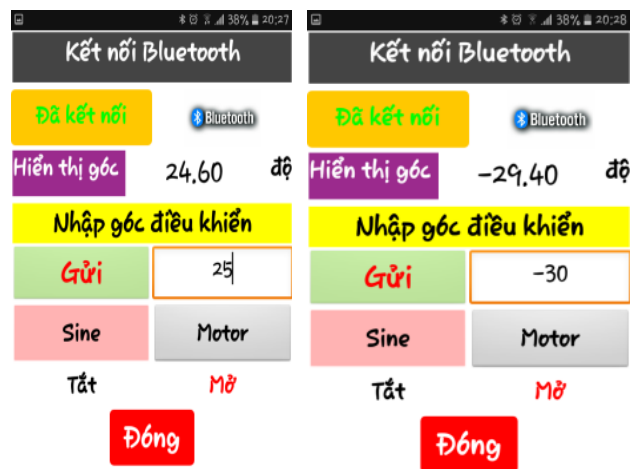
a) Current angle at 24.60 degree. b) Current angle at -29.40 degree.

Fig. 8. Result of current angle is updated on mobile application.

From the results through many test runs, in general, the system calculates and dis-plays the device's rotation angle, controls the device with the mechanical push button, and controls arbitrarily angle by smartphone.

The system at the moment still consists of some limitations. First, the difference between the read angle from Encoder and the angle to be rotated is still large and unstable. This originates from the arranging functions in the main program as well as the order of statements. It would be better to alter the algorithm more logically.

Second, the synchronizing between the state of the physical button and of the button on the application is still slow and sometimes crashes. Its key reason is the exchange data between Arduino and module HC-05 as well as mobile app and Blue-tooth network. This could be overcome by modifying the synchronizing algorithm.

V. CONCLUSION

System has been operated by impacts from foot or buttons. A mobile application on smartphone is an useful combination for controlling system and monitoring information of it. This deserves to be a considerable choice for the ankle joint training with an affordable expenditure. The proposed system operates exactly as design. This brings a promise for application with clinical trials on patients in the following studies.

ACKNOWLEDGMENT

REFERENCES

[1] K C C Chan, Lap Ki Chan, Daniel Tik-Pui Fong, Youlian Hong, Patrick Shu Hang Yung, "A Systematic Review on Ankle Injury and Ankle Sprain in Sports", Sports Medicine 37(1):73-94, February 2007.

[2] Bonasia DE, Dettoni F, Femino JE, Phisitkul P, Germano M, Amendola A. "Total ankle replacement: why, when and how?". Iowa Orthop J. 2010;30:119-130.

[3] Chinn L, Hertel J. "Rehabilitation of ankle and foot injuries in athletes". Clin Sports Med. 2010;29(1):157-167. Doi:10.1016/j.csm.2009.09.006.

[4] Bleakley CM, Taylor JB, Dischiavi SL, Doherty C, Delahunt E. "Rehabilitation Exercises Reduce Reinjury Post Ankle Sprain, But the Content and Parameters of an Optimal Exercise Program Have Yet to Be Established: A Systematic Review and Meta-analysis." Arch Phys Med Rehabil. 2019 Jul;100(7):1367-1375. doi: 10.1016/j.apmr.2018.10.005. Epub 2018 Oct 26. PMID: 30612980.

[5] Zhang, M., Davies, T.C. & Xie, S. "Effectiveness of robot-assisted therapy on ankle rehabilitation – a systematic review". J NeuroEngineering Rehabil 10, 30 (2013). https://doi.org/10.1186/1743-0003-10-30.

[6] Tsoi, Y. "Modelling and adaptive interaction control of a parallel robot for ankle rehabilita-tion." (2011).

[7] Oña ED, Garcia-Haro JM, Jardón A, Balaguer C. "Robotics in Health Care: Perspectives of Robot-Aided Interventions in Clinical Practice for Rehabilitation of Upper Limbs." Ap-plied Sciences. 2019; 9(13):2586. https://doi.org/10.3390/app9132586.

[8] Stein Joel, Hughes Richard, Fasoli Susan, Krebs Hermano Igo, Hogan Neville, "Clinical Applications of Robots in Rehabilitation", Critical Reviews in Physical and Rehabilitation Medicine, 17(3):217-230.

[9] D. Minh Duc, L. T. Thuy Tram, P. Dang Phuoc and T. Xuan Tuy, "Study on Ankle Reha-bilitation Device Using Linear Motor," 2019 International Conference on System Science and Engineering (ICSSE), Dong Hoi, Vietnam, 2019, pp. 573-576, doi: 10.1109/ICSSE.2019.8823118.

[10] Bonnel F., Bonnin M., Canovas F., Chamoun M., Bouysset M. (1998) "Anatomy of the foot and ankle". In: Bouysset M. (eds) Bone and Joint Disorders of the Foot and Ankle. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-06132-9_1

[11] Brockett, Claire L, and Graham J Chapman. "Biomechanics of the ankle." Orthopaedics and trauma vol. 30,3 (2016): 232-238. doi:10.1016/j.mporth.2016.04.015

[12] Michel De Maeseneer, Hardi Madani, Leon Lenchik, Monica Kalume Brigido, Maryam Shahabpour, Stefaan Marcelis, Johan de Mey, Aldo Scafoglieri, "Normal Anatomy and Compres-sion Areas of Nerves of the Foot and Ankle: US and MR Imaging with Anatom-ic Correlation", RadioGraphics 2015; 35:1469–1482.

[13] Brockett CL, Chapman GJ. "Biomechanics of the ankle". Orthop Trauma. 2016 Jun; 30(3):232-238. doi: 10.1016/j.mporth.2016.04.015. PMID: 27594929.

Classification-Segmentation Pipeline for MRI via Transfer Learning and Residual Networks

Nghia Duong-Trung✉
Technische Universität Berlin
Straße des 17. Juni 135, 10623 Berlin, Germany
Email: nghia.duong-trung@tu-berlin.de

Dung Ngoc Le Ha
Can Tho University of Technology
74000 Can Tho city, Vietnam
Email: hlndung@ctuet.edu.vn

Hiep Xuan Huynh
Can Tho University
74000 Can Tho city, Vietnam
Email: hxhiep@ctu.edu.vn

Abstract—**Artificial intelligence association into brain magnetic resonance imaging (MRI) and clinical practices embrace substantial cancer diagnosis improvement. The advancement of deep learning has improved the processing and analysis of MRI, boosting models' performance, decreasing the destructive effects of data sources overload, and increasing accurate detection and time efficacy. However, that specific dataset leads to diverse research fields such as image processing and analysis, detection, registration, segmentation, and classification. This paper proposes a decision-making pipeline for MRI data by combining image classification and segmentation. First, the pipeline should correctly produce a correct decision given an MRI image. If the figure is classified as defective, the pipeline can extract defect regions and highlight them accordingly. We have implemented several advanced convolutional neural networks with transfer learning and residual techniques to address two broad clinical concerns in one decision-making workflow.**

I. Introduction and Motivation

TODAY, clinical practice is an area of interest and research where extensive research and technical recommendations have been developed in response to increasingly complex challenges [1], [2], [3]. Identifying and analyzing diseases is increasingly difficult because they are ever more sophisticated. Fortunately, artificial intelligence has revolutionized clinical practice in many areas such as cancer diagnosis with medical imaging [4], automatic classification diseases based on descriptions [5], [6], and maximizing hospital efficiency [7]. Among many approaches, deep learning has been proven superior in a wide range of clinical data and practice scenarios. Regarding MRI, the complex feature can be represented effectively by utilizing deep learning-based models in detection, registration, classification, and segmentation problems.

Employing a convolutional neural network (CNN) for image classification is one of the reasonable rises, and it is an essential model in developing an automatic disease diagnosis [8], [9]. Among competitors of the ImageNet challenge in 2012, the deep learning-based model AlexNet proposed by Krizhevsky et al.[10] won the championship. CNN has become the backbone architecture for addressing almost all problems in computer science. Many CNN-based approaches have been investigated for addressing MRI image classification [11].

Due to CNN's dominant performance in the MRI classification domain, people began exploiting CNN for MRI segmentation. More specifically, MRI diagnosis is the subdivision of different brain regions to detect brain diseases, such as cancer and Parkinson's syndrome. Consequently, automatic segmentation of defect regions in MRI is significantly essential in everyday clinical routines and medical research [12]. With the performance of CNN, excellent segmentation approaches have been developed based on CNN and continuously become front tier in particular segmentation competitions [13].

However, one interesting research question that someone might consider is that we should combine several research domains and develop a practical workflow that supports medical analysis and recommendation. This article aims to propose a decision-making pipeline for MRI data by combining image classification and image segmentation. First, given an MRI image, the classification part of the pipeline should make a correct decision. If the brain MRI is classified as defective, the segmentation part can extract defect regions and highlight them accordingly. Thus, the Class-Seg workflow is designed and implemented by leveraging transfer learning, residual network [14] design, and several state-of-the-art CNN models.

II. Blueprint of the MRI Class-Seg Pipeline

As mentioned in the previous section, we propose a machine learning pipeline to support MRI diagnosis by (i) applying transfer learning techniques to select the best detection model and (ii) integrating classification with segmentation in one channel to improve MRI diagnosis and treatment. The blueprint of our proposed pipeline is presented in Figure 1. Here, the MRI dataset is randomly divided without replacement into several portions. Five primarily used CNN models are selected for the task of classification. Technically, the CNN part's weights are transferred from pre-trained models on ImageNet. The CNN part is freezing out of the backpropagation process. Then it is flattened and fed into our proposed dense layers, see Figure 2, where the weights are learned. The authors implement ResUNet based on UNet architecture comprising several residual blocks to overcome the vanishing gradients problems in deep CNN architecture regarding the segmentation task. We present how classification and segmentation tasks can be combined in Algorithm 1. Note in the pseudo-code that phases 1 and 2 correspond to points (i) and (ii) in this section, respectively. Transfer learning is applied in phase 1, where multiple models are reused, while in phase 2, we train the segmentation model from scratch.

Figure 1.   The Proposed Classification-Segmentation Pipeline for Brain MRI.

```
Layer (type)              Output Shape         Param #
===========================================================
dense (Dense)             (None, 1024)         (from flatten)

dropout (Dropout)         (None, 1024)         0

dense (Dense)             (None, 1024)         1049600

dropout (Dropout)         (None, 1024)         0

dense (Dense)             (None, 1024)         1049600

dropout (Dropout)         (None, 1024)         0

dense (Dense)             (None, 256)          262400

dropout (Dropout)         (None, 256)          0

dense (Dense)             (None, 2)            514
===========================================================
```

Figure 2.   Our proposed dense layers to be trained with MRI dataset.

III. TRANSFER LEARNING

Transfer learning technique is to reuse the pre-trained models learned in one or more different domains and utilize the knowledge to enhance learning in any other domain [15], [16]. Reusing a trained model that solves a problem similar to your data source is very versatile. It allows a machine learning approach to be applied to data drawn from a wide range of different sources from the one upon which it has initially been trained [17], [18], [19]. It might take weeks to train modern CNN models with millions of parameters fully. Transfer learning proposed many re-trained options such as fine-tuning model weight, freezing layers, and even re-train from scratch. It shortcuts a lot of network design and training

by transferring a trained set of parameters from a predefined category and re-weights the model's parameters from new data. This technique, called inductive transfer learning, aims to effectively improve training in the target domain by practicing knowledge transferability from many other sources. By re-updating weights, the effects of dissimilar observations will be reduced, and it might thus produce a more objective approach. Every CNN-based model consists of two parts, a series of convolutional-pooling layers as the tail and sequences of dense layers as the head. One transferability strategy is to transfer the learned parameters to the tail and freeze it. Flatten is applied at the tail's last layer before integrating it into the head. Backpropagation and parameters update is done within the head. In Figure 3, we present the key difference between traditional machine learning and the adoption of transfer learning.

IV. EXPERIMENTS

A. Experimental MRI Dataset

We have exploited our proposed classification-segmentation pipeline on the most reputable brain MRI segmentation dataset[1]. By the time we conducted this paper, there had been 42 code solutions and six discussions on the dataset. However, none developed a systematic pipeline for the dual task of binary classification and image segmentation. The 1GB contains 7860 brain MRI figures manually annotated with fluid-attenuated inversion recovery (FLAIR) abnormality

[1]https://www.kaggle.com/mateuszbuda/lgg-mri-segmentation

Algorithm 1 The proposed classification-segmentation pipeline.

1: **Phase 1**: Train classification model
2: Splitting MRI dataset into 70%-15%-15% training, validation (val.), test sets
3: **for** each model to apply transfer learing **do**
4:     Transfer pretrained weights from ImageNet to the CNN part and freeze it
5:     Integrate the proposed dense layers, see Figure 2
6:     Optimize Equation 2
7: **end for**
8: Get the best classification model, called *model 1*
9:
10: **Phase 2**: Train segmentation model
11: Defect MRI images are split into 70%-15%-15% training, val., test(*) sets
12: Optimize Equation 3
13: Get the optimal segmentation model, called *model 2*
14:
15: **Phase 3**: Class-Seg combination
16: **for** Each image i in test(*) **do**
17:     **if** *model 1* predicts i as non-defect **then**
18:         Continue
19:     **else**
20:         *model 2* segments the defect region
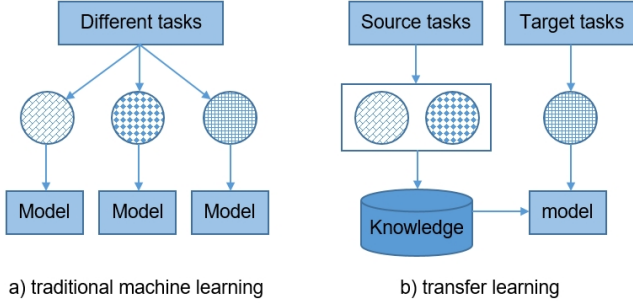21:   **end if**
22: **end for**

Figure 3. In a), different models for seperated tasks. In b) one apporach can be re-used for many tasks.

segmentation masks. First, the dataset was collected from 110 patients included in The Cancer Genome Atlas lower-grade glioma collection. Then, the images with at least one possible FLAIR sequence and genomic cluster are selected.

B. Loss Functions

In the previous section, the authors have mentioned that the pipeline consists of two distinct tasks: first, the model should correctly classify defective images from normal ones; second, after the image is classified as defected, it is fed into segmentation to reveal the region of abnormality. Hence, the pipeline applies two different loss functions: categorical cross-entropy loss to address binary classification scenario and Focal

Tversky [20] loss which is highly recommended for handling imbalanced data and small regions-of-interest segmentation.

Categorical cross-entropy (CCE) loss is a softmax activation plus a cross-entropy (CE) loss. The CE is defined as follows in a binary classification problem.

$$CE = \sum_{i=1}^{C=2} y_i \log(s_i) = -y_1 \log(s_1) - (1-y_1) \log(1-s_1), \quad (1)$$

where y_i, s_i, and C are the groundtruth, the predicted CNN score and the number of class respectively. Then CE plus softmax activation which yields CCE as follows.

$$CCE = -\sum_{i}^{C} y_i \log(f(s)_i), \quad (2)$$

where $f(s)_i = \frac{e^{s_i}}{\sum_j^C e^{s_j}}$.

The focal Tversky (FT) loss is defined as follows.

$$FT = \sum_{i}^{C} (1 - T_i)^{\frac{1}{\gamma}}, \quad (3)$$

where $\gamma = \frac{4}{3}$ by default as described in [20]. T_i is denoted as follows.

$$T_i = \frac{\sum_{j=1}^{N} p_{ji} g_{ji} + \epsilon}{\sum_{j=1}^{N} p_{ji} g_{ji} + \alpha \sum_{j=1}^{N} p_{j\bar{i}} g_{ji} + \beta \sum_{j=1}^{N} p_{ji} g_{j\bar{i}} + \epsilon}, \quad (4)$$

where N provides the total number of pixels in an image. ϵ is numerical stability. p_{ji} is the probability that pixel j is of the lesion class i, while $p_{j\bar{i}}$ is that of non-lesion class i. The same meaning is applied to g_{ji} and $g_{j\bar{i}}$. α and β are tunable hyperparameters to shift the recall emphasis in case of large class imbalance. In our experiments, we set the value of $\alpha = 0.7$ and $\beta = 0.75$.

C. Experimental Results

For the task of classification, the authors have deployed 5 well-known CNN models in computer vision community, e.g. ResNet50 [21], ResNet50V2 [21], InceptionResNetV2 [22], EfficientNet [23], and MobileNetV2 [24]. These models are easily called using TensorFlow API [25]. While in the segmentation task, the authors implement ResUNet [26], one of the most state-of-the-art segmentation models. The authors run three times for each approach and report the average scores. The classification and segmentation performance of all models have presented in Table I. We illustrate several segmentation samples in Figure 4.

In Table I, the first five models are used to select the best candidate for the classification task. The best solution is the InceptionResNetV2 model, which achieves 95% of accuracy. The authors report the number of trainable parameters, the average training duration, the accuracy score on the test set, and two basic F1-score schemes [27]. Turning to segmentation,

Table I
CLASSIFICATION AND SEGMENTATION PERFORMANCE OF ALL MODELS.

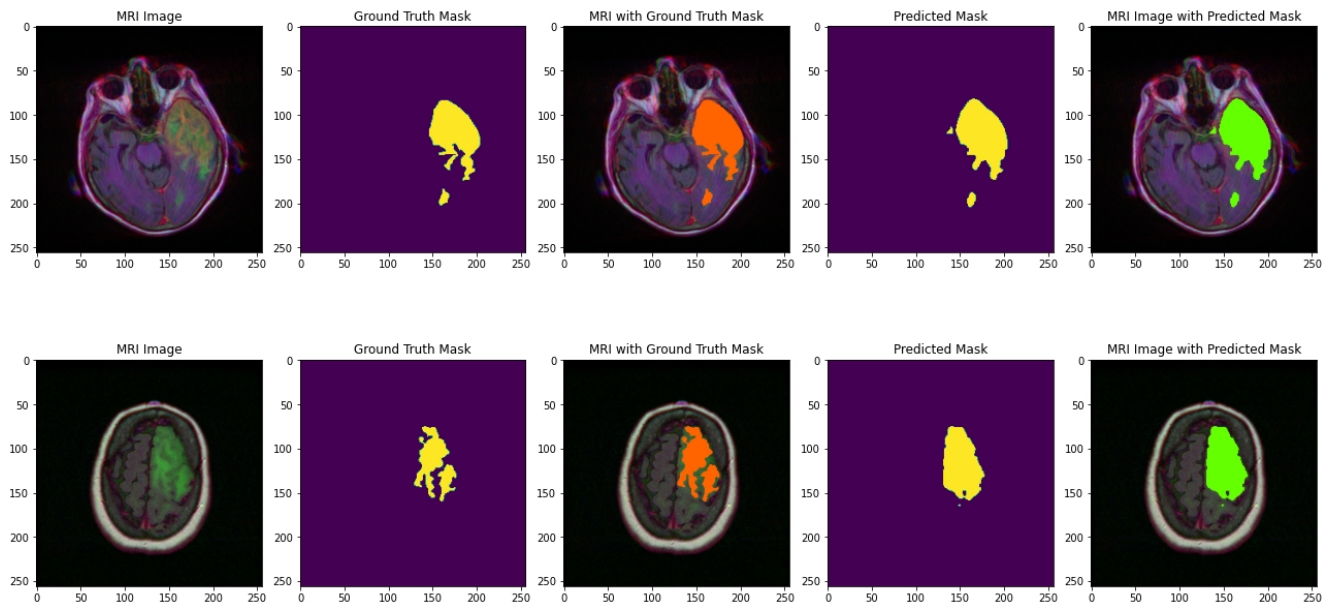| Model	Trainable params (million)	Training time (minutes)	Test Acc.	Micro avg F1-score	Macro avg F1-score
ResNet50	59.452.162	15.04	0.93	0.93	0.92
ResNet50V2	59.436.930	13.36	0.88	0.88	0.86
InceptionResNetV2	70.795.106	27.78	**0.95**	**0.95**	**0.95**
EfficientNetB0	27.342.206	13.48	0.66	0.67	0.43
MobileNetV2	25.558.530	09.33	0.41	0.42	0.37
			Val. Tversky		
ResUNet	1.206.129	0.63	0.89	-	-

Figure 4. The Result of the Segmentation Process. From left to right in each row, the original MRI figure, annotated groundtruth mask on defect region, the overlapping of them, the predicted mask provided by the segmentation process, and the predicted overlapping.

the authors implement ResUNet because of the speed, high performance.We report the Tversky score for this task.

D. Reproducibility

The authors have conducted all experiments on Google Colab. GPU runtime type has been activated by default. We encourage further reproducibility by engaging readers by revealing all models' weights and architecture. Regarding the paper's length restriction, the authors add additional materials and other experiment resources in our GitHub repository[2].

V. CONCLUSION

We have described and implemented the proposed Class-Seg workflow by leveraging transfer learning and residual network together with several state-of-the-art convolutional neural models. We aim to combine the classification and segmentation of brain MRI into a single clinical practice. To our knowledge, we are the first to combine two research

[2]https://github.com/duongtrung/Class-Seg-Brain-MRI

directions on the well-known Kaggle brain MRI dataset, in which more than 40 code solutions have been investigated. Intensive experiments have been conducted to develop a clinically acceptable automatic workflow for better brain MRI diagnosis.

REFERENCES

[1] L. da Cruz, C. Sierra-Franco, G. Silva-Calpa, and A. Raposo, "Enabling autonomous medical image data annotation: A human-in-the-loop reinforcement learning approach," in *Proceedings of the 16th Conference on Computer Science and Intelligence Systems*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. Maciaszek, M. Paprzycki, and D. Ślęzak, Eds., vol. 25. IEEE, 2021, p. 271–279.

[2] J. Dörpinghaus, S. Schaaf, V. Weil, and T. Hübenthal, "An efficient approach towards the generation and analysis of interoperable clinical data in a knowledge graph," in *Proceedings of the 16th Conference on Computer Science and Intelligence Systems*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. Maciaszek, M. Paprzycki, and D. Ślęzak, Eds., vol. 25. IEEE, 2021, p. 59–68.

[3] R. Damaševičius, O. Abayomi-Alli, R. Maskeliūnas, and A. Abayomi-Alli, "Bilstm with data augmentation using interpolation methods to improve early detection of parkinson disease," in *Proceedings of the 2020 Federated Conference on Computer Science and Information*

Systems, ser. Annals of Computer Science and Information Systems, S. Agarwal, D. N. Barrell, and V. K. Solanki, Eds. IEEE, 2020, pp. 371–380.

[4] M. L. Giger, "Machine learning in medical imaging," *Journal of the American College of Radiology*, vol. 15, no. 3, pp. 512–520, 2018.

[5] N. Duong-Trung, X. N. Hoang, T. B. T. Tu, K. N. Minh, V. U. Tran, and T.-D. Luu, "Blueprinting the workflow of medical diagnosis through the lens of machine learning perspective," in *2019 International Conference on Advanced Computing and Applications (ACOMP)*. IEEE, 2019, pp. 23–26.

[6] N. Duong-trung, N. Quynh, T. Tang, and X. S. Ha, "Interpretation of Machine Learning Models for Medical Diagnosis," *Advances in Science, Technology and Engineering Systems Journal*, vol. 5, no. 5, pp. 469–477, 2020.

[7] M. Fairley, D. Scheinker, and M. L. Brandeau, "Improving the efficiency of the operating room environment with an optimization and machine learning model," *Health care management science*, vol. 22, no. 4, pp. 756–767, 2019.

[8] H. Benbrahim, H. Hachimi, and A. Amine, "Deep convolutional neural network with tensorflow and keras to classify skin cancer images," *Scalable Computing: Practice and Experience*, vol. 21, no. 3, pp. 379–390, 2020.

[9] S. Kusuma and J. D. Udayan, "Analysis on deep learning methods for ecg based cardiovascular disease prediction," *Scalable Computing: Practice and Experience*, vol. 21, no. 1, pp. 127–136, 2020.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.

[11] T. Kaur and T. K. Gandhi, "Deep convolutional neural networks with transfer learning for automated brain image classification," *Machine Vision and Applications*, vol. 31, no. 3, pp. 1–16, 2020.

[12] M. Mittal, M. Arora, T. Pandey, and L. M. Goyal, "Image segmentation using deep learning techniques in medical images," in *Advancement of machine intelligence in interactive medical image analysis*. Springer, 2020, pp. 41–63.

[13] Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin, and B. J. Erickson, "Deep learning for brain mri segmentation: state of the art and future directions," *Journal of digital imaging*, vol. 30, no. 4, pp. 449–459, 2017.

[14] G. Liang and L. Zheng, "A transfer learning method with deep residual network for pediatric pneumonia diagnosis," *Computer methods and programs in biomedicine*, vol. 187, p. 104964, 2020.

[15] W. Ying, Y. Zhang, J. Huang, and Q. Yang, "Transfer learning via

[16] Q. Yang, Y. Zhang, W. Dai, and S. J. Pan, *Transfer learning*. Cambridge University Press, 2020.

[17] N. Duong-Trung, L.-D. Quach, and C.-N. Nguyen, "Learning deep transferability for several agricultural classification problems," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 1, 2019.

[18] N. Duong-Trung, L.-D. Quach, M.-H. Nguyen, and C.-N. Nguyen, "A combination of transfer learning and deep learning for medicinal plant classification," in *Proceedings of the 2019 4th International Conference on Intelligent Information Technology*, 2019, pp. 83–90.

[19] ——, "Classification of grain discoloration via transfer learning and convolutional neural networks," in *Proceedings of the 3rd International Conference on Machine Learning and Soft Computing*, 2019, pp. 27–32.

[20] N. Abraham and N. M. Khan, "A novel focal tversky loss function with improved attention u-net for lesion segmentation," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 2019, pp. 683–687.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[22] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, 2017.

[23] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning*. PMLR, 2019, pp. 6105–6114.

[24] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.

[25] D. Sarkar, R. Bali, and T. Ghosh, *Hands-On Transfer Learning with Python: Implement advanced deep learning and neural network models using TensorFlow and Keras*. Packt Publishing Ltd, 2018.

[26] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "Resunet-a: a deep learning framework for semantic segmentation of remotely sensed data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 162, pp. 94–114, 2020.

[27] N. Duong-Trung, *Social Media Learning: Novel Text Analytics for Geolocation and Topic Modeling*. Cuvillier Verlag, 2017.

learning to transfer," in *International Conference on Machine Learning*. PMLR, 2018, pp. 5085–5094.

LoRa Based Sensor and Actuator Networks in Smart Livestock Farming Applications

Duc Chinh Hoang
Hanoi University of Science and Technology
Hanoi, Vietnam
chinh.hoangduc@hust.edu.vn

Van Minh Pham
Hanoi University of
Science and Technology
Hanoi, Vietnam

Quoc Khanh Tran
Hanoi University of Science
and Technology
Hanoi, Vietnam

Krishnanand K. R
Berkeley Education Alliance for Research
Singapore, Singapore

Anh Hoang
Hanoi University of Science and Technology
Hanoi, Vietnam

Abstract—**Fast growth of farming activities in Vietnam requires employment of new technologies in data acquisition, control, and communication to make them operate more efficiently. In livestock farming applications, environmental condition monitoring and control are essential to maintain good comfort for the animal, reduce the farmworker's workload, and minimize the risk of disease spread. In this work, a wireless sensor and actuator system based on LoRa protocol is developed to deploy in hog farms. The ambient temperature and air quality information is acquired and processed. The data is then utilized for automated control of the cooling fans to comfort the animal in the barns. Besides, long term data collection can be also used to understand the system comprehensively and thus facilitate the farm owner to enhance the farming operation subsequently. A graphical user interface is also provided for the operator to supervise the system and intervene if necessary. Experimental results are provided to demonstrate the performance of the whole system.**

Index Terms—**LoRa, Wireless Sensor and Actuator Networks, Smart Farming**

I. Introduction

INDUSTRIAL automation in developing countries such as Vietnam is accelerating rapidly. Its agriculture, once the main contributor to the economy, still plays a crucial role. However, the natural resources have become expensive, and the aggressive urbanization results in the decrement of farmland. Modernization of the agriculture is a must in order to optimize the limited resources and maximize the profit. Recent advanced technologies in communication and computation enable new trend of smart farming applications. A number of Wireless Sensor and Actuator Networks (WSANs) have been developed for the last decade and shown a great potential in agriculture as stated in [1-3]. Wireless Sensor Networks (WSNs) consists of a number of nodes which can be deployed easily in various types of environment and co-operate well with each other to perceive required information [2, 4]. Actuators equipped with wireless communication modules and integrated in the WSNs make them WSANs in order to adjust the environmental conditions as desired. An example of ZigBee based WSANs is presented in for heating and cooling loads in buildings [5]. The areas of interest in agriculture applications are usually much larger than office buildings or residential building systems. Authors in [6] developed a WSAN system to manage the cattle on the open field. In close area such as greenhouses or livestock buildings, environmental quality monitoring and control are essential to provide healthy living space for the plant or animal [7].

Our work focuses on developing LoRa based WSANs to manage indoor conditions in livestock farming applications. LoRa is a low power wide area wireless network protocol which is often used in large scale applications. Cost effectiveness, long range and low energy consumption features of LoRa make it one of the most suitable candidates for smart city or smart farm applications [8-11]. When being compared to other popular wireless protocols such as ZigBee, Bluetooth, and WiFi, the main advantages of LoRa are the lowest power consumption and the longest communication range [9]. Although, a tradeoff is its low data rate, it is still sufficient for the slow process applications like farming systems. In these applications, parameters such as temperature, carbon dioxide (CO_2) concentration, etc., which have great influence on the farming animal habitat [12], should be observed and managed carefully. The developed system would help to perceive the CO_2 concentration and temperature values, and adjust the ventilation system when it is required, thus realizing low-cost automation.

The structure of this paper is as following: Section 1 introduced the background of this work, the overall system and requirements are provided in Section 2, Section 3 presents the system design and development, Section 4 provides the experiment results and discussion, and the paper is concluded in Section 5.

