Rough Sets Turn 40:
From Information Systems to Intelligent Systems

Andrzej Skowron
Systems Research Institute
Polish Academy of Sciences, Warsaw, Poland
Cardinal Stefan Wyszyński University, Warsaw, Poland
Email: andrzejj.skowron@gmail.com

Dominik Ślęzak
Institute of Informatics
University of Warsaw, Warsaw, Poland
QED Software Sp. z o.o., Warsaw, Poland
Email: slezak@mimuw.edu.pl

Abstract—The theory of rough sets was founded by Zdzisław Pawlak as a framework for data and knowledge exploration. His seminal paper titled "Rough Sets" was published in 1982, in International Journal of Computer and Information Sciences. One of the key aspects that lets us use rough sets in practical scenarios is the notion of information system, which comes from even earlier Professor Pawlak's works. Information systems are the means for data and knowledge representation. They constitute the input to rough set mechanisms aimed at computing approximations of concepts and deriving compacted, interpretable decision models. In particular, the fundamental notion of the indiscernibility relation is defined on the basis of a given information system. Accordingly, we discuss to what extent information systems can serve as the basis for intelligent systems. We claim that in many cases it is not enough to treat a data set – represented as an information system – as a purely abstract object with no linkage to the data origins. Oppositely, we should give ourselves a technical possibility to construct information systems dynamically, taking into account interaction with physical environments where the data comes from. With this respect, we refer to the notions of interactive granular computing and we generally consider together the paradigms of rough sets, information systems, and information granulation.

Index Terms—Rough Sets; Information Systems; Data Mining; Big Data; Interactive Granular Computing; Intelligent Systems

I. INTRODUCTION AND BASIC CONCEPTS

ZDZISŁAW Pawlak (1926-2006) founded the theory of rough sets with the aim of analyzing incomplete information by means of approximations [1]. His book published in 1991 [2] established worldwide recognition of rough sets as an approach which can model complex problems using simple constructs. The idea of rough sets originated from earlier Professor Pawlak's research on knowledge representation and information retrieval [3], [4]. The key observation was that we often operate with objects (cases, instances) which are indistinguishable from each other, so approximating the sets of objects by using indistinguishable "blocks" is the most reasonable thing we can do. A need for reaching such simple solutions was important for Professor Pawlak during his entire scientific career which included far more achievements than just rough sets [5]. It is also interesting to note that the basic models he used working in different fields, such as conflict analysis [6] or concurrency [7], were based on information systems. In the same time, his personal interests were going far beyond computer science and science in general [8].

The updated (after 25 years) viewpoint on rough sets was presented in [9], [10], [11]. As it happens with every theory, it took time to understand the actual contribution of rough sets with respect to other approaches. Firstly, rough sets were compared with the theory of fuzzy sets [12]. Over the years, it turned out that these two theories can be successfully combined because they offer complementary granules understood as computational building blocks for approximations [13], [14]. In the next sections, one can see a number of examples of such combinations. Another thread of comparisons was devoted to the principles of information granulation and granular computing (GrC) [15], [16], whereby it is generally assumed to operate on groups of objects / instances / items gathered together into granules. In this case, the relationship turned out to be even more natural because such granules (or generally, various frameworks and layers of granular information) are a natural input for calculations related to rough sets.

Nowadays, rough sets are particularly popular in the area of learning decision models from the data. There are many rough set methods aiming at feature selection and interpretable decision model construction [17], [18], [19]. However, the principles of rough sets have much broader influence. They have an impact on various methodologies of decision making, *e.g.* multi-criteria decision making [20] or three-way decision making which strongly relies on rough set positive, negative and boundary regions [21]. Moreover, rough sets are employed to enhance expressive and algorithmic capabilities of many approaches to data mining and knowledge discovery. A good example here is an extension of standard data clustering toward rough clustering (whereby we search and operate with lower and upper approximations of rough data clusters) and its fuzzy hybridizations [22], [23], [24]. Another example corresponds to formal concept analysis [25] which has this year its 40th anniversary exactly like rough sets. With that respect, there are many approaches to constituting rough set approximations of formal concepts or in other words, formal concepts which comprise of rough set approximations [26]. Rough sets turned out to be useful also in other fields of science and industry. For instance, they were adopted in design of database engines and solutions, with respect to query language extensions (which can be referred as rough querying) and acceleration of commercial database software performance [27], [28].

In all above scenarios, rough set methods (or mechanisms that adopt rough set principles) require an input to derive approximations of concepts, relations, etc. In [1], Professor Pawlak discussed several examples of domains, whereby the inputs to rough set methods could take different forms. One of those forms corresponded to the notion of an information system [3]. In the framework proposed by Professor Pawlak, information systems are aimed at representing the underlying data, information or knowledge that we want to use to describe (and approximate) the concepts of practical interest. An information system comprises of objects, attributes and the values of attributes over those objects. Information systems resemble data tables in the theory of relational databases. It should be noted that data tables, in particular decision tables, have been studied and used in many applications since 60-ties of the XX-century [29], [30]. Still, we need to remember that Professor Pawlak's goal to operate with information systems was to represent information (including information derived from the data) rather than focus on the data itself [2]. This topic is actually wider and one can find analogies between information systems and data / information / knowledge representation frameworks in some other theories [25], [31], [32].

The complete idea of information systems has been created thanks to cooperation of Professor Pawlak with other scientists [4], [33]. At that time, there was a great demand for constituting the foundations for information representation and retrieval [34]. All those works were reaching beyond a standard understanding of data storage and processing, particularly with respect to incompleteness, imprecision and indeterminism of information that one needs to handle in practice. Over the years, the concept of non-deterministic information systems evolved in many interesting directions [35], [36]. Going back to the principles of information granulation [16], one may say that such information systems comprise of descriptions – also called signatures – of granules (composed in different ways, specific to different applications) in terms of available attributes, where those signatures are not always precise.

Operating with granules introduces a useful abstraction, a kind of border between granules' signatures (being the inputs to further computations) and granules' internals, whereby one can locate guidelines about *how* and / or *what for* those signatures are derived from physical data sources. This is why in this paper, we are interested equally in: (i) rough sets as the methodology for deriving concept approximations and compacted decision models, (ii) information systems which provide the input to those derivations, and (iii) the paradigms of GrC and information granulation, as various forms of granules can be "hidden right under" the abstract descriptions that information systems consist of. We focus particularly on the methodology of interactive granular computing (IGrC) [37] and interactive information systems [38], where the above-mentioned abstraction was explicitly introduced in terms of complex granules (c-granules) with the embedded control mechanisms that decide how their signatures are derived.

IGrC raised from the observation that traditional ways of computing do not take into account how the process of

perceiving attribute values is realized, *where* and *when* to access the concerned objects in a physical space, and *why* particular attributes are selected. This kind of awareness / attention / agency [39] is important for designing intelligent systems which are supposed to deal with complex phenomena in the real physical world [40]. This becomes even more important when one realizes that unexperienced data scientists often treat the available data sets as an ultimate baseline without investigating how those sets were collected.

In the above considerations, we evolved from information systems toward intelligent systems. We referred to the concepts of awareness and perception (perceiving attribute values), we can also refer to the concept of cognition. With this respect, let us cite the following statement of Leslie Valiant[1], which is particularly relevant when extending GrC toward IGrC:

> *A fundamental question for artificial intelligence is to characterize the computational building blocks that are necessary for cognition.*

The above computational building blocks can be treated as a generalization of granules known from GrC [15]. (In particular, indistinguishable blocks / indiscernibility classes known from rough sets can be treated as atomic granules.) Naturally, such granules / blocks / classes have been already studied within GrC in the context of cognition [14]. However, when moving from blocks to computational building blocks, it is indeed useful to rely on IGrC because therein, the aforementioned c-granules are aimed at more tasks than just storing their signatures. Such c-granules build the relevant configurations of physical objects, initiate and modify interactions between them, so they are generally responsible for perceiving the physical world. They link physical objects with abstract objects used to represent the instances of decision making from the viewpoint of models working on information systems.

In the next sections, we consider some examples of challenges with respect to which IGrC can be worth adapting. For now, let us mention just one of them, namely hierarchical learning [41], [42]. From a logical viewpoint, one can think about it as learning satisfiability relations at different levels of hierarchy [43], [44]. This includes learning logical structures (e.g. relational structures or models), as well as logical formulas and their semantics expressed using those structures. The current methods of hierarchical learning are often based on GrC, with a special emphasis on designing hierarchies of information systems by basing on domain knowledge [45], [46]. However, granules on which the corresponding reasoning pipelines are performed, cannot neglect the underlying hierarchies of physical objects that are crucial for perception processes. Also, different layers of hierarchy can be connected to different types of sensors and actuators. Thus, the IGrC framework can be helpful indeed to embrace both, the relationships between information systems at different levels of hierarchy and the relationships between particular systems and the associated physical-object-related information sources.

---

[1]people.seas.harvard.edu/~valiant/researchinterests.htm

In the rest of the paper, in Section II, we refer to some selected literature on rough sets. This section is quite extensive given the anniversary flavor of the paper. In Section III, we go back to the discussion about the importance of information systems. We emphasize a need of operating with information systems (and the results of rough set computations over information systems) considered in a wider context of interactions between abstract and physical objects of different sorts. We go through several aspects of applications, whereby this kind of interaction is needed. We show to what extent the paradigms of IGrC can be helpful. We also refer to some concepts known from the domain of big data in order to put our discussion into other contexts. In Section IV, we conclude the paper.

Let us reemphasize the retrospective context of this paper. Besides the 40th anniversary of rough sets (which is our major focus) and the 40th anniversary of formal concept analysis (which was mentioned above), there are two more celebrations worth mentioning. The first of them refers to the rough set workshop series which "visited" the FedCSIS conferences for the first time 10 years ago[2] and which is now back to the program of technical sessions[3]. Secondly, this year's FedCSIS hosts the 30th International Symposium on Concurrency, Specification and Programming (CS&P 2022)[4]. The topics related to rough sets and information systems have been always visible at the CS&P events. In particular, the above-cited papers [36], [40], [44] come from CS&P.

II. SELECTED RELATED WORK ON ROUGH SETS

In order to provide a better viewpoint of the theory and applications of rough sets, we refer to two events from the past. These references will also constitute a better background for our major goal in this paper, which is the review of new advances on rough-set-related information systems.

A. Rough Sets at FedCSIS 2012

The first considered event is FedCSIS 2012 held 10 years ago in Wrocław, Poland[5]. As already mentioned, that was the first time when rough sets occurred so intensively at a FedCSIS conference. Let us start outlining the FedCSIS 2012 rough-set-related publications from [47], [48]. The first paper equips the Variable Precision Rough Set (VPRS) approach [49] with a Bayesian background [50]. The second paper combines VPRS with fuzzy rough set methods [24] in order to produce flexible decision rules. In summary, both papers deal with information imprecision – modeled by probabilities (which is the domain of VPRS) and fuzziness (which can be used to work *e.g.* with partial matching of rules' antecedents) – and attempt to extract interpretable decision models from the data [11].

The topic of rough-set-driven decision rules is considered also in one more FedCSIS 2012 publication [51]. In general, one will see throughout our whole paper that rough set principles fit the field of rule induction very well [18], [31].

[2]fedcsis.org/2012/rsa.html
[3]fedcsis.org/2022/rsta
[4]fedcsis.org/2022/csp
[5]fedcsis.org/2012/

This relationship is evident not only at a technical algorithmic level but also with respect to the common assumption of looking at the data through the glasses of information granules [16]. For more examples of connections between the worlds of rules and rough sets, let us refer *e.g.* to [17], [52].

Going further, papers [53], [54] introduce new heuristic measures that can be used during attribute reduction. It is worth noting that attribute reduction – or in other words algorithmic elimination of redundant attributes from the constructed set of attributes – is an important contribution of rough set research to knowledge discovery and in particular to its phase of feature selection [55]. As a complement to typical feature selection algorithms which attempt to add the most useful attributes, rough set methods take as input the sets of attributes produced by those typical algorithms and attempt to additionally compact them by eliminating unnecessary or approximately unnecessary elements. The additional aspect of attribute elimination occurs in just a few machine learning methodologies worldwide [56], so we can indeed say that this is an important rough sets' contribution to this area.

Papers [57], [58] continue the topic of attribute reduction. The first of them proposes greedy algorithms for deriving so-called superreducts from data sets with multivalued decision attributes (target variables). It is one more example of dealing with information imprecision in rough set frameworks. Superreducts are the subsets of attributes which are sufficient to induce values (or as in this case, the sets of possible values) of decision attributes. It is also important to note that the notion of superreduct is equivalent to the notion of test in the test theory [31]. The second paper compares the notions of decision bireduct (aimed at deriving both the sets of attributes and the sets of data objects for which those attributes are sufficient to construct rule-based decision models) and approximate reduct (aimed at eliminating as many attributes as possible, even if the ability to induce decisions is not fully preserved). This comparison was later extended in [17].

The next two papers extend the topic of feature selection toward some of modern data challenges, namely high-dimensionality and large data volumes. Paper [59] combines attribute reduction with attribute clustering. Attributes are first grouped using some rough-set-inspired measures and then the methods of attribute reduction work iteratively on cluster representatives. This allows for decreasing the complexity of attribute reduction for high numbers of attributes and it also improves interpretability of results. These methods were later extended to let them work with attribute groups which can be set up for many reasons, including heterogeneity of data sources that are required to derive attribute values [60].

Paper [61] copes with big data volumes by putting attribute reduction and decision tree induction into a relational database framework, whereby the corresponding algorithms are implemented in SQL. The authors extend some previous ideas in this field [19] and, in particular, employ an open source database engine called Infobright Community Edition to run experiments. Infobright Community Edition is an example of using rough sets to optimize other types of data computations,

in this case – query execution in relational databases[6]. This emphasizes that rough sets can be successfully used not only for machine learning and data mining but also for other tasks of big data processing. We refer to [62] for current developments related to Infobright Community Edition. We also refer to [63], where the Infobright's technology performance is explained in terms of rough set operations on specifically aggregated (granulated) multivalued information systems.

The last two rough-set-related publications are interesting from the information systems' viewpoint as well. In [64], the source of building an information system is a transformed ontological graph which encodes our knowledge about a given area [65]. The rules derived using the Dominance Rough Set Approach (DRSA) [20] express useful regularities within the original graph. This is actually an illustration of the fundamental idea behind information systems, namely, that such systems may contain not only the empirical data but they may also integrate it with domain knowledge [11], [43].

Finally, paper [66] presents the real-world application of rough sets to explore medical data. Herein, the information system – the input for rough-set-based model learning methods – does not correspond directly to the original data measurements. It is rather a result of a sequence of time-window-driven data aggregations which are typical for building hierarchical information systems describing complex objects [45], [67]. This work applied in particular a rough-set-based software system for machine learning and data mining – called RSES – which is now available in a library format [68] (see also the RSES extension targeted at spatio-temporal concepts[7]). It is also one more practical use case of deploying the Infobright Community Edition database engine to run the underlying operations over granulated and compressed data sets.

B. Rough Set Contest at PP-RAI 2022

The second considered event is the PP-RAI 2022 conference held this year in Gdynia, Poland[8]. The chairs of PP-RAI 2022 decided to celebrate the 40th anniversary of rough sets by organizing the contest for the most influential article on rough sets co-authored by Polish researchers in 2020 or later[9]. Let us discuss below the articles submitted to this contest.

Papers [69], [70] operate at the edge of rough sets and formal concept analysis [25]. The first paper adopts the principles of attribute reduction (or more generally, model compaction) to simplify so-called fuzzy concept lattices, introduced as the means for representing patterns and regularities hidden in numerical data [71]. The second paper is actually an extension of the previously-cited work [36]. The authors attempt to put classical rough sets, formal concept analysis and the DRSA-style extensions of rough sets [20] into a unified conceptual pipeline aimed at transforming the data – through various forms of (possibly multivalued) information systems [33] – to knowledge. Within such a universal framework, the authors

[6]en.wikipedia.org/wiki/Infobright
[7]mimuw.edu.pl/~bazan/roughice/?sLang=en
[8]pp-rai2022.umg.edu.pl/
[9]roughsets.org/newspage/events/

reconsider special cases of rough set operators known from different approaches. Therefore, one may say that this paper is a direct continuation of the ideas introduced in [1].

Papers [72], [73] link rough sets with logical foundations. The first paper shows how to express reasoning based on the VPRS-style extensions of rough sets [49] within the framework provided by a probabilistic extension of PROLOG [74]. The second paper shows how to reason about the properties of various types of rough set approximations within the framework provided by Mizar – a powerful system for automated proving [75]. Needless to say, such foundations are crucial for every theory, including reasoning within the theory and reasoning about the theory. We refer to [76] for more information about logical background of rough sets.

Papers [77], [78] present further advances in the previously-discussed popular rough set approaches such as the above-mentioned DRSA and fuzzy rough sets, respectively. The first paper uses the statistical learning machinery [79] to give new insights into parameters of probabilistic extensions of DRSA. The second paper, somewhat analogously, attempts to provide new interpretation of fuzzy rough set parameters. This is done by considering a new form of fuzzy granules [80], which consequently leads toward more intuitive derivation of fuzzy rough decision rules. One can say that these two articles fall into the same thematic categories as the previously-considered FedCSIS 2012 publications [47] and [48], respectively.

Papers [81], [82] continue the topic of feature selection. The first paper refers to heuristic attribute evaluation measures and data discretization techniques analogous to those reported in [10], [19]. The second paper seems to be particularly interesting as it extends the already-discussed topic of rough set software packages and libraries [24], [68] toward hardware optimizations that are specific for high performance computing. Such optimizations should be further compared and integrated with other acceleration opportunities, *e.g.* adaptation of MapReduce [60] and analytical database engines [61].

Papers [83], [84] refer to rough set software too. The first paper reports one more package delivering rough set methods for data mining and knowledge discovery. The second paper is about the application of that package to biomedical data mining. This second paper – besides its important experimental results – touches the aspects of visual data analytics [85], [86] and a need of understanding both, the analytical processes and their outcomes by subject matter experts [87], [88].

Papers [89], [90], [91] illustrate more real-world applications of rough set methods in the area of biomedicine. The first paper uses rough set approximations built over neighborhood-based information granules [92]. The remaining two papers confirm the expressive power of the DRSA-based decision rules. They also compare the accuracy of rule-based models with other approaches (such as random forests and logistic regression [93]) and show how to derive the attribute importance (see *e.g.* [94]) from the considered rules.

Papers [95], [96] continue the topic of rule induction. The first paper can be compared to [35], as both of them deal with deriving probabilistic rules from incomplete information sys-

tems, assuming several types of incompleteness. The second paper applies both rough-set-based [18] and fuzzy-set-based [12] rules in the task of posture detection. This is an example of real-world application, whereby the multi-stage solution needs to integrate sensor calibration, sensor data acquisition, inducing rules from the acquired data, as well as rule-based inference. With respect to making all such layers working together, this work can be compared to [43], [66], [67].

Papers [97], [98] deal with ensembles of decision models. The first paper employs so-called Dominance-based Rough Set Balanced Rule Ensemble for fraud detection. Herein, it is worth adding that rough set methods and applications include also examples of operating with ensembles of the aforementioned approximate reducts [99] and bireducts [17], [58] which correspond to bigger collections of rules. The second paper shows how to negotiate between classifiers and actually refers to the aforementioned conflict analysis model proposed by Professor Pawlak [6], [100]. On the other hand, the mechanism of voting in the third paper relies on the aforementioned three-way decision making [21]. It is worth emphasizing that solutions described in both papers attempt to provide a deeper insight into the ensemble decisions.

Paper [101] remains in the area of ensembles of decision models but it also touches an important aspect of incremental learning in dynamic data environments [102]. Herein, it is worth recalling a gentle difference between reasoning about objects or states in a repetitive fashion (whereby the values of attributes in information systems need to be cyclically updated) and reasoning about temporal objects or phenomena (which require different construction of information systems with attributes reflecting changes and trends) [103], [104].