II. Systems descriptions

Overall system architecture is illustrated in Fig. 1. The WSAN is organized in sub networks. Each sub system consisting of several LoRa based sensor and actuator nodes is deployed in a barn. There are two types of nodes: a local control node which is unique in a barn and sensor nodes. Sensor nodes are equipped with sensing module to acquired environmental information. The local control node acts as a coordinator to collect the data from sensor nodes within the same barn, processes, and forwards to the central server. The local control node is also connected to a variable frequency drive (VFD) for controlling the ventilation fans. These fans help to pump fresh air into the barn as well as to

cool down the indoor temperature during hot days. In some farms where the biogas is available for heating, the fans can also help to bring hot air inside the barn in winter faster. A supervisory software is employed at the control center to look over and provide the initial settings of the whole system.
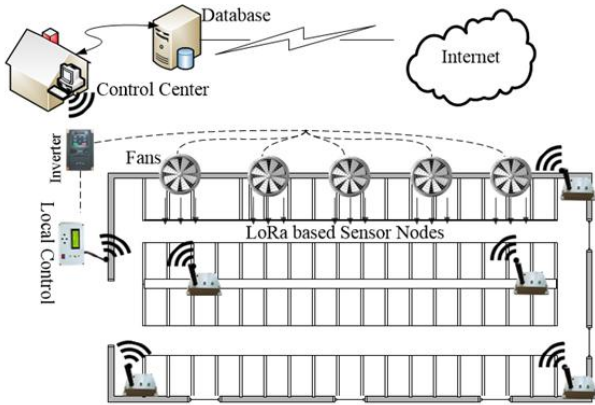

Fig 1. LoRa based sensor and actuator system in hog barns

III. SMART FARM MORNITORING AND CONTROL SYSTEM
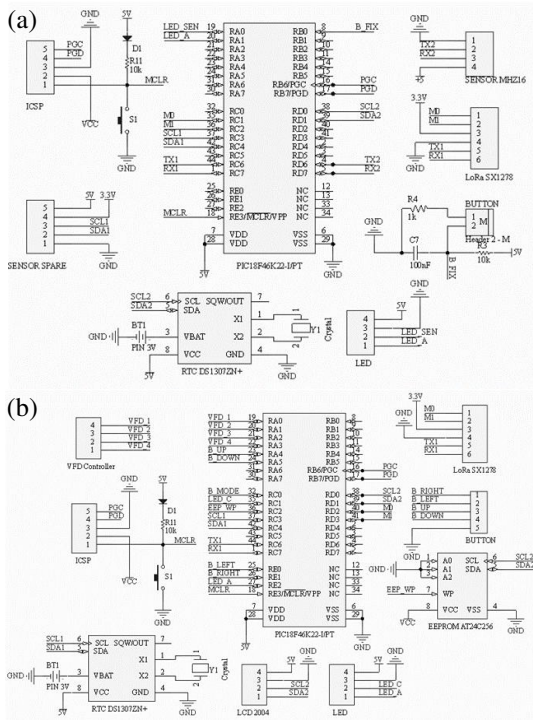
A. LoRa based hardware platforms

Fig 2. Schematic of (a) sensor node and (b) the local control node

The designed system mainly consists of 3 main components: Sensor Node, Local Control Node and Control Center. These components connect wirelessly to each other, transmit data via the Lora wireless communication protocol. Each cage consists of multi sensor circuits collects data from sensors through the RS485 communication standard and send it to the Local Control Circuits located in the cage. The local control circuits process the data and send control signals to the drive to control fan speed. In addition, the Local Control also has the function to send the collected parameters to the Control Center. Control Center can be a

personal computer or embedded computer, it is responsible for monitoring the parameters of interest and sending the setting values to the local control unit at each barn.

Sensor Node Hardware. The sensor nodes are equipped with a PIC microcontroller as a processing unit, a MHZ-16 sensors which can measure ambient temperature and CO_2 concentration. It also includes a real-time clock (RTC) unit and a LoRa module for sending data to the local control nodes. Fig.2a shows the schematic of a sensor nodes without the CO_2 sensor which is connected to the UART pins via a short cable.

Local Control Node Hardware. Similar to the sensor nodes, the local control node also contains a RTC unit, a PIC microcontroller and a LoRa module. Besides, this node is equipped with a LCD and buttons for getting user inputs locally and displaying the measurement on demand. An EEPROM is integrated to store the operation settings. The local control node is able to change fan speed set point by setting via 4 I/O digital pins. The mainboard schematic of this node is shown in Fig. 2b and has the options of being equipped with a LCD interface and buttons placed on its casing.
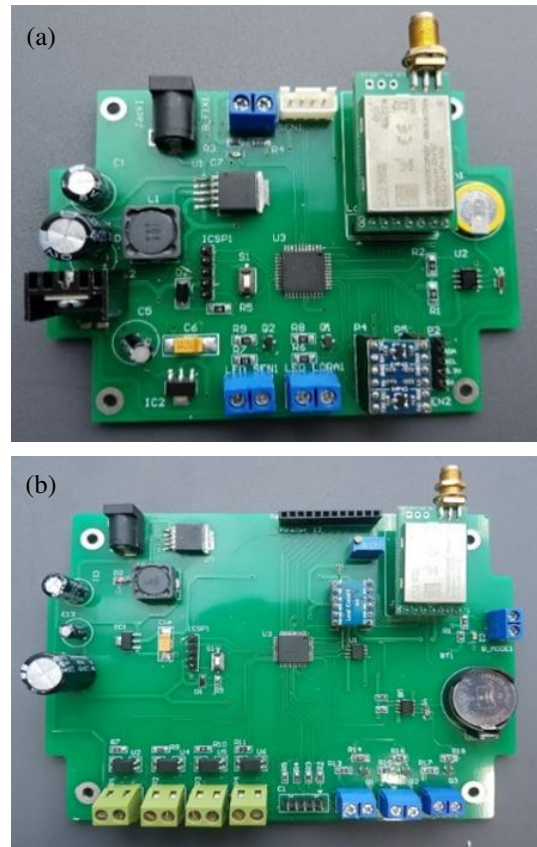
Fig 3. Hardware platform of (a) sensor node and (b) the local control node

Key components of the two nodes are listed in Table 1. The high energy consuming components are the PIC microcontroller, MHZ-16 sensor modules and the LCD at the local control node. In the deployment area, the nodes are supplied by main power via AC/DC converters, thus energy is not yet a concern. However, to reduce energy consumption, especially when main power is hard to reach at some location inside the barn and nodes are supplied by batteries, the LCD is kept turned off most of the time and is energized only if a user touches the buttons on the local control node. Similarly, other components are put in sleep or

energy saving mode if not in use. Implementation of sensor nodes and local control nodes is illustrated in Fig. 3a and 3b respectively.

TABLE I.   MAIN COMPONENTS OF THE HARDWARE PLATFORMS

Components	Function	Interface	Power Consumption
PIC18F46K22	Microcontroller	UART, I2C, SPI	50 - 75 mW
DS1307ZN+	RTC	I2C	10 mW
Module Lora E32	Communication	UART	100 mW
MHZ-16	Sensor	UART	500 mW
AT24C256	EEPROM	I2C	15 mW
LCD 2004	Display	I2C	5 W

B. Supervisory and actuation schemes

Local Control Node Information perceives by all the sensor nodes in the same barn is aggregated at the local controller and only the average values are forwarded to control center so that the traffic load in the network is reduced. In order to avoid packet collision within the sub network, a communication scheme similar to time division multiple access (TDMA) is adopted. Initially, local control node broadcasts a message to perform time synchronization amongst the nodes and set value of the RTC. The sensor nodes are also allocated time slots to send data to the local control node. Operation settings of the sub network can be provided by users locally via LCD and buttons on the local control node or sent from the control center software. Based on this input information and the measurements from the sensor node, the local control node makes decision on how the ventilation fans should be controlled.

Control Center Software. The software includes a graphical user interface (GUI) and a MySQL database which stores all the historical data. The GUI is designed and developed with PyQT5 to supervise the whole system. It enables the user to monitor the farm parameters as well as to input settings remotely at the management room in real-time. Initially, the settings of the desire conditions in the farm such as range of temperature or CO_2 concentration are sent from control center room. The settings are then recorded in local control node's EEPROM. In case the system is restarted and no input from the control center is provided, those store values in EEPROM will be used. All the system information as well as acquired data are recorded in the MySQL database with a schema consisting the barn information, the settings and the collected measurements as shown in Fig. 4.
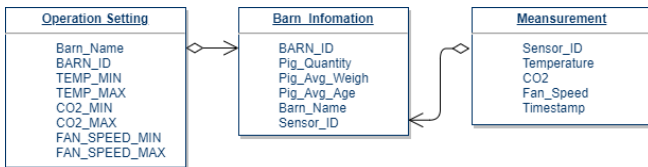

Fig 4. Database scheme diagram

Actuation Scheme. The system is constantly monitored and controlled remotely to ensure the proper living environment for animals. Key operation settings include threshold of parameters such as maximum and minimum feasible fan speed, SP_{max} and SP_{min}; maximum and minimum temperature, T_{max} and T_{min}; maximum and minimum CO_2 concentration $CO2_{max}$ and $CO2_{min}$. A

simplified control scheme is adopted to regulate the fan speed. The fan speed is calculated as the following:

$$\Delta_{Fan_{CO_2}} = \frac{(CO2_{measured} - CO2_{min})(SP_{max} - SP_{min})}{(CO2_{max} - CO2_{min})} \quad (1)$$

$$\Delta_{Fan_T} = \frac{(T_{measured} - T_{min})(SP_{max} - SP_{min})}{(T_{max} - T_{min})} \quad (2)$$

From (1) and (2), we get:

$$\Delta_{Fan} = \max(\Delta_{Fan_{CO_2}}, \Delta_{Fan_T}) \quad (3)$$

The final fan speed value is computed as:

$$SP_{cal} = SP_{min} + \Delta_{Fan} \quad (4)$$

The calculated value is then set to the VFD to adjust the fans via 4 digital input pins. This setting mechanism allow 16 levels of the speed from 0 to 15 with respected to 0% to 100% full speed linearly. The closest level to the calculated value is selected as VFD set point. If both $CO2_{measured}$ and $T_{measured}$ are lower than the $CO2_{min}$ and T_{min}, fans are turned off. On the other hand, if any measurement is greater than the highest threshold, fan speed is set at full speed.

IV. SYSTEM OPERATION AND EXPERIMENTAL RESULTS

The system in this work is designed for but not limited to a farm of 9 hog barns with different area as shown in Fig. 5. The average dimension of each barn is 50m×18m. A management room is located around 150-200m away from the furthest barn. There are 4 or 6 ventilation fans with rated power is around 0.8-1.2kW installed in each barn.

Fig 5.   The livestock farm layout

At testing phase, a sub network of 2-5 nodes is deployed in each barn. Each node is set to perceive measurement values and send to the local control node every minute in sequence. The local control node then gathers the information, process and forward the average values to the control center in the next one minute.

The GUI displays are illustrated in Fig 6. and Fig. 7. Various parameters of the system are shown in Fig. 6 include real-time measurements of average temperature and CO2 concentration in each barn as well as extra information of the livestock in the barn like the quantity, average age and weight. Historical data of the CO2 and temperature can be extracted from the database and displayed as shown in Fig. 7 together with the operation settings.
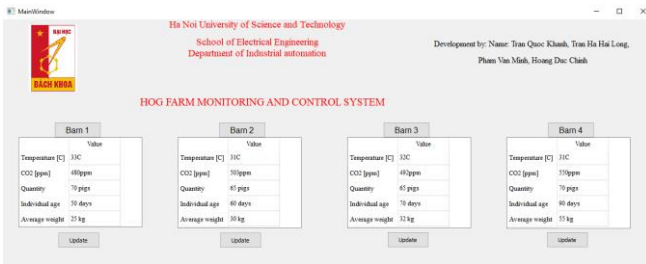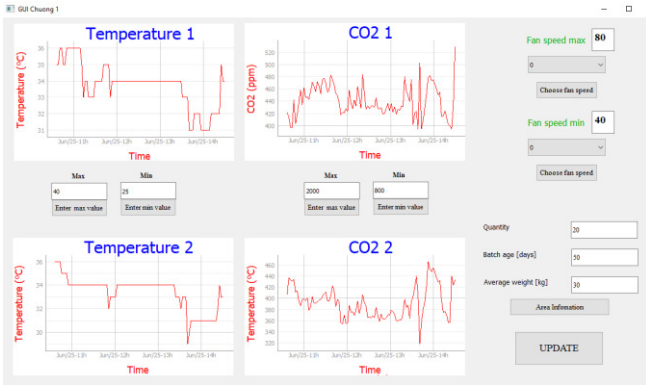
Fig 6. Main screen of the GUI

Fig 7. Historical data and setting values in the GUI

Fig. 8. shows the data during daytime on a typical summer day. The temperature can rise as high as 37°C and is kept above 29°C throughout the day. The variation of fan speed set point is shown in Fig. 9. The set point is recalculated over the time whenever the local control node receives sensing values from all the sensor nodes, the period is around 5 minutes.
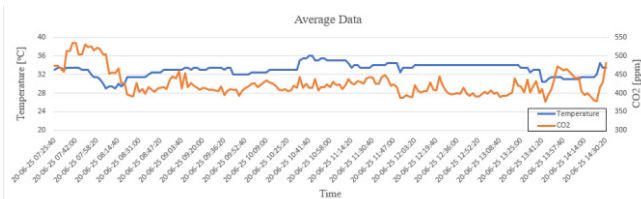

Fig 8. Average temperature and CO2 measurements aggregated by the local control node
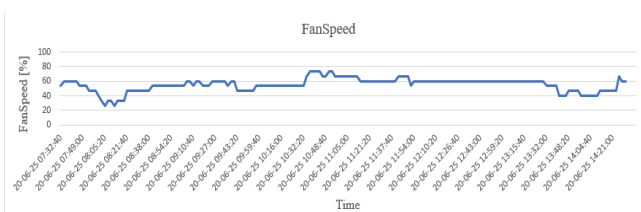

Fig 9. Fan speed set point as calculated

V. CONCLUSIONS

An energy efficient long range low cost LoRa based wireless sensor and actuator network is presented for the application of smartly managing indoor conditions in livestock farming application. Temperature and carbon dioxide measurement data are continuously collected and displayed at the control center room. These real-time data are used to compute the required variation in the fan speed setting in a farm with multiple barns, so as to pump fresh air into the barn as well as to cool down the indoors to improve comfort of the livestock. Our practical results from deployment show that the systems are able to facilitate the farm operation reliably and efficiently. Future work includes

enhancements to customize the control scheme using the livestock information such as quantity, age or weight of the livestock in each barn. Internet of Things (IoT) edge computing and blockchain technologies [13, 14] can also be explored and integrated. Thus, they can provide holistic solutions for modernization of livestock applications.

REFERENCES

[1] Tamoghna Ojha, Sudip Misra, N. Singh Raghuwanshi: "Wireless sensor networks for agriculture: The state of the art in practice and future challenges". In: Computers and Electronics in Agriculture, vol. 118, pp. 66-84 (2015)

[2] T. Wark, D. Swain, C. Crossman, P. Valencia, G. Bishop Hurley and R. Handcock: " Sensor and Actuator Networks: Protecting Environmentally Sensitive Areas ". In: IEEE Pervasive Computing, vol. 8, no. 1, pp. 30-36 (2009)

[3] Usmonov, M., Gregoretti, F.: "Design and implementation of a LoRa based wireless control for drip irrigation systems". In: 2017 2nd International Conference on Robotics and Automation Engineering (ICRAE), pp. 248–253 (2017).

[4] Jennifer Yick, Biswanath Mukherjee, Dipak Ghosal: "Wireless sensor network survey". In: Computer Networks, vol. 52, issue 12, pp. 2292-2330 (2008)

[5] A. Molina Garcia, J. A. Fuentes, E. Gomez Lazaro, A. Bonastre, J. C. Campelo and J. J. Serrano: "Development and Assessment of a Wireless Sensor and Actuator Network for Heating and Cooling Loads". In: IEEE Transactions on Smart Grid, vol. 3, no. 3, pp. 1192-1202 (2012)

[6] Antonio Franco Salas, Diego Luis Valera, Araceli Peña: "Energy Efficiency in Greenhouse Evaporative Cooling Techniques: Cooling Boxes versus Cellulose Pads". In: Energies 7(3):1427-1447 (2014)

[7] Martin C. Bor, Utz Roedig, Thiemo Voigt, and Juan M. Alonso: "Do LoRa Low Power Wide Area Networks Scale". In: Proceedings of the 19th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM '16). Association for Computing Machinery, NY, USA, pp. 59–67 (2016)

[8] U. Noreen, A. Bounceur and L. Clavier: "A study of LoRa low power and wide area network technology" . In: 2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Fez, pp. 1-6 (2017)

[9] Meonghun Lee, Haengkon Kim, Ha Jin Hwang, Hyun Yoe: "IoT Based Management System for Livestock Farming". In: Computer Science and Ubiquitous Computing, CUTE 2018, CSA 2018, LNEE, vol. 536, pp. 195-200, Springer Singapore (2019).

[10] Qiulan Wu, Chuanqi Zhao, Yong Liang, Dalei Zhang, Junmeng Hao: "Farmland Information Acquisition System Based on LoRa Wireless Sensor Network". In: Computer and Computing Technologies in Agriculture XI (CCTA 2017), IFIP Advances in Information and Communication Technology, vol. 545, pp. 529-539 Springer (2017)

[11] St. Penkov, A. Taneva, M. Petrov, On-line Monitoring System with LoRaWAN, IFAC-PapersOnLine, vol. 52, 25, 2019, pp. 533-538

[12] Meonghun Lee, Haengkon Kim, Ha Jin Hwang, Hyun Yoe: "IoT Based Management System for Livestock Farming". In: Computer Science and Ubiquitous Computing, CUTE 2018, CSA 2018, LNEE, vol. 536, pp. 195-200, Springer Singapore (2019).

[13] Hajder, Piotr & Hajder, Mirosław & Liput, Mateusz & Nycz, Mariusz. : "Direct communication of edge elements in the Industrial Internet of Things". In Communication Papers of the 2020 Federated Conference on Computer Science and Information Systems, ACSIS, vol. 23, pp. 35–42 (2020)

[14] Puri, V., Priyadarshini, I., Kumar, R. et al.: "Smart contract based policies for the Internet of Things". In Cluster Computing, vol. 24, pp. 1675–1694 (2021)

A Multi-criteria Fuzzy Random Crop Planning Problem using Evolutionary Optimization

Dao Minh Hoang
School of Applied Mathematics and Informatics
Hanoi University of Science and Technology
Hanoi, Vietnam
Hoang.DM202895M@sis.hust.edu.vn

Tran Van Xuan
School of Applied Mathematics and Informatics
Hanoi University of Science and Technology
Hanoi, Vietnam
Xuan.TV212049M@sis.hust.edu.vn

Tran Ngoc Thang
School of Applied Mathematics and Informatics
Hanoi University of Science and Technology
Hanoi, Vietnam
thang.tranngoc@hust.edu.vn

Abstract—In this paper, we deal with fuzzy random objectives in a multi-criteria crop planning problem considered as a multi-objective linear programming problem. These fuzzy random factors are related to decision making processes in practice, especially the uncertainty and synthesized objectives of experts. The problem is transformed into a multi-objective nonlinear programming problem by a step of evaluating expectation value. Instead of using classical methods, we use a multi-objective evolutionary algorithms called NSGA-II to solve the equivalent problem. This helps finding many approximate solutions concurrently with a low time consumption. In computational experiments, we create a specific fuzzy random crop planning problem with the data synthesized from government's reports and show convergence of the algorithm for proposed model.

Index Terms—Fuzzy random coefficients, Crop planning problem, Multi-objective programming, Evolutionary algorithms

I. Introduction

IN PRACTICAL optimization problems, the unknown and uncertain factors are inevitable. They present the errors in measuring processes, the randomness or vagueness of data or expert's knowledge, the approximation or uncertainty in decision making and inference processes,... These factors affect the parameters of both objective functions and constraints in programming problems. There are some approaches for single objective problem of stochastic programming, for instance probabilistic programming (Vajda [18]), fuzzy programming (Sakawa [13]), fuzzy random programming (Luhandjula and Gupta [9]). Then these studies are extended for multi-objective stochastic programming problem (see Sakawa [13], Katagiri et al. [8], Yano et al. [21], [22] and references therein). In these works, the authors consider the multi-objective linear programming problem including fuzzy/ random or both fuzzy random coefficients. The stochastic problems are transformed by probabilistic models or fuzzy models to the deterministic variants. These problems are usually nonlinear and nonconvex problems. They are solved by deterministic optimization algorithms in interactive approaches to find some efficient solutions. In this research, we also consider multi-objective

fuzzy random linear programming modeled by the expectation model [8] and solve it by efficient multi-objective evolutionary algorithms.

Crop planning problem is a very popular and highly concerned in agricultural countries. Besides regularly maximizing the net revenue, the managers may deal with other objectives like gross revenue, water consumption, erosion, labor,... with restrictions of land, labor, min-max yield requirement. A wide survey of optimization techniques for crop planning was studied by Jain et al. [7] and provides views on previous research about this problem. Derived from reality, its coefficients are hard to be specified exactly, therefore fuzzy approaches including fuzzy goals are usually engaged [10], [14] and also fuzzy random approaches [17], [21], [22]. Our stochastic model for crop planning problem has some differences from the models of Yano et al. [21] and Toyonaga et al. [17]. While these models dealt with possibility measure directly, our main focus is discrete randomness of random fuzzy variables. It comes from the decision making process of experts in practice which is presented specifically in Section 2.2.

Due to the advantages of parallel computing as well as compatibility with various types of objectives and constraints, multi-objective evolutionary algorithms (MOEA) which are based on evolutionary algorithms have been advanced for decades [6], [19] to deal with multi-objective optimization problems. The essential attribute of these algorithms is population-based nature, which grants the ability to generate many non-dominated solutions simultaneously, therefore they can gradually converge toward Pareto solutions and present an approximate Pareto front at each iteration loop if well-established. With the significant performance when compared to later algorithms [4], NSGA-II [2] has been a standard algorithm used to evaluate new methods. For that reason, we choose NSGA-II represented for MOEA to solve our crop planning problem, which is different from the deterministic methods in [17], [21].

The paper content is summed up as follows: Section 2 for-

mulates our crop planning problem in form of a fuzzy random multi-objective linear programming which is transferred to an equivalent form of degree of possibility. An expectation model is defined in Section 3 as a means to handle the stochastic problem. The algorithm NSGA-II is presented in fourth section with the aim of solving the expectation model. Computational experiments are shown in the next one and the last one presents a conclusion.

II. CROP PLANNING AS A FUZZY RANDOM MULTI-OBJECTIVE LINEAR PROGRAMMING PROBLEM

A. Crop planning problem

Crop planning problem (CPP) is a problem of considering what kind of crops should be planned on a predetermined cultivated area in order to archive optimal benefits of predefined objectives. Suppose that n is the number of disparate crops whose planned areas denoted by x_i, $i = 1, \ldots, n$. In practice, agricultural managers often determine several essential objectives like revenue, time consumption, labor, water consumption, erosion [7], we assume that k is the number of objectives. Then the problem is formulated as follows

$$\text{Min } Cx = (c_1 x, \ldots, c_k x) \qquad (1)$$
$$\text{s.t. } x \in X$$

where $x = (x_1, \ldots, x_n)^T$ and $c_i = (c_{i1}, \ldots, c_{in})$, $i = 1, \ldots, k$ are the coefficients of ith objective function. The feasible set X may contain constraints that are relevant to cultivation like land, labor, water, min-max yield,...

B. Crop Planning Problem with fuzzy random coefficients

We formulate crop planning problem in the same way likes Yano et al. [20] with the random and fuzzy transformation processes dealing with practical needs. Specifically, in an expert-based decision system, we need to quantify reliability of each expert after gathering estimated coefficients in optimization model. Also, we have to evaluate modeling error as well, therefore it is essential to associate both random and fuzzy approaches in order to similarly fit the actual model.

We assume that there is not only one set of values for coefficients vectors of objective functions but an expert group's decision with their own distinct reliability. Denoting E_i as the number of experts who specify coefficients vector of ith objective function, p_{ie} $(e \in \{1, 2, ..., E_i\})$ is the quantified reliability of expert e which satisfies

$$\sum_{e=1}^{E_i} p_{ie} = 1, \quad i = 1, 2, ..., k. \qquad (2)$$

Subsequently, by denoting that random attribute by symbol "$-$", linear programming problem (1) becomes

$$\text{Min } \overline{C}x = (\bar{c}_1 x, \ldots, \bar{c}_k x) \qquad (3)$$
$$\text{s.t. } x \in X,$$

where $\bar{c}_i = (\bar{c}_{i1}, \ldots, \bar{c}_{in})$, are estimated by expert E_i, $i = 1, 2, ..., k$. Each the value $\bar{c}_{ij}$ has its own values c_{ije} evaluated by expert e $(e \in \{1, 2, ..., E_i\})$ with the reliability p_{ie}.

By examining each estimated c_{ije} $(e \in \{1, 2, ..., E_i\})$ values's bias caused by miscalculation of human, we regard estimated coefficients by each expert as symmetric triangular fuzzy variables with membership function in form

$$\mu_{\tilde{c}_{ije}}(t) = \max\left\{0, 1 - \frac{|t - c_{ije}|}{\gamma_{ij}}\right\}, \qquad (4)$$

where $\tilde{c}_{ije}$ are fuzzy extension variables of c_{ije}, γ_{ij} are positive constant which denote the spread of fuzzy numbers. With implementation of Zadeh's extension principle, each objective function becomes fuzzy random variable and has membership function as follows

$$\mu_{\tilde{c}_i x}(y) = \max\left\{0, 1 - \frac{|y - \bar{c}_i x|}{\gamma_i x}\right\}, \quad i = 1, 2, ..., k, \qquad (5)$$

where $\gamma_i = (\gamma_{i1}, \gamma_{i2}, ..., \gamma_{in}) > 0$. Therefore, problem (3) transforms into

$$\text{Min } \widetilde{\overline{C}}x = (\tilde{c}_1 x, \ldots, \tilde{c}_k x) \qquad (6)$$
$$\text{s.t. } x \in X,$$

where $\tilde{\bar{c}}_i = (\tilde{\bar{c}}_{i1}, \ldots, \tilde{\bar{c}}_{in})$, $i = 1, \ldots, k$ and $\tilde{\bar{c}}_{ij}$ is a fuzzy random variable with its own fuzzy numbers set which represents expert estimation.

Furthermore, decision makers often expect objective functions to reach desired values as close as possible in practice. Consequently, we can define a set of fuzzy goal functions corresponding to objective functions to assess the quality of each solution, the formula is expressed by

$$\mu_{\tilde{G}_i}(y) = \begin{cases} 1 & y < \delta_i^1 \\ \frac{y - \delta_i^0}{\delta_i^1 - \delta_i^0} & \delta_i^1 \leq y \leq \delta_i^0 \\ 0 & y > \delta_i^0, \end{cases} \qquad (7)$$

where δ_i^0 is the maximum acceptable value that i^{th} objective function are not expected to exceed and δ_i^1 is the upper bound of most effective range that i^{th} objective function is required to reach. By utilizing the notion of degree of possibility, the level that objective function $\tilde{\bar{c}}_i x$ satisfies the fuzzy goal $\tilde{G}_i$ is presented as

$$\Pi_{\tilde{\bar{c}}_i x}\left(\tilde{G}_i\right) = \sup_y \min\left\{\mu_{\tilde{\bar{c}}_i x}(y), \mu_{\tilde{G}_i}(y)\right\}, i = 1, ..., k. \qquad (8)$$

In consequence, problem (6) can be regarded as

$$\text{Max } \Pi_{\tilde{\bar{c}}_i x}\left(\tilde{G}_i\right), \quad i = 1, \ldots, k \qquad (9)$$
$$\text{s.t. } x \in X$$

III. EXPECTATION MODEL

To deal with problem (9), Katagiri et al. [8] introduced E-model which maximizes the expectation of possibility measure. Considering the ith objective function estimation of expert e $(e \in \{1, 2, ..., E_i\})$ in (5) and fuzzy goal (7), possibility measure of (8) in that scenario becomes

$$\Pi_{\tilde{c}_{ie} x}\left(\tilde{G}_i\right) = \frac{(\gamma_i - c_{ie}) x + \delta_i^0}{\gamma_i x - \delta_i^1 + \delta_i^0}, \quad i = 1, 2, ..., k \qquad (10)$$

Thus, the expectation evaluation of (8) is calculated as follows

$$E\left[\Pi_{\tilde{\tilde{c}}_i x}\left(\tilde{G}_i\right)\right] = \sum_{e=1}^{E_i} p_{ie}\Pi_{\tilde{c}_{ie}x}\left(\tilde{G}_i\right)$$

$$= \frac{\left(\gamma_i - \sum_{e=1}^{E_i} p_{ie}c_{ie}\right)x + \delta_i^0}{\gamma_i x - \delta_i^1 + \delta_i^0}. \quad (11)$$

By notating $Q_i^E = E\left[\Pi_{\tilde{\tilde{c}}_i x}\left(\tilde{G}_i\right)\right]$, we examine problem (9) by an expectation maximizing model as follows

$$\text{Max } Q_i^E(x), \quad i = 1, \dots, k, \quad (12)$$
$$\text{s.t. } x \in X.$$

IV. THE MULTI-OBJECTIVE EVOLUTIONARY ALGORITHM SOLVING CROP PLANNING PROBLEM

In this section, we use the evolutionary algorithm (EA) approach NSGA-II [2] (Non-dominated sorting genetic algorithm II) to deal with the equivalent crop planning problem as a multi-objective programming problem (12), instead of using the deterministic algorithms (see [15], [16] and references therein) or interactive methods (see [17], [21]). The reason is that EAs simultaneously find many approximate solutions (see [1], [11]) without analysis of objective functions and have speedy performance when compared with classical methods. Decision makers can easily choose the optimal solution from this approximated solution set, for instance, by adding a sub-criteria as optimizing a function over the finite set of approximated solutions. And NSGA-II has been one of the state-of-the-art evolutionary algorithms for years. The NSGA-II's principal idea is presented briefly as follows

After initializing a random solution set which is a N sized parent population P_0, the algorithm launches into a iterations loop with t^{th} parent population P_t as input. Firstly, P_t creates a same size offspring population Q_t by implementing genetic operators: selection, crossover, mutation. Next, P_t and Q_t are merged to construct the $2N$ sized population R_t in order to maintain the elitism of populations. A procedure to classify R_t into a set of ranked non-dominated fronts $\mathcal{F}$ is performed afterwards, before the second one selects N best solutions from $\mathcal{F}$ as the next generation population P_{t+1} due to low-rank priority and proposed crowding distances measurement which partly produces solutions diversity in objective space. This iteration loop ends when termination conditions are met.

V. COMPUTATIONAL EXPERIMENTS

In this section, implementation of NSGA-II is illustrated in a particular example of crop planning problem.

TABLE I
MIN-MAX QUANTITY LIMITS IN 4 KINDS OF CULTIVATED FARMS.

Type	Min quantity(tons)	Max quantity(tons)
Rice	$22,000,000$	∞
Vegetable	$6,000,000$	$12,000,000$
Rice	$3,872,000$	$7,000,000$
Shrimp	$850,000$	$6,000,000$

TABLE II
PARAMETERS OF OBJECTIVE FUNCTIONS.