Finally, papers [105], [106] combine the principles of rough sets and GrC with popular machine learning methods, referring to decision model ensembles as well. The idea is to prepare compacted data inputs – called granular reflections [15] – for the algorithms responsible for learning decision models such as *e.g.* neural networks or random forests (see [93] again). From a conceptual perspective, it corresponds to the afore-mentioned studies on aggregated / granulated / summarized information systems [63], [66]. This topic has also interesting relationships with some branches of approximate computing [107] and compressed image recognition [108].

III. INFORMATION SYSTEMS AND IGrC

As we have already emphasized, rough sets are based on data / information granulation. Both the original rough set approach and its extensions, need granules (and their descrip-tions / signatures) as inputs to compute approximations. The same applies to rough set methods of constructing decision models, *e.g.* rule-based models [52], [95]. On the other hand, granules can take different forms and have different origins. They can be partition blocks (induced by combinations of at-tribute values or ranges), dominance classes or neighborhoods [20], [89], relationships based on fuzzy (dis)similarity and (in)discernibility [78], [80] and so on. In information systems, granules can take different information signatures such as

precise values, value sets, ranges and distributions [35], [57]. Those signatures can be computed using different aggrega-tion mechanisms, often assuming non-trivial interdependencies with processes and devices that produce the data [63], [66]. The reliability and accessibility of information – therefore also reliability and accessibility of the outcomes of calculations over information systems – requires a careful analysis of all phases of forming the contents of such systems.

In Section I, we highlighted that the IGrC framework [37], [40] could be helpful to keep information systems aligned with respect to practical needs of operations on them in different contexts. In the next subsections, we will elaborate on several aspects of such alignment. As already discussed, IGrC uses so-called c-granules in order to create configurations of physical objects and control interactions between them so as to achieve computational objectives. Now let us add that the control mechanisms embedded within c-granules rely on one more type of granules – informational c-granules (ic-granules) which include both abstract (informational) and physical layers. They contain specifications how to link the abstract and physical worlds, whereby the abstract world corresponds in particular to (the networks of) information systems. The perceived properties of physical objects can be used to transform the current configurations of ic-granules, *i.e.* to modify interactions between objects. Such mechanisms require a design of new methods of reasoning about *where*, *when*, *what*, and *how* to perceive using different sensors or actuators. New methods for judging membership (alignment, matching) of the perceived situations in (with) rough set approximations of complex concepts are needed too.

A. Reliability of Information

This kind of reliability is studied in many fields. In the domain of big data, it is referred as one of the "V's" – Veracity [109]. Actually, we have already dealt with some of other "V's" in the previous sections, *e.g.* Volume [61], Velocity [101] and Variety [60]. However, without addressing Veracity, i.e. assuring data quality that is transformed into information reliability, any solutions focused on those other "V's" cannot guarantee anything useful. Another popular term related to this problem is "garbage data". It refers to the fact that if a machine learning method is executed on improper data, then the resulting models cannot be expected to work successfully. The causes of data being garbage data may be connected to problems with *e.g.* sensor measurements, data parsing, or even data labels acquired from human experts [67], [110].

A technical solution to cope with garbage data is often to filter them out by using validation procedures (*e.g.* checking sensor scales) [99]. However, in many applications – such as [66] (Subsection II-A), [96] (Subsection II-B) or just-mentioned [67] – it would lead toward disqualifying too broad data fragments, if any formalized validation is possible at all. Another approach is to *live* with the unreliable data and moreover, to take such unreliability into account while con-ducting any computations. With this respect, non-deterministic information systems have some tools to express uncertainty

by replacing precise values with sets, intervals, etc. [35], [63]. However, (a degree of) reliability remains something different, as it refers to the way the data was acquired from the physical world rather than the specification of attribute values.

The IGrC framework is quite natural when it comes to reasoning about such an additional layer of information. Interactive granular computations can be actually extended toward adaptive searching strategies for the most relevant and reliable data, spatio-temporal windows pointing out to fragments of the physical world where the most reliable measurements and / or actions should be performed, and so on. This kind of reasoning may be also associated with the domain of data governance which extends towards data and information security, accessibility, as well as the protocols of interactions between intelligent systems and humans [111], [112]. On the other hand, the discussed physical-world-related aspects can be an additional contribution of IGrC to data governance. Moreover, IGrC can be helpful to operate with often softly expressed regulations about data integrity and timeliness.

It is also worth referring the above discussion to the meanings of aleatoric and epistemic uncertainties in machine learning [113]. From this perspective, a limited reliability of the contents of information systems can be treated as one of ingredients of the epistemic uncertainty, as it puts together both, the model and the data deficiencies. However, we believe that these two sources of deficiencies should be kept separately, with the third type of experimental / physical uncertainty explicitly considered. The analysis of this third type of uncertainty should be taken into account when assessing the efficiency and stability of machine learning models, especially given the fact that in some practical scenarios the inputs to the learning algorithms can be unreliably extracted for the purpose of *e.g.* accelerating computations [114].

B. Acquisition of Information

In order to talk about information reliability, we first need to assure that information can be gathered at all. In practice, there is often a great variety of data available but it does not mean yet that the corresponding information is sufficiently complete to perform any kind of analysis. (This relates to one more "V": Value.) Some promising approaches to data enrichment refer to the paradigm of active learning [115], which can be further extended toward establishing an interactive loop within which subject matter experts label data objects that are of the highest interest to the machine learning algorithms. One just needs to think about controlling the quality of such labels [110].

Similarly, the data enrichment processes can rely on connecting information systems with physical systems [116]. Actually, it is worth pointing out that humans can be considered as a special kind of physical objects that interact with decision support systems and / or intelligent systems. This refers to a broader topic of the information and communication technologies (ICT) systems [39] which put together the aspects of hardware (*e.g.* sensors), software (*e.g.* machine learning methods), the data (including domain knowledge), and the system users (in particular subject matter experts).

The above ideas require a firm layer that connects information systems with the physical world where the data comes from. In IGrC, every granule should have an access to instructions how to compose the values of particular attributes for particular objects [38], [40]. Moreover, it is important for this layer to log a history of attempts to calculate particular fragments of an information system. Such history may let us avoid mistakes and misinterpretations related to the data acquisition processes. That history may be also useful while assessing reliability of the current contents of an information system. Such mechanisms can be adopted also from the architectures of granular database engines [62], [107], whereby the aspects of information completeness and reliability are equally important as in the field of machine learning.

Going back to the framework of active learning, let us claim that subject matter experts can assist us not only in enriching the data with labels but also enriching the data mining algorithms with domain insights. As an example, let us think one more time about the task of feature selection. There are various techniques of measuring and visualizing attribute importance [88], [90], [94] but they are usually applied to report to humans the final results instead of "inviting" them into a more interactive dialogue on feature selection process. In this regard, we refer to [85] where incrementally constructed information systems are employed to guide subject matter experts through such an interactive process, letting them share their recommendations about the most relevant attributes.

Last but not least, when it comes to decision problems related to complex phenomena, it is worth attempting – using the elements of active learning and human-computer interaction – to acquire from subject matter experts even more advanced knowledge, expressed in terms of hierarchical structures and dependencies. This fits the paradigm of computing with words [117] (which also corresponds to the foundations of information granulation with respect to decomposing complex problems onto their smaller components) and, in particular, the following challenge formulated by Judea Pearl [118]:

> *Traditional statistics is strong in devising ways of describing data and inferring distributional parameters from sample. Causal inference requires two additional ingredients: a science-friendly language for articulating causal knowledge, and a mathematical machinery for processing that knowledge, combining it with data and drawing new causal conclusions about a phenomenon.*

C. Accessibility and Cost of Information

In practical deployments, there is always a risk that some of data sources – which are needed to calculate some of attributes being inputs to a decision model – will be temporarily unavailable (because of *e.g.* physical connection problems or dissatisfaction of some data governance rules) or unreliable (as discussed in Subsection III-A). Feature selection [56], including contribution of rough set methods to elimination / reduction of redundant attributes [10], can be a remedium to this problem – less attributes require less aspects of data to be

calculated. Moreover, it is possible to diversify data sources needed to derive attributes that are used by particular models in an ensemble [17], [55]. This increases a chance that at least some of models would be usable in a given situation.

However, it is not only about the accessibility, sometimes it is also about the cost. For instance, in the recent contests organized at one of online data mining competition platforms[10] [119], [120], the participants purposefully did not take into account some of the available data sources (modalities) because – according to them – derivation of attributes from those sources would be too expensive computationally. Going further along this path, one may say that even if such "expensive" attributes are included into a model, its deployed version should have a choice to decide dynamically which of them (and at what level of precision [114]) are necessary to be calculated. This is important the more so as in intelligent systems dealing with complex phenomena, the most adequate selections and meanings of attributes can be changed over time [40].

Going even further, if appropriate metadata is maintained on the side of an information system, the same attribute values can be derived from different information sources or using different data modalities, perhaps subject to different cost and precision [43], [67]. This is well-aligned with the IGrC assumptions discussed in the previous subsections. The point is to pass to information granules the decision power in regard to how they produce information that they are responsible for, and make decision models responsible for timely asking those granules for particular information pieces [38], [116].

The above discussion can be extended toward a broader topic of whether the data / information updates should be rather "pushed" or "pulled" in the computational pipelines which involve learning and applying the learnt models. A common assumption is that any change in the underlying information should me more or less quickly transmitted to the inputs of a model, causing its recalculation or at least modification of its behavior [101], [102]. However, in big data scenarios it is not so obvious – it may be safer to leave such decisions in hands of information granules equipped with well-designed triggers and internal cost models [37], [107].