	$e=1$	$e=2$	$e=3$	γ_{ij}
c_{11e}	-66	-63	-64	5
c_{12e}	-180	-185	-177	4
c_{13e}	-294	-300	-295	5
c_{14e}	-384	-386	-380	3
p_{1e}	0.5	0.3	0.2	
c_{21e}	83	88	85	3
c_{22e}	231	225	234	4
c_{23e}	65	64	68	3
c_{24e}	127	125	125	3
p_{2e}	0.6	0.2	0.2	

In Mekong Delta of Vietnam, there are many types of cultivate farms, but for a concise example we only consider four main kinds including rice, vegetable, fruit and shrimp farm. By denoting their farming land areas as $x = (x_1, x_2, x_3, x_4)^T$, we deal with an multi-objective optimization problem modeled in form of (6) with two objectives: revenue and labor. With 2019 data collected from Vietnam MARD[1]'s reports, we convert restriction of land areas and min-max quantity limits described in Table I to problem's constraints also the coefficients of objective functions after dividing them by same numbers in order to get small values without changing the ratios. Therefore, the crop planning problem is represented as follows

$$
\begin{aligned}
\min \quad & \tilde{\tilde{c}}_1 x = \tilde{\tilde{c}}_{11}x_1 + \tilde{\tilde{c}}_{12}x_2 + \tilde{\tilde{c}}_{13}x_3 + \tilde{\tilde{c}}_{14}x_4 \\
\min \quad & \tilde{\tilde{c}}_2 x = \tilde{\tilde{c}}_{21}x_1 + \tilde{\tilde{c}}_{22}x_2 + \tilde{\tilde{c}}_{23}x_3 + \tilde{\tilde{c}}_{24}x_4 \\
\text{s.t.} \quad & x_1, x_2, x_3, x_4 \in \mathbb{R}, \\
& x_1 + x_2 + x_3 + x_4 \leq 35000, \\
& -6.11x_1 \leq -22000, \\
& -1.42x_2 \leq -6000, \\
& 1.42x_2 \leq 12000, \quad (13) \\
& -1.06x_3 \leq -3872, \\
& 1.06x_3 \leq 7000, \\
& -0.12x_4 \leq -850, \\
& 0.12x_4 \leq 6000,
\end{aligned}
$$

where $\tilde{\tilde{c}}_{ij}$ $(i=1,2; j=1,2,3,4)$ are fuzzy random coefficients specified by three experts with corresponding possibility, spread values given in Table II and the parameters of fuzzy goal functions are shown in Table III. With denoting feasible set in (13) as X for short, we transform (13) into an E-model described in (12)

$$\text{Min } \left(-Q_1^E(x), -Q_2^E(x)\right) \quad (14)$$
$$\text{s.t. } x \in X.$$

For solving multi-objective optimization problem (14) by NSGA-II, we initialize genetic parameters including: crossover probability is 0.5, mutation probability is 0.5, population size is 100, simulated binary crossover, random reset mutation, max iteration is 200. To observe improvement of NSGA-II, we collect populations of 4 iterations and visualize their objective values in Fig. 1, owing to the convergence of population that can't get better after iteration 200. We can

[1]Ministry of Agriculture and Rural Development: https://mard.gov.vn/.

TABLE III
PARAMETERS OF FUZZY GOAL FUNCTIONS.

	δ_i^0	δ_i^1
$i = 1$	-5000000	-11000000
$i = 2$	5000000	2000000

TABLE IV
HYPERVOLUME OF POPULATIONS.

Iteration	Hypervolume
25	0.1499
50	0.1649
100	0.1976
200	0.2293

notice that after converging to Pareto optimal front at iteration 100, NSGA-II enhances diversity of population and eventually creates last population marked by black pentagram in large range distribution which provides wide range strategies for making decisions with different circumstances. We also use the hypervolume metric [12] to quantify the result which shown in Table IV. The algorithm is deployed in MATLAB and consumed 68.6 seconds for 200 iterations on Intel Core i7-5600U which is a bit slow cause of recreation of feasible solutions in genetic phases.
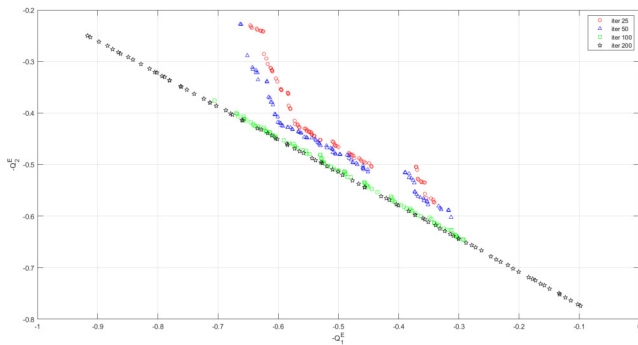

Fig. 1. NSGA-II generations on the proposed crop planning problem.

VI. CONCLUSION

In this paper, we use a multi-objective evolutionary algorithms - NSGA-II to solve the fuzzy random crop planning problem with the data analysed from ministry's reports. We construct a crop planning problem based on uncertainty of experts for collecting data, synthesizing objective suggestions and making decisions. The final result presents the convergence of the algorithm and the diversity's improvement afterwards giving decision makers multiple plans to choose while balancing examined objectives. Based on considering the stochastic factors, the cropping arrangement is more flexible in different conditions. In future, we shall model the fuzzy random multi-objective linear programming problem by other stochastic models for example variance or probability one, as well as establish more efficient algorithms to solve.

REFERENCES

[1] S. Fidanova, M. Ganzha, O. Roeva, "InterCriteria Analyzis of Hybrid Ant Colony Optimization Algorithm for Multiple Knapsack Problem," 16th Conference on Computer Science and Intelligence Systems, vol. 25, 2021, pp.173–180.

[2] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," IEEE Transactions on Evolutionary Computation, vol. 6, 2002, pp.182–197.

[3] Z. Du, K. Chen, "Enhanced Artificial Bee Colony with Novel Search Strategy and Dynamic Parameter," Computer Science and Information Systems, vol. 16, no. 3, 2019, pp. 939–957.

[4] S. Garcia, C. Trinh, "Comparison of Multi-objective Evolutionary Algorithms to Solve the Modular Cell Design Problem for Novel Biocatalysis," in Processes, vol. 7, 2019, p. 361.

[5] A.O. Hamadameen, Z. M. Zainuddin, "Multiobjective fuzzy stochastic linear programming problems in the 21st century," Life Science Journal, vol. 10, 2013, pp. 616–647.

[6] W. Huang, Y. Zhang, L. Li, "Survey on Multi-Objective Evolutionary Algorithms," Journal of Physics: Conference Series, vol. 1288, 2019.

[7] R. Jain, L. Malangmeih, S. S. Raju, S. K. Srivastava, K. Immaneulraj, A. P. Kaur, "Optimization techniques for crop planning: A review," Indian Journal of Agricultural Sciences, vol. 88, 2018, pp. 1826–1835.

[8] H. Katagiri, M. Sakawa, H. Ishii, "Multiobjective fuzzy random linear programming using E-model and probability measure," Joint 9-th IFSA World Congress and 20-th NAFIPS International Conference, vol. 4, 2001, pp. 2295–2300.

[9] M. K. Luhandjula, M. M. Gupta, "On fuzzy stochastic optimization," Fuzzy Sets and Systems, vol. 81, 1996, pp. 47–55.

[10] S. A. Mortazavi, R. Hezareh, S. A. Kaliji, S. S. Mehr, "Application of Linear and Non-linear Programming Model to Assess the Sustainability of Water Resources in Agricultural Patterns," International Journal of Agricultural Management and Development, vol. 4, 2014, pp. 27–32.

[11] J. L. Pachuau, A. Roy, G. Krishna, H. K. Saha, "Estimation of traffic matrix from links load using genetic algorithm," Scalable Computing: Practice and Experience, vol. 22, no. 1, 2021, pp. 29–38.

[12] N. Riquelme, C. V. Lücken, B. Baran, "Performance metrics in multi-objective optimization," Latin American Computing Conference, 2015, pp. 1–11.

[13] M. Sakawa, "Fuzzy Sets and Interactive Multiobjective Optimization," in Plenum, 1993.

[14] J. Soltani, A. R. Karbasi, S. M. Fahimifard, "Determining optimum cropping pattern using Fuzzy Goal Programming (FGP) model," African Journal of Agricultural Research, vol. 6, 2011, pp. 3305–3310.

[15] T. N. Thang, D. T. Luc, N. T. B. Kim, "Solving generalized convex multiobjective programming problems by a normal direction method," A Journal of Mathematical Programming and Operations Research, vol. 65, 2016, pp. 2269-2292.

[16] T. N. Thang, V. K. Solanki, D. T. Anh, N. T. N. Anh, P. V. Hai, "A monotonic optimization approach for solving strictly quasiconvex multiobjective programming problems," Journal of Intelligent & Fuzzy Systems, vol. 38, 2020, pp. 6053-6063.

[17] T. Toyonaga, T. Itoh, H. Ishii, "A Crop Planning Problem with Fuzzy Random Profit Coefficients," Fuzzy Optimization and Decision Making, vol. 4, 2005, pp. 51–69.

[18] S. Vajda, "Probabilistic Programming," Academic Press, 1972.

[19] Q. Xu, Z. Xu, T. Ma, "A Survey of Multiobjective Evolutionary Algorithms Based on Decomposition: Variants, Challenges and Future Directions," IEEE Access, vol. 8, 2020, pp. 41588–41614.

[20] H. Yano, K. Matsui, M. Furuhashi, "Multiobjective fuzzy random linear programming problems based on coefficients of variation," International Journal of Applied Mathematics, vol. 44, 2014, pp. 137–143.

[21] H. Yano, M. Sakawa, "Interactive fuzzy decision making for multiobjective fuzzy random linear programming problems and its application to a crop planning problem," Computational Intelligence, vol. 577, 2015, pp. 143–157.

[22] H. Yano, "Interactive Multiobjective Decision Making Under Uncertainty," CRC Press, 2017.

Hardware Trojan Detection Based on Side-Channel Analysis Using Power Traces and Machine Learning

Van-Phuc Hoang
Le Quy Don Technical University, 236 Hoang Quoc Viet Str., Hanoi, Vietnam
Email: phuchv@lqdtu.edu.vn

Abstract—**With the continuous development of the Integrated Circuit (IC) manufacturing where international outsourcing is one of the main trends, hardware Trojan (HT) has been considered as a serious problem for hardware security in modern electronic systems. This paper presents a novel HT detection method based on the side-channel analysis with power traces and the machine learning (ML) technique. Side-channel information of the AES encryption core was acquired by the power consumption measurement equipment and then classified with Softmax regression. The ML technique was applied to classify and detect the HT. The experimental results have clarified the efficiency of the proposed method.**

IndexTerms—**Hardware Trojan, power traces, SCA, machine learning**

I. Introduction

RECENTLY, the issues of cybersecurity and hardware oriented security become very critical [1], [2], [3]. Specifically, in the field of semiconductor, almost Integrated Circuit (IC) vendors aim to outsource different steps in the chip production cycle to different companies from different countries so that the production cost and time can be reduced. On the other hand, this business model of semiconductor industry also leads to the threat of hardware security including hardware Trojan (HT).

By definition, a HT is a malicious hardware module inserted in the ICs during any step of the IC design or fabrication cycle [4]. An HT consists of two components: Trigger and Payload, as shown in Fig. 1. The Trigger is the condition (such as the value of S_2S_1 in this example) so that the HT becomes active from the inactive state. On the other hand, the Payload performs the function of the HT. Once inserted in an IC, the HT can execute number of dangerous operations such as Denial of Service (DoS), extracting the secret information (for example, private cipher key) or changing the circuit behavior, etc. HT designs are often difficult for detecting either accidentally through production testing or deliberately using specially designed tests which can activate and detect HTs. The advanced HT insertion methods also allow the resistance to popular HT detection techniques using high-resolution side-channel signals such as power dissipation data, electromagnetic emission (EM), computation latency and temperature. As shown in Fig. 2, the attacker can insert HTs in some steps of the IC fabrication cycle so that different HT detection methods are required, respectively.
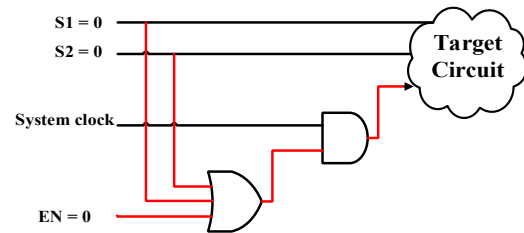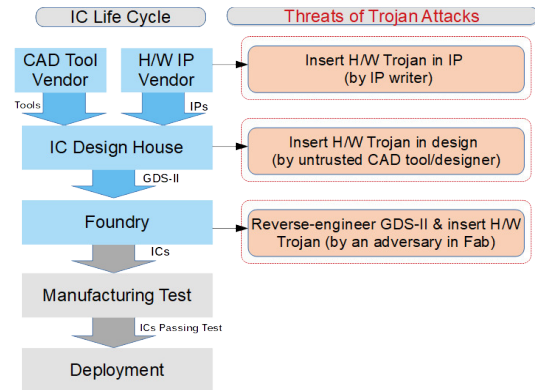

Fig. 1. A minimalist HT example.

Fig. 2. The threat of HT at different steps in the IC fabrication cycle, adapted from [4].

Basically, the IC manufacturers also perform the chip testing when the production cycle is completed. However, the HT is often designed with too small size and activated by very specific conditions. Hence, the normal testing methods can not detect the HT effectively, this leads to the requirement of efficient HT detection methods. There are number of papers in literature mentioning HT detection techniques. However, there are very few papers concerning the use of machine learning (ML) techniques for HT detection. On the other hand, the recent advancements of ML techniques have inspired the researchers to apply these techniques in a broad range of applications. Therefore, this paper targets the feasibility study and experiments of ML assisted techniques for HT detection.

The remainder of this paper is organized as follows. In Section 2, the existing HT detection techniques will be introduced briefly. Then, the proposed HT detection based on

power traces and machine learning technique is described in Section 3, together with the implementation results. Finally, Section 4 concludes the paper and proposes some ideas for our future work.

II.   EXISTING TECHNIQUES FOR HARDWARE TROJAN DETECTION

Currently, detection and prevention are two main categories to protect the embedded systems from the risk of HTs [4]. Prevention consists of modifying the original circuit during the conception phase to provide a secure design (against HT), to support one particular HT detection method or to make a trusted IC/chip production chain. On the other hand, detection includes techniques to clarify the presence of HTs in the design. The summary of the existing techniques to detect the HT is shown in Fig. 3. In the destructive methods, the reverse engineering is often employed to extract the circuit netlist or layout. The reconstruction of circuit layers is performed with the methods using chemical or optical principles. Hence, the destructive techniques can detect HT in the circuit with very high accuracy. However, the disadvantages of these destructive techniques are the high cost, long time and that the requirement of destroyed tested circuits so that these circuits cannot be re-used. Therefore, the non-destructive techniques are attracting most of researchers in literature. Consequently, two types of test-time and run-time techniques are often applied for non-destructive techniques.
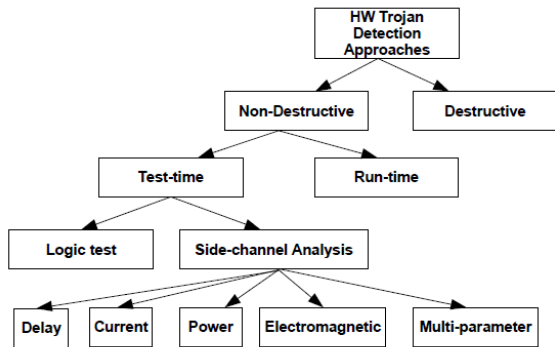

Fig. 3. Classification existing techniques for HT detection.

Side-channel analysis (SCA) is a non-destructive technique providing many potentials to improve the HT detection performance. SCA techniques use the side-channel information from the HT infected circuit and compare with side-channel information from the HT free circuits (or reference circuits). Side-channel information is divided into two groups: Energy (current, power consumption, emission energy, etc.) and signal path delay. The main advantage of the SCA approach is its ability to detect HTs even when HTs are not activated. The larger the HT, the more effective this method is and the simpler testing process [4-5]. Currently, SCA is considered among the most efficient techniques for HT detection [5]. Since the changes of the IC design will also lead to corresponding changes in the physical parameters, SCA techniques can detect many types of HTs with different sizes

and structures. Moreover, recently, the research on the application of different artificial intelligence (AI) techniques including ML and deep learning (DL) in hardware security has also shown promising results [6-7]. Hence, in this work, we aim to propose a new HT detection technique with power traces and ML.

III.   PROPOSED HT DETECTION METHOD AND RESULTS

In this paper, the main design is the 128-bit AES encryption core (AES_128) with the block diagram as shown in Fig. 4. The AES core uses the secret key to encrypt the plain text (Msg) and provides the cipher text after 10 rounds [8]. The HT based on the well-known Trust-Hub library [9] is inserted in this 128-bit AES encryption design.
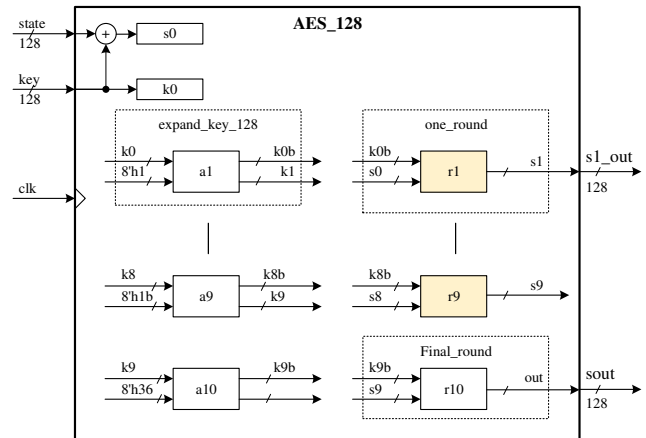

Fig. 4. AES-128 block diagram.

Table 1 presents the power consumption analysis results with the Xilinx XPower Analyzer tool for the 128-bit AES encryption core in FPGA hardware. It can be seen that the HT-infected circuit has a power consumption higher about 1% than the HT-free circuit. Based on this analysis, we propose the use of power traces of the AES core and advanced ML based signal processing techniques to detect the HT.

TABLE 1. POWER CONSUMPTION (MW) OF THE CIRCUITS USING XILINX XPOWER ANALYZER TOOL.

Power consumption type	Without HT	With HT
Dynamic power	25.84	26.54
Static power	64.69	64.71
Total power	90.53	91.26

In this work, a hardware security evaluation board (SAKURA-G) was used to implement the AES encryption core in Xilinx Spartan-6 FPGA device. This board is embedded with a specific design for SCA, especially with the power consumption data. Figure 5 presents the power traces acquisition process in our experiments. On the other hand, the measurement and experimental configuration setup for this work is shown in Fig. 6. The digital oscilloscope can capture the power traces and save in CSV files which can be displayed as in Fig. 7 for an example of the 10-round AES encryption

operation. The effect of HT in the AES operation is presented in Fig. 8 where the cipher output was wrong due the activation of the inserted HT.
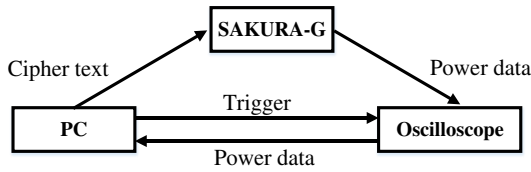

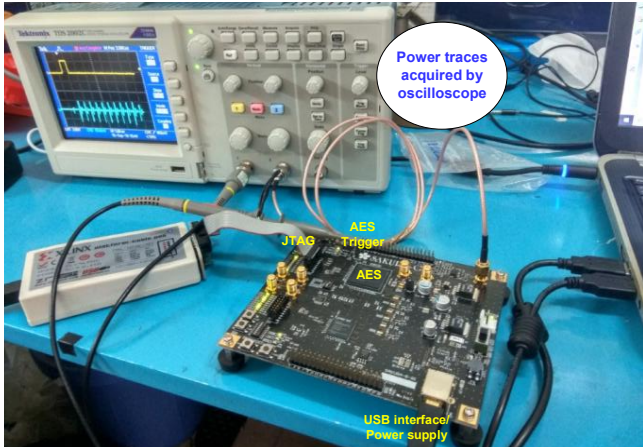Fig. 5. Power consumption acquisition process for HT detection experiments.

Fig. 6. Measurement setup and experimental configuration for the proposed HT detection method.
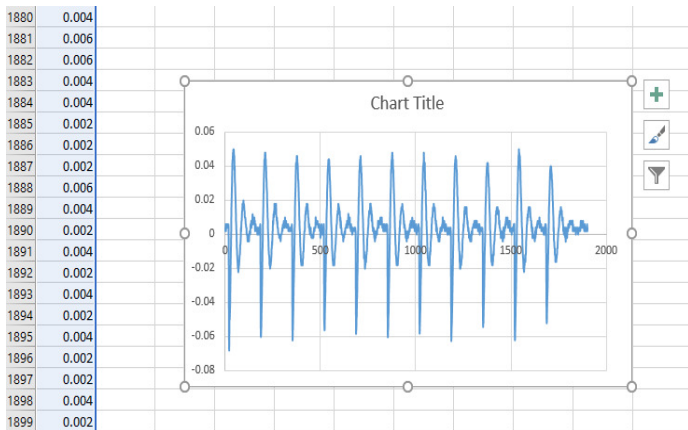

Fig. 7. The plot from CSV file captured by oscilloscope in the 10-round AES encryption process.

In this preliminary work, four different values of the AES encryption key were used to collect the power consumption data (power traces). For each key, we perform a power sampling by 30 times. In this experiment, the data with forth key is used as the reference design. The high resolution oscilloscope (Tektronix TDS2002C) was used with the support of NI LABVIEW software to build the dataset of power traces. After removing the unwanted data, we obtained 1837 power

traces. From this dataset, Softmax Regression algorithm was used to classify the circuits with and without HT.
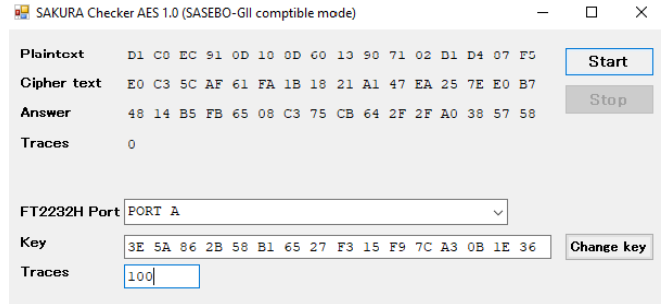

Fig. 8. The effect of HT in the wrong AES operation.

The results in Fig. 9-11 show that for the fourth key (running in an HT circuit) corresponding to the class 3, the Receiver Operation Characteristic (ROC) value is significantly different from the remaining three keys (corresponding to the class 0, 1, 2) running in the AES circuit without HT, thereby showing the ability to detect HT. Be noted that, the labels L1, L2, L3 and L4 in Figs. 9 - 10 correspond to the classes 0, 1, 2 and 3 in Fig. 11. Figure 9 is also the results of data visualization by the 2-dimension principal component analysis (PCA). With the result of HT detection accuracy of up to 97%, we can see the feasibility of applying ML techniques in classifying the power signals obtained from IC designs to detect HT in the chip (an FPGA device in this work).
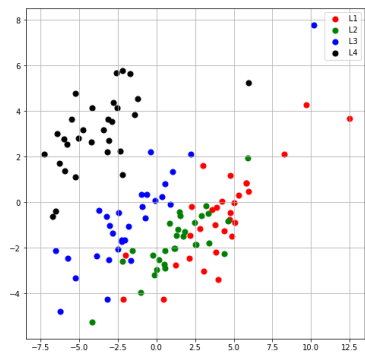

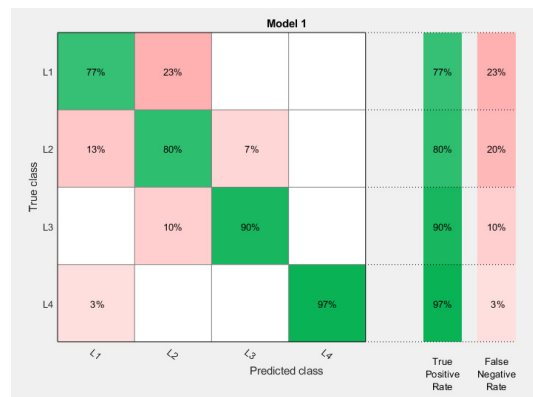Fig. 9. Data visualization by 2-dimension PCA.

Fig. 10. Confusion Matrix of the proposed ML based HT detection method with two cases of True Positive and False Negative.
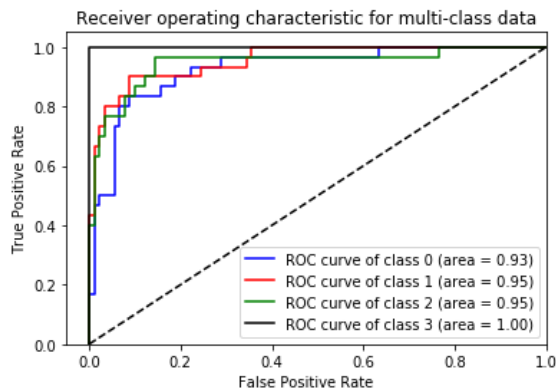
Fig. 11. ROC curve for the proposed HT detection technique.

IV. CONCLUSION

This paper has proposed a new method to detect HT using the SCA using the power traces and ML techniques. The implementation and analysis results with SAKURA-G hardware security board have shown the feasibility of the proposed solution. Moreover, with the detection accuracy of up to 97%, the proposed ML based HT detection method will have a high potential for practical applications. Furthermore, we will investigate, evaluate the proposed technique with more data, add effects of technology variations and complete the implementations on ASIC designs to provide more efficient HT detection methods.

REFERENCES

[1] W. Ou, J. Zeng, Z. Guo, W. Yan, D. Liu, S. Fuentes, "A Homomorphic-encryption-based Vertical Federated Learning Scheme for Rick Management," Computer Science and Information Systems, vol. 17, no. 3,pp. 819–834, 2020.

[2] Faisal Alotaibi, Alexei Lisitsa, "Matrix profile for DDoS attacks detection," Proceedings of the 16th Conference on Computer Science and Intelligence Systems, Annals of Computer Science and Information Systems, vol. 25, pp. 357–361, 2021.

[3] Zhifeng Hu, Feng Zhao, Lina Qin, Hongkai Lin, "Network Virus and Computer Network Security Detection Technology Optimization," Scalable Computing: Practice and Experience, vol. 22. no. 2, 2021.

[4] S. Bhunia and M. M. Tehranipoor, The Hardware Trojan War: Attacks, Myths, and Defenses, Springer, pp. 15-51, 2018.

[5] A. Amelian and S.E. Borujeni, "A Side-Channel Analysis for Hardware Trojan detection based on Path Delay Measurement," Journal of Circuits, Systems, and Computers Vol. 27, No. 9, (2018).

[6] Elnaggar, R. & Chakrabarty, "Machine Learning for Hardware Security-Opportunities and Risks," K. J Electron Test (2018) 34: 183.

[7] N. -T. Do, V. -P. Hoang and V. -S. Doan, "Performance Analysis of Non-Profiled Side Channel Attacks Based on Convolutional Neural Networks," 2020 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), Ha Long, Vietnam, 2020, pp. 66-69.

[8] M. Dao, V. Hoang, V. Dao and X. Tran, "An Energy Efficient AES Encryption Core for Hardware Security Implementation in IoT Systems," Proc. 2018 International Conference on Advanced Technologies for Communications (ATC), Ho Chi Minh City, Vietnam, 2018, pp. 301-304.

[9] Trojan Benchmarks, Available: https://www.trust-hub.org/resource/benchmarks.

Authentication and Encryption algorithms for data security in Cloud computing: A comprehensive review

Thanh Ngoc Nguyen
Department of Computer Networks
and Data Communications
Eastern International University
Binh Duong, Vietnam
thanh.nguyenngoc.set16@eiu.edu.vn

Thien T. T. Le
Faculty of Electronics
and Telecommunication
Saigon University
Ho Chi Minh City, Vietnam
thanhthien2003@gmail.com

Abstract—**With the growth of data stored in cloud, data may become the target of attackers in the Internet. Therefore, the end users require high confidentiality, integrity and authentication in order to protect their data in cloud. In this paper, we aim at a comprehensive studying about the data security in cloud computing. The paper will discuss the details of cloud computing data security challenges and find out which are the most important challenges as well as the efficient solutions. The existing authentication and encryption algorithms are compared in terms of users' scenarios, outstanding features and the limitation. We also review the advantages and drawbacks of the algorithms for data security in terms of cloud computing services.**

Index Terms—**cloud computing, data security, encryption, authentication.**

I. Introduction

CLOUD computing is one of the most popular technologies which allow many users using the resource without hardware implementation. Cloud computing is employed widely by both organizations and individuals because of its flexibility and mobility [1]. Cloud computing is a network model which is on-demand, efficient and users can access to shared resources, such as: storage, networks, ser-vice. The concept of cloud computing is X-as-a-Service or shortly called XaaS in which "X" represented to three major cloud service models as follows: (1) "S" for software which represented as Software-as-a-Service (SaaS); (2) "I" for infra-structure which represented as Infrastructure-as-a-Service (IaaS); (3) "P" for platform which represented as Platform-as-a-Service (PaaS). This concept allows many system components such as: databases, IT infrastructure can be distributed as a service. In IaaS, the cloud service provider provides users virtual machines and storage for higher business capabilities. In PaaS, the service provider delivers an environment to deploy, test and run applications. SaaS is the most used service, which allows user to use applications deployed by service provider. The Internet provides environment or platform for cloud computing through which many services are delivered to users. An important aspect of cloud computing is Utility-oriented which its strategy is "pay-per-use" to charge users.

Cloud computing has many advantages such as money saving, no up-front commitments, efficient resource allocation, and on-demand accessibility. However, there are some challenges for cloud computing which need to be taken into account such as data security, technical issues related to infrastructure management, legal issues due to different policies in distinct nations [1-3]. Security is the most important issue in cloud computing and it is growing dramatically every year. The main cause of security in cloud is its service models [4]. According to [5], each kind of cloud service models have its own security issues, which are listed as follows: (1) Security issues of SaaS: data theft by malicious attacks, unable to monitor data transfer amongst cloud users, etc.; (2) Cloud security issues experienced IaaS: unable to completely control the access to sensitive data, attackers can host a data theft attack in cloud infrastructure, security staff shortage, etc.; (3) PaaS's security issues: require a higher cost and bigger effort to implement and maintain the infrastructure, new types of cyber attacks. Therefore, the protection of data stored in cloud is urgently necessary; it can be achieved by employing authentication, virtualization and encryption, which help prevent unauthorized access [6]. It is necessary to develop the data security method in order to ensure the confidentiality, integrity, and availability of data in any circumstances [7]. There are many methods such as authentication, encryption, third-party auditing, identity and access management techniques which can be deployed in different cloud computing types [3, 8]. For example, in private cloud, the cloud owners use the form of Active Directory to store all of the credentials in the server and the authentication process are done via virtual private network. In public cloud, users can connect to service providers through Internet simultaneously and they can use any devices to access to cloud resources anywhere. This is the reason why public cloud exposes to be more vulnerable than private cloud, so the cloud providers have to include highly secured authentication methods.

In this paper, we will focus on the data security challenge, which is the hottest issues in cloud computing. We also explore distinct security challenges and also study about the methods used to remedy those issues. The data encryption and decryption will be described in terms of data security in cloud. The authentication method is also investigated in details.

This paper is divided into six sections. In the next section, we explain some threats to data in cloud computing. In section III, we will discuss some countermeasures to protect the data and provide some comparisons between

different solutions. We describe other existing algorithms for data security from relevant literature surveys in Section IV. In section V, the comparisons among distinctive algorithms are discussed in terms of encryption and authentication. Finally, the conclusion and the further challenges are given in Section VI.