D. Networks of Information Systems

Continuing with the topic of intelligent systems aimed at reasoning about complex phenomena, we already know that data sources required to learn the underlying decision models cannot be acquired in a single-step process. It is necessary to provide such systems with permanent links to relevant fragments of the physical world and keep adapting (actively but also reasonably, from the computational perspective) the induced models following changes in the perceived situation. Recalling our comments about logical reasoning related to hierarchical systems [43] (see the end of Section I), one should also be aware that these hierarchical structures are dynamically changing in time and the relevant reasoning methods should allow to the system to perform the necessary reasoning about

such dynamical structures. This seems to be aligned with the following opinion expressed by Frederick Brooks [121]:

> *Mathematics and the physical sciences made great strides for three centuries by constructing simplified models of complex phenomena, deriving, properties from the models, and verifying those properties experimentally. This worked because the complexities ignored in the models were not the essential properties of the phenomena. It does not work when the complexities are the essence.*

The starting point is to work with environments which create, maintain and synchronize multiple dynamic information systems. Such networks of information systems would be still a kind of abstraction of the real world but on the other hand, they would reflect it more accurately than single systems.

Let us first focus on the aforementioned hierarchical systems. We have already referred to the approaches whereby domain knowledge – expressed in terms of ontologies of concepts associated with a particular decision problem – is utilized to decompose that problem onto simpler components located within a hierarchical schema and then, to aggregate perceived information along that schema [45], [117]. To facilitate such aggregation process, it is indeed convenient to design a hierarchy of information systems whose objects (and therefore also attributes) correspond to different levels of conceptual granularity. This idea is actually analogous to modeling the data by means of multi-table relational database structures [114], [119], and it can be observed in quite a few applications mentioned earlier [46], [66], [67], [103].

Somewhat "orthogonal" aspect of thinking about multiple information systems refers to concurrency and distributed computations. From this perspective, at each level of the above-discussed hierarchies, we may actually imagine a group of systems working collectively and exchanging information. Herein, it is important to refer to the models proposed by Professor Pawlak [7], as well as the history of the aforementioned conferences on Concurrency, Specification and Programming (CS&P). Furthermore, it is useful to refer also to the works on the networks of information systems linked by so-called infomorphisms [32], [122]. Some relevant realizations can be found also in other domains. For instance, the already-considered granular database engine [107] contained a mechanism of distributed execution of analytical queries, whereby particular computational nodes could exchange with each other some approximate partial answers and, basing on such understood rough set approximations, decide autonomously whether it is worth requesting for the precise results.

Once we have a hierarchy / distribution of information systems, we can extend their network with the IGrC-based connections to the physical world [38], [123]. This implies a number of challenges, as the above-discussed coordination between particular information systems needs to be combined with coordination of each single system with its physical "alter ego". For instance, we can consider a more active version of the tasks of attribute selection and extraction [56], [60], whereby it is required to develop new methods of

selection and construction of sensors. At a more general level, the whole idea requires a distributed control of c-granules, whereby specific reasoning methods (related to cooperation / competition between granules) need to express the expected behavioral patterns of the whole "society" of granules. Herein, one can seek for inspirations in the previously discussed conflict analysis [6]. The requirements of the aforementioned ICT systems [39], web intelligence [65] or *e.g.* IoT analytics [124] – whereby there are a number of federated learning scenarios involving distributed agents (and their underlying information systems) – can be a useful analogy as well.

IV. CONCLUSIONS AND FUTURE DIRECTIONS

The first goal of this paper was to expose the current progress of the theory and applications of rough sets – the methodology founded by Zdzisław Pawlak with the aim of deriving and expressing important patterns and dependencies subject to limited (incomplete, imprecise) information about the concepts of practical interest [1], [9]. We examined connections of rough sets with decision making [20], [21], logics [73], [76], probability [47], [72], statistical / machine learning [77], [105], data mining [19], [23], fuzzy sets [13], [48], formal concept analysis [26], [69], and other data / information / knowledge representation methodologies [31], [32]. We discussed some of rough set techniques aimed at attribute selection / reduction treated as a component of knowledge discovery processes [10], [17], with particular emphasis on computational scalability challenges [60], [82]. We paid special attention to rough set approaches to construction of interpretable (explainable by design) rule-based decision models [18], [48], [52], [97]. We referred to rough set software packages for data mining and machine learning [24], [68], [83], as well as other technologies which utilize rough set approximation principles for their internal purposes [62]. We also recalled several (out of many) applications of rough set methods in real-world data analysis, including biomedical and healthcare applications whereby interpretability of decision models is of special importance [66], [67], [84], [91].

Our second goal was to address the progress in the area of information systems [3], [4]. We referred to their extensions [33], [70] and we outlined a number of applications which use specifically formed information systems as the means for representing (granulated / aggregated) data, (uncertain / imprecise) information, and (appropriately transformed) knowledge structures [35], [43], [63], [89]. We pointed out that information systems – especially their hierarchies and networks – constitute the means for reasoning about complex spatio-temporal phenomena [45], [104]. We also claimed that information systems can be a medium to conduct interactive data analytics involving subject matter experts [85] and support interactions between multiple data exploration processes [123]. That led us toward discussing the current challenges (often referred as the big data "V's") in front of information systems understood as the means for representing and delivering data required for the learning processes [102], [109], [110], [119]. Accordingly, we examined whether the principles of so-called interactive

granular computing (IGrC) [37], [116] can help us to face those challenges and to what extent they are aligned with some of emerging trends in machine learning [115], [124].

It was important for us to discuss the principles of granular computing – including IGrC – together with rough sets and information systems, as these three domains interfere with each other in many interesting ways [11], [14], [15], [40]. In particular, IGrC may have future implications for the design of intelligent systems, *e.g.* when it comes to so-called perceptual rough sets[11]. If one wants to build rough set approximations of complex concepts in real-world environments, then it is required to design a dynamic space of granules which are able to reason about complex approximation constructions. The corresponding reasoning methods will need to be far richer than the ones considered so far in rough set applications.

Some other future directions for rough sets and information systems refer to continuation of development of real-world applications, focused on *e.g.* images and video recordings [22], [46], [120], as well as signals and sensor measurements [99], [96], [103]. This kind of development should emphasize strong assets of rough sets, such as straightforward interpretability of the derived decision models, even when it comes to modeling very complex and dynamic situations [11], [101]. Needless to say, interpretability is now the key objective for a great majority of machine learning applications [87], [88].

When thinking about the future it is also worth referring to the history. That reflected one more objective: acknowledging the 40th anniversary of rough sets [1], their founder [5], as well as some of relevant past and present events such as FedCSIS 2012 (rough set papers published exactly 10 years ago) [47], [48], [51], [53], [54], [57], [58], [59], [61], [64], [66], the PP-RAI 2022 rough set contest [69], [70], [72], [73], [77], [78], [81], [82], [83], [84], [89], [90], [91], [95], [96], [97], [98], [101], [105], [106] and celebration of the 30th CS&P – the event series whereby this paper's topics have been regularly addressed [36], [40], [44], [103], [116], [123].

In the end, let us recall that this is not the first anniversary corresponding to rough sets in the history of the FedCSIS conferences. Indeed, FedCSIS 2016 (Gdańsk, Poland) hosted the international panel discussion in memoriam of the 90th anniversary of the birth and the 10th anniversary of the death of Professor Pawlak[12]. The previously-cited publications [4], [8], [100] were prepared specially for that panel.

REFERENCES

[1] Z. Pawlak, "Rough Sets," *International Journal of Computer and Information Sciences*, vol. 11, no. 5, pp. 341–356, 1982. [Online]. Available: doi.org/10.1007/BF01001956

[2] ——, *Rough Sets – Theoretical Aspects of Reasoning about Data*, ser. Theory and Decision Library D. Springer, 1991. [Online]. Available: doi.org/10.1007/978-94-011-3534-4

---

[11] See *e.g.* A. Skowron: Perceptual Rough Set Approach in Interactive Granular Computing (keynote). Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2022), July 11-15, 2022, Milan, Italy. roughsets.org/bin/1f094938771802d02dec99c66823875c.PDF

[12] fedcsis.org/2016/plenary_panel

[3] ——, "Information Systems – Theoretical Foundations," *Information Systems*, vol. 6, no. 3, pp. 205–218, 1981. [Online]. Available: doi.org/10.1016/0306-4379(81)90023-5

[4] V. W. Marek, "Working with Zdzisław Pawlak – Personal Reminiscences," in *Proceedings of the 2016 Federated Conference on Computer Science and Information Systems, FedCSIS 2016, Gdańsk, Poland, September 11-14, 2016*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., vol. 8. IEEE, 2016, pp. 189–190. [Online]. Available: doi.org/10.15439/2016F002

[5] A. Skowron, M. K. Chakraborty, J. W. Grzymała-Busse, V. W. Marek, S. K. Pal, J. F. Peters, G. Rozenberg, D. Ślęzak, R. Słowiński, S. Tsumoto, A. Wakulicz-Deja, G. Wang, and W. Ziarko, "Professor Zdzisław Pawlak (1926-2006): Founder of the Polish School of Artificial Intelligence," in *Rough Sets and Intelligent Systems – Professor Zdzisław Pawlak in Memoriam – Volume 1*, ser. Intelligent Systems Reference Library, A. Skowron and Z. Suraj, Eds. Springer, 2013, vol. 42, pp. 1–56. [Online]. Available: doi.org/10.1007/978-3-642-30344-9_1

[6] Z. Pawlak, "An Inquiry into Anatomy of Conflicts," *Journal of Information Sciences*, vol. 109, pp. 65–78, 1998. [Online]. Available: doi.org/10.1016/S0020-0255(97)10072-X

[7] ——, "Concurrent versus Sequential – the Rough Sets Perspective," *Bulletin of the EATCS*, vol. 48, pp. 178–190, 1992.