II.   THREATS IN CLOUD COMPUTING

Although cloud computing provides customers a great deal of benefits, it still has some drawbacks which cause the loss of data. Therefore, the customers cannot access the data stored in cloud or the data has been changed. These drawbacks include data breach, data loss, and insecure APIs. The threats in cloud computing are explained in Table I as follows.

TABLE I. MAIN DATA THREATS IN CLOUD COMPUTING

Types of threats	Causes	Consequences	Countermeasures
Data breach	Outside attacks, low-secured encryption and encryption key loss, malicious insider	Data can be viewed, stolen or destroyed	Strong encryption mechanisms, strong firewall, multifactor authentication methods
Data loss	Natural disaster or incidents caused by human, outside hackers	The loss of your data of the entire cloud system	Backup data frequently, strong encryption mechanisms, strong firewall, multifactor authentication methods
Insecure APIs	Interact with the programmes using the APIs in the wrong way or not safely; use or install insecure third-party programmes or unauthorized programmes	Data can be viewed, stolen or destroyed	Use only programmes from trusted sources

III.   DATA SECURITY COUNTERMEASURES IN CLOUD COMPUTING

To prohibit and reduce the impacts of threats mentioned in previous section, some solutions have been employed by companies and cloud service providers. In this section, we will discuss about some popular countermeasures such as authentication and data encryption.

A.   Authentication

Authentication is the process of determining the identity of users who want to access to the resources in cloud [3]. Four methods of authentication can be listed as follows: (1) something the individual knows, for example, a password; (2) something the individual possesses such as electronic or physical keys, which are commonly called token; (3) something the individual is, for instance, the finger-print, face or retina; (4) something the individual does, such as voice and hand-writing and so on. Authentication also provides access control service to compare the credentials of users with credentials stored in server. Some most popular types of authentication method are discussed as follow.

1)   Biometric authentication.

In [9], biometric authentication is employed in the aspect of identification and authentication in cloud data security. Biometric refers to the biological sciences which can be listed in two main classes:

(1) Physiological and behavioral. Physiological is variable because it is different between people and it is relevant to human's physical body parts such as: fingerprints, facial recognition and so on.

(2) Behavioral is about the behavior of people such as signature and voice.

Details of biometric authentication technologies may consist of one or many features such as: finger print, face recognition, IRIS technology, hand geometry technology, retina geometry technology, speaker recognition, signature verification technique.

2)   Multifactor authentication

Multifactor authentication employs not only one but also more than one factor to verify the users [10,11]. In the two-factor authentication, username and password are used to determine the user while accessing the cloud. In order to secure the authentication method, more than one factor is used such as voice recognition, facial recognition, or mobile identity number.

Beside credentials, some other techniques can be used as a secondary factor such as captchas, one-time password (OTP). Because of using more than one au-thentication factor, the security level of multifactor authentication is enhanced. OTP is often used in online transaction which can be explained as follows: the server creates an OTP which can be used only one time, and this OTP is sent to the users via two main means, mobile phone text or email [11]. Another mechanism can be used as secondary factor is captcha, this technique is quite popular and we can see it quite often when surfing the Internet. Captcha is employed to prevent web applications from attacks of malwares, there are a few types of captcha such as numbers, alphabet letters, images or a combination of numbers and letters [9,11].

In multifactor authentication, the more factors employed, the more secure the system is and to prove this, we make a comparison between two-factor authentication, which used handwriting recognition in addition to normal password, and five-factor authentication, which include password, voice recognition, facial recognition, Mobile identity number (IMEI) and International Mobile Subscriber Identity (IMSI) [8]. The detail comparison can be found in Table II below.

B.   Encryption

Encryption is employed a lot these days to secure the information sent over the Internet and the secured information can only be seen by the intended recipient [12]. The overall process of encryption is described as follow: (1) encrypting plain text into cipher, (2) receivers who have a secret key can decrypt the cipher text into readable text; in both stages, a unique key is used by senders and receivers to

encrypt or decrypt. Depend on the number of keys used, we can separate encryption into two different types, symmetric and asymmetric [11-13].

1) Symmetric encryption

Symmetric encryption is also known as single-key encryption, which uses only one key for encryption and decryption. The use of symmetric must meet two requirements: (1) strong encryption algorithm; (2) the secret key transfer must be in a secure fashion. There are some popular symmetric algorithms such as DES, 3DES, Advanced Encryption Standard (AES). AES is a symmetric encryption algorithm which has a higher security level than DES and 3DES. Because there is not any successful attack against AES, it becomes one of the most popular encryption algorithms which is employed by major organizations such as banks, governments [15]. The AES Encryption algorithm is shown in Fig. 1.

In Fig. 1, the AES allows the key lengths of 16 bytes and 10 rounds in the en-cryption process. AES takes a 16-byte block as an input for encryption and de-cryption algorithm, this input block is used to make a 4x4-byte matrix, which was referred to as a State array. Most of the rounds, except the last round, consist of four distinct steps as follow: (1) SubBytes, where each byte is substituted based on a S-box; (2) ShiftRows, in which each row is shifted cyclically; (3) MixCol-umns, where each four bytes of each column is combined; (4) AddRoundKey is the step of combination of each byte of the state with the round key. The last round comprises only three stages, which are the same as in previous rounds, except the MixColumns step.
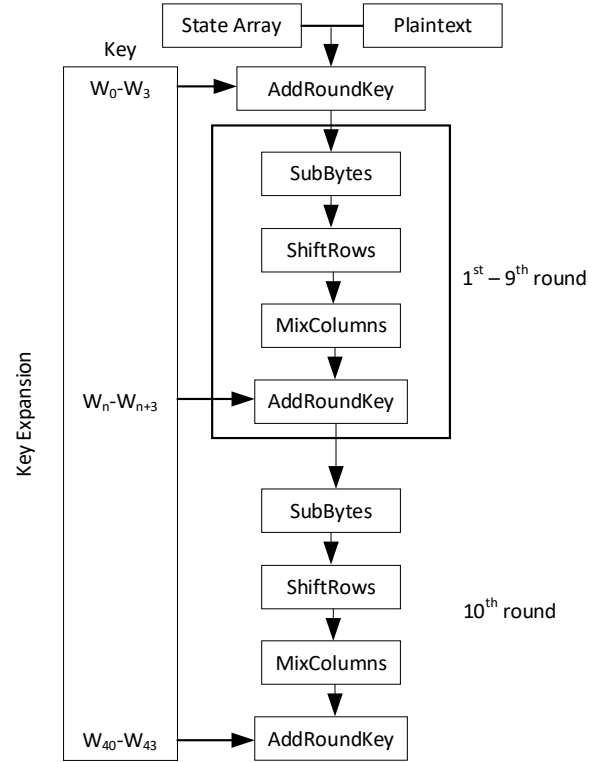
TABLE II. COMPARISON BETWEEN TWO-FACTOR AND FIVE-FACTOR AUTHENTICATION

Types of Authentications	Two-factor authentication	Multi-factor Authentication
Techniques	Handwriting recognition. Username and password	Username and password, voice recognition, facial recognition, Mobile identity number (IMEI), International Mobile Subscriber Identity (IMSI)
Strengths	The process of signing in is simple. Processing the handwriting figure on the cloud help reduces time consumption	All of the factors are processed in cloud, so the speed, time and efficiency are improved. The communication is secured by using transport security layer or secure socket layer. IMEI and IMSI are used to protect the system against the devices' loss.
Weaknesses	Some errors can occur because the change in users' writing style. There is no mutual authentication between the parties and also no protection against attacks.	It is too complicated for end users because too many factors are employed and there is also no mutual authentication.

Fig. 1. AES Encryption [12, 15]

2) Asymmetric encryption

Asymmetric encryption or also called public-key encryption which uses two separated keys, named as public key and secret key [12, 15]. RSA is an acronym of its inventors' names, Rivest, Shamir and Adleman, it is used worldwide in encryption and digital signature [11]. The security level of this algorithm depends on the difficulty of the disintegration of large numbers and the public and private key are made from two large prime numbers. There are three main steps in RSA algorithm and all of them are listed as below: (1) key generation; (2) encryption; (3) decryption [16]. The key generation is shown Fig. 2.
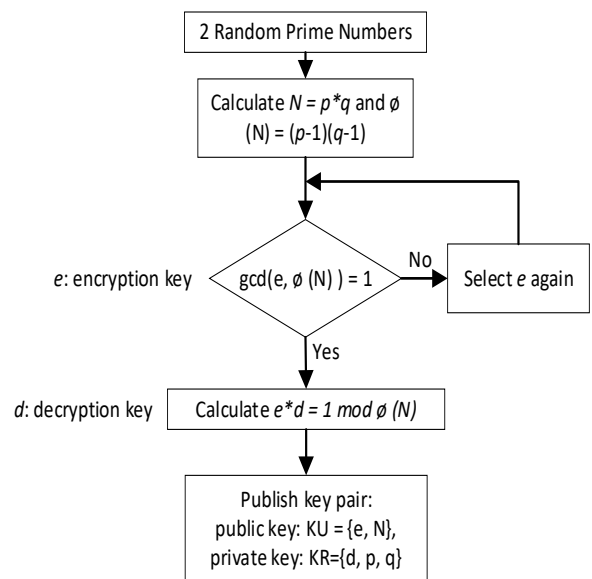

Fig. 2. RSA key generation [12,16]

The details of these three steps are described as below:

(1) Key generation: first, select two random prime numbers p and q. Then, compute the system modulus $N = p \times q$ and $\varnothing(N) = (p-1) \times (q-1)$. Afterwards, select a random encryption key e, where $1 < e \varnothing(N)$, $gcd(e, \varnothing(N)) = 1$. In the next step, compute decryption key d by solving the equation: $e \times d = 1 \bmod \varnothing(N)$ and $0 \le d \le N$. Finally, the public key $KU = \{e, N\}$ is published while the private key $KR = \{d, p, q\}$ is still kept.

(2) Data encryption: after acquiring the public key $KU = \{e, N\}$ from the first stage, we compute $C = M^e \bmod N$, where $0 \le M < N$.

(3) Data decryption: we use the private key $KR = \{d, p, q\}$ to decrypt, we compute $M = C^d \bmod N$.

IV. EXISTING ALGORITHMS FOR DATA SECURITY IN CLOUD COMPUTING

A. Data security in cloud computing using three-factor authentication

In [17], the data security in cloud computing is protected by the three-factor authentications which consists of login phase, authentication phase, and biometric phase. The three factors are user credentials, fingerprints and passwords. In the initialization phase, the server chooses the private key and public key by using elliptical cryptography. Then, the user verifies himself with the server using these factors: identity, password, and the biometric imprints.

B. Scrambling and descrambling of document image for data security in cloud computing

In [18], the users may store their document image on cloud which requires high confidentiality, identity and authentication. The Arnold transform scrambling and descrambling is used to protect users' data. Before the users store their image in the cloud, the document images are applied by the Arnold Transformation. This process ensures that any unauthorized persons cannot be able to read the document image information. The users must apply the descrambling process to retrieve the original image.

C. Towards DNA based data security in the cloud computing environment

According to [19], many cryptography methods have been applied to protect data of the users in cloud. The DNA Based Data Security (DNABDS) encryption scheme is applied to calculate the secret key or the public key. The data is encrypted by the DNA based public key, then the users decrypt data by applying the DO's private key. The 1024-bit secret key is generated based on DNA computing, user's attributes and Media Access Control (MAC) address of the user, and decimal encoding rule, American Standard Code for Information Interchange (ASCII) value. The server and the users exchange data and keys by using the secure communication medium such as secure socket layer.

D. Cognitive cryptography for data security in cloud computing

In order to protect data of users, the advanced multilevel user authentication protocol is applied by using hybrid CAPTCHA codes. This type of authentication only is distributed for providing data access amongst the experts or trusted users of specific areas. These codes will define a new class of cognitive CAPTCHAs, which based on the recognition or interpretation of noisy pattern [20].

E. Public-key encryption secure against related randomness attacks (RRA) for improved end-to-end security of cloud/edge computing

In [21], the data security is provided by secure communication channel which requires the public key and private key at both users and the server. The author focuses on constructing secure public key encryption scheme against related randomness attacks. The RRA security under chosen plaintext attack (RRA-CPA) secure public key encryption scheme is derived from any publicly deniable encryption; and the RRA security under Chosen Ciphertext Attack (RRA-CCA) secure public key encryption scheme is derived from standard Indistinguishability under Chosen Ciphertext Attack (IND-CCA) public key encryption scheme with a hardcore function for arbitrarily correlated input [21].

F. Cyber security risks in robotics

In modern era, robotics is developing at a rapid pace and appears in every aspect of people around the world, this also leads to some new security challenges in this area. Therefore, some solutions were recommended to remedy the security issues, which include: Communication Robustness, Data Distribution Service in ROS, Authentication Mechanism in YARP, Securing the Cloud, Communication Buses. Robotics is expected to be one of the most important area in the future, and with the implementation of suggested solutions, the security risks in robotics can be reduced to the minimum [22].

G. Internet memes: A novel approach to distinguish humans and bots for authentication

In [23], bots have been overwhelmed the Internet and they have also evolved themselves to bypass some types of CAPTCHA authentication, making it difficult to distinguish human and bots. In order to eliminate automatic activities of bots, Internet Memes, a new type of authentication is employed. This type of authentication cannot be learnt by bots because Memes are dynamic and changed frequently, so it can identify which are human activities and which are bots' activities.

H. Some cyberpsychology techniques to distinguish humans and bots for authentication

These days, with the growth of Artificial Intelligence (AI), bots can imitate humans' behaviors to bypass traditional types of authentication and break into computer systems. Therefore, a psychology-based method is implemented to distinguish between humans and bots, which is called cyber-

psychological authentication. The deployment of cyber-psychological authentication includes lots of means, such as: pronunciation and translation checking, event classification and so on, and it has successfully differentiated between humans' behaviours and bots' behaviours [24].

I. Secure cloud computing authentication

In [25], the elliptic curve cryptography algorithm is applied to secure data transmission from a user to a server. In the system consists of two users and one server, the sender encrypts the plain text using the receiver's public key before sending the cipher text to the server. Then, the receiver uses its private key to decrypt data. The elliptic curve cryptography (EEC) is the public encryption which requires public key and private key. The EEC is less complexity than the RSA [12] but it provides the same security. However, the EEC requires higher in terms of size and computation requirements.

J. The elliptic curve cryptography- (ECC-) based three-factor authentication scheme

In [26], the cloud computing server combines IoT technologies which provide a huge amount of data from the IoT devices sending to the server of the cloud. The cloud server can be accessed by some authenticated users. The user's IoT devices are already registered by applying the ECC encryption. The user requests the registration phase by using the identity, password, and biometric information. The user logins the server and processes authentication steps through the public channel.

K. Secure authentication protocol for remote healthcare system

In [27], the secured authentication protocol is developed for remote healthcare systems where the server is connected to the IoT bio-devices monitor the patient. The IoT bio-devices will collect the data of patients then send to the server; the doctor can access the information of patient at the server through Internet. In [27], the authentication consists of five phases which are (1) system setup phase; (2) new sensor registration phase; (3) user registration phase; (4) login and authentication phase; (5) password changing phase.

L. Fingerprint-based authentication using single user credential in IoT

In [28], the user can access the cloud computing server's services by using the concept of Single Sign-On. The first step is registration and verification which uses the users' fingerprint scan, hash password and registration details. This information is encrypted by using the encryption key. The cloud server checks the registration and verification of the user; then setup a secure channel to the user. In the second step, the user can login by using his register's biometric registration.

V. COMPARISON AND DISCUSSION

A. Comparison

In this section, the comparison of current existing works on data security is summarized as in Table III. We compare the existing algorithms in terms of data encryption and authentication methods. The outstanding features and the limitations of these algorithms are also discussed and presented in Table III.

B. Discussion

The existing works focus on the authentication process which requires many parameters. The most common parameter is the biometric information such as DNA, fingerprint because they are unique. Therefore, the authentication and verification process will provide secure channel between the sender and the receiver or the user to the server. However, the public encryption scheme is widely used to protect the data because high security. It uses public key for encryption and private key for decryption.

TABLE III. COMPARISON BETWEEN CURRENT EXISTING WORKS ON DATA SECURITY

Current existing works	Scenarios	Methods		Outstanding features	Limitation
		Data encryption	Authentication		
Three-Factor Authentication [17]	Providing security for data in cloud	none	Credentials, fingerprints and passwords	Using elliptical cryptography	Simple process
Scrambling and Descrambling of Document Image [18]	Provide authentication process for users' document images	Arnold transform scrambling and descrambling	none	high confidentiality, identity for document image	Not mention other types of document
DNA based data security [19]	Protecting data of the users in cloud	DNA Based Data Security DNABDS) encryption scheme	none	The secret key is generated based on DNA computing and user's attributes results in high confidential of data	Slow performance
Multilevel user authentication protocol [20]	Providing data access amongst the experts	none	Hybrid CAPTCHA codes is based on the noisy pattern	Secures the confidential data	Limited applications which are only used for scientific data or experts
Public-key encryption secure against related randomness attacks (RRA) [21]	Improving end-to-end security of cloud computing	Secure public key encryption scheme	none	Avoid randomness attacks	Lack of authentication of the users

Internet memes [22]	Distinguish humans and bots for authentication	none	Internet Memes-based authentication, Memes are dynamic and changed frequently	Bots cannot learn to authenticate any service	Lack of datasets, extracting data from images may not be simple because the lack of tools
Cyberpsychology techniques to distinguish humans and bots for authentication [23]	Bots can imitate humans' behaviors to bypass traditional types of authentication	none	cyber-psychological authentication	Employ the psychological characteristics such as: pronunciation and translation checking, event classification	Reduce the users' experience because it takes time for users to answer the questions
Secure cloud computing authentication [25]	Data transmission from a user to a server	elliptic curve cryptography	none	Provide high secure channel	High complexity and calculation
The elliptic curve cryptography- (ECC-) based three-factor authentication scheme [26]	Data transmission from the IoT devices to the server of the cloud	elliptic curve cryptography	identity, password, and biometric information	Establish high secure channel between IoT devices and the server	High complexity and calculation of EEC crypto may require high memory and power at the IoT devices
Secure authentication protocol for remote healthcare system [27]	IoT and healthcare system	none	Identification of the sensor	Prevent replay attack and insider attack	The healthcare data should be encrypted because only the doctor can access the data
Fingerprint-based authentication using single user credential in IoT [28]	Single user in IoT – server transmission	none	fingerprint scan, hash password and registration details	The single account for multiple server may remove the fake user account	Data may require encryption at the receiver

VI. CURRENT ISSUES AND CHALLENGES

A. Current issues

In the view of users, cycle of data consists of six stages: create, store, use, share, archive and destroy. If the user store data in the cloud, it is needed to authenticate the users before carrying out any activities such as using, sharing, or deleting. Data should be protected to ensure the CIA triad which includes confidentiality, integrity, and availability. The cryptography algorithm is used to secure communication channel in which it needs having the key management mechanism. In RSA, two keys are used for establishing the secure channel between two end users require the mechanism to secretly distribute keys to the users.

B. Challenges

The users may be a mobile device, a laptop or any internet-connecting device which has limitation resources. Therefore, authentication mechanisms should be developed taken the resource of user into account which requires less computing, memory, and storage requirements. In addition, we also need to consider the devices of Internet-of-Things that are placed in the house, garden, or cars to collect the environment data. The sensing data may be stored in the cloud, then, only the authenticated users can access and modify data. The cryptographic and authentication method can be deployed together in order to prevent any attackers modifying data.

Even though encryption algorithm can secure data during transmission, the key management should be considered. In addition, the time complexity and computation of encryption should be low because of high energy consumption and high memory calculation. The elliptic curve cryptography is widely used to establish high data security during transmission.

In the recent years, a number of IoT devices increase which lead to a massive amount of data stored in cloud's server. It is necessary to provide a secure-direct communication from IoT devices to server in cloud [29]. The group users in cloud will be the new challenges because many applications require more than one user to access the same data in the cloud server. The authentication steps must collect biometric of different users then setup a secure channel. In the future, fog computing architecture and cloud computing will require the machine learning approach to manage the users, communication, and data of users [30-32].

VII. CONCLUSIONS

In this paper, we have gone through some concepts in Cloud computing which introduce the Cloud computing platform and different types of cloud. Some advantages and disadvantages of Cloud computing have been discussed. Therefore, we can help everyone make a choice easily if they are considering implementing a Cloud platform. We have done a research about a specific type of challenges that many Cloud computing platforms are facing these days, which are Data security threats.

We have made a study about how to prevent hackers to approach sensitive data of users in which the Cloud service providers are using Authentication. This solution requires the user to identify themselves when they want to log in the accounts. However, this method has some drawbacks, such

as the careless users can accidentally give their identifications for the hackers. Therefore, we have chosen to study about another solution with higher security level, Encryption. We have provided some information about encryption methods and the algorithms, which represent for those methods. In the future, we will develop and implement an authentication mechanism to evaluate the performance of different authentication methods.

REFERENCES

[1]. R. Buyya, C. Vecchiola, and S. T. Selvi, "Mastering cloud computing: foundations and applications programming". Elsevier, 2013.

[2]. P. R Kumar, P. H. Raj, and P. Jelciana, "Exploring data security issues and solutions in cloud computing" in Procedia Computer Science, vol. 125, pp. 691-697, January 2018.

[3]. Yunchuan Sun, Junsheng Zhang, Yongping Xiong and Guangyu Zhu, "Data security and privacy in cloud computing" in International Journal of Distributed Sensor Networks, July 2014.

[4]. B. Alouffi, M. Hasnain, A. Alharbi, W. Alosaimi, H. Alyami and M. Ayaz, "A Systematic Literature Review on Cloud Computing Security: Threats and Mitigation Strategies" in IEEE Access, vol. 9, pp. 57792-57807, April 2021.

[5]. M. K. Sasubilli and V. R, "Cloud computing security challenges, threats and vulnerabilities", 2021 6th International Conference on Inventive Computation Technologies (ICICT), pp. 476-480, 2021.

[6]. Isma Zulifqar, Sadia Anayat, Imtiaz Kharal, "A review of data security challenges and their solutions in cloud computing" in I.J. Information Engineering and Electronic Business, March 2021.

[7]. M. De Donno, A. Giaretta, N. Dragoni, A. Bucchiarone and M. Mazzara, "Cyber-storms come from clouds: Security of cloud computing in the IoT era", May 2019.

[8]. T. S. Chou, "Security threats on cloud computing vulnerabilities" in International Journal of Computer Science & Information Technology, vol. 5, no. 3, June 2013.

[9]. D. Bhattacharyya, R. Ranjan, F. Alisherov and M. Choi, "Biometric authentication: A review" in International Journal of u-and e-Service, Science and Technology, vol. 2, no. 3, pp. 13-28, September 2009.

[10]. I. Indu, P. R. Anand, and V. Bhaskar, "Identity and access management in cloud envi-ronment: Mechanisms and challenges" in Engineering science and technology, an international journal, vol. 21, no. 4, pp. 574-588, May 2018.

[11]. J. K. Mohsin, L. Han, M. Hammoudeh, and R. Hegarty, "Two factor vs multi-factor, an authentication battle in mobile cloud computing environments" in Proceedings of the International Conference on Future Networks and Distributed Systems, pp. 1-10, July 2017.

[12]. W. Stallings, "Cryptography and network security principles and practice". 7th ed., Pearson, England, 2017.

[13]. R. Bhanot, R. Hans, "A review and comparative analysis of various encryption algorithms" in International Journal of Security and Its Applications, vol. 9, no. 4, pp. 289-306, April 2015.

[14]. N. Khanezaei and Z. M. Hanapi, "A framework based on RSA and AES encryption algorithms for cloud computing services" in 2014 IEEE Conference on Systems, Process and Control, pp. 58-62, Kuala Lumpur, Malaysia, December, 2014.

[15]. A. Sachdev and M. Bhansali, "Enhancing cloud computing security using AES algorithm" in International Journal of Computer Applications, vol. 67, no. 9, April 2013.

[16]. X. Zhou and X. Tang, "Research and implementation of RSA algorithm for encryption and decryption" in IEEE Proceedings of 2011 6th international forum on strategic technology, vol. 2, pp. 1118-1121, August 2011.

[17]. S. Nalajala, B. Moukthika, M. Kaivalya, K. Samyuktha and N. L. Pratap, "Data security in cloud computing using three-factor authentication" in International Conference on Communication, Computing and Electronics Systems, pp. 343-354, Springer, Singapore, 2020.

[18]. N. Salimath, S. Mallappa, N. Padhy and J. Sheetlani, "Scrambling and descrambling of document image for data security in cloud computing" in Smart Intelligent Computing and Applications, pp. 283-290, Springer, Singapore, 2020.

[19]. S. Namasudra, D. Devi, S. Kadry, R. Sundarasekar and A. Shanthini, "Towards DNA based data security in the cloud computing environment" in Computer Communications, vol. 151, pp. 539-547, February 2020.

[20]. U. Ogiela, "Cognitive cryptography for data security in cloud computing" in Concurrency and Computation: Practice and Experience, vol. 32, no. 18, September 2020.

[21]. P. Liu, "Public-key encryption secure against related randomness attacks for improved end-to-end security of cloud/edge computing" IEEE Access, vol. 8, pp. 16750-16759, January 2020.

[22]. I. Priyadarshini, "Cyber security risks in robotics" in Cyber security and threats: concepts, methodologies, tools, and applications, pp. 1235-1250, IGI Global, May 2018.

[23]. I. Priyadarshini and C. Cotton, "Internet memes: A novel approach to distinguish hu-mans and bots for authentication" in Proceedings of the future technologies conference, pp. 204-222, Springer, Cham, October 2019.

[24]. I. Priyadarshini, H. Wang and C. Cotton, "Some cyberpsychology techniques to distinguish humans and bots for authentication" in Proceedings of the Future Technologies Conference, pp. 306-323, Springer, Cham, October 2019.

[25]. M. Chakraborty, B. Jana and T. Mandal, "A Secure Cloud Computing Authentication Using Cryptography," in Proceedings of International Conference on Emerging Trends and Innovations In Engineering And Technological Research (ICETIETR), 2018, pp. 1-4, doi: 10.1109/ICETIETR.2018.8529100..

[26]. H. Lee, D. Kang, Y. Lee, and D. Won, "Secure three-factor anonymous user authentication scheme for cloud computing environment". Wireless 2021Communications and Mobile Computing, vol. 2021, July 2021.

[27]. M. Azrour, J. Mabrouki and R. Chaganti, "New Efficient and Secured Authentication Protocol for Remote Healthcare Systems in Cloud-IoT", Security and Communication Networks, 2021.

[28]. B. Alemu, R. Kumar, D. Sinwar, G. Raghuwanshi, "Fingerprint Based Authentication Architecture for Accessing Multiple Cloud Computing Services using Single User Credential in IOT Environments", In Journal of Physics: Conference Series, Vol. 1714, No. 1, p. 012016.

[29]. P. Hajder, M. Hajder, M. Liput, M. Nycz, S. Agarwal, D. N. Barrell, & V. K. Solanki, "Direct communication of edge elements in the Industrial Internet of Things". In FedCSIS (Communication Paper) ACSIS, Vol. 23, pages 35–42 (2020).

[30]. C. Mechalikh, H. Taktak, F. Moussa, "PureEdgeSim: A Simulation Framework for Performance Evaluation of Cloud, Edge and Mist Computing Environments" Computer Science and Information Systems, Vol. 18, No. 1, 43–66. (2021).

[31]. B. N. Barreto, A. R. de Sa, & A. D. R. L. Ribeiro, "A Fog Computing Architecture for Security and Quality of Service". In FedCSIS (Position Papers), ACSIS, Vol. 19, pages 69–73, 2019.

[32]. M. Saleem, M. R. , Warsi, S. Islam, A. Anjum, & N. Siddiquii, "Trust Management in the World of Cloud Computing. Past Trends and Some New Directions", Scalable Computing: Practice and Experience, 22(4), 425-444, 2021.

Cancer Prediction Using Cascade Generalization and Duo Output Neural Network

Chatree Nilnumpetch
Department of Computer Science
Ramkhamhaeng University
Bangkok, Thailand
nchatree@ru.ac.th

Somkid Amornsamankul
Department of Mathematics
Mahidol University and Centre of
Excellence in Mathematics
Bangkok, Thailand
somkid.amo@mahidol.ac.th

Pawalai Kraipeerapun
Department of Computer Science
Ramkhamhaeng University
Bangkok, Thailand
pawalai@rumail.ru.ac.th

Abstract—This paper proposes the combination of cascade generalization and duo output neural network based on feedforward backpropagation neural networks for cancer prediction. Duo output neural network is a neural network that is created based on two opposite targets in order to predict two opposite results. Cascade generalization is a technique that consists of a set of machines that are sorted together in which the predicted output produced from the previous machine plus the original training input are used for the creation of each machine. In this study, cascade generalization is organized in two levels: the base level and the meta level. In this research, duo output neural network is trained in each level of cascade generalization. Two outputs produced from the base level which are truth output and non-falsity output are averaged. The average result plus the original input are used for training a machine in meta level. The proposed technique is tested using two cancer datasets from UCI machine learning repository and found that our technique provides the best overall results when compared with three individual techniques.

Index Terms—Cascade generalization, Duo output neural network, Feedforward neural network, Binary classification, Breast cancer, Cervical cancer.

I. Introduction

NEURAL networks are one of the most widely used classification techniques in medicine [1, 2, 3]. This paper proposes a new technique based on neural network used for cancer prediction for small datasets. Breast and cervical cancer are common cancers in women, which develop in the cells of the breasts and cervix, respectively [4, 5]. In this study, two cancer datasets are tested which are breast cancer Coimbra [6] and cervical cancer behavior risk [7]. Both datasets consisted of a small number of instances which are 116 and 72, respectively. The breast cancer Coimbra dataset [6] is very popular, in contrast to cervical cancer behavior risk dataset [7], which has not been tested much in research. For breast cancer Coimbra, several techniques have been used for classification. For example, it was found that extreme learning machine can provide the highest result with 80% accuracy when compared to neural network, k-nearest neighbor and support vector machine (SVM) [8]. The gradient boosting was found to provide the best accuracy which is 74.14% when compared to logistic regression, decision tree, random forest, k-nearest neighbor, SVM and naïve Bayes [9]. However, it was found that the highest accuracy belongs to random forest when compared to logistic regression, decision tree, k-nearest neighbor, naïve Bayes, SVM, and neural network [10]. Instead of using individual technique, bagging is used for the classification [11]. Bagging of decision tree and bagging of k-nearest neighbor were found to give the highest results when compared to bagging of other techniques which are SVM, multilayer perceptron, logistics regression, and random forest. Deep neural network

with two hidden layers was found to provide the highest result with 75.94% accuracy when compared to SVM and decision tree [12]. In [13], deep neural network with four hidden layers was found to give better accuracy results than random forests and logistic regression. In [14], the backpropagation neural network was improved using Nelder Mead and it was found that this technique can provide result with 76.5217% accuracy which is better than the original backpropagation neural network. Polat and Sentürk [15] proposed a three-step hybrid structure in which MAD normalization was used to normalized data in the first step. After that, k-means clustering based feature weighing was used in the second step. AdaBoostM1 was used in the third step for the classification. Their technique provided results with 91.37% accuracy. Yavuz and Eyupoglu [16] used principal component analysis cascaded by median filtering for data transformation and used generalized regression neural network for classification. It was found that their accuracy rate was 0.9773. Chiu *et al*. [17] proposed a novel method consisting of transfer learning, principal component analysis, multilayer perceptron, and SVM. Their technique can give 86.97% accuracy which is the highest accuracy when compared to other six machine learning algorithms. Idris and Ismail [18] proposed a combination of fuzzy-ID3 and FUZZYDBD in order to classify breast cancer. Their technique was found to provide 70.69% accuracy which is better than ID3, SVM, c4.5, naïve Bayes, k-nearest neighbor, and random forest.

The focus of this paper is to improve the cancer prediction technique based on feedforward backpropagation neural network for small datasets. The proposed approach is to use duo output neural network [19] and cascade generalization [20] to classify data.

The remainder of this paper is organized as follows. In Section II, two existing techniques are explained which are duo output neural network and cascade generalization. Section III describes our proposed approach which is the integration of duo output neural network and cascade generalization. In Section IV, two datasets used in this study are illustrated and their experimental results are explained. The conclusion of this paper is explained in Section V.

II. Individual Techniques

A. Duo Output Neural Network

Duo output neural network is a technique that uses two opposite targets for training a neural network in order to solve binary classification or regression problems [19]. In general, a neural network is trained using one target for binary classification. Instead of using only the original target,

duo output neural network also applies the opposite value of the original target for training a neural network. These both targets are called truth target and falsity target, respectively. The falsity target value is computed as 1 minus the original target value. A neural network is trained using both targets and produces two outputs which are truth output and falsity output. Fig.1 shows the model of duo output neural network (DONN) in which a neural network, NN1, is trained using truth target (T) and falsity target (F). Two outputs are produced which are truth output (T1) and falsity output (F1). The classification result will be 1 if T1 is greater than F1, otherwise the classification result will be 0.
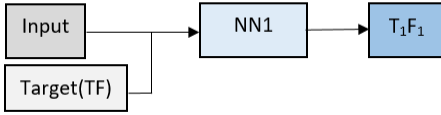
Fig. 1.   Duo Output Neural Network (DONN)

B. Cascade Generalization

Cascade generalization [20] is a technique that consists of several classifiers sequentially connected in which the next classifier is trained based on blending the original input features with the output predicted from the current classifier. Therefore, several levels of classifiers can be created. In our study, two levels named base level and meta level are implemented using a feedforward backpropagation neural network. Fig.2 describes the cascade generalization using neural network (CG-NN). In the base level, the original input features and the original truth target (T) are used to train the feedforward neural network (NN1). The result is the truth output T1. In the meta level, T1 and the original input features are used to train the feedforward neural network (NN2). The truth output T2 is then predicted. The classification result is 1 if T2 is greater than 0.5, otherwise the classification result is 0.
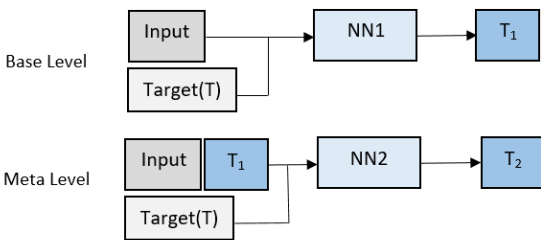

Fig. 2.   Cascade Generalization using Neural Network (CG-NN)

III. CASCADE GENERALIZATION USING DUO OUTPUT NEURAL NETWORK

Fig. 3 illustrates our proposed approach which is cascade generalization using duo output neural network (CG-DONN). In the base level, the duo output neural network (NN1) is trained to produce T1 and F1. The average of the truth output (T1) and the non-falsity output (1-F1) obtained from the base level is used in the meta level. In the meta level, the blending of the original input and the average result are used for training the feedforward neural network (NN2). The truth output (T2) and the falsity output (F2) are produced from the meta level. The classification result will be 1 if T2 is greater than F2, otherwise the classification result will be 0.
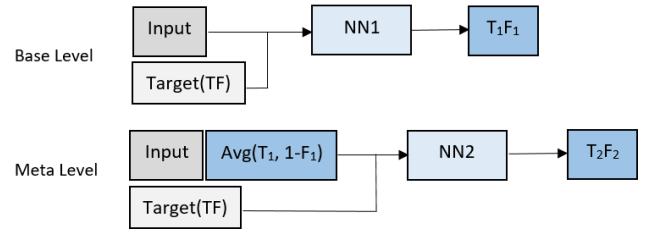

Fig. 3.   Cascade Generalization using Duo Output Neural Network (CG-DONN)

IV. EXPERIMENTAL METHODOLOGY AND RESULTS

A. Dataset

Two cancer datasets which are breast cancer Coimbra [6] and cervical cancer behavior risk [7] from UCI machine learning repository [21] are tested on our proposed technique. Breast cancer Coimbra contains nine input features, which are anthropometric data gathered from blood analysis. These features are used to classify 116 people into two groups, which are patient and healthy people. Cervical cancer behavior risk consists of eighteen input features gathered from behavior, intention, attitude, norm, perception, motivation, social support, and empowerment. Seventy-two instances are divided into two classes, which are "has cervical cancer" and "no cervical cancer".

B. Experiment and Results

Four techniques shown in Table I are compared in this paper. The fourth technique is our proposed technique. Feedforward backpropagation neural network is applied to each technique. For all neural networks, the number of nodes in the hidden layer is set to double the number of input features. Other parameters are set as default except the initial weight which is set to be different. Each dataset is randomly divided into three folds so that three-fold cross validation is applied.

TABLE I.   FOUR TECHNIQUES USED IN THIS PAPER (TECHNIQUE 4 IS THE PROPOSED APPROACH)

No.	Technique
1	Neural network (NN)
2	Cascade generalization using neural network (CG-NN)
3	Duo output neural network (DONN)
4	Cascade generalization using duo output neural network (CG-DONN)

For each round of three-fold cross validation, fifteen sets of machines are built so that we can see the average and the best results. In the first technique, fifteen feedforward back-propagation neural networks are trained in each round. In order to control the environment, these three sets of neural networks are used in the second technique.

In the second technique, all neural networks created in the first technique are used in the base level. In each round of meta level, fifteen neural networks are created using the blending of original input features and output predicted from each neural network obtained from the base level.

In the third technique, the environment is also controlled by using the same parameter values of the first technique except the number of targets and outputs which is set to two. Therefore, fifteen duo output feedforward backpropagation neural networks are created.

In the fourth technique, all duo output neural networks created in the third technique are used in the base level of cascade generalization. In each round of meta level, fifteen duo output neural networks are created based on the blending of original input features and the average of truth and non-falsity values obtained from each machine in the base level.

For each round of three-fold cross validation, the best and the average of fifteen results are computed. For breast cancer Coimbra dataset, Table II and Fig.4 show the best results whereas Table III and Fig.5 show the average results obtained from each fold of each technique.

TABLE II.   THE BEST PERCENT CORRECT OBTAINED FROM EACH FOLD OF EACH TECHNIQUE FOR BREAST CANCER COIMBRA DATASET

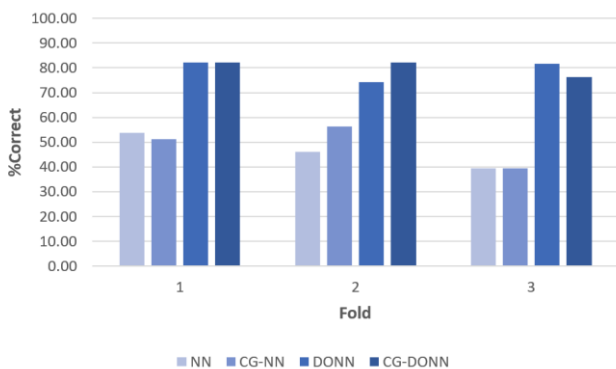| Fold	Technique (%correct)			
	NN	CG-NN	DONN	CG-DONN
1	53.85	51.28	82.05	82.05
2	46.15	56.41	74.36	82.05
3	39.47	39.47	81.58	76.32

Fig. 4.   The best result obtained from each fold of each technique for breast cancer Coimbra dataset

TABLE III.   THE AVERAGE PERCENT CORRECT OBTAINED FROM EACH FOLD OF EACH TECHNIQUE FOR BREAST CANCER COIMBRA DATASET

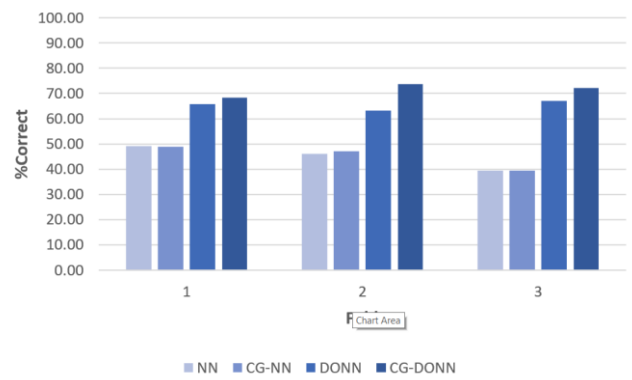| Fold	Technique (%correct)			
	NN	CG-NN	DONN	CG-DONN
1	49.05	48.90	65.82	68.21
2	46.15	47.18	63.08	73.67
3	39.47	39.47	67.03	72.29

Fig. 5.   The average result obtained from each fold of each technique for breast cancer Coimbra dataset

The average of best results for breast cancer Coimbra dataset are shown in Table IV and Fig.6 whereas the overall average are shown in Table V and Fig.7. It can be seen that CG-DONN can give the highest correct percentage.

TABLE IV.   THE AVERAGE OF BEST PERCENT CORRECT VALUES FOR BREAST CANCER COIMBRA DATASET

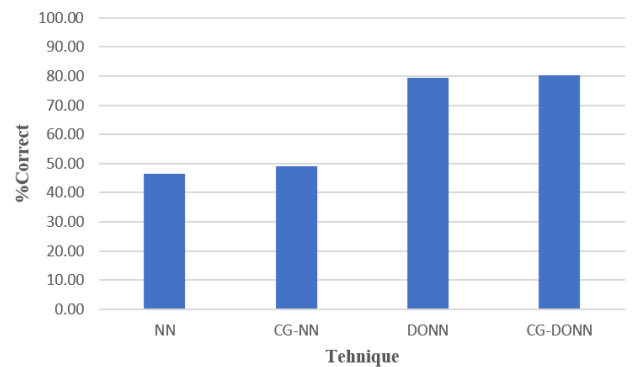| Technique (%correct)			
NN	CG-NN	DONN	CG-DONN
46.49	49.06	79.33	80.14

Fig. 6.   The average of best results for breast cancer Coimbra dataset

TABLE V.   THE OVERALL AVERAGE PERCENT CORRECT VALUES FOR BREAST CANCER COIMBRA DATASET

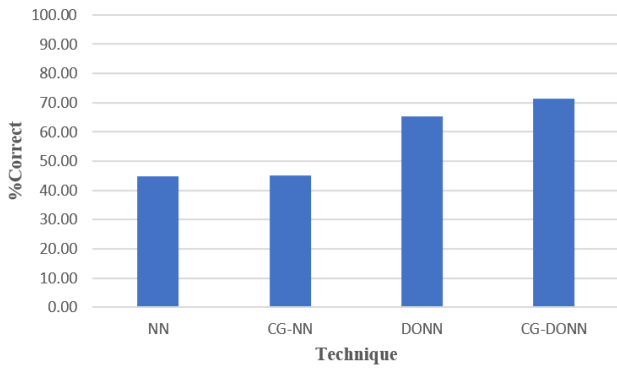| Technique (%correct)			
NN	CG-NN	DONN	CG-DONN
44.89	45.18	65.31	71.39

Fig. 7.   The overall average for breast cancer Coimbra dataset

Table VI shows the percentage increase in the percent correct of the overall average values from the existing techniques to our proposed technique CG-DONN for breast cancer Coimbra dataset.

TABLE VI.   THE PERCENTAGE INCREASE WHEN COMPARED TO CG-DONN FOR BREAST CANCER COIMBRA DATASET

Percentage increase compared to CG-DONN		
NN	CG-NN	DONN
59.02	57.99	9.31

For cervical cancer behavior risk dataset, the best and the average of fifteen results are computed. Table VII and Fig.8 show the best results whereas Table VIII and Fig.9 show the average results obtained from each fold of each technique.

TABLE VII.   THE BEST PERCENT CORRECT OBTAINED FROM EACH FOLD OF EACH TECHNIQUE FOR CERVICAL CANCER BEHAVIOR RISK DATASET

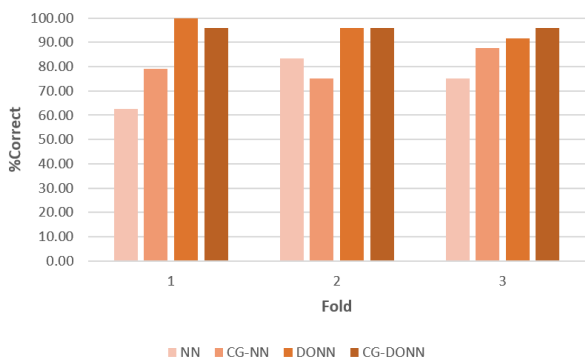| Fold	Technique (%correct)			
	NN	CG-NN	DONN	CG-DONN
1	62.50	79.17	100.00	95.83
2	83.33	75.00	95.83	95.83
3	75.00	87.50	91.67	95.83

Fig. 8.   The best result obtained from each fold of each technique for cervical cancer behavior risk dataset

TABLE VIII.   THE AVERAGE PERCENT CORRECT OBTAINED FROM EACH FOLD OF EACH TECHNIQUE FOR CERVICAL CANCER BEHAVIOR RISK DATASET

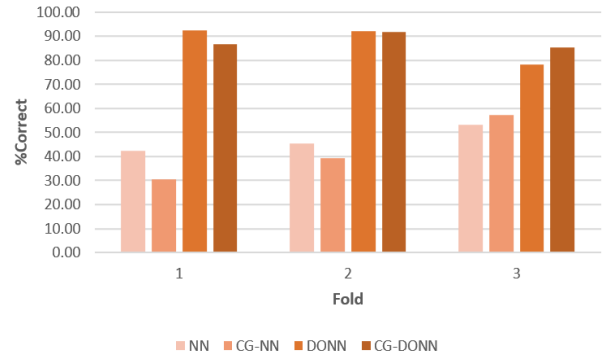| Fold	Technique (%correct)			
	NN	CG-NN	DONN	CG-DONN
1	42.50	30.54	92.50	86.67
2	45.29	39.46	92.21	91.67
3	53.04	57.21	78.33	85.29

Fig. 9.   The average result obtained from each fold of each technique for cervical cancer behavior risk dataset

The average of best results for cervical cancer behavior risk dataset are shown in Table IX and Fig.10 whereas the overall average are shown in Table X and Fig.11.

TABLE IX.   THE AVERAGE OF BEST PERCENT CORRECT VALUES FOR CERVICAL CANCER BEHAVIOR RISK DATASET

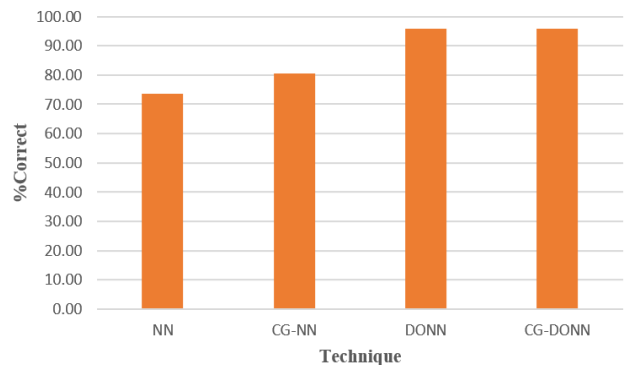| Technique (%correct)			
NN	CG-NN	DONN	CG-DONN
73.61	80.56	95.83	95.83

Fig. 10. The average of best results for cervical cancer behavior risk dataset

TABLE X.   THE OVERALL AVERAGE PERCENT CORRECT VALUES FOR CERVICAL CANCER BEHAVIOR RISK DATASET

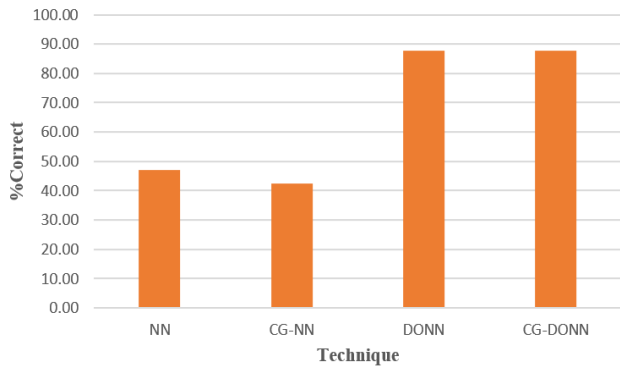| Technique (%correct)			
NN	CG-NN	DONN	CG-DONN
46.94	42.40	87.68	87.88

Fig. 11. The overall average for cervical cancer behavior risk dataset
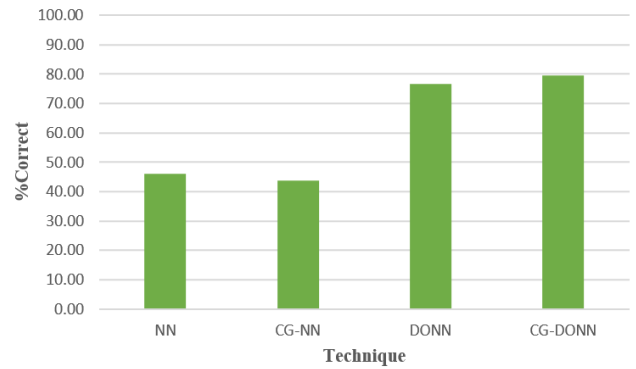

Fig. 12. The average results of both cancer datasets

Table XI shows the percentage increase in the percent correct of the overall average values from the existing techniques to our proposed technique CG-DONN for cervical cancer behavior risk dataset.

TABLE XI.   THE PERCENTAGE INCREASE WHEN COMPARED TO CG-DONN FOR CERVICAL CANCER BEHAVIOR RISK DATASET

Percentage increase compared to CG-DONN		
NN	CG-NN	DONN
87.19	107.24	0.22

From Table II, it can be seen that the best accuracy of breast cancer Coimbra dataset obtained from NN, CG-NN, DONN, and CG-DONN are 53.85, 56.41, 82.05, and 82.05, respectively. DONN and CG-DONN give the same highest accuracy. However, the technique that gives the highest average of best results shown in Table IV is CG-DONN which is 80.14.

From Table VII, the best accuracy of cervical cancer behavior risk dataset obtained from NN, CG-NN, DONN, and CG-DONN are 83.33, 87.50, 100, and 95.83, respectively. It is found that DONN provides the highest accuracy. However, the highest average of best results shown in Table IX is 95.83 which belongs to both DONN and CG-DONN.

From the overall average percent correct shown in Table V and X, it can be noticed that CG-DONN provides the maximum overall average for breast cancer Coimbra and cervical cancer behavior risk datasets which are 71.39 and 87.88, respectively. It is also found that DONN provides better average results than CG-NN and NN.

From both datasets, the average percent correct of both datasets can be computed as shown in Table XII and Fig.12. Table XIII shows the percentage increase in the percent correct from the existing techniques to the proposed technique CG-DONN for both cancer datasets. It can be seen that CG-DONN provides the highest accuracy.

TABLE XII.   THE AVERAGE PERCENT CORRECT VALUES OF BOTH DATASETS

Technique (%correct)			
NN	CG-NN	DONN	CG-DONN
45.92	43.79	76.49	79.63

TABLE XIII.   THE PERCENTAGE INCREASE WHEN COMPARED TO CG-DONN FOR BOTH CANCER DATASETS

Percentage increase compared to CG-DONN		
NN	CG-NN	DONN
73.42	81.83	4.10

V. CONCLUSION

CG-DONN is a combination of duo output neural network and cascade generalization. It is found that CG-DONN can improve learning accuracy of feedforward neural networks since it applies various kinds of learning data and utilizes multiple classifiers that are connected systematically. The proposed technique is compared to three existing individual techniques which are NN, CG-NN, and DONN. Two cancer datasets are experimented based on three-fold cross validation. Fifteen machines are run in each round of cross validation for each technique. It is found that CG-DONN is able to produce the average results better than DONN, CG-NN, and NN.

REFERENCES

[1] T. E. Idriss, A. Idri, I. Abnane and Z. Bakkoury, "Predicting Blood Glucose using an LSTM Neural Network," 2019 Federated Conference on Computer Science and Information Systems (FedCSIS), 2019, pp. 35-41.

[2] H. Benbrahim, H. Hachimi and A. Amine, "Deep Convolutional Neural Network with TensorFlow and Keras to Classify Skin Cancer Images," Scalable Computing: Practice and Experience, vol. 21, no. 3, pp. 379-389, 2020.

[3] Y. Dai, B. Xu, S. Yan and J. Xu, "Study of cardiac arrhythmia classification based on Convolutional Neural Network," Computer Science and Information Systems. vol. 17. no. 2, pp. 445-458, 2020.

[4] A. Birsen, I. A. Aydin, S. Rızalar, H. Oz, D. Meral, "Breast and Cervical Cancer Knowledge and Awareness among University Students," Asian Pacific journal of cancer prevention: APJCP, vol. 16, pp. 1719-1724, 2015.

[5] R. Agrawal, "Predictive Analysis Of Breast Cancer Using Machine Learning Techniques", ing. Solidar, vol. 15, no. 3, pp. 1-23, Sep. 2019.

[6] M. Patrício, J. Pereira, J. Crisóstomo, P. Matafome, M. Gomes, R. Seiça, F. Caramelo, "Using Resistin, glucose, age and BMI to predict the presence of breast cancer," BMC Cancer, vol. 18, 2018.

[7] Sobar, R. Machmud, A. Wijaya, "Behavior Determinant Based Cervical Cancer Early Detection with Machine Learning Algorithm," Advanced Science Letters, vol. 22, pp. 3120-3123, 2016.

[8]   M. F. Aslan, Y. Celik, K. Sabanci, and A. Durdu, "Breast Cancer Diagnosis by Different Machine Learning Methods Using Blood Analysis Data", IJISAE, vol. 6, no. 4, pp. 289-293, Dec. 2018.

[9]   Y. Austria, M. Goh, L. Jr, J. Lalata, J. Goh, H. Vicente, "Comparison of Machine Learning Algorithms in Breast Cancer Prediction Using the Coimbra Dataset," International journal of simulation: systems, science & technology, vol 20, 2019.

[10] G. Ullah, HaiYan, "Comparative performance analysis of machine learning models for breast cancer diagnosis," International Journal of Scientific and Research Publications (IJSRP), vol. 10, no. 1, 2020.

[11] Naveen, R. K. Sharma, A. Ramachandran Nair, "Efficient Breast Cancer Prediction Using Ensemble Machine Learning Models," In Proceedings of the 4th International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT), 2019, pp. 100-104.

[12] S. Poorani, P. Balasubramanie, "Deep Neural Network Classifier in Breast Cancer Prediction," International Journal of Engineering and Advanced Technology (IJEAT), vol. 9, pp. 2106-2109, 2019.

[13] A. Karaci, "Predicting Breast Cancer with Deep Neural Networks," In: Hemanth D., Kose U. (eds) Artificial Intelligence and Applied Mathematics in Engineering Problems. ICAIAME 2019. Lecture Notes on Data Engineering and Communications Technologies, vol. 43, pp. 996-1003, 2020.

[14] E., J. Kusuma, G. F. Shidik, R. A. Pramunendar, "Optimization of Neural Network using Nelder Mead in Breast Cancer Classification," International Journal of Intelligent Engineering & Systems, vol. 13 , no. 6, pp. 330-337, 2020.

[15] K. Polat and U. Sentürk, "A Novel ML Approach to Prediction of Breast Cancer: Combining of mad normalization, KMC based feature weighting and AdaBoostM1 classifier," 2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), 2018, pp. 1-4.

[16] E. Yavuz and C. Eyupoglu, "An effective approach for breast cancer diagnosis based on routine blood analysis features," Medical & Biological Engineering & Computing, vol. 58, pp. 1583–1601, 2020.

[17] H. -J. Chiu, T. -H. S. Li and P. -H. Kuo, "Breast Cancer–Detection System Using PCA, Multilayer Perceptron, Transfer Learning, and Support Vector Machine," in IEEE Access, vol. 8, pp. 204309-204324, 2020.

[18] N. F. Idris and M. A. Ismail, "Breast cancer disease classification using fuzzy-ID3 algorithm with FUZZYDBD method: automatic fuzzy database definition," PeerJ Computer Science, vol. 7, e427, 2021.

[19] P. Kraipeerapun, S. Amornsamankul, C.C. Fung, S. Nakkrasae, "Applying Duo Output Neural Networks to Solve Single Output Regression Problem," In: Leung C.S., Lee M., Chan J.H. (eds) Neural Information Processing. ICONIP 2009. Lecture Notes in Computer Science, vol. 5863, pp. 554-561, 2009.

[20] J. Gama, P. Brazdil, "Cascade Generalization," Machine Learning, vol. 41, no. 3, pp. 315–343, 2000.

[21] D. Dua, C. Graff, UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science, 2019.

Densely Populated Regions Face Masks Localization and Classification Using Deep Learning Models

Anh Pham-Hoang-Nam
School of Computer Science and Engineering
International University
Ho Chi Minh City, Vietnam
phna0220@gmail.com

Linh Phung-Khanh
School of Computer Science and Engineering
International University
Ho Chi Minh City, Vietnam
phungkhanhlinh.iu@gmail.com

Vi Le-Thi-Tuong
School of Computer Science and Engineering
International University
Ho Chi Minh City, Vietnam
lttvi1822@gmail.com

Nga Ly-Tu
School of Computer Science and Engineering
International University
Ho Chi Minh City, Vietnam
ltnga@hcmiu.edu.vn

Abstract—Over the last year, the correct wearing of facial masks in public is still a relevant matter in the fight against the COVID-19 pandemic. A popular approach that helps regulate the situation by global researchers is building smart systems for face mask detection. Following such spirit, this paper will contribute to the literature in two main aspects:
(1) We first propose a new face mask detector model using the state-of-the-art RetinaFace for face localization in populous regions and the ResNet50V1 classifier to group the faces under 3 categories: correctly-worn, incorrectly-worn and no-masks-worn.
(2) In order to select the ResNet50V1 as the backbone for the final model, we also analyzed its performance in accordance with another 3 classifiers on a face mask dataset beforehand. Performance metrics from the test phase have shown that our detector achieved the best accuracy among all the works compared, with $94,59\%$ on one test dataset and a less satisfactory 69.6% on another due to certain characteristics of the set. The code is available at: https://github.com/barbatoz0220/ Densely-populated-FMD.git

Index Terms—Dense Population Regions, Face Mask, Localization, Classification, Covid-19, Deep Learning, MobileNet, ResNet, AIZOO, Neuralet

I. Introduction

Unexpected as it may seem, the discovery of the first COVID-19 case has tragically started a series of ongoing depressive episodes for many people across the world, while at the same time presenting a global reordering moment in many aspects [1]. The pandemic has been listed as an extreme global crisis when the 1.4 million infected cases in April 2020 [2] have risen to more than 250 million [3]. These statistics would have been worse had it not been for the intensive implementation and conformance to many suggested preventive measures [2], from which the subject of our research - the

correct wearing of face masks in public - is withdrawn since it is the most recommended, widely applicable and highly effective in reducing transmission rate with or without the implementation of other intervention methods [4].

While there exists a vast body of other technologies that are being utilized to help relieve the situation, within the scope of our research, we will solely focus on the use of Deep Learning in the field of Face Recognition to detect correct face-masks-wearing in public. Following such premise, the technology has proven to remain a trending topic as the latest review of Wang and Deng has introduced and discussed thoroughly the past, present and future various concepts as well as researches [5]. Even more so, with respect to growing concerns regarding the COVID-19 global pandemic up to date, the technology has definitely gathered enough traction and gained significant interest when research teams around the world rushed to develop and propose Deep Learning models to help protect the health of public communities [6]. Whether the degree of time was back in the early stages around which the outbreak occurred or varied through the current year, many research papers were studied and published in dedication to the means of carefully monitoring the usage of facial masks in public place [7]. For example, a three-component model has been proposed by Loey et al. in their research to supervise the wearing of medical face masks in public [8]. By using YOLO-v2 with ResNet-50, the hybrid model was then exclusively trained on 2 datasets featuring medical face masks and achieved a higher average precision rate (81%) in comparison to one of its related works' model, which was trained on a different dataset with mixed types of face masks (76.1%). Another approach from Loey et al. is to focus on the detection of people who are not wearing face masks in public [9]. Using ResNet-50 combined with Support Vector Machine
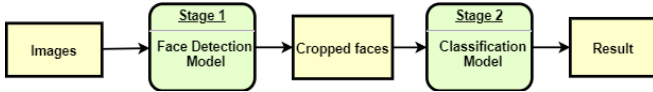
Fig. 1: Face Mask Detection Sequential Model

Fig. 2: Architecture of Our Classification Head Model

(SVM) and ensemble algorithm, their results have shown that the SVM classifier generally performed better than the others in comparison and also achieved higher accuracy scores using the same datasets. One interesting and realistic contribution, as expressed by Rudraraju et al. [10], was to consider the monitoring of entry and access control as another function to the face mask detection system. This was also one of the few works surveyed that also performed extensive research on the topic of correctness in mask wearing through the architecture of their application.

However, given how the wearing of different masks types as well as the validity of mask-wearing itself in crowded areas has not been broadly reviewed in the literature, we will attempt to fill in that gap with the proposal of a two-stage face mask detection model in this research paper. Our contributions can be summarized as follows:

- We will propose a face mask classification model that employs transfer learning by combining our head model with a backbone model.
- In order to opt for the most prominent backbone, from the architectures of MobileNetV1 [11], MobileNetV2 [12], ResNet50V1 [13] and ResNet50V2 [14], we respectively evaluate their versions of classification models on a custom dataset that we created using the two sources Kaggle-12K [15] and the MaskedFace-Net [16].
- Following the integration of our classification model and RetinaFace, the proposed model's capabilities will then be assessed against the AIZOO Face Mask Detector [17] and Neuralet Face Mask Detector [18] on the two datasets Face Mask Detection [19] and the MAsked FAce [20].

Following this introduction, Section II will briefly describe our methodology; then, the specifications of our setups behind the experiences will be detailed in Section III, whereas Section IV will elaborate further on how each steps in those experiments are conducted with their corresponding results to justify our proposed model abilities; and ultimately, our research paper will be concluded alongside a brief view at possible future improvements at the end of Section V.