[8] J. F. Peters and S. Ramanna, "Maximal Nucleus Clusters in Pawlak Paintings. Nerves as Approximating Tools in Visual Arts," in *Proceedings of the 2016 Federated Conference on Computer Science and Information Systems, FedCSIS 2016, Gdańsk, Poland, September 11-14, 2016*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., vol. 8. IEEE, 2016, pp. 199–202. [Online]. Available: doi.org/10.15439/2016F004

[9] Z. Pawlak and A. Skowron, "Rudiments of Rough Sets," *Information Sciences*, vol. 177, no. 1, pp. 3–27, 2007. [Online]. Available: doi.org/10.1016/j.ins.2006.06.003

[10] ——, "Rough Sets and Boolean Reasoning," *Information Sciences*, vol. 177, no. 1, pp. 41–73, 2007. [Online]. Available: doi.org/10.1016/j.ins.2006.06.007

[11] ——, "Rough Sets: Some Extensions," *Information Sciences*, vol. 177, no. 1, pp. 28–40, 2007. [Online]. Available: doi.org/10.1016/j.ins.2006.06.006

[12] L. A. Zadeh, "Fuzzy Sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965. [Online]. Available: doi.org/10.1016/S0019-9958(65)90241-X

[13] C. Cornelis, "Hybridization of Fuzzy Sets and Rough Sets: Achievements and Opportunities," in *Proceedings of the 2022 Federated Conference on Computer Science and Intelligence Systems, Sofia, Bulgaria, September 4-7, 2022*, ser. Annals of Computer Science and Information Systems, M. Ganzha, M. Paprzycki, and D. Ślęzak, Eds., vol. 30, 2022.

[14] S. K. Pal, "Soft Data Mining, Computational Theory of Perceptions, and Rough-Fuzzy Approach," *Information Sciences*, vol. 163, no. 1-3, pp. 5–12, 2004. [Online]. Available: doi.org/10.1016/j.ins.2003.03.014

[15] L. Polkowski and P. Artiemjew, *Granular Computing in Decision Approximation – An Application of Rough Mereology*, ser. Intelligent Systems Reference Library. Springer, 2015, vol. 77. [Online]. Available: doi.org/10.1007/978-3-319-12880-1

[16] A. Skowron and J. Stepaniuk, "Information Granules: Towards Foundations of Granular Computing," *International Journal of Intelligent Systems*, vol. 16, no. 1, pp. 57–85, 2001.

[17] S. Stawicki, D. Ślęzak, A. Janusz, and S. Widz, "Decision Bireducts and Decision Reducts – A Comparison," *International Journal of Approximate Reasoning*, vol. 84, pp. 75–109, 2017. [Online]. Available: doi.org/10.1016/j.ijar.2017.02.007

[18] J. W. Grzymała-Busse, "Rule Induction," in *Data Mining and Knowledge Discovery Handbook, 2nd ed*, O. Maimon and L. Rokach, Eds. Springer, 2010, pp. 249–265. [Online]. Available: doi.org/10.1007/978-0-387-09823-4_13

[19] H. S. Nguyen, "Approximate Boolean Reasoning: Foundations and Applications in Data Mining," *Transactions on Rough Sets*, vol. 5, pp. 334–506, 2006. [Online]. Available: doi.org/10.1007/11847465_16

[20] S. Greco, B. Matarazzo, and R. Słowiński, "Rough Sets Theory for Multicriteria Decision Analysis," *European Journal of Operational Research*, vol. 129, no. 1, pp. 1–47, 2001. [Online]. Available: doi.org/10.1016/S0377-2217(00)00167-3

[21] Y. Yao, "Three-Way Decisions and Cognitive Computing," *Cognitive Computation*, vol. 8, no. 4, pp. 543–554, 2016. [Online]. Available: doi.org/10.1007/s12559-016-9397-5

[22] P. Maji and S. K. Pal, "Maximum Class Separability for Rough-Fuzzy C-Means Based Brain MR Image Segmentation," *Transactions on Rough Sets*, vol. 9, pp. 114–134, 2008. [Online]. Available: doi.org/10.1007/978-3-540-89876-4_7

[23] G. Peters, F. A. Crespo, P. Lingras, and R. Weber, "Soft Clustering – Fuzzy and Rough Approaches and Their Extensions and Derivatives," *International Journal of Approximate Reasoning*, vol. 54, no. 2, pp. 307–322, 2013. [Online]. Available: doi.org/10.1016/j.ijar.2012.10.003

[24] L. S. Riza, A. Janusz, C. Bergmeir, C. Cornelis, F. Herrera, D. Ślęzak, and J. M. Benítez, "Implementing Algorithms of Rough Set Theory and Fuzzy Rough Set Theory in the R Package "RoughSets"," *Information Sciences*, vol. 287, pp. 68–89, 2014. [Online]. Available: doi.org/10.1016/j.ins.2014.07.029

[25] R. Wille, "Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts," in *Ordered Sets, Proceedings*, ser. NATO Advanced Studies Institute, I. Rival, Ed., vol. 83. Dordrecht: Springer, 1982, pp. 445–470.

[26] R. E. Kent, "Rough Concept Analysis: A Synthesis of Rough Sets and Formal Concept Analysis," *Fundam. Informaticae*, vol. 27, no. 2/3, pp. 169–181, 1996. [Online]. Available: doi.org/10.3233/FI-1996-272305

[27] S. Naouali and R. Missaoui, "Flexible Query Answering in Data Cubes," in *Data Warehousing and Knowledge Discovery, 7th International Conference, DaWaK 2005, Copenhagen, Denmark, August 22-26, 2005, Proceedings*, ser. Lecture Notes in Computer Science, A. M. Tjoa and J. Trujillo, Eds., vol. 3589. Springer, 2005, pp. 221–232. [Online]. Available: doi.org/10.1007/11546849_22

[28] D. Ślęzak and V. Eastwood, "Data Warehouse Technology by Infobright," in *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2009, Providence, Rhode Island, USA, June 29 – July 2, 2009*, U. Çetintemel, S. B. Zdonik, D. Kossmann, and N. Tatbul, Eds. ACM, 2009, pp. 841–846. [Online]. Available: doi.org/10.1145/1559845.1559933

[29] M. L. Hughes, R. M. Shank, and E. S. Stein, *Decision Tables*. MDI Publications, 1968.

[30] A. M. Moreno Garcia, M. Verhelle, and J. Vanthienen, "An Overview of Decision Table Literature 1982-2000," in *The Fifth International Conference on Artificial Intelligence and Emerging Technologies in Accounting, Finance and Tax, Huelva, Spain, November 2-3*, 2000. [Online]. Available: feb.kuleuven.be/prologa/download/overview82-2000.pdf

[31] I. Chikalov, V. V. Lozin, I. Lozina, M. Moshkov, H. S. Nguyen, A. Skowron, and B. Zielosko, *Three Approaches to Data Analysis – Test Theory, Rough Sets and Logical Analysis of Data*, ser. Intelligent Systems Reference Library. Springer, 2013, vol. 41. [Online]. Available: doi.org/10.1007/978-3-642-28667-4

[32] A. Skowron, J. Stepaniuk, and J. F. Peters, "Rough Sets and Infomorphisms: Towards Approximation of Relations in Distributed Environments," *Fundamenta Informaticae*, vol. 54, no. 2-3, pp. 263–277, 2003. [Online]. Available: content.iospress.com/articles/fundamenta-informaticae/fi54-2-3-12

[33] E. Orłowska and Z. Pawlak, "Representation of Nondeterministic Information," *Theoretical Computer Science*, vol. 29, pp. 27–39, 1984. [Online]. Available: doi.org/10.1016/0304-3975(84)90010-0

[34] W. Lipski Jr., "On Databases with Incomplete Information," *Journal of the ACM*, vol. 28, no. 1, pp. 41–70, 1981. [Online]. Available: doi.org/10.1145/322234.322239

[35] H. Sakai and M. Nakata, "Rough Set-based Rule Generation and Apriori-based Rule Generation from Table Data Sets: A Survey and a Combination," *CAAI Transactions on Intelligence Technology*, vol. 4, no. 4, pp. 203–213, 2019. [Online]. Available: doi.org/10.1049/trit.2019.0001

[36] M. Wolski and A. Gomolińska, "Semantic Rendering of Data Tables – Multivalued Information Systems Revisited," in *Proceedings of the 25th International Workshop on Concurrency, Specification and Programming, Rostock, Germany, September 28-30, 2016*, ser. CEUR Workshop Proceedings, B. Schlingloff, Ed., vol. 1698. CEUR-WS.org, 2016, pp. 113–124. [Online]. Available: ceur-ws.org/Vol-1698/CS&P2016_11_Wolski&Gomolinska_Semantic-Rendering-of-Data-Tables-Multivalued-Information-Systems-Revisited.pdf

[37] A. Skowron and A. Jankowski, "Rough Sets and Interactive Granular

Computing," *Fundamenta Informaticae*, vol. 147, no. 2-3, pp. 371–385, 2016. [Online]. Available: doi.org/10.3233/FI-2016-1413

[38] A. Skowron and P. Wasilewski, "Interactive Information Systems: Toward Perception Based Computing," *Theoretical Computer Science*, vol. 454, pp. 240–260, 2012. [Online]. Available: doi.org/10.1016/j.tcs.2012.04.019

[39] M. Lippi, S. Mariani, M. Martinelli, and F. Zambonelli, "Individual and Collective Self-Development: Concepts and Challenges," in *Proceedings of the 2022 Federated Conference on Computer Science and Intelligence Systems, Sofia, Bulgaria, September 4-7, 2022*, ser. Annals of Computer Science and Information Systems, M. Ganzha, M. Paprzycki, and D. Ślęzak, Eds., vol. 30, 2022.

[40] A. Jankowski, A. Skowron, and R. W. Świniarski, "Interactive Complex Granules," in *Proceedings of the 22nd International Workshop on Concurrency, Specification and Programming, Warsaw, Poland*, ser. CEUR Workshop Proceedings, M. S. Szczuka, L. Czaja, and M. Kacprzak, Eds., vol. 1032. CEUR-WS.org, 2013, pp. 206–218. [Online]. Available: ceur-ws.org/Vol-1032/paper-18.pdf

[41] T. A. Poggio and S. Smale, "The Mathematics of Learning: Dealing with Data," *Notices of the American Mathematical Society*, vol. 50, no. 5, pp. 537–544, 2003.

[42] P. Stone, *Layered Learning in Multi-Agent Systems: A Winning Approach to Robotic Soccer*. MIT Press, 2000.