II. METHODOLOGY

The face mask detector that we are proposing will be composed of two stages: the first stage is the identification of facial regions from an image or a frame and the second one is the classification of detected faces into pre-defined subcategories. The stages are demonstrated following the orders in Figure 1. The design of such sequential model brings 2 benefits to our detector.

Firstly, the state-of-the-art RetinaFace will be applied to identify the Region of Interest (RoI) or faces available in one
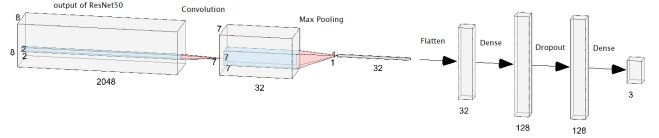
image. Being a single-stage dense face localization model, it can achieve an average precision score of 91.286% [21] on the hard subset of the WIDER FACE dataset. In other words, the mentioned face detection model was trained given a dataset of crowded, diverse types of faces, including normal, masked and make-up faces. Such process is believed to be beneficial to our model in terms of increasing detection accuracy.

Secondly, during the second stage of classifying all the detected faces, transfer learning is applied using the four backbone models: MobileNetV1, MobileNetV2, ResNet50V1, ResNet50V2. The experiments to opt for the backbone models are described in Section IV. Our head model is simple and light-weighted, including a convolution layer, a max-pooling layer, and a series of fully connected layers and drop-out. Details about the head of the classification model can be found in Fig. 2. The probabilistic output is (p_1, p_2, p_3), where p_i is simply the corresponding probability that the classified image belongs to one of the three classes discussed later in Section IV-A. Such architecture allows us to freely opt for the most prominent backbone model without having to modify our entire model.

III. IMPLEMENTATION

A. Training details

The model is trained using the binary loss function and Adam optimizer with the initial learning rate at 10^{-4}, dropout rate at 0.5 and batch size of 32 on NVIDIA Tesla V100-SXM2 ($16GB$) using Google Colab environment. Therefore, we will also upload all the datasets to accessible Google Drive folders on the same account for ease of execution and reusability. The training process terminates after 50 *epochs*.

B. Validating and testing details

1) Validating details: There are three purposes to the validating process. Evidently, the first is to evaluate the

TABLE I: Testing Dataset Information

Dataset	No. images	No. faces	No. class 1	No. class 2	No. class 3
Kaggle	853	4072	3232	717	123
MAFA	4935	10033	6354	996	-

where *No. images, No. faces, No. class 1, No. class 2, No. class 3* are defined as the number of images, of faces in total and of faces belonging to each of the correctly-worn, no-mask-worn and incorrectly-worn class, respectively.

performance of our classification model. Additional to that, such process will discover how different base models, or the so-called backbones, might affect the efficiency of the classification model once applied. Last but not least, due to the various sources of our training and validation datasets, their heterogeneity, and imbalance should be carefully revised to see whether they are strongly involved in the creation of our model. The metrics mentioned in this subsection are determined following said reasons.

We first define the accuracy of the classification model on the validation set after 50 epochs as follows:

$$\psi_{classify}(y, \hat{y}) = \frac{1}{m} \sum_{i=0}^{m-1} 1(\hat{y}_i = y_i), \tag{1}$$

where m is the size of the validation set and y, $\hat{y}$ is the set of classified and ground-truth label, accordingly.

To illustrate the model stability, the variance σ^2 of the classification model for both training and validation sets is defined in (2).

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (\psi_i - \mu)^2, \tag{2}$$

where ψ is the accuracy obtained after each epoch i and n is the total number of epochs.

Precision will be used to estimate the classification abilities of our model to not falsely label negative samples. The calculation of recall will also be adopted to estimate the sensitivity in finding all positive samples. The general formula of precision $P(y, \hat{y})$ and recall $R(y, \hat{y})$ are given in (3) and (4).

$$P(y, \hat{y}) = \frac{TP}{TP + FP} \tag{3}$$

$$R(y, \hat{y}) = \frac{TP}{TP + FN}, \tag{4}$$

where TP, TN, FP, FN are respectively true positives, true negative, false positive, false negative number of classified cases.

With the imbalance of the dataset under consideration, a computation is made for the two types of precision and recall: macro and weighted. While macro metrics ignore the imbalance of the dataset, weighted metrics take into account how it can alter the final results. The formula for the macro and weighted precision and recall are given in (5), (6), (7) and (8) respectively.

Let L be the set of available classes, y and $\hat{y}$ are defined as in (1), we have:

$$P_{macro}(y, \hat{y}) = \frac{1}{|L|} \sum_{l \in L} P(y_l, \hat{y}_l) \tag{5}$$

$$P_{weighted}(y, \hat{y}) = \frac{1}{\sum_{l \in L} |\hat{y}_l|} \sum_{l \in L} |\hat{y}_l| P(y_l, \hat{y}_l) \tag{6}$$

$$R_{macro}(y, \hat{y}) = \frac{1}{|L|} \sum_{l \in L} R(y_l, \hat{y}_l) \tag{7}$$

$$R_{weighted}(y, \hat{y}) = \frac{1}{\sum_{l \in L} |\hat{y}_l|} \sum_{l \in L} |\hat{y}_l| R(y_l, \hat{y}_l) \tag{8}$$

2) Testing details: In order to reasonably evaluate the model with the test set, three metrics, namely average confidence, accuracy of face detection model, and ordinary accuracy are defined in (9), (11), (10).

Let D be the set of all faces detected by our model, then the average confidence β, known as its ability to correctly classify a given face, can be described as:

$$\beta = \overline{1 - \alpha} = \frac{1}{|D|} \sum_{d \in D} \psi_d, \tag{9}$$

where α is the significance level and ψ is the accuracy or the confidence of each detected face.

Assume that T represents the set of all ground-truth faces, we define Ψ_{RoI} as the accuracy of face detection model computed by the formula:

$$\Psi_{RoI} = \frac{|D|}{|T|} \tag{10}$$

The last metric to be used in the testing process is the final accuracy, representing the ability of the model to correctly classify a given face.

Let $A \subset D$, where A is the set of correctly localized and classified faces, the final accuracy Ψ_{final} is defined as

$$\Psi_{final} = \frac{|A|}{|D|} \tag{11}$$

The evaluation method is given in Algorithm 1, in which the threshold is set at $\theta_{lower} = 0.5$. This parameter is used in Line 7 under the form of pseudocode.

Algorithm 1 Evaluation using IoU

1: **for** $image = 1, 2 \ldots, N$ **do**
2:   $detectedFaces \leftarrow$ Detect faces in that image and classify them
3:   $totalDetectedFaces+ = len(detectedFaces)$
4:   $totalConfidence+ = \sum detectedFacesConfidence$
5:   **for** $realFace = 1, 2, \ldots$ **do**
6:     **for** $detectedFace = 1, 2, \ldots$ **do**
7:       **if** $\theta_{lower} < IoU(realFaceBoundingBox,$
        $detectedFaceBoundingBox)$ **then**
8:         **if** $realFaceClass == detectedFaceClass$ **then**
9:           $trueVal+ = 1$
10:           Break the outer for loop since the correct detected face has been found
11:         **end if**
12:       **end if**
13:     **end for**
14:   **end for**
15: **end for**
16: Calculate $\beta, \Psi_{RoI}, \Psi_{final}$ in (9), (10), (11) respectively.

IV. EXPERIMENT AND RESULT

Given the 4 datasets used in our research, namely Kaggle-12K [15], MaskedFace-Net [16], Kaggle Face Mask [19] and MAsked FAces (MAFA) [20], we have decided to select them accordingly for the tasks and to attain even distribution among the 3 classes of mask-wearing.

A. Dataset

1) Training and validation dataset: The training dataset for the classification stage is a human face dataset composed of all the images collected from two sources, namely, Kaggle-12K [15] and MaskedFace-Net [16]. These images are then classified into 3 classes, as seen in Table I. Class 1 represents the correctly-worn-mask faces, Class 2 being the incorrectly-worn-mask faces, and Class 3 for the no-mask-worn faces. The masks used in this dataset are not necessarily medical masks. The class correctly-worn-mask faces is designated to faces that are fully covered with fabric or medical masks. On the contrary, faces are classified into incorrectly-worn-mask when the people in question are wearing their masks in such ways that vital parts like noses or mouths are left uncovered. There are 13338 images in total, in which 3594 of them are in the incorrectly-worn-mask class, 4816 are for the correctly-worn class, and 4928 for the no-mask faces. The distribution of images for each class are shown in Fig. 3. To appropriately train and validate the classification models, we have divided this dataset into 2 subsets, in which 90% of the images are used as training set and the other 10% are used as validation set. All images contain only 1 face.

2) Testing dataset: Our model will be tested and evaluated on the two datasets Kaggle Face Mask [19] and MAsked FAces (MAFA) [20] with their details listed in Table I. Kindly note that each image in these sets will contain multiple faces for the sake of testing the entire model, and that the annotation file in MAFA has declared the "invalid" class to hold various cases that do not follow our definition of incorrectly-worn-mask class (for example, the faces in question were occluded, blurred, or partially shown in their corresponding frames, etc.); ergo, they have been excluded from our consideration and left untouched.

B. Comparison of our classification model backbone

In this subsection, we will validate the four classification backbone models, namely MobileNetV1, MobileNetV2, ResNet50V1 and ResNet50V2, and filter out the model with the best performance possible.

1) MobileNets: For practical purposes, it is ideal that our face mask detection model can be embedded in cameras or smaller digital devices. Therefore, small yet powerful architectures as MobileNets are considered here in our works. MobileNetV1, proposed by Google in 2017, is a rare model which combines depth wise separable convolutions and 11
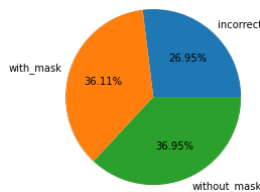

Fig. 3: Training dataset distribution

standard convolutions for its input channels filtering and linear computation. MobileNetV1 only has 88 layers in total and the size of $16MB$ [11]. MobileNetV2, proposed by Google a year after the introduction of MobileNetV1, is an updated version of its first version with highlights in the two new features: linear bottlenecks between the layers, and shortcut connections between the bottlenecks [12]. It was reported to be lighter and $30 - 40\%$ faster on a Google Pixel phone than MobileNetV1 while having an increase in ImageNet Top 1 accuracy [12, 22].

With MobileNets as our backbones to support the head model described earlier, we are able to provide the information about the accuracy and loss of the model during 50 epochs in Table II. In general, the implementation of both MobileNets as backbones is synonymous with the increase in accuracy and decrease in loss. We can acknowledge that the validation accuracy of the model using MobileNetV2 as the backbone is less stable than the one with MobileNetV1. Withstanding some minor fluctuations, the loss of MobileNetV2-supported model on the validation set tends to increase in the later epochs, whereas its training loss remains the same. For further evaluation of our usage of MobileNets as backbone, the mentioned metrics in Section III are calculated in Table II.

Despite the better performance of MobileNetV2 compared to its ancestor in ImageNet [12, 22], on our dataset, particularly the validation set, MobileNetV1-supported model results in a higher accuracy of 0.9981 after 50 epochs with the minimum accuracy being 0.9940 and the maximum 1.0. In contrast, the accuracy of MobileNetV2-based model ranges from 0.9910 to 0.9985 with the final accuracy at 0.9970. The classification model based on MobileNetV1 is also more stable than the one on MobileNetV2. It should be noted that, in spite of being 1.4 times less in training accuracy variance, the MobileNetV2-supported model scores more in validation accuracy than the MobileNetV1-based. All the precision and recall metrics of the classification model with MobileNetV1 are slightly higher than those of that using MobileNetV2. Additionally, within the margins of our experiments, the MobileNetV1-supported model tends to classify in a shorter duration of 1.139 seconds in comparison with 1.445 seconds of MobileNetV2. Lastly, it is observed from the weighted and macro precision and recall metrics that the imbalance of our dataset is trivial, and thus the validation process strictly follows the performance of the models. Conclusively, MobileNetV1-supported classification model is able to achieve higher results on our dataset and within a shorter period of run time as opposed to the one employing MobileNetV2 as its backbone.

2) ResNets: Being one of the most groundbreaking architecture developed in 2015 by Microsoft and won 1st place in the ILSVRC classification competition with top-5 error rate of 3.57%, ResNet uses residual blocks, which applied the idea of skip connections, to overcome the problem of vanishing gradients while the depth of the convolution network increases [13]. A few months after the birth of ResNet, the second version of it was proposed also by Microsoft. ResNetV2 improved the residual unit, which facilitates the training process and improves generalization [14].

TABLE II: Evaluation on Our Classification Model Using MobileNets

Backbone	σ^2_{train}	σ^2_{val}	$\psi_{classify}$	P_{macro}	$P_{weighted}$	R_{macro}	$R_{weighted}$	Run Time(s)
MobileNetV1	$1.507*10^{-4}$	$1.759*10^{-6}$	0.9981	0.9982	0.9981	0.9980	0.9981	1.139
MobileNetV2	$1.053*10^{-4}$	$1.896*10^{-6}$	0.9970	0.9972	0.9970	0.9967	0.9970	1.445

TABLE III: Evaluation on Our Classification Model Using ResNets

Backbone	σ^2_{train}	σ^2_{val}	$\psi_{classify}$	P_{macro}	$P_{weighted}$	R_{macro}	$R_{weighted}$	Run Time (s)
ResNet50V1	$4.895*10^{-5}$	$4.183*10^{-7}$	0.9982	0.9984	0.9983	0.9981	0.9982	2.467
ResNet50V2	$2.037*10^{-4}$	$4.003*10^{-6}$	0.9985	0.9986	0.9985	0.9981	0.9985	2.274

TABLE IV: Summary of Models

Model	Dataset	β	Ψ_{RoI}	Ψ_{final}
Proposed	Kaggle	0.9960	0.8347	0.9459
	MAFA	0.9950	0.9580	0.6960
AIZOO FMD	Kaggle	0.8671	0.5454	0.8249
	MAFA	0.9325	0.7938	0.6453
Neuralet FMD	Kaggle	0.9723	0.4050	0.3220
	MAFA	0.9383	0.1886	0.0587

Information concerning the loss and accuracy of ResNet50V1-based and ResNet50V2-based models within 50 epochs is shown in Table. III. It is visible that the accuracy of the training process for these two models rises sharply and then stabilizes. Inversely, on the validation set, both models suffer from minor fluctuations. There are also similarities in their loss functions as the training loss values of both swiftly decrease in the beginning then remain constant until some later epochs where they are to rise one more time. Their evaluation metrics can be found in Table III. By comparison, the scoring results obtained from ResNet50V2-supported model are 0.0002 unit higher than those from other model roughly. While the run time of the model based on ResNet50V1 is longer than that on ResNet50V2, the variance in both train and validation sets' accuracy scores of ResNet50V1 based are approximately 4 to 10 times smaller than of ResNet50V2.

Considering all results, the model based on ResNet50V2 backbone scores the highest in accuracy, precision and recall (both weighted and unweighted). Meanwhile, the ResNet50V1-based model emerged with the lowest variance for both the train and validation sets but at the same time the slowest classifier. Nevertheless, the duration of run time of ResNets are about 2 times longer than that of MobileNets. For the last assessment, the final model, including both localization and classification stages, is tested on the Kaggle dataset. There, we were able to localize 3399 faces over 4072 faces, 3215 faces of which are accurately classified by the model using ResNet50V1. The other three, namely ResNet50V2, MobileNetV1 and MobileNetV2, respectively made 2992, 2992, 2991 classifications. Based on the small variance on the previous dataset and the high accuracy on this dataset, the ResNet50V1-supported model was the reasonably ideal backbone for the classification model that will be used in the following subsection.

C. Result

We now compare our proposed model with another two models, specifically, AIZOO and Neuralet face mask detectors (AIZOO FMD and Neuralet FMD for short, respectively). AIZOO FMD developed by AIZOOTech, is a light-weighted, single-stage detector with only 1.01 million parameters and 24 layers for location and classification. AIZOO supports all popular deep learning frameworks model and inference code. In this comparison, we use only the Tensorflow version [17]. Neuralet FMD is supported by Neuralet company and is sponsored by Lanthorn Solutions. It is a two-stage detector that can both propose region of interests and provide the needed classification on those regions. Its first stage, known as the detector stage, uses one of two models for the ×86 configuration, openpifpaf model (model to estimate human pose estimation) and tinyface model. In this subsection, we run the model using openpifpaf for detector and OFM Classifier for the classifier of Neuralet FMD [18].

As shown in Fig. 4 and Fig. 5, the results for accuracy and confidence of our proposed model are higher than those of AIZOO and Neuralet on both datasets. By using RetinaFace for face localization, which was able to detect 900 faces out of reportedly 1151 people, our proposed model managed to recognize more faces per photo and returned higher score of Ψ_{RoI} than the other mentioned models on both test datasets.

With ResNet50V1, the accuracy of our model to appropriately detect the faces across more than 800 images of the Kaggle dataset is at 0.9459 and is relatively higher than the compared models. On the other hand, for the MAFA dataset, perhaps as a consequence of the mentioned omission of incorrectly-worn masks previously mentioned in section III, the accuracy we managed to achieve was an unsatisfactory value of 0.696. Still, this result is the highest one achieved among the three models, despite the fact that ours was trained to label the same faces under the three aforementioned categories, while the conventionally used annotation in the dataset contained only two labels that both AIZOO and Neuralet could follow. The proposed model also gave competent results on the scores, or the so-called confidence, of precisely classifying a given region of interest.

By achieving good results in Ψ_{RoI} and Ψ_{final}, our model has therefore showcased its ability to (1) convincingly detect and classify faces in dense population regions as well as (2) to separate the detected faces into 3 classes, which enabled a more appropriate classification of the incorrectly worn mask faces. Please refer to the resulted illustrations for both cases that were featured on our GitHub repository using this link: https://github.com/barbatoz0220/Densely-populated-FMD.git
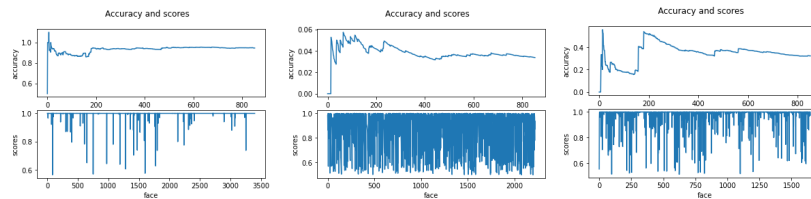
Fig. 4: Comparison of Accuracy and Score (Confidence) Among the Proposed Model - AIZOO - Neuralet on Kaggle Dataset
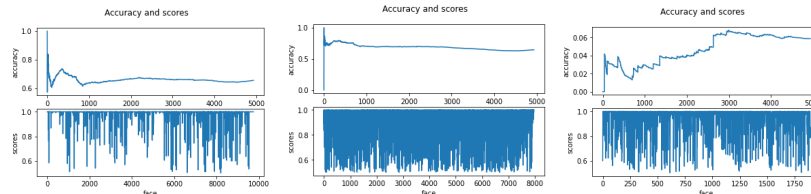

Fig. 5: Comparison of Accuracy and Score (Confidence) Among the Proposed Model - AIZOO - Neuralet on MAFA Dataset

V. CONCLUSION

Given the lack of research on mask detection for crowded regions, as well as on the classification of face masks with respect to variety and validity, we have proposed a new face mask detector model for detection in densely populated regions and validation of masks wearing following how they are worn: correctly, incorrectly and without. Our proposed model correctly localized 83.47% faces and classified 94.59% of the confined set. While there are still certain limitations to some of the class variance, the performance metrics have justified our effectiveness in the combination of ResNet50V1 and RetinaFace. It is certainly possible for our model to be better optimized and utilized in the foreseeable future. In terms of data, we firmly believe that extensive attempts to improve the imbalance in current sets and, perhaps, to modify our own set from renowned sources will allow us to achieve more optimistic results. Further researches to exploit more capable architectures would definitely be considered and integration with tools, such as OpenCV, will ensure more opportunities for public usage through real-world application.

REFERENCES

[1] Warwick McKibbin and Roshen Fernando. The economic impact of covid-19. *Economics in the Time of COVID-19*, 45, 2020.
[2] World Health Organization et al. World health organization coronavirus disease 2019 (covid-19) situation report, 2020.
[3] Worldometers. Coronavirus updates (live) - covid-19 coronavirus pandemic. Available at https://www.worldometers.info/coronavirus/.
[4] Steffen E Eikenberry, Marina Mancuso, Enahoro Iboi, Tin Phan, Keenan Eikenberry, Yang Kuang, Eric Kostelich, and Abba B Gumel. To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the covid-19 pandemic. *Infectious Disease Modelling*, 5:293–308, 2020.
[5] Mei Wang and Weihong Deng. Deep face recognition: A survey. *Neurocomputing*, 429:215–244, 2021.
[6] Connor Shorten, Taghi M Khoshgoftaar, and Borko Furht. Deep learning applications for covid-19. *Journal of big Data*, 8(1):1–54, 2021.
[7] Elliot Mbunge, Sakhile Simelane, Stephen G Fashoto, Boluwaji Akinnuwesi, and Andile S Metfula. Application of deep learning and machine learning models to detect covid-19 face masks-a review. *Sustainable Operations and Computers*, 2021.

[8] Mohamed Loey, Gunasekaran Manogaran, Mohamed Hamed N Taha, and Nour Eldeen M Khalifa. Fighting against covid-19: A novel deep learning model based on yolo-v2 with resnet-50 for medical face mask detection. *Sustainable Cities and Society*, page 102600, 2020.
[9] Mohamed Loey, Gunasekaran Manogaran, Mohamed Hamed N Taha, and Nour Eldeen M Khalifa. A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the covid-19 pandemic. *Measurement*, 167:108288, 2021.
[10] Srinivasa Raju Rudraraju, Nagender Kumar Suryadevara, and Atul Negi. Face mask detection at the fog computing gateway. In *2020 15th Conference on Computer Science and Information Systems (FedCSIS)*, pages 521–524. IEEE, 2020.
[11] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.
[12] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018.
[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *European conference on computer vision*, pages 630–645. Springer, 2016.
[15] Ashish Jangra. Face mask 12k images dataset - 12k images divided in training testing and validation directories, May 2020. Available at https://www.kaggle.com/ashishjangra27/face-mask-12k-images-dataset.
[16] Adnane Cabani, Karim Hammoudi, Halim Benhabiles, and Mahmoud Melkemi. Maskedface-net – a dataset of correctly/incorrectly masked face images in the context of covid-19. *Smart Health*, 19:100144, 2021. Available at https://github.com/cabani/MaskedFace-Net.
[17] Detect faces and determine whether people are wearing mask. Available at https://github.com/AIZOOTech/FaceMaskDetection.
[18] Face mask detection at the edge. Available at https://neuralet.com/face-mask-detection-at-the-edge/#showcase-section-4.
[19] Larxel. Face mask detection, May 2020. Available at https://www.kaggle.com/andrewmvd/face-mask-detection.
[20] Rahul. Masked face data, May 2020. Available at https://www.kaggle.com/rahulmangalampalli/mafa-data.
[21] Jiankang Deng, Jia Guo, Evangelos Ververas, Irene Kotsia, and Stefanos Zafeiriou. Retinaface: Single-shot multi-level face localisation in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5203–5212, 2020.
[22] Howard Sandler. Mobilenetv2: The next generation of on-device computer vision networks. Available at https://ai.googleblog.com/2018/04/mobilenetv2-next-generation-of-on-html.

State Observer-Based Backstepping Sliding Mode Control for Electro-Hydraulic Systems

Phan Viet Tan
Hanoi, Vietnam

Nguyen Quang Duc
Hanoi, Vietnam

Le Ngoc Minh
Hanoi, Vietnam
minh.ln181646@sis.hust.edu.vn

Cuong Nguyen Manh
Hanoi, Vietnam

Danh Huy Nguyen
Hanoi, Vietnam
huy.nguyendanh@hust.edu.vn

Tung Lam Nguyen
Hanoi, Vietnam
lam.nguyentung@hust.edu.vn

Abstract—In the paper, a control method using backstepping integrated with sliding mode control based on high-gain observer is proposed for an Electro-Hydraulic system (EHS). The observer is dedicated to estimate full-state variables by using output position feedback. In the initial step, the control design formula is constructed via backstepping technique. Following that, the control signal is designed based on sliding mode control (SMC) algorithm, and thus lead to the controller that has the advantages of these two control laws.

Index Terms—component, backstepping, high-gain observer, electro–hydraulic system (EHS), sliding mode.

I. Introduction

NOWADAYS, various useful applications related to electro-hydraulic system (EHS), including manipulators [1], anti-lock braking systems [2], hydraulic excavator [3] and plate hot rolling [4] are extremely important in modern industrial automation due to its distinct advantages: higher load stiffness, superior load efficiency and high power-to-weight ratio.

However, EHSs exhibit significant dynamical nonlinearities such the complicated flow properties of servo valves. Numerous control methods have been studied and proposed to achieve the acceptable control quality – such as the local linearization of nonlinear dynamics in [5]. Practically, to overcome the control design's challenges in precisely measuring the state parameters or the disturbance through sensors due to the limitation of space and cost constraint, there is a trend in using a variety of different observers, mentioned in [6] and [7]. Recently, the high-gain state observer is proposed to estimate the state parameters, and especially, one can tune the observer bandwidth to obtain the desired stability/robustness properties [6]. For the control technique, many have used PID [8], backstepping [7], and sliding mode [9]. One of the most common approaches is backstepping method, it was applied to EHS [6] for achieving exponential convergence of the system position and load pressure. But in some cases, the control law using this technique is sensitive to noise and the higher the order of the system, the more complex the calculation steps, hence reduces the response speed of the system.

Sliding mode control approach is also a suitable alternative due to the significant efficiency in stabilizing the position controlling in the changes of the system [9]. It gives us a much more quality performance when there was appearance of disturbance but meets difficulties with massive calculation and "the chattering effect" leading to considerable wear of moving mechanical parts and high heat lost in power circuits [10].

In this study, a high-gain state observer integrated backstepping control with an improvement in combining the sliding mode control is designed to enhance the quality in position tracking. By using the state parameters estimated from the observer and applying a hybrid of backstepping and sliding mode control, the proposed control achieves good system performances.

II. Electro-Hydraulic System model formulation

The physical equations describing the EHSs is (Figure 1) presented in detail in [7] and briefly given as

$$\begin{cases} \dfrac{d\underline{x}}{dt} = A\underline{x} + \underline{\gamma}(\underline{x}, u, y) \\ y = C\underline{x} \end{cases} \quad (1)$$

Where: $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ and

$$\underline{\gamma}(\underline{x}, u, y) = \begin{bmatrix} 0 \\ k_1 x_1 + k_2 x_2 + (k_3 - 1)x_3 + d_2 \\ g_1 x_2 + g_2 x_3 + d_3 + g_3 u \end{bmatrix}$$
$$= [0 \quad \gamma_2 \quad \gamma_3]^T$$

$\underline{\gamma}(\underline{x}, u, y)$ satisfies Locally Lipschitz Condition in the following compact set:

$$\left\| \underline{\gamma}(\underline{x}, u, y) - \underline{\gamma}(\hat{\underline{x}}, u, y) \right\| \le a \| \underline{x} - \hat{\underline{x}} \| \quad \forall \underline{x}, \hat{\underline{x}}, u, y \quad (2)$$

where a is called Lipschitz constant and $C = [1\ 0\ 0]$ and:

$$k_1 = -\frac{k}{m}; k_2 = -\frac{b}{m}; k_3 = \frac{A_p}{m}; g_1 = -h_1; g_2 = -h_2; g_3 = h_3\sqrt{P_s - sgn(u)x_3}$$

$$d_2 = -\frac{F_L}{m} - \frac{F_F}{m} + \Delta f_2; d_3 = \Delta f_3$$
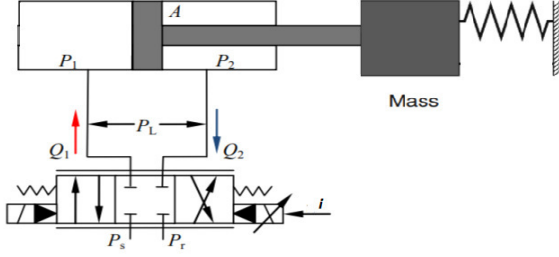

Fig. 1. A Representative EHS

III. HGGSOB – BASED BACKSTEPPING SLIDING MODE CONTROL

A. High-gain States Observer Design

To approximate $\underline{x}$ we propose a state observer as follow:

$$\frac{d\hat{\underline{x}}}{dt} = A\hat{\underline{x}} + \gamma(\hat{\underline{x}}, u, y) + L(y - \hat{y}) \qquad (3)$$

where $\hat{\underline{x}}$ is the state estimation and $L = [l_1 \quad l_2 \quad l_3]^T$ is the observer gain, chosen such that $(A - LC)$ has all real parts of eigenvalues in the left half plane, constant matrix L is given as follows $L = [h_1\varepsilon^{-1} \quad h_2\varepsilon^{-2} \quad h_3\varepsilon^{-3}]^T$, where ε is small positive number and h_1, h_2, h_3 are coefficient of the Hurwitz polynomial.