[43] P. Doherty, W. Łukaszewicz, A. Skowron, and A. Szałas, *Knowledge Representation Techniques – A Rough Set Approach*, ser. Studies in Fuzziness and Soft Computing. Springer, 2006, vol. 202. [Online]. Available: doi.org/10.1007/3-540-33519-6

[44] S. Dutta and P. Wasilewski, "Dialogue in Hierarchical Learning of a Concept Using Prototypes and Counterexamples," in *Proceedings of the 24th International Workshop on Concurrency, Specification and Programming, Rzeszów, Poland, September 28-30, 2015*, ser. CEUR Workshop Proceedings, Z. Suraj and L. Czaja, Eds., vol. 1492. CEUR-WS.org, 2015, pp. 126–133. [Online]. Available: ceur-ws.org/Vol-1492/Paper_12.pdf

[45] J. G. Bazan, "Hierarchical Classifiers for Complex Spatio-Temporal Concepts," *Transactions on Rough Sets*, vol. 9, pp. 474–750, 2008. [Online]. Available: doi.org/10.1007/978-3-540-89876-4_26

[46] S. H. Nguyen, T. T. Nguyen, M. S. Szczuka, and H. S. Nguyen, "An Approach to Pattern Recognition Based on Hierarchical Granular Computing," *Fundamenta Informaticae*, vol. 127, no. 1-4, pp. 369–384, 2013. [Online]. Available: doi.org/10.3233/FI-2013-915

[47] I. Düntsch and G. Gediga, "Weighted Lambda Precision Models in Rough Set Data Analysis," in *Federated Conference on Computer Science and Information Systems – FedCSIS 2012, Wrocław, Poland, 9-12 September 2012, Proceedings*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., 2012, pp. 287–294. [Online]. Available: fedcsis.org/proceedings/2012/pliks/89.pdf

[48] T. Fan, C. Liau, and D. Liu, "Variable Precision Fuzzy Rough Set Based on Relative Cardinality," in *Federated Conference on Computer Science and Information Systems – FedCSIS 2012, Wrocław, Poland, 9-12 September 2012, Proceedings*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., 2012, pp. 43–47. [Online]. Available: fedcsis.org/proceedings/2012/pliks/398.pdf

[49] W. Ziarko, "Variable Precision Rough Set Model," *Journal of Computer and System Sciences*, vol. 46, no. 1, pp. 39–59, 1993. [Online]. Available: doi.org/10.1016/0022-0000(93)90048-2

[50] Z. Pawlak, "Rough Sets, Decision Algorithms and Bayes' Theorem," *European Journal of Operational Research*, vol. 136, no. 1, pp. 181–189, 2002. [Online]. Available: doi.org/10.1016/S0377-2217(01)00029-7

[51] B. Zielosko, "Sequential Optimization of γ-Decision Rules," in *Federated Conference on Computer Science and Information Systems – FedCSIS 2012, Wrocław, Poland, 9-12 September 2012, Proceedings*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., 2012, pp. 339–346. [Online]. Available: fedcsis.org/proceedings/2012/pliks/87.pdf

[52] M. Kryszkiewicz, "ACBC-Adequate Association and Decision Rules Versus Key Generators and Rough Sets Approximations," *Fundamenta Informaticae*, vol. 148, no. 1-2, pp. 65–85, 2016. [Online]. Available: doi.org/10.3233/FI-2016-1423

[53] L. G. Nguyen, "Metric Based Attribute Reduction in Decision Tables," in *Federated Conference on Computer Science and Information Systems – FedCSIS 2012, Wrocław, Poland, 9-12 September 2012, Proceedings*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., 2012, pp. 311–316. [Online]. Available: fedcsis.org/proceedings/2012/pliks/311.pdf

[54] L. G. Nguyen and H. S. Nguyen, "On Elimination of Redundant Attributes from Decision Table," in *Federated Conference on Computer Science and Information Systems – FedCSIS 2012, Wrocław, Poland, 9-12 September 2012, Proceedings*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., 2012, pp. 317–322. [Online]. Available: fedcsis.org/proceedings/2012/pliks/324.pdf

[55] R. Polikar, J. DePasquale, H. S. Mohammed, G. Brown, and L. I. Kuncheva, "Learn^{++}.MF: A Random Subspace Approach for the Missing Feature Problem," *Pattern Recognition*, vol. 43, no. 11, pp. 3817–3832, 2010. [Online]. Available: doi.org/10.1016/j.patcog.2010.05.028

[56] M. Dash and H. Liu, "Consistency-based Search in Feature Selection," *Artificial Intelligence*, vol. 151, no. 1-2, pp. 155–176, 2003. [Online]. Available: doi.org/10.1016/S0004-3702(03)00079-1

[57] M. Azad, I. Chikalov, M. Moshkov, and B. Zielosko, "Tests for Decision Tables with Many-Valued Decisions – Comparative Study," in *Federated Conference on Computer Science and Information Systems – FedCSIS 2012, Wrocław, Poland, 9-12 September 2012, Proceedings*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., 2012, pp. 271–277. [Online]. Available: fedcsis.org/proceedings/2012/pliks/140.pdf

[58] S. Stawicki and S. Widz, "Decision Bireducts and Approximate Decision Reducts: Comparison of Two Approaches to Attribute Subset Ensemble Construction," in *Federated Conference on Computer Science and Information Systems – FedCSIS 2012, Wrocław, Poland, 9-12 September 2012, Proceedings*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., 2012, pp. 331–338. [Online]. Available: fedcsis.org/proceedings/2012/pliks/348.pdf

[59] A. Janusz and D. Ślęzak, "Utilization of Attribute Clustering Methods for Scalable Computation of Reducts from High-Dimensional Data," in *Federated Conference on Computer Science and Information Systems – FedCSIS 2012, Wrocław, Poland, 9-12 September 2012, Proceedings*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., 2012, pp. 295–302. [Online]. Available: fedcsis.org/proceedings/2012/pliks/330.pdf

[60] M. Grzegorowski, A. Janusz, D. Ślęzak, and M. S. Szczuka, "On the Role of Feature Space Granulation in Feature Selection Processes," in *2017 IEEE International Conference on Big Data (IEEE BigData 2017), Boston, MA, USA, December 11-14, 2017*, J. Nie, Z. Obradovic, T. Suzumura, R. Ghosh, R. Nambiar, C. Wang, H. Zang, R. Baeza-Yates, X. Hu, J. Kepner, A. Cuzzocrea, J. Tang, and M. Toyoda, Eds. IEEE Computer Society, 2017, pp. 1806–1815. [Online]. Available: doi.org/10.1109/BigData.2017.8258124

[61] M. Kowalski and S. Stawicki, "SQL-based Heuristics for Selected KDD Tasks over Large Data Sets," in *Federated Conference on Computer Science and Information Systems – FedCSIS 2012, Wrocław, Poland, 9-12 September 2012, Proceedings*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., 2012, pp. 303–310. [Online]. Available: fedcsis.org/proceedings/2012/pliks/395.pdf

[62] M. Wnuk, S. Stawicki, and D. Ślęzak, "Reinventing Infobright's Concept of Rough Calculations on Granulated Tables for the Purpose of Accelerating Modern Data Processing Frameworks," in *2020 IEEE International Conference on Big Data (IEEE BigData 2020), Atlanta, GA, USA, December 10-13, 2020*, X. Wu, C. Jermaine, L. Xiong, X. Hu, O. Kotevska, S. Lu, W. Xu, S. Aluru, C. Zhai, E. Al-Masri, Z. Chen, and J. Saltz, Eds. IEEE, 2020, pp. 5405–5412. [Online]. Available: doi.org/10.1109/BigData50022.2020.9378233

[63] D. Ślęzak, P. Synak, A. Wojna, and J. Wróblewski, "Two Database Related Interpretations of Rough Approximations: Data Organization and Query Execution," *Fundam. Informaticae*, vol. 127, no. 1-4, pp. 445–459, 2013. [Online]. Available: doi.org/10.3233/FI-2013-920

[64] K. Pancerz, "Dominance-based rough set approach for decision systems over ontological graphs," in *Federated Conference on Computer Science and Information Systems – FedCSIS 2012, Wrocław, Poland, 9-12 September 2012, Proceedings*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., 2012, pp. 323–330. [Online]. Available: fedcsis.org/proceedings/2012/pliks/366.pdf

[65] N. Zhong, J. Ma, R. Huang, J. Liu, Y. Yao, Y. Zhang, and J. Chen, "Research Challenges and Perspectives on Wisdom Web of Things (W2T)," *The Journal of Supercomputing*, vol. 64, no. 3, pp. 862–882, 2013. [Online]. Available: doi.org/10.1007/s11227-010-0518-8

[66] J. G. Bazan, S. Bazan-Socha, S. Buregwa-Czuma, P. W. Pardel, and B. Sokołowska, "Predicting the Presence of Serious Coronary Artery Disease Based on 24 hour Holter ECG Monitoring," in *Federated Conference on Computer Science and Information Systems – FedCSIS 2012, Wrocław, Poland, 9-12 September 2012, Proceedings*,

M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., 2012, pp. 279–286. [Online]. Available: fedcsis.org/proceedings/2012/pliks/227.pdf

[67] Ł. Sosnowski and J. Wróblewski, "Toward Automatic Assessment of a Risk of Women's Health Disorders Based on Ontology Decision Models and Menstrual Cycle Analysis," in *2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, December 15-18, 2021*, Y. Chen, H. Ludwig, Y. Tu, U. M. Fayyad, X. Zhu, X. Hu, S. Byna, X. Liu, J. Zhang, S. Pan, V. Papalexakis, J. Wang, A. Cuzzocrea, and C. Ordonez, Eds. IEEE, 2021, pp. 5544–5552. [Online]. Available: doi.org/10.1109/BigData52589.2021.9671481

[68] A. Wojna and R. Latkowski, "Rseslib 3: Library of Rough Set and Machine Learning Methods with Extensible Architecture," *Transactions on Rough Sets*, vol. 21, pp. 301–323, 2019. [Online]. Available: doi.org/10.1007/978-3-662-58768-3_7

[69] M. J. Benítez-Caballero, J. Medina, E. Ramírez-Poussa, and D. Ślęzak, "Rough-set-driven Approach for Attribute Reduction in Fuzzy Formal Concept Analysis," *Fuzzy Sets and Systems*, vol. 391, pp. 117–138, 2020. [Online]. Available: doi.org/10.1016/j.fss.2019.11.009

[70] M. Wolski and A. Gomolińska, "Data Meaning and Knowledge Discovery: Semantical Aspects of Information Systems," *International Journal of Approximate Reasoning*, vol. 119, pp. 40–57, 2020. [Online]. Available: doi.org/10.1016/j.ijar.2020.01.002

[71] R. Belohlávek and V. Vychodil, "What is a Fuzzy Concept Lattice?," in *Proceedings of the CLA 2005 International Workshop on Concept Lattices and their Applications Olomouc, Czech Republic, September 7-9, 2005*, ser. CEUR Workshop Proceedings, R. Belohlávek and V. Snásel, Eds., vol. 162. CEUR-WS.org, 2005. [Online]. Available: ceur-ws.org/Vol-162/paper4.pdf

[72] P. Doherty and A. Szałas, "A Landscape and Implementation Framework for Probabilistic Rough Sets Using ProbLog," *Information Sciences*, vol. 593, pp. 546–576, 2022. [Online]. Available: doi.org/10.1016/j.ins.2021.12.062

[73] A. Grabowski, "Automated Comparative Study of Some Generalized Rough Approximations," *Fundamenta Informaticae*, vol. 179, no. 2, pp. 165–182, 2021. [Online]. Available: doi.org/10.3233/FI-2021-2019

[74] L. De Raedt, A. Kimmig, and H. Toivonen, "ProbLog: A Probabilistic Prolog and Its Application in Link Discovery," in *IJCAI 2007, Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad, India, January 6-12, 2007*, M. M. Veloso, Ed., 2007, pp. 2462–2467. [Online]. Available: ijcai.org/Proceedings/07/Papers/396.pdf

[75] A. Grabowski, A. Korniłowicz, and A. Naumowicz, "Four Decades of Mizar – Foreword," *Journal of Automated Reasoning*, vol. 55, no. 3, pp. 191–198, 2015. [Online]. Available: doi.org/10.1007/s10817-015-9345-1

[76] P. Pagliani and M. K. Chakraborty, *A Geometry of Approximation – Rough Set Theory: Logic, Algebra and Topology of Conceptual Patterns*, ser. Trends in Logic. Springer, 2008. [Online]. Available: doi.org/10.1007/978-1-4020-8622-9

[77] Y. Kusunoki, J. Błaszczyński, M. Inuiguchi, and R. Słowiński, "Empirical Risk Minimization for Dominance-based Rough Set Approaches," *Information Sciences*, vol. 567, pp. 395–417, 2021. [Online]. Available: doi.org/10.1016/j.ins.2021.02.043

[78] M. Palangetić, C. Cornelis, S. Greco, and R. Słowiński, "Granular Representation of OWA-based Fuzzy Rough Sets," *Fuzzy Sets and Systems*, vol. 440, pp. 112–130, 2022. [Online]. Available: doi.org/10.1016/j.fss.2021.04.018

[79] V. Vapnik, "Principles of Risk Minimization for Learning Theory," in *Advances in Neural Information Processing Systems 4, [NIPS Conference, Denver, Colorado, USA, December 2-5, 1991]*, J. E. Moody, S. J. Hanson, and R. Lippmann, Eds. Morgan Kaufmann, 1991, pp. 831–838. [Online]. Available: papers.nips.cc/paper/506-principles-of-risk-minimization-for-learning-theory

[80] L. A. Zadeh, "Toward a Theory of Fuzzy Information Granulation and its Centrality in Human Reasoning and Fuzzy Logic," *Fuzzy Sets and Systems*, vol. 90, no. 2, pp. 111–127, 1997. [Online]. Available: doi.org/10.1016/S0165-0114(97)00077-8

[81] U. Stańczyk and B. Zielosko, "Heuristic-based Feature Selection for Rough Set Approach," *International Journal of Approximate Reasoning*, vol. 125, pp. 187–202, 2020. [Online]. Available: doi.org/10.1016/j.ijar.2020.07.005

[82] M. Kopczyński and T. Grześ, "Hardware Rough Set Processor Parallel Architecture in FPGA for Finding Core in Big Datasets," *Journal of Artificial Intelligence and Soft Computing Research*, vol. 11, no. 2, pp. 99–110, 2021. [Online]. Available: doi.org/10.2478/jaiscr-2021-0007

[83] M. Garbulowski, K. Diamanti, K. Smolińska, N. Baltzer, P. Stoll, S. Bornelöv, A. Øhrn, L. Feuk, and J. Komorowski, "R.ROSETTA: An Interpretable Machine Learning Framework," *BMC Bioinformatics*, vol. 22, no. 1, p. 110, 2021. [Online]. Available: doi.org/10.1186/s12859-021-04049-z

[84] S. Stratmann, S. A. Yones, M. Garbulowski, J. Sun, A. Skaftason, M. Mayrhofer, N. Norgren, M. K. Herlin, C. Sundström, A. Eriksson, M. Höglund, J. Palle, J. Abrahamsson, K. Jahnukainen, M. C. Munthe-Kaas, B. Zeller, K. Pokrovskaja Tamm, L. Cavelier, J. Komorowski, and L. Holmfeldt, "Transcriptomic Analysis Reveals Proinflammatory Signatures Associated with Acute Myeloid Leukemia Progression," *Blood Advances*, vol. 6, no. 1, pp. 152–164, 2022. [Online]. Available: doi.org/10.1182/bloodadvances.2021004962

[85] D. Ślęzak, M. Grzegorowski, A. Janusz, and S. Stawicki, "Toward Interactive Attribute Selection with Infolattices," in *Rough Sets – International Joint Conference, IJCRS 2017, Olsztyn, Poland, July 3-7, 2017, Proceedings, Part II*, ser. Lecture Notes in Computer Science, L. Polkowski, Y. Yao, P. Artiemjew, D. Ciucci, D. Liu, D. Ślęzak, and B. Zielosko, Eds., vol. 10314. Springer, pp. 526–539. [Online]. Available: doi.org/10.1007/978-3-319-60840-2_38

[86] F. Sperrle, M. El-Assady, G. Guo, R. Borgo, D. H. Chau, A. Endert, and D. A. Keim, "A Survey of Human-centered Evaluations in Human-centered Machine Learning," *Computer Graphics Forum*, vol. 40, no. 3, pp. 543–567, 2021. [Online]. Available: doi.org/10.1111/cgf.14329

[87] P. P. Angelov and X. Gu, "Toward Anthropomorphic Machine Learning," *Computer*, vol. 51, no. 9, pp. 18–27, 2018. [Online]. Available: doi.org/10.1109/MC.2018.3620973

[88] S. M. Lundberg and S. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, and R. Garnett, Eds., 2017, pp. 4765–4774. [Online]. Available: proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

[89] K. Pancerz, "Rough Set Based Description of Plasmodium Propagation," *International Journal of Unconventional Computing*, vol. 15, no. 4, pp. 287–299, 2020. [Online]. Available: oldcitypublishing.com/journals/ijuc-home/ijuc-issue-contents/ijuc-volume-15-number-4-2020/ijuc-15-4-p-287-299/

[90] Ł. Pałkowski, M. Karolak, J. Błaszczyński, J. Krysiński, and R. Słowiński, "Structure-Activity Relationships of the Imidazolium Compounds as Antibacterials of Staphylococcus aureus and Pseudomonas Aeruginosa," *International Journal of Molecular Sciences*, vol. 22, no. 15, 2021. [Online]. Available: mdpi.com/1422-0067/22/15/7997

[91] M. Karolak, Ł. Pałkowski, B. Kubiak, J. Błaszczyński, R. Łunio, W. Sawicki, R. Słowiński, and J. Krysiński, "Application of Dominance-based Rough Set Approach for Optimization of Pellets Tableting Process," *Pharmaceutics*, vol. 12, no. 11, p. 1024, 2020. [Online]. Available: doi.org/10.3390/pharmaceutics12111024

[92] T. Y. Lin, "Neighborhood Systems: Mathematical Models of Information Granulations," in *Proceedings of the IEEE International Conference on Systems, Man & Cybernetics: Washington, D.C., USA, 5-8 October 2003*. IEEE, 2003, pp. 3188–3193. [Online]. Available: doi.org/10.1109/ICSMC.2003.1244381

[93] E. Frank, M. A. Hall, and I. H. Witten, "The WEKA Workbench," Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.