B. Backstepping- Sliding Mode Control design

We define the position of piston/output, velocity of piston and pressure difference tracking error as e_1, e_2, e_3. Let us start by defying the first Lyapunov candidate function as:

$$V_1 = \frac{1}{2}e_1^2 \qquad (4)$$

The derivative of the V_1 is given by:

$$\dot{V} = e_1\dot{e}_1 = e_1(x_2 - \dot{x}_{1d}) \qquad (5)$$

We choose $x_{2d} = \dot{x}_{1d} - p_1e_1$ where p_1 is a positive number, then:

$$\dot{V} = -p_1e_1^2 + e_1e_2 \qquad (6)$$

Let us define V_2 as:

$$V_2 = V_1 + \frac{1}{2}e_2^2 \qquad (7)$$

Then

$$\dot{V}_2 = -p_1e_1^2 + e_1e_2 + e_2\dot{e}_2$$
$$= -p_1e_1^2 + e_2(e_1 + \dot{e}_2) \qquad (8)$$

From (7), $\dot{V}_2$ can be rewritten as:

$$\dot{V}_2 = -p_1e_1^2 + e_2(e_1 + k_1x_1 + k_2x_2 + k_3x_3 + d_2 - \dot{x}_{2d}) \qquad (9)$$

Let choose:

$$x_{3d} = \frac{1}{k_3}(-e_1 - k_1x_1 - k_2x_2 - d_2 + \dot{x}_{2d} - p_2e_2) \qquad (10)$$

where p_2 is a strict positive gain, this results in:

$$\dot{V}_2 = -p_1e_1^2 - p_2e_2^2 + k_3e_2e_3 \qquad (11)$$

Assuming e_3 will be driven to zero, it can be shown that:

$$\dot{V}_2 = -p_1e_1^2 - p_2e_2^2 \leq 0 \qquad (12)$$

Now u is chosen to drive e_3 to zero, we define sliding surface S_3 as follow:

$$S_3 = x_3 - x_{3d} \qquad (13)$$

Differentiating sliding surface S_3 yields:

$$\dot{S}_3 = \dot{x}_3 - \dot{x}_{3d}$$
$$= g_1x_2 + g_2x_3 + g_3u + d_3 - \dot{x}_{3d} \qquad (14)$$

The control signal u can be divided to two components: u_{eq} will drive sliding surface to zero and u_{sw} will lead the system states on the sliding surface thus the control signal can be written as:

$$u = u_{eq} + u_{sw} \qquad (15)$$

From (28), it is straightforward to show that:

$$u_{eq} = \frac{1}{g_3}(-g_1x_2 - g_2x_3 - d_3 + \dot{x}_{3d}) \qquad (16)$$

In order to have $S_3 \rightarrow 0$, we need to design signal u_{sw} renders $S_3\dot{S}_3 < 0$, so u_{sw} can be selected as follow:

$$u_{sw} = \frac{1}{g_3}(-m_1\text{sign}(S_3) - m_2S_3) \qquad (17)$$

where m_1, m_2 are positive constants. With the designed control signal, the stability of the system will be considered according to the Lyapunov stable standards. The third Lyapunov candidate function is proposed to prove the stability of the whole system:

$$V_3 = \frac{1}{2}S_3^2 \qquad (18)$$

Taking time derivative of V_3 results in:

$$\dot{V}_3 = -m_1S_3\text{sign}(S_3) - m_2S_3^2 \leq 0 \qquad (19)$$

Finally, the control law can be written as:

$$\begin{cases} u = \frac{1}{g_3}(g_1x_2 + g_2x_3 - d_3 + \dot{x}_{3d}) + \frac{1}{g_3}(-m_1\text{sign}(S_3) - m_2S_3) \\ x_{2d} = \dot{x}_{1d} - p_1e_1 \\ x_{3d} = \frac{1}{k_3}(-e_1 - k_1x_1 - k_2x_2 - d_2 + \dot{x}_{2d} - p_2e_2) \end{cases}$$
(20)

IV. SIMULATION RESULTS

In the simulation and experiments (using MATLAB version 2018a), the following parameters were used: $m_1 = 300$, $m_2 = 300$, $p_1 = 150$, $p_2 = 600$, $\varepsilon = $ 1e-6 $L = [3\varepsilon^{-1} \quad 3\varepsilon^{-2} \quad 1\varepsilon^{-3}]^T$. In the simulation, to examine the performance of the closed-loop system with the proposed method, the following case is considered with disturbances are parameter uncertainties, $d_2 = -0.3 \times f_2$ and $d_3 = -0.3 \times f_3$

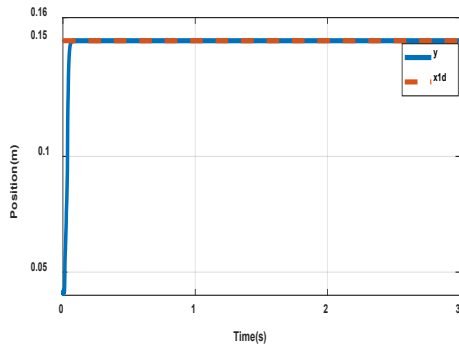### A. Constant reference tracking performance

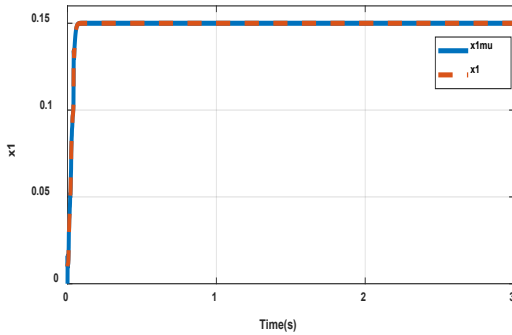Fig. 4.1. Backstepping integrated with sliding mode control
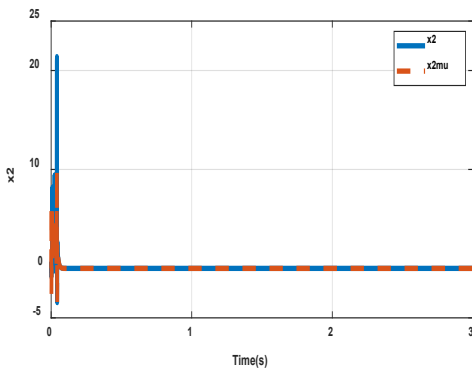

Fig. 4.2. Estimation performance of x_1

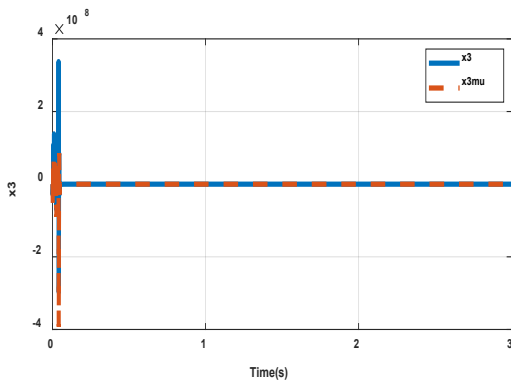Fig. 4.3. Estimation performance of x_2

Fig. 4.4. Estimation error of x_3

Fig. 4.1 exhibit the position tracking errors. As we see, the position of the piston approaches quickly and well-tracking to the desired output signal. The proposed method exhibits zero

position tracking error. Fig. 4.2, 4.3, and 4.4 show that the state variables were well estimated in a moment.

B. *Comparison to conventional backstepping control*
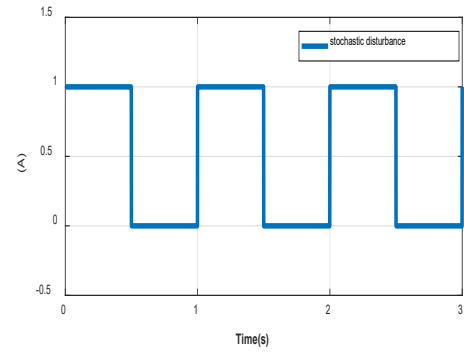
Fig. 4.5. Stochastic disturbance with pulse appearance
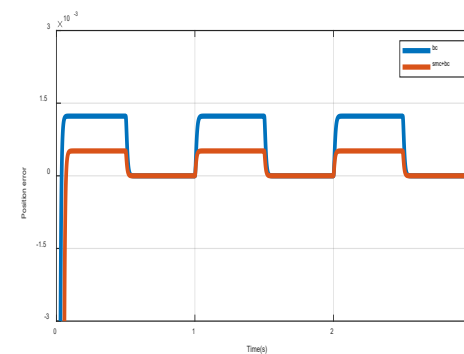
Fig. 4.6. Tracking error of two method with high gain state observer.
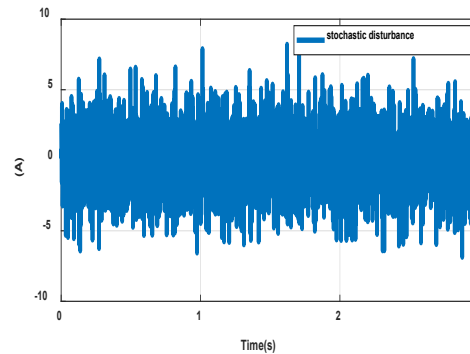
Fig 4.7. Stochastic disturbance in white noise appearance
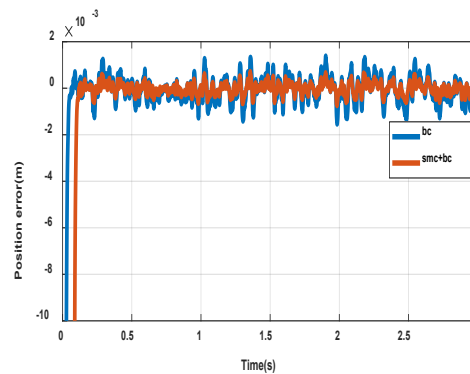
Fig. 4.8. Tracking error of two method with high gain state observer and constant input

In two cases, the proposed method always has a smaller position error characteristic than conventional backstepping controller. It shows that the proposed method helps stabilize effectively when the stochastic disturbances present in some part of EHSs.

V. CONCLUSION

The HGSOB-based backstepping aggregated sliding mode control for EHSs was proposed to guarantee the position tracking performance in the face of stochastic disturbances. The HGSOB was designed to estimate full-state variables with very high accuracy. The sliding mode control help EHSs stable with stochastic disturbances in several part of the system, while backstepping force the output signal rapidly approaching to desired signal. The simulation indicates that the proposed methods improve the position tracking performance. In the future, we are looking forward to conducting more research on parallel robots.

ACKNOWLEDGMENT

REFERENCES

[1] Zhu, W.-H., & Piedboeuf, J.-C. (2005). Adaptive Output Force Tracking Control of Hydraulic Cylinders with Applications to Robot Manipulators. Journal of Dynamic Systems, Measurement, and Control.

[2] Liu, T., Yu, Z., Xiong, L., & HAN, W. (2017). Anti-Lock Braking System Control Design on An Integrated-Electro-Hydraulic Braking System. SAE International Journal of Vehicle Dynamics, Stability, and NVH.

[3] Le Hanh, D., Ahn, K. K., Kha, N. B., & Jo, W. K. (2009). Trajectory control of electro-hydraulic excavator using fuzzy self-tuning algorithm with neural network. Journal of Mechanical Science and Technology, 23(1), 149–160.

[4] Kim, M. Y., & Lee, C.-O. (2006). An experimental study on the optimization of controller gains for an electro-hydraulic servo system using evolution strategies. Control Engineering Practice, 14(2), 137–147.

[5] Kim, W., Won, D., Shin, D., & Chung, C. C. (2012). Output feedback nonlinear control for electro-hydraulic systems. Mechatronics, 22(6), 766–777.

[6] Won, D., Kim, W., Shin, D., & Chung, C. C. (2015). High-Gain Disturbance Observer-Based Backstepping Control with Output Tracking Error Constraint for Electro-Hydraulic Systems. IEEE Transactions on Control Systems Technology, 23(2), 787–795.

[7] Guo, Q., Zhang, Y., Celler, B. G., & Su, S. W. (2016). Backstepping Control of Electro-Hydraulic System Based on Extended-State-Observer with Plant Dynamics Largely Unknown. IEEE Transactions on Industrial Electronics, 63(11), 6909–6920.

[8] Samakwong, T., & Assawinchaichote, W. (2016). PID Controller Design for Electro-hydraulic Servo Valve System with Genetic Algorithm. Procedia Computer Science, 86, 91–94.

[9] Guan, C., & Pan, S. (2008). Adaptive sliding mode control of electro-hydraulic system with nonlinear unknown parameters. Control Engineering Practice, 16(11), 1275–1284.

[10] Utkin, V., & Hoon Lee. (n.d.). Chattering Problem in Sliding Mode Control Systems. International Workshop on Variable Structure Systems, 2006. VSS'06.

LQR Controller Design for Mini Motion Package Electro-Hydraulic Actuator Control

NguyenVan Tan
School of Engineering – Technique Thu Dau Mot
University Binh Duong, Vietnam
tannv@tdmu.edu.vn or 0000-0001-5929-6000

Huy Q. Tran*
Faculty of Engineering – Technology Nguyen Tat
Thanh University Ho Chi Minh City, Vietnam
tqhuy@ntt.edu.vn

Phan Viet Hung
School of Engineering – Technique Thu Dau Mot
University Binh Duong, Vietnam
hungpv@tdmu.edu.vn

Huynh Minh Phu
School of Engineering – Technique Thu Dau Mot
University Binh Duong, Vietnam
phuhm@tdmu.edu.vn

Abstract—**Electro-hydraulic actuators have been widely applied in the industry because they have several major advantages. In this paper, we focused on controlling the mini motion package electro-hydraulics actuator. First, a mathematical model of the electro-hydraulic actuator (EHA) was implemented to apply the control process to the proposed system. Second, we applied the linear quadratic regulator (LQR) controller to a linear model that is converted from the nonlinear EHA system. Finally, the numerical simulation results were performed in which the results obtained from the LQR controller were compared with the PID controller to show the superiority of the proposed solution.**

Index Terms—**Linear control system; LQR controller; PID controller, Electro-hydraulics actuator.**

I. INTRODUCTION

IN THE past decades, different types of electro-hydraulic systems have been developed with different purposes for industrial systems, such as heavy-duty equipment and some machining equipment. These systems have many great advantages such as fast response, high control precision and high output power based on position, force and pressure control signals. It is considered one of the most promising choices for modern industry and has become an indispensable control method in mechanical manufacturing, construction machinery and defense. Moreover, the development of information technology has also greatly improved the quality of control process.

To improve the performance of the electro-hydraulic actuator (EHA) system, several EHA types were designed as shown in [1-3], and the use of conventional and advanced controllers to improve control accuracy, including proportional–integral–derivative (PID) controller [4-7], sliding mode controller and adaptive back-stepping ccontrol [8,9], has been studied. Recently, many researchers have been focused on new approaches, such as fault-tolerant control and fault-tolerant control compensation based on fault estimation using observer technology [4,10-12]. Research on linear quadratic regulator (LQR) controller were also implemented for the EHA system [13-15].

In this paper, an LQR controller is applied to control the linear EHA system. The gain K in the LQR controller is de-

signed based on the algebraic Riccati Equation with the quadratic performance functional and the cost function. The cost function corresponding to R and Q is equally important for the control and the state variables as the output (piston position, piston velocity and piston acceleration). The controller can be tuned by varying the positive elements in the Q, R matrix to achieve the desired response. A PID controller is used for the numerical simulation process of the EHA system. The response signal in the LQR case is compared with the response signal in the PID case.

The important contributions of paper can be shortened as follows:

- The mathematical model of the EHA system was converted to the linear system for the application of this controller

- The gain K was calculated using Matlab software via Q, R matrices, the algebraic Riccati Equation with the quadratic performance functional, and the cost function

- A scale value factor N is also computed thanks to the EHA's parameters.

- The successful numerical simulation results of the EAH system was respectively presented.

II. EHA MODEL FORMULATION

The dynamics of EHA can be expressed the motion of the positioning piston via the mass of object M_p [4]:

$$m_p \ddot{\ell}_p + \Upsilon \dot{\ell}_p + F_{sp} + F_{frc} + \upsilon = S_1 \Pi_1 - S_2 \Pi_2 \qquad (1)$$

where m_p, ℓ_p, $\dot{\ell}_p$ and $\ddot{\ell}_p$ are the mass, the position, the velocity, and the acceleration of object M_p respectively. S_1, Π_1, and S_2, Π_2 are the area, the pressure in two chambers of the cylinder respectively.

F_{sp}, Υ, F_{frc}, and υ are the external load force of the spring impact on the piston, the viscosity damping

* corresponding author

coefficient, the friction force, and unknown disturbance respectively.

The spring force F_{sp} can be computed as:

$$F_{sp} = \delta_{sp} \ell_p \tag{2}$$

where δ_{sp} is the stiffness of the spring.

The friction force F_{frc} can be presented as [4]

$$F_{frc} = \sqrt{2e}\left(F_{brk} - F_C\right) e^{-\left(\frac{v_p}{v_{st}}\right)^2} \frac{v_p}{v_{st}} + F_C \tanh\left(\frac{v_p}{v_{cl}}\right) \tag{3}$$

where F_{brk}, F_C, and B_v are breakaway friction, Coulomb friction, and viscous friction coefficient, respectively. v_p, and v_{st} are position velocity, Stribeck velocity threshold.

The hydraulic continuity equations for the EHA system can be expressed as [4]:

$$\dot{\Pi}_1 = \varphi_1 \left(\Phi_{1i} - S_1 \dot{\ell}_p\right) \tag{4}$$

$$\dot{\Pi}_2 = \varphi_2 \left(\Phi_{2i} + S_2 \dot{\ell}_p\right) \tag{5}$$

$$\Phi_{pump} = \Delta_p \omega \tag{6}$$

where

$$\Phi_{13i} = \Phi_{pump} + \Phi_{1v} - \Phi_{3v} - \Phi_i$$
$$\Phi_{24i} = -\Phi_{pump} + \Phi_{2v} - \Phi_{4v} + \Phi_i$$
$$\varphi_1 = \frac{\beta_e}{\left(V_{01} + S_1 \ell_p\right)}, \text{ and } \varphi_2 = \frac{\beta_e}{\left(V_{02} - S_2 \ell_p\right)}$$

β_e, Φ_i, V_{01}, and V_{02} are the effective bulk modulus in each chamber, the internal leakage flow rate of the cylinder, the initial total control volumes of the first and the second chamber respectively [4].

$\Phi_{1v}, \Phi_{2v}, \Phi_{3v}, \Phi_{4v}$, and Φ_{pump} are the flow rate through the pilot operated check valve on the left, on the right, the flow rate through the pressure relief valve on the left, on the right, and the pump flow rate respectively. Δ_p, and ω are the displacement and the speed of the servo pump [4]

From equation (1) to (6) and the derivative both sides of equation (1), we have:

$$m_p \dddot{\ell}_p + \Psi \ddot{\ell}_p + \delta_{sp} \dot{\ell}_p = S_1 \Pi_1 - S_2 \Pi_2 \tag{7}$$

where

$$\Psi = -\frac{F_C}{v_{cl}} \tanh\left(\frac{\dot{x}_p}{v_{cl}}\right)^2 + \frac{\sqrt{2e}\left(F_{brk} - F_C\right) e^{-\left(\frac{\dot{x}_p}{v_{st}}\right)^2}}{v_{st}} - \frac{2\sqrt{2e}\dot{x}_p^2 \left(F_{brk} - F_C\right) e^{-\left(\frac{\dot{x}_p}{v_{st}}\right)^2}}{v_{st}^3} + \frac{F_C}{v_{cl}} + \Upsilon$$

Substitution equation (4), (5) into (7), we obtain as:

$$\dddot{\ell}_p = -\frac{\Psi}{m_p} \ddot{\ell}_p - \frac{\delta_{sp}}{m_p} \dot{\ell}_p + \frac{\left(S_1 \varphi_1 + S_2 \varphi_2\right)}{m_p} \Delta_p u$$
$$- \frac{\left(S_1^2 \varphi_1 + S_2^2 \varphi_2\right)}{m_p} \dot{\ell}_p + \frac{\Omega}{m_p} \tag{8}$$
$$= f\left(\ell, u\right)$$

where

$$\Omega = S_1 \varphi_1 \Phi_{13i} - S_2 \varphi_2 \Phi_{24i}$$

Equation (8) can be written as

$$\dot{x} = f(x, u) \tag{9}$$

where
$x = \begin{bmatrix} x_1^T & x_2^T & x_3^T \end{bmatrix}$, and $x_1 = \ell_p; x_2 = \dot{\ell}_p; x_3 = \ddot{\ell}_p$
Equation (9), can be rewritten as

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} x_2 \\ x_3 \\ -\frac{\Psi}{m_p} \ddot{\ell}_p - \frac{\delta_{sp}}{m_p} \dot{\ell}_p + \Gamma_1 \Delta_p u - \Gamma_2 \dot{\ell}_p + \frac{\Omega}{m_p} \end{bmatrix} \tag{10}$$
$$= \begin{bmatrix} f_1(x, u) \\ f_2(x, u) \\ f_3(x, u) \end{bmatrix}$$

$$\Gamma_1 = \frac{\left(S_1 \varphi_1 + S_2 \varphi_2\right)}{m_p}; \Gamma_2 = \frac{\left(S_1^2 \varphi_1 + S_2^2 \varphi_2\right)}{m_p}$$

The appropriate equation written in a state-space format of EHA system based on a Taylor expansion is given as:

$$\dot{x}(t) = Ax(t) + Bu(t) \tag{11}$$

where A, B matrices are calculated as:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}; B = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}; \ u = \omega$$

where the parameters of the EHA system are provided in Table 1

Table 1. Basic parameters of the EHA system.

Parameters	Symbols	Values	Units
Piston cross-sectional area on the left	S_1	0.0013	m^2
Piston cross-sectional area on the right	S_2	$9.4 \ 10^{-4}$	m^2
The initial total control volumes on the left	V_{01}	$2.09 \ 10^{-4}$	m^3
The initial total control volumes on the right	V_{02}	$4.006 \ 10^{-5}$	m^3
Mass of the control object	m_p	10	kg
Bulk Modulus	β_e	$2.9 \ 10^8$	Pa
The stiffness of the spring	δ_{sp}	2383	Nm
The displacement of the servo pump	Δ_p	$3.5 \ 10^{-6}$	m^3

The elements of matrix B is computed as:

$$b_1 = \frac{\partial f_1(x,u)}{\partial u}; \quad b_2 = \frac{\partial f_2(x,u)}{\partial u};$$

$$b_3 = \frac{\partial f_3(x,u)}{\partial u} = \frac{S_1^2 \beta_e}{V_{01} + S_1 x_1} + \frac{S_2^2 \beta_e}{V_{02} - S_2 x_1}$$

The elements of matrix A is computed as:

$$a_{11} = \frac{\partial f_1(x,u)}{\partial x_1} = 0; \quad a_{12} = \frac{\partial f_1(x,u)}{\partial x_2} = 1; \quad a_{13} = \frac{\partial f_1(x,u)}{\partial x_3} = 0$$

$$a_{21} = \frac{\partial f_2(x,u)}{\partial x_1} = 0; \quad a_{12} = \frac{\partial f_2(x,u)}{\partial x_2} = 0; \quad a_{13} = \frac{\partial f_2(x,u)}{\partial x_3} = 1$$

$$a_{31} = \frac{\partial f_3(x,u)}{\partial x_1} \Bigg|_{\substack{x_1=0 \\ x_2=0 \\ u=0}}$$

$$= \left(\frac{S_1^2 \beta_e}{(V_{01} + S_1 x_1)^2} - \frac{S_2^2 \beta_e}{(V_{02} - S_2 x_1)^2} \right) x_2 \Bigg|_{\substack{x_1=0 \\ x_2=0 \\ u=0}}$$

$$- \left(\frac{S_1^3 \beta_e}{(V_{01} + S_1 x_1)^2} - \frac{S_2^3 \beta_e}{(V_{02} - S_2 x_1)^2} \right) * u \Bigg|_{\substack{x_1=0 \\ x_2=0 \\ u=0}}$$

$$a_{32} = \frac{\partial f_3(x,u)}{\partial x_2}$$

$$= \left(\frac{2F_c \tanh\left(\frac{x_2}{v_{cl}}\right)\left(\tanh\left(\frac{x_2}{v_{cl}}\right)^2 - 1\right)}{v_{cl}^2} \right) x_3 - 6\sqrt{2e} \ e^{-\left(\frac{x_2}{v_{st}}\right)^2} \left(\frac{F_{brk} - F_c}{v_{st}^3} \right) x_2$$

$$+ 4\sqrt{2ee} \ e^{-\left(\frac{x_2}{v_{st}}\right)^2} \left(\frac{F_{brk} - F_c}{v_{st}^5} \right) x_2^3 - \gamma_2 - \frac{S_1 \beta_e}{V_{01} + S_1 x_1} - \frac{S_2 \beta_e}{V_{02} - S_2 x_1} \Bigg|_{\substack{x_1=0 \\ x_2=0 \\ u=0}}$$

$$a_{33} = \frac{\partial f_3(x,u)}{\partial x_3}$$

$$= \sqrt{2e} \ e^{-\left(\frac{x_2}{v_{st}}\right)^2} \frac{F_{brk} - F_C}{v_{st}} - \frac{F_C\left(\tanh\left(\frac{x_2}{v_{cl}}\right)^2 - 1\right)}{v_{cl}} - \gamma_1 \Bigg|_{\substack{x_1=0 \\ x_2=0 \\ u=0}}$$

$$- (2\sqrt{2e} \ e^{-\left(\frac{x_2}{v_{st}}\right)^2} \left(\frac{F_{brk} - F_C}{v_{st}^3} \right) x_2^2 \Bigg|_{\substack{x_1=0 \\ x_2=0 \\ u=0}}$$

III. LQR CONTROLLER DESIGN

LQR is a powerful method for designing controllers based on the optimal algorithm that minimizes a given cost function. The parameter of the cost function is determined by two matrices Q and R which is calculated via the algebraic Riccati Equation with the quadratic performance functional and the cost function (12):

$$J(K) = \frac{1}{2} \int_0^\infty \left(x^T(t)Qx(t) + u^T(t)Ru(t) \right) dt \tag{12}$$

The optimal control signal is represented as

$$u^*(t) = -Kx(t) \tag{13}$$

Gain K is described as (14)

$$K = R^{-1} B^T P \tag{14}$$

where P is determined by the algebraic Riccati Equation (15):

$$PA + A^T P + Q - PBR^{-1}B^T P = 0 \tag{15}$$

The state-space feedback Equation (16) can be built as shown in Fig. 1.

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) \\ y(t) = Cx(t) + Du(t) \end{cases} \tag{16}$$

We choose matrices C, and D is shown as:
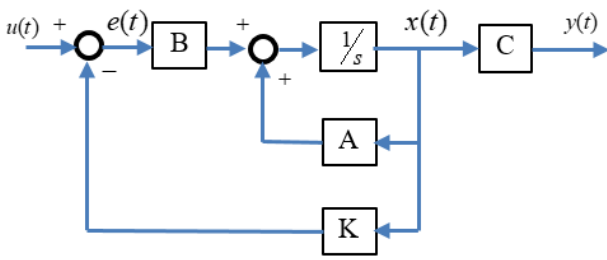
$$C = I_{3\times3}; D = 0 \qquad (17)$$

Fig. 1. Scheme diagram of the EHA system in state-space

Finally, matrices Q and R can choose as:

$$Q = C^T C = I_{3\times3} \; ; \; R = 1 \qquad (18)$$

IV. SIMULATION RESULTS

The obtained results from the simulation process of the EHA system with parameters (Table 1) are shown in Table 2 to verify the effectiveness of the suggested techniques. The numerical simulation processes are carried out and then the performance comparison between LQR and PID controllers with the input u(t) =0.2*ones(size(t)) was drawn.

First, simulation of the EHA system is done by the LQR controller with chosen matrices Q and R. Step response of piston position and velocity position are shown in Fig. 2.

$Q = I_{3\times3}$; and R = 1 and gain: K_1 =[1.000008   1.5746 10^{-3} 0.1]

Then, we continued obtain other simulations for cases 2, 3, and 4 in Table 2. These cases correspond to figures 3, 4, 5, respectively.
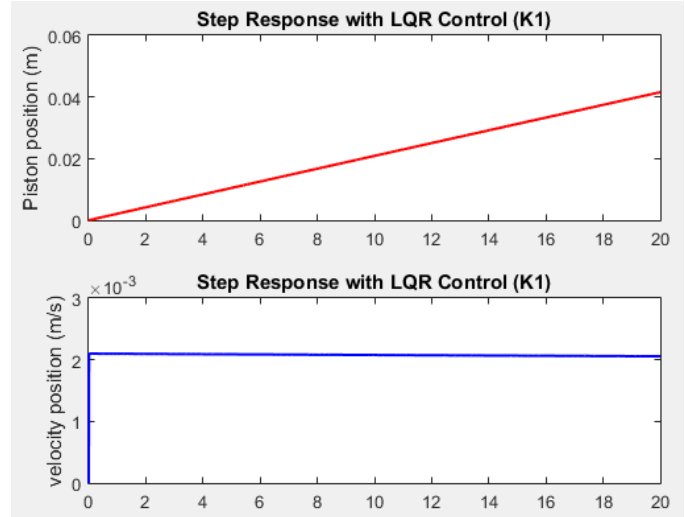
Fig. 2. Position and velocity response with LQR controller (K_1)

Table 2. Experimental results.

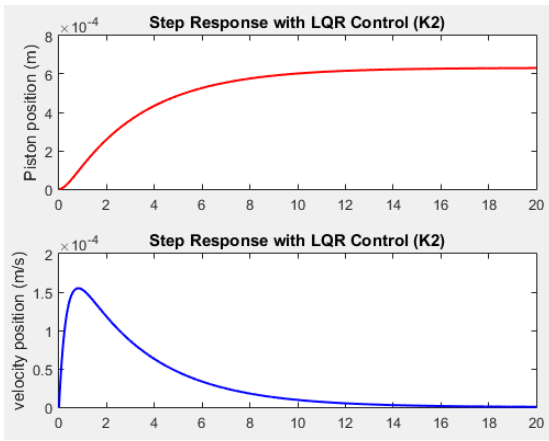| Cases	K, Q, R, N	Status	Setting time (s)
1	$Q = I_{3\times3}$; R=1, N=1 K_1 =1.0000008; 1.5746 10^{-3} ; 0.1	Unstable	None
2	$Q = 100000.I_{3\times3}$; R=1, N=1 K_2 =316.2278 146.2361 316.2275	Stable	18
3	$Q = \begin{bmatrix} 100000 & 0 & 0 \\ 0 & 0.001 & 0 \\ 0 & 0 & 0.005 \end{bmatrix}$ R = 0.005; N = 4.472 10^3 K_3 = 4.472 10^3 ; 4.683; 0.9997527	Stable	1.35
4	$Q = \begin{bmatrix} 100000 & 0 & 0 \\ 0 & 0.005 & 0 \\ 0 & 0 & 0.005 \end{bmatrix}$ R = 0.0001; N = 3.16 10^4 K_4 = 3.162 10^4 ; 211.316; 7.0708	Stable	0.25

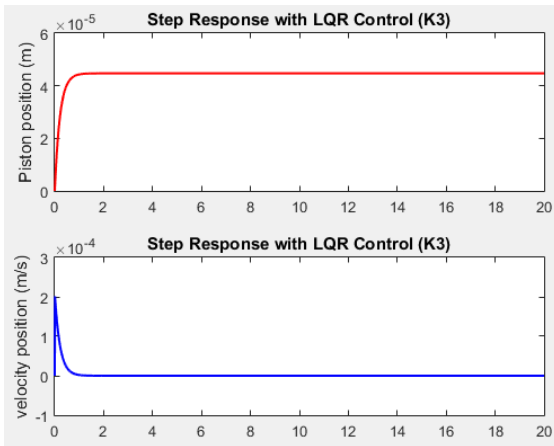Fig. 3. Position and velocity response with LQR controller (K_2)

Fig. 4. Position and velocity response with LQR controller (K_3)
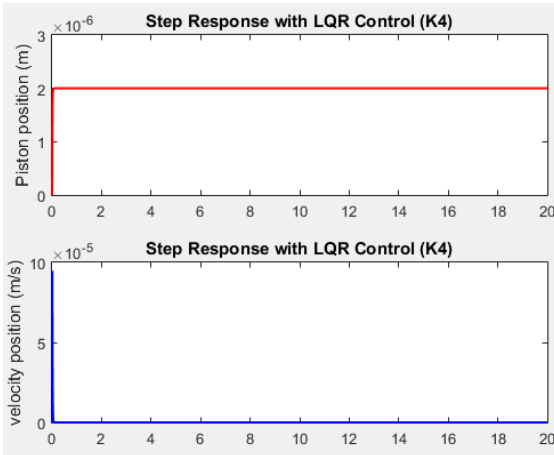

Fig. 5. Position and velocity response with LQR controller (K_4)

The scheme diagram of the EHA model in Simulink with N factor using the LQR controller can be described in Fig. 6. The simulation result in the third case is shown in Fig. 7 which is compared with position response for case without change (K_p.= 27.52272; K_I = 148.0119; K_D = 0.2) in the PID controller.
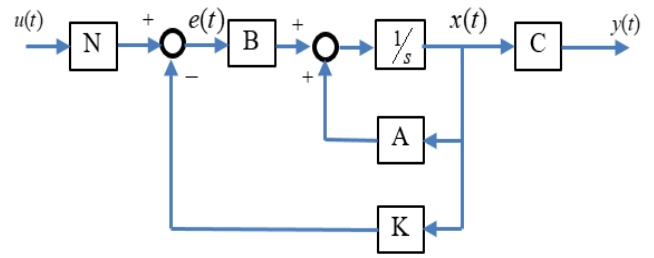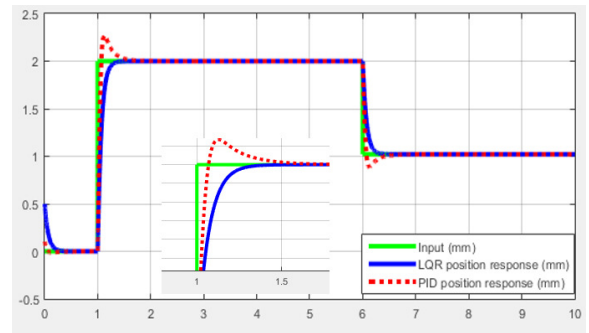

Fig. 6. Scheme diagram of the EHA system in state-space

The simulation results in the fourth case shown in Fig. 8 which is also compared with position response for case without change (K_p.= 27.52272; K_I = 148.0119; K_D = 0.2) in the PID controller.