[94] W. R. Rudnicki, M. Kierczak, J. Koronacki, and H. J. Komorowski, "A Statistical Method for Determining Importance of Variables in an Information System," in *Rough Sets and Current Trends in Computing, 5th International Conference, RSCTC 2006, Kobe, Japan, November 6-8, 2006, Proceedings*, ser. Lecture Notes in Computer Science, S. Greco, Y. Hata, S. Hirano, M. Inuiguchi, S. Miyamoto, H. S. Nguyen, and R. Slowinski, Eds., vol. 4259. Springer, 2006, pp. 557–566. [Online]. Available: doi.org/10.1007/11908029_58

[95] P. G. Clark, C. Gao, J. W. Grzymała-Busse, T. Mroczek, and R. Niemiec, "Complexity of Rule Sets in Mining Incomplete Data Using Characteristic Sets and Generalized Maximal Consistent Blocks," *Logic Journal of the IGPL*, vol. 29, no. 2, pp. 124–137, 2021. [Online]. Available: doi.org/10.1093/jigpal/jzaa041

[96] B. Pękała, T. Mroczek, D. Gil, and M. Kępski, "Application of Fuzzy and Rough Logic to Posture Recognition in Fall Detection System," *Sensors*, vol. 22, no. 4, p. 1602, 2022. [Online]. Available: doi.org/10.3390/s22041602

[97] J. Błaszczyński, A. T. de Almeida Filho, A. Matuszyk, M. Szeląg, and R. Słowiński, "Auto Loan Fraud Detection Using Dominance-based Rough Set Approach versus Machine Learning Methods," *Expert Systems with Applications*, vol. 163, p. 113740, 2021. [Online]. Available: doi.org/10.1016/j.eswa.2020.113740

[98] M. Przybyła-Kasperek, "Coalitions' Weights in a Dispersed System with Pawlak Conflict Model," *Group Decision and Negotiation*, vol. 3, pp. 549–591, 2020. [Online]. Available: hdl.handle.net/20.500.12128/13891

[99] D. Ślęzak, M. Grzegorowski, A. Janusz, M. Kozielski, S. H. Nguyen, M. Sikora, S. Stawicki, and Ł. Wróbel, "A Framework for Learning and Embedding Multi-Sensor Forecasting Models into a Decision Support System: A Case Study of Methane Concentration in Coal Mines," *Information Sciences*, vol. 451-452, pp. 112–133, 2018. [Online]. Available: doi.org/10.1016/j.ins.2018.04.026

[100] A. Wakulicz-Deja and M. Przybyła-Kasperek, "Pawlak's Conflict Model: Directions of Development," in *Proceedings of the 2016 Federated Conference on Computer Science and Information Systems, FedCSIS 2016, Gdańsk, Poland, September 11-14, 2016*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., vol. 8. IEEE, 2016, pp. 191–197. [Online]. Available: doi.org/10.15439/2016F003

[101] Y. Zhang, D. Miao, W. Pedrycz, T. Zhao, J. Xu, and Y. Yu, "Granular Structure-based Incremental Updating for Multi-Label Classification," *Knowledge Based Systems*, vol. 189, 2020. [Online]. Available: doi.org/10.1016/j.knosys.2019.105066

[102] D. Brzeziński, J. Stefanowski, R. Susmaga, and I. Szczęch, "On the Dynamics of Classification Measures for Imbalanced and Streaming Data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 8, pp. 2868–2878, 2020. [Online]. Available: doi.org/10.1109/TNNLS.2019.2899061

[103] K. Pancerz, W. Paja, M. Wrzesień, and J. Warchoł, "Classification of Voice Signals through Mining Unique Episodes in Temporal Information Systems: A Rough Set Approach," in *Proceedings of the 21th International Workshop on Concurrency, Specification and Programming, Berlin, Germany, September 26-28, 2012*, ser. CEUR Workshop Proceedings, L. Popova-Zeugmann, Ed., vol. 928. CEUR-WS.org, 2012, pp. 280–291. [Online]. Available: ceur-ws.org/Vol-928/0280.pdf

[104] A. Skowron and P. Synak, "Reasoning in Information Maps," *Fundamenta Informaticae*, vol. 59, no. 2-3, pp. 241–259, 2004. [Online]. Available: content.iospress.com/articles/fundamenta-informaticae/fi59-2-3-10

[105] K. Ropiak and P. Artiemjew, "On a Hybridization of Deep Learning and Rough Set Based Granular Computing," *Algorithms*, vol. 13, no. 3, p. 63, 2020. [Online]. Available: doi.org/10.3390/a13030063

[106] P. Artiemjew and K. Ropiak, "A Novel Ensemble Model – The Random Granular Reflections," *Fundamenta Informaticae*, vol. 179, no. 2, pp. 183–203, 2021. [Online]. Available: doi.org/10.3233/FI-2021-2020

[107] G. Toppin, J. Borkowski, D. Ślęzak, S. Shi, P. Synak, J. Wróblewski, T. J. Wongkee, and G. Charalabopoulos, "System and Method for Granular Scalability in Analytical Data Processing," US Patent Application 20150088807, 2014.

[108] M. Przyborowski, T. Tajmajer, Ł. Grad, A. Janusz, P. Biczyk, and D. Ślęzak, "Toward Machine Learning on Granulated Data – a Case of Compact Autoencoder-based Representations of Satellite Images," in *IEEE International Conference on Big Data (IEEE BigData 2018), Seattle, WA, USA, December 10-13, 2018*, N. Abe, H. Liu, C. Pu, X. Hu, N. K. Ahmed, M. Qiao, Y. Song, D. Kossmann, B. Liu, K. Lee, J. Tang, J. He, and J. S. Saltz, Eds. IEEE, 2018, pp. 2657–2662. [Online]. Available: doi.org/10.1109/BigData.2018.8622562

[109] G. D. Tré, T. Boeckling, Y. Timmerman, and S. Zadrożny, "Handling Veracity of Nominal Data in Big Data: A Multipolar Approach," in *Flexible Query Answering Systems – 13th International Conference, FQAS 2019, Amantea, Italy, July 2-5, 2019, Proceedings*, ser. Lecture Notes in Computer Science, A. Cuzzocrea, S. Greco, H. L. Larsen, D. Saccà, T. Andreasen, and H. Christiansen, Eds.,

vol. 11529. Springer, 2019, pp. 317–328. [Online]. Available: doi.org/10.1007/978-3-030-27629-4_29

[110] R. S. Geiger, D. Cope, J. Ip, M. Lotosh, A. Shah, J. Weng, and R. Tang, ""Garbage In, Garbage Out" Revisited: What Do Machine Learning Application Papers Report about Human-Labeled Training Data?" *Quantitative Science Studies*, vol. 2, no. 3, pp. 795–827, 2021. [Online]. Available: doi.org/10.1162/qss_a_00144

[111] M. Kassen, *Open Data Governance and Its Actors – Theory and Practice*, ser. Studies in National Governance and Emerging Technologies. Palgrave Macmillan, 2022. [Online]. Available: doi.org/10.1007/978-3-030-92065-4

[112] D. Plotkin, *Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance, 2nd Edition*. Academic Press, 2020.

[113] A. Der Kiureghian and O. Ditlevsen, "Aleatory or Epistemic? Does It Matter?" *Structural Safety*, vol. 31, no. 2, pp. 105–112, 2009. [Online]. Available: sciencedirect.com/science/article/pii/S0167473008000556

[114] D. Ślęzak and A. Chądzyńska-Krasowska, "Approximate Decision Tree Induction over Approximately Engineered Data Features," in *Rough Sets – International Joint Conference, IJCRS 2020, Havana, Cuba, June 29 – July 3, 2020, Proceedings*, ser. Lecture Notes in Computer Science, R. Bello, D. Miao, R. Falcon, M. Nakata, A. Rosete, and D. Ciucci, Eds., vol. 12179. Springer, 2020, pp. 376–384. [Online]. Available: doi.org/10.1007/978-3-030-52705-1_28

[115] B. Settles, "Active Learning Literature Survey," University of Wisconsin–Madison, Computer Sciences Technical Report 1648, 2009.

[116] S. Dutta and A. Skowron, "Interactive Granular Computing Connecting Abstract and Physical Worlds: An Example," in *Proceedings of the 29th International Workshop on Concurrency, Specification and Programming (CS&P 2021), Berlin, Germany, September 27-28, 2021*, ser. CEUR Workshop Proceedings, H. Schlingloff and T. Vogel, Eds., vol. 2951. CEUR-WS.org, 2021, pp. 46–59. [Online]. Available: ceur-ws.org/Vol-2951/paper18.pdf

[117] L. A. Zadeh, Ed., *Computing with Words: Principal Concepts and Ideas*, ser. Studies in Fuzziness and Soft Computing. Springer, 2012, vol. 277. [Online]. Available: doi.org/10.1007/978-3-642-27473-2

[118] J. Pearl, "Causal Inference in Statistics: An Overview," *Statistics Surveys*, vol. 3, pp. 96–146, 2009. [Online]. Available: doi.org/10.1214/09-SS057

[119] A. Janusz, D. Kałuża, A. Chądzyńska-Krasowska, B. Konarski, J. Holland, and D. Ślęzak, "IEEE BigData 2019 Cup: Suspicious Network Event Recognition," in *2019 IEEE International Conference on Big Data (IEEE BigData), Los Angeles, CA, USA, December 9-12, 2019*, C. K. Baru, J. Huan, L. Khan, X. Hu, R. Ak, Y. Tian, R. S. Barga, C. Zaniolo, K. Lee, and Y. F. Ye, Eds. IEEE, 2019, pp. 5881–5887. [Online]. Available: doi.org/10.1109/BigData47090.2019.9005668

[120] M. Matraszek, A. Janusz, M. Świechowski, and D. Ślęzak, "Predicting victories in video games – IEEE bigdata 2021 cup report," in *2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, December 15-18, 2021*, Y. Chen, H. Ludwig, Y. Tu, U. M. Fayyad, X. Zhu, X. Hu, S. Byna, X. Liu, J. Zhang, S. Pan, V. Papalexakis, J. Wang, A. Cuzzocrea, and C. Ordonez, Eds. IEEE, 2021, pp. 5664–5671. [Online]. Available: doi.org/10.1109/BigData52589.2021.9671650

[121] F. P. Brooks, Jr., *The Mythical Man-Month: Essays on Software Engineering, Anniversary Edition*. Addison-Wesley, 1995.

[122] S. Dutta, A. Skowron, and M. K. Chakraborty, "Information Flow in Logic for Distributed Systems: Extending Graded Consequence," *Information Sciences*, vol. 491, pp. 232–250, 2019. [Online]. Available: doi.org/10.1016/j.ins.2019.03.057

[123] A. Skowron, A. Jankowski, and P. Wasilewski, "Interactive Computational Systems: Rough Granular Approach," in *Proceedings of the 21th International Workshop on Concurrency, Specification and Programming, Berlin, Germany, September 26-28, 2012*, ser. CEUR Workshop Proceedings, L. Popova-Zeugmann, Ed., vol. 928. CEUR-WS.org, 2012, pp. 358–369. [Online]. Available: ceur-ws.org/Vol-928/0358.pdf

[124] C. Savaglio, M. Ganzha, M. Paprzycki, C. Badica, M. Ivanovic, and G. Fortino, "Agent-based Internet of Things: State-of-the-Art and Research Challenges," *Future Generation Computer Systems*, vol. 102, pp. 1038–1053, 2020. [Online]. Available: doi.org/10.1016/j.future.2019.09.016