Fig. 7. Position and velocity response with LQR controller (K_3)
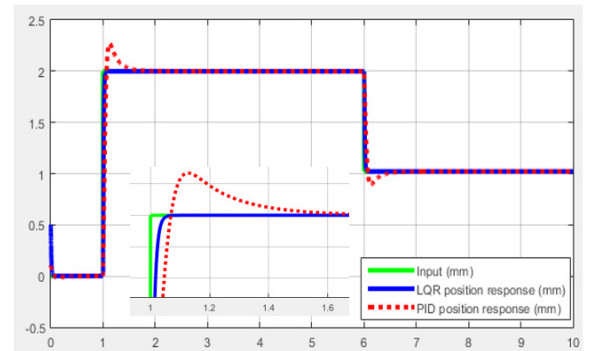
Fig. 8 Position and velocity response with LQR and PID controller (K_4)

In summary, the simulation results in the first case shown unstable position and velocity responses as shown in Fig. 2. In the second case (Fig. 3), the response is more stable when the Q value increases. The third and fourth cases showed the great effect of the Q and R parameters through the obtained response signal as illustrated in Fig. 4 and Fig. 5. We also see this influence in Fig.7 and Fig.8. These figures shown the superiority of the proposed controller thanks to achieved position responses.

V. CONCLUSIONS

The numerical simulation when applying the LQR controller show that the feedback signal of piston position in the EHA system achieved the expected results compared to the PID controller. The overshoot of the PID

controller is greatly reduced and the response tim e of the feedback signal is also improved as shown in the previous section. However, to achieve the desired control signal when using an LQR controller, it is important to be experienced in the selection of parameters Q, R and parameter N.

REFERENCES

[1] Gabriele . A, Andrea . V, "A design solution for efficient and compact electro-hydraulic actuators", ScienceDirect, Procedia Engineering 106 (**2015**) 8 – 16

[2] Juliang . X , Qinyu . L, and et al "Theoretical and Experimental Analysis of the Hydraulic Actuator Used in the Active Reflector System" *Hindawi Mathematical Problems in* Engineering. Volume *2018, Article ID 8503628, 13 pages, doi.org/10.1155/2018/8503628*

[3] Shaoyang . Q, David . F, and et al "A Closed Circuit Electro-Hydraulics Actuator with Energy Recuperation Capability", *Novel system solutions, 12th International Fluid Power Conference, Dresden (2020)*

[4] Tan, N. V.; and Cheolkeun, H.; "Experimental Study of Sensor Fault-Tolerant Control for an Electro-Hydraulic Actuator Based on a Robust Nonlinear Observer". *Energies,* MDPI, Open Access Journal, Energies (**2019**), 12, 4337; doi:10.3390/en12224337

[5] Tan, V. N.; and Cheolkuen, H.; "The Actuator and Sensor Fault Estimation Using Robust Observer Based Reconstruction for Mini Motion Package Electro-Hydraulic Actuator". *Intelligent Computing Methodologies, Proceeding of 15th International Conference, ICIC 2019.* Nanchang, China, August 3–6. Part III, Pp. 244-256. (**2019**).

[6] Norlela . I, Mazidah . T, and et al, "PID Studies on Position Tracking Control of an Electro-Hydraulic Actuator", *International Journal of Control Science and Engineering 2012, 2(5): 120-126, DOI: 10.5923/j.control.20120205.04*

[7] Skarpetis .M. G, and Koumboulis. F. N, "Robust PID controller for electro — Hydraulic actuators," *2013 IEEE 18th Conference on Emerging Technologies & Factory Automation (ETFA)*, Cagliari, Italy, (**2013**), pp. 1-5, doi: 10.1109/ETFA.2013.6648165.

[8] Ahn, K.K.; Nam, C.N.D.; Jin, M. Adaptive Back-stepping Control of an Electrohydraulic Actuator. *IEEE/ASME Trans. Mechatron. (***2014***)*, 19, 987–995.

[9] Thomas. A. T, Parameshwaran. R, and et al "Improved Position Tracking Performance of Electro-Hydraulic Actuator Using PID and Sliding Mode Controller", *IETE Journal of Research*, (2019), DOI: 10.1080/03772063.2019.1664341

[10] Ahian, S.A., Truong, D.Q., Chowdhury, P. et al. Modeling and fault-tolerant control of an electro-hydraulic actuator. Int. J. Precis. Eng. Manuf. 17, 1285–1297 (**2016**). https://doi.org/10.1007/s12541-016-0153-2

[11] Mahulkar .V, Adams D. E, and Derriso .M, "Adaptive fault-tolerant control for hydraulic actuators," *2015 American Control Conference (ACC)*, Chicago, IL, USA, (**2015**), pp. 2242-2247, doi: 10.1109/ACC.2015.7171066.

[12] Phan .V. D, Vo .C. P., Dao .H. V. and Ahn .K. K., "Robust Fault-Tolerant Control of an Electro-Hydraulic Actuator With a Novel Nonlinear Unknown Input Observer," *in IEEE Access*, vol. 9, pp. 30750-30760, 2021, doi: 10.1109/ACCESS.2021.3059947.

[13] Pourebrahim .M, Ghafari ,A. S, and Pourebrahim .M, "Designing a LQR controller for an electro-hydraulic-actuated-clutch model," *2016 2nd International Conference on Control Science and Systems Engineering (ICCSSE)*, Singapore, (**2016**), pp. 82-87, doi: 10.1109/CCSSE.2016.7784358.

[14] Tan .V .V, Olivier .S, Luc .D, Enhancing roll stability of heavy vehicle by LQR active anti-roll bar control using electronic servo-valve hydraulic actuators. Vehicle System Dynamics, Taylor & Francis, (**2017**), 55 (9), pp.1405-1429. Doi: 10.1080/00423114.2017.1317822.

[15] Priyanka .E, Maheswari .C, Thangavel .S. "Remote monitoring and control of LQR-PI controller parameters for an oil pipeline transport system". *Proceedings of the Institution of Mechanical eers, Part I: Journal of Systems and Control Engineering*. (**2019**); 233(6):597-608. doi:10.1177/0959651818803183

Artificial Intelligence applications in anomaly identification detection of big database

*

Phan Huy Thang
National Credit Information Center (CIC)
Hanoi, Vietnam
thangph@creditinfo.org.vn

Nguyen Thi Ngoc Anh
School of Applied Mathematics and Informatics
Hanoi University of Science and Technology
CMC institute of science and technology
Hanoi, Vietnam
anh.nguyenthingoc@hust.edu.vn,
ORCID 0000-0002-6555-9740

Abstract—**Data matching is the process of finding, matching, and combining records from many databases or even within one database that belong to the same entities. All parts of the data matching process have been improved during the previous decade as a result of research in various disciplines such as applied statistics, data mining, machine learning, database administration, and digital libraries.Indeed, with the significant advance in artificial intelligence over the past decade, all aspects of the data identification process, especially on how to improve the accuracy of data matching. Firstly, this paper presents the process of comparing data, detailing the steps to perform pre-processing data, comparing the data fields of each record, classification, and quality assessment. Secondly, the paper introduces a method to expand the problem of identifying duplicate objects with big data. Third, the paper also provides specific aspects of unstructured data matching times. Moreover, the methodology of solving big data matching problems by machine learning is proposed. Finally, the proposed method is applied to the problem of database cleanup and identification of identifier abnormalities at the national credit centre CIC with correct results from 96% to 98%. The achieved results are not only theoretical but also practical in business operations at CIC.**

Index Terms—**big data, abnormality detection, duplicate profiles, similarity, artificial intelligence**

I. Introduction

HUGE volumes of data are generated every day as a result of the ever-increasing sharing of information on the Internet. Big data really starts when we understand the value of the information hidden in the data. When we have enough resources, Machine Learning is the key to that technology. The relation between machine learning and big data has been developing thanks to the proliferation of data and vice versa. The value of big data depends on the data's ability to exploit knowledge. In the past 20 years, Bigdata and AI has strongly developed. The development of AI has come from three different factors: (1) advancement in the study of algorithms,

(2) an increase in computing ability, and (3) an explosion of data.

Machine learning is capable of self-learning based on input data without programming specifically. The aim of machine learning is to make the system intelligent. Machine learning helps humans in processing huge amounts of information (Big data) that we face every day. Machine learning have three common types: supervised, unsupervised and semi-supervised.

However, data labelling for big data is expensive and takes time consuming, most huge datasets are unlabelled and contain a variety of attributes from various types of data [10], and as a result, unsupervised learning on big data has recently resurfaced. Furthermore, the scale of data has a substantial impact on the complexity of unsupervised machine learning algorithms, making real-time unsupervised learning on huge data a serious challenge. Clustering problem is one of the most often used unsupervised learning strategies in data mining. It enables analysts to automatically generate groups in datasets based on the similarity records, which can be of many types such as texts, numbers, category, or a combination of these types [10]. Cluster analysis is widely employed in a variety of domains, including economics, science and technology.

Currently, the research topic of detecting abnormality based on AI in processing big data has not been officially published in scientific papers in Vietnam. The practical application of AI has not been studied and applied to abnormality detection.

Until this time, other publications are mainly about mathematics tools, assessing the potential of big data analytics with businesses, applications in the field of genetic classification, assessing customer behavior in the purchasing process, Image Processing, Language, and Speech Processing. Many researchers have paid attention to the solution of detecting duplicate records (based on similarity) in big database and merging them into a unique identifier [1] [3] [4] [6]. Detecting duplicate records is meaningful in:

* Searching and removing duplicate records in the entire

database;
* Merging the data from other sources with an internal database;
* Improving data quality and consistency;
* Creating more adequate, in-depth, and comprehensive data tables;
* Data cleaning

In recent years, machine learning algorithms for record duplicate detection have been studied widely. Some machine learning methods such as K-means, Adaptive neuro-fuzzy inference system (ANFIS) [2], Support Vector Machine, Naïve Bayes, Decision Tree [5], DSC++ algorithm (Arfa Skandar et al. [8]), Semantic - Syntactic Method (Djulaga Hadzic and Nermin Sarajlic [9]). This paper aims to detect duplicate records in the entire database. When the new code calculates the new record duplicate similarity compared with the existing records in the database, duplicated records are matched to a unique identifier record.

The paper's layout consists of 4 sections: Section 1 -An overview of cleaning up the database by finding similar identifiers. Section 2 shows the methodology of big data processing application to solve the problem with large computational complexity; Section 3 - Practical applications at CIC; Section 4 - conclusions and discussion.

II. METHODOLOGY

A. Processing the data for the similarity records in the general model

The general model of data processing is done in 4 steps, the illustration below:

Step 1: Data collection: The data were collected from CIC's databases and will be stored for further processing steps. Data includes many different types: numerical data, categorical data, or text data. The data on detected duplicate records will be used to evaluate the next steps.
Step 2: Data processing: This is an important step, as a premise for the next processing steps. In this step, there will be different processing methods that depend on the type of data. Then, the importance of collecting data is evaluated. After that, important fields of data are extracted for developing a model.
Step 3: Data exploration: The data will be divided into two parts of the training dataset and test dataset in the previous step, they will be used to develop the model. Data will be assessed according to two methods:

+ Similarity detection has based on the rules that CIC developed, using similar measurement methods.
+ Similarity detection has based on machine learning models. The model is continuously trained based on historical data (instances in which duplicate records have been detected) and user reviews to make the model more and more optimal.

Step 4: Evaluation: New profiles need to be assessed through an established model to give similarity evaluation scores. The higher the score of these records is, the greater the similarity is, and vice versa.

B. Locality sensitive hashing

Locality sensitive hashing (LSH) is a probabilistic algorithm that hashes similar input items into the same buckets. Data-independent hashing approaches, such as locality-sensitive hashing, are commonly used in hashing-based approximate closest neighbour search algorithms. This algorithm solves the problem is that given a large number N in the millions or even billions of text documents, find pairs that are near duplicates document of text. The main LSH $Document_i, Document_j$ in dataset are considered:

A min hash function $hast$ such that:

- if $Similarity(Document_i, Document_j)$ is high, then with high probability $hash(Document_i) = hash(Document_j)$
- if $Similarity(Document_i, Document_j)$ is low, then with high probability $hash(Document_i) \neq hash(Document_j)$.

Hashing documents into buckets, pairs of near duplicate documents hash into the same bucket. This is the key method to solve the big dataset that see Figure 2.

C. Split records

In the problem of data reduction and matching big databases, the complexity of the algorithm is the factor considered at first. This helps ensure that the solution processing times are within an acceptable threshold. With similarity matching algorithms for data reduction, when performing a comparison on a database consisting of n records, the complexity is $O(n^2)$. However, it is more optimal to perform a similarity comparison throughout this database when comparing only a record with records that are likely to be similar to this record. A split record is used for this purpose.

Splitting records is clustering the database into small clusters (called bucket). With the characteristics of the problem of removing duplicate records and having the feature of finding similar records, these buckets must fulfill similarity requirements as follows:

$$\begin{cases} sim(r_i, r_j) \geq t & \forall r_i, r_j \in bucket_k \\ sim(r_i, r_j) \leq t & \forall r_i \in bucket_m, r_j \in bucket_n, m \neq n \end{cases}$$
$$(1)$$

Where:

$sim(r_i, r_j)$: the degree of similarity between 2 records r_i and r_j.
t: is a given similarity threshold.
$bucket_k$: the cluster of records k.

From the above requirements, the records after being classified into small clusters must fulfill: records in the same cluster must have similarity level above a given threshold, records in other clusters must have similarity less than the threshold given. There are many algorithms used for database clustering purposes, which have based on the used type of data. The
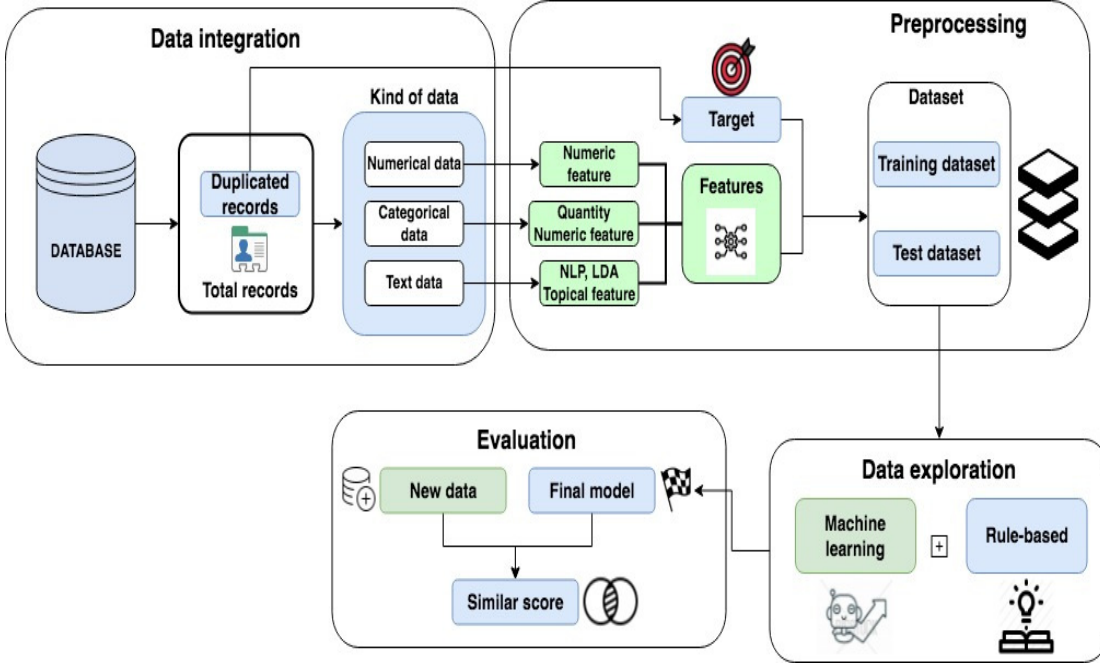
Fig. 1. General model of data processing for similar records.

algorithms are into two groups as follows: Group of Algorithms used for numerical data: K-mean Clustering, DBScan, PAM,... Group of Algorithms used for string data: Multibit Tree, Succinct Multibit Tree. In this problem, identity data including string data (such as Name, ID Number) and numeric data (such as Date of Birth, Address), we have combined both two groups of algorithms for the purpose of splitting records.

D. Similarity

The similarity is a numerical measure of how different two or more data objects are. This measure is used to reflect the intensity of the relationship between data objects. The problem of similarity calculation is illustrated as follows: Two attribute values are d_i and d_j. The aim is to find a value $S(d_i, d_j), S \in [0; 1]$, that shows the similarity between d_i and d_j. The higher the value is, the greater the similarity between d_i and d_j. The similarity value is used in some functions: Jaccard index, Jaro distance, Haversine formula for points.

1) Jaccard index: The Jaccard index measures similarity between sample sets based on statistical methods. Accordingly, the similarity value between two A and B strings can be calculated as:

$$sim(A, B) = \frac{A \cap B}{A \cup B}, \qquad 0 \le sim(A, B) \le 1 \qquad (2)$$

2) Jaro distance: The Jaro distance defines the measure of similarity between two strings. Given two s_1 and s_2 string, the Jaro distance d between two strings is defined as:

$$d = \frac{1}{3} \left(\frac{m}{|s_1|} + \frac{m}{|s_2|} + \frac{m-t}{m} \right) \qquad (3)$$

Where: m: the number of matching characters between two strings; t: $\frac{1}{2}$ of the number of transposition steps (transposition). This function counts the number of characters that are common in two strings and the distance is not greater than the following value: $\frac{\max(|s_1|,|s_2|)}{2} - 1$. Transpositions are defined as the number of common characters in two strings (but the order in the sequence is different) divided by 2.

3) Haversine distance: Assuming P_1 and P_2 points are converted into the corresponding latitude and longitude $long_1$, lat_1 and $long_2$, lat_2. The Haversine distance d is calculated as:

$$a = \sin^2 \left(\frac{\Delta lat}{2} \right) + \cos(lat_1).\cos(lat_2).\sin^2 \left(\frac{\Delta long}{2} \right) \qquad (4)$$

$$c = 2.\arctan 2(\sqrt{a}, \sqrt{1-a}); d = R.c \qquad (5)$$

Where: $\Delta lat = lat_2 - lat_1$; $\Delta long = long_2 - long_1$; R: the radius of the Earth (R=6731km).

Finally, in order to calculate the similarity between two points, it is possible to use the Haversine distance to a value in the segment $[0, 1]$ by using functions such as sigmoid, exponential function,...

E. Support Vector Machine method

Support Vector Machine (SVM) is a statistical and computer science concept for a set of associated supervised learning methods for classification and regression analysis. The standard SVM multiply inputs and classifies them into two different classes. Therefore, SVM is a binary classification algorithm. SVM was developed by Vapnik, and his colleagues

in the 1970s in Russia then became popular and famous in the 1990s. SVM has many practical applications such as genetic analysis, marketing, or facial recognition [7].

Support Vector Machine (SVM) is a Supervised Learning algorithm that has been used to divide data (Classification) into separate groups. A Support Vector Machine builds a hyperplane or a set of hyperplanes in a multidimensional or infinite-dimensional space, which can be used for classification, regression, or other tasks. For the best classification, the hyperplanes are intuitively located as far away from the data points of all layers (called margins) as possible. In general, the larger the margin is, the smaller the generalization error of the classification algorithm is.

With the binary classification problem, the SVM constructs a separating hyperplane located as far away from the data points of all layers (called margin) as possible. Assuming that there is training dataset D with n points, each point x_i belongs to one of two class labels: $y_i : D = \{(x_i, y_i) \forall x_i \in \mathcal{R}, y_i \in \{-1, 1\}\}$. Each hyperplane is described by the equation and to maximize the margin with the following requirements:

- $wx_i - b \leq 1$ if the data point belongs to the positive class.
- $wx_i - b \geq 1$ for other classes.
- Or $y_i(wx_i - b) \geq 1 \quad \forall 1 \leq i \leq n$

The optimal problem for SVM can follow quadratic programming formulation:

$$\max \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{i=1}^{n} \alpha_i \alpha_j y_i y_j x_i^t x_j \qquad (6)$$

Subject to:

$$\begin{cases} \lambda_i \geq 0 \\ \sum_{i=1}^{n} \alpha_i y_i = 0 \end{cases}$$

In the case of data points that are non-linear, a soft margin SVM or a non-linear function can be used to map the input data space to a larger number space that this space can split with a hyperplane. This technique is called a kernel trick.

F. Proposed model

The proposed model is shown below:

The original raw data are available in database A and Database B. In the case of the data reduction problem, database A and database B represent a set of data. In the matching problem, database A and database B are two different data sets. Implementation steps include:

- *Step 1*: Data pre-processing: In this step, the data fields in both databases A and database B are normalized to a common format, such as the names written in the right order of Surname – Middle name - Name, or the date of birth in the format of DD-MM-YYYY,...
- *Step 2*: Indexing: The normalized data will be put into the indexing system, and then the system calculates the records based on the data of the record. The indexing

TABLE I
Table 1: Description of data

Column name	Description	Data types
TENKH	Full name	Varchar2(150)
SOCMT	Identity card number	Varchar2(30)
ADDRESS	Address	Varchar2(250)
GIOITINH	Gender	Varchar(1)
DTHOAI	Phone number	Varchar(50)
MSTHUE	Tax code	Varchar(30)

algorithm is selected to assure that adjacent records also have similar data content.

- *Step 3*: Creating suspected pairs of similarity: Based on indexing the records which are generated from step 2, the system collects records into the bucket (containing adjacent indexing records) and creates suspected pairs of similarity according to the method of exhaustion.
- *Step 4*: Comparing similarities: To compare the similarity on the data content of the record pairs which have similar doubt. The rules for evaluating compared information can be used. For example, the rule of matching the records have entire common names, dates of birth, ID number with a 1-character difference. Besides, the machine learning model is also applied for calculating similarity. After comparing the records, they are grouped into three categories: records with a high similarity, records with disparity, and records with suspected similarity (corresponding to the achieved scores). Similar records should be checked by CIC's specialized division.
- *Step 5*: Evaluating: After pairs of similar suspected records are directly checked about similarity ability, the system will record labels of pairs of records whether they are in common or not. Since it provides a review of the effectiveness of the solution.
- *Step 6*: Training the machine learning model for similarity scoring: Based on the user rating on pairs of duplicated suspected records, the system uses these ratings as labels as input to the model training process. This training-evaluation process will be run in a loop. It helps this model improve accuracy over time.

III. APPLICATION AND RESULTS

Data description: The data is extracted from CIC's customer database, that will be used to perform the duplicate record detection problem. The information of customers used for the problem includes Full name, ID number, Address, Gender, Phone number, Tax code. The data are described in Table 1.

Result In this paper, with the experimental scale, we have used a data volume of 10 million records with addresses in the North of Vietnam. In addition to the current data of the credit data system, we also include a data set of records that have been assessed by the specialized divisions at CIC as duplicates (5,000 records).

This data set will be included in the duplicate detection tool given by the authors in the proposed model and evaluated
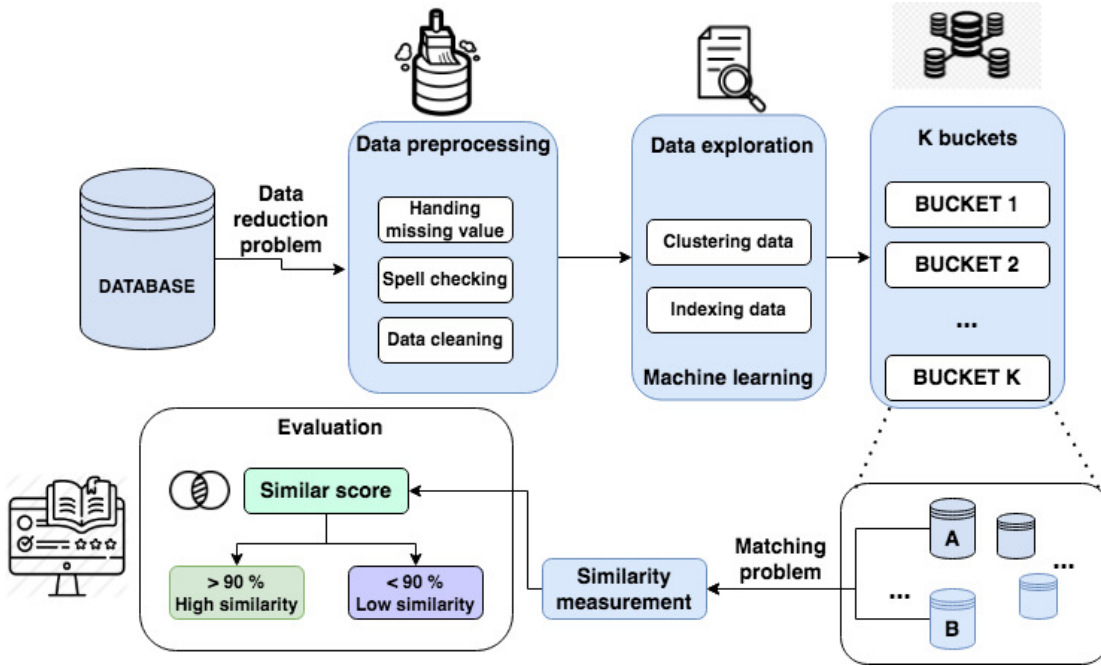
Fig. 2. General model of data processing for similar records.

TABLE II
NUMBER OF SUSPECTED DUPLICATE RECORDS DETECTED

Similarity threshold	Number of suspicious records
0.98	7,800)
0.95	9,400
0.90	15,200

TABLE III
COVERAGE RATES WITH DUPLICATE DATA

Similarity threshold	Coverage rate
0.98	73%
0.95	81%
0.90	97%

TABLE IV
EVALUATION METRIC OF MACHINE LEARNING MODEL FOR PROPOSED

Metric	Result
Accuracy	0.97
Precision	0.96
Recall	0.98
F1 score	0.97

through indicators including the number of suspicious records detected and the coverage rate with duplicate labeled records.

After two hours blocking process, the solution return a data set of 10 million records to 800,000 blocks (approximately 12 times), each block contains about 12 records on average, thereby reducing the number of comparisons needed from $10,000,000 \times (10,000,000 - 1)/2\ 50 \times 10^{12}$ to $800,000 \times (12 \times 11)/2\ 52 \times 10^{6}$ nearly 1 million times.

The similarity is calculated on the records in the same data block in 3 hours to evaluate the similarity of the records and the efficiency through the number of detected similar pairs of records when the similarity exceeds the permitted threshold, the results are as Table 3. The records have been labeled similarly by specialized division at CIC, the authors also evaluated the coverage of the solution with this data group, and the specific results are shown in the Table 3. The quality of the machine learning model for duplicate record detection was also assessed after training the model with 80% of the labeled records (4000 pairs of records with duplicate labels) on a test dataset of 1000 pair of records, for the specific result see Table 4.

IV. CONCLUSION AND DISCUSSION

Duplicate Record Identification is a solution to detect duplicate records (based on similarity) in a big database system and combine them into a unique identifier.

Detecting and removing duplicates helps search and remove duplicate records in one or more databases. Identifying duplicate records contributes to data cleaning, and cleaning usually takes up to 80% of the data-maker's time to use. The solution we propose will help reduce at least 40% of the time for this step.

Duplicate detection and elimination contribute to data preparation for in-depth intelligence analysis. The quality of in-depth intelligence analysis reports depends entirely on the reliability of the data source. Eliminating duplicate data is a prerequisite in building highly reliable data sources.

In particular, the identification of duplicate records contributes to automatic fraud prevention. This work allows companies to trace the steps, enabling investigations to arrive at the source of the problem.

The paper points out the similarity assessment by calculating similarity scores to assess the degree of duplication between records.

The proposed methods were applied by the authors to the problem of the National Credit Information Center CIC and gave correct results from 96% to 98%.

REFERENCES

[1] Ahmed K. Elmagarmid, Panagiotis G. Ipeirotis, Vassilios S. Verykios, Duplicate record detection: A Survey, In: IEEE Transactions on knowledge and data engineering 2007, Vol.19.

[2] G. Ranganathan, V.Bindhu,. Jenifer Raj, Duplicate record detection using intelligent approaches, In: International Journal of Pure and Applied Mathematics 2018, Vol.119, No.12, pp.13077–13087.

[3] Peter Christen, Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection; Springer (2012).

[4] Batini, C., Scannapieco, and M.: Data quality: Concepts, methodologies and techniques. Data-Centric Systems and Applications. Springer (2006)

[5] Arasu, A., Götz, M., Kaushik el at: On active learning of record matching packages.In: ACM SIGMOD, pp.783—794. Indianapolis (2010).

[6] Alvarez, R., Jonas, J., Winkler, W., Wright, R .: Interstate voter registration database matching: the Oregon-Washington 2008 pilot project. In: Workshop on Trustworthy Elections, pp.17—17. USENIX Association (2009).

[7] Roya Hassanian-esfahani, Mohammad-javad Kargar , Sectional MinHash for near-duplicate detection, In: Expert Systems with Applications, Volume 99, 1 June 2018, pp.203–212.

[8] Arfa Skandar, Mariam Rehman,Maria Anjum, An Efficient Duplication Record Detection Algorithm for Data Cleansing, In: International Journal of Computer Applications, Volume 127, October 2015, pp.28-37.

[9] Djulaga Hadzic and Nermin Sarajlic, Methodology for fuzzy duplicate record identification based on the semantic-syntactic information of similarity, In Journal of King Saud University - Computer and Information Sciences, Volume 32, 2020, pp.126-136.

[10] Toan Nguyen Mau and Van-Nam Huynh, An LSH-based k-representatives clustering method for large categorical data, Neurocomputing, volume 463, pages 29-44, year 2021,

Author Index