JAISCR, 2027, Vol. 17, No. 1, pp. 61– 84
10.2478/jaiscr-2026-0024

A NOVEL METHOD FOR IMPROVING DEEP LEARNING PERFORMANCE USING ROUGH SET THEORY

Piotr Pięta, Tomasz Szmuc*, Rafał Mrówka

Department of Computer Science and Artificial Intelligence, AGH University of Kraków, al. A. Mickiewicza 30, 30-059 Kraków, Poland

E-mail: tsz@agh.edu.pl

Submitted: 13th March 2026; Accepted: 19th May 2026

Abstract

Dimensionality reduction is a key stage in data processing. Eliminating irrelevant attributes from the input dataset enables more efficient processing and supports the development of robust Machine Learning models. In particular, when dealing with large datasets, attribute reduction algorithms can significantly reduce computational complexity while minimizing the loss of valuable information contained in the data. This paper explores the potential application of a Feature selection method based on Rough Set Theory for multivariate data. The experiments were conducted on the publicly available Taiwan Credit dataset. The results show that limiting the input feature space to a minimal set of attributes generated by a reduct selection algorithm not only reduces training and inference time for Machine Learning or Deep Learning models but may also improve their classification performance. The models achieved higher Accuracy and better adaptability to the input data, as confirmed by increased or comparable Precision and Recall scores. Additionally, the findings indicate that applying Feature selection using a reduct set can lead to more efficient data processing. These results suggest that integrating Rough Set Theory with Deep Learning models still represents a promising approach to analyzing attribute-value data.

Keywords: rough sets, deep learning, dimension reduction, minimal reduct

1 Introduction

Modern data analysis is characterized by the dynamic advancement of methods and techniques for processing various types of information [90]. Data typically appear in homogeneous formats with a different dimensionality, such as: images, audio, sequences of video frames and increasingly data point clouds [65]. Numerous algorithms and methods [4, 6, 15, 16, 30] enable the discovery of useful patterns hidden within the dataset, which can significantly impact the success of business initiatives and key investments [25]. In practical applications, which encompass a wide range of scenarios – from Image analysis [5, 82] and customer segmentation [66], Clustering [12] to Trend analysis on market [48, 81], Sentiment and social analysis [11, 39, 49, 54], Recommendation systems [41] and Natural Language Processing (NLP) [67], Machine Learning (ML) offers the most versatile solutions in dealing with them in an optimized way [5, 61, 64, 79]. As Big Data continues to grow [40], traditional statistical inference-based methods prove to be definitely insufficient, not only failing to cap-

ture critical dependencies but also lacking the efficiency needed to process vast amounts of data simultaneously.

Data mining [25] can be regarded as a multistage process, aimed at extracting valuable insights (hidden patterns), where the accuracy and validity of interpretations remain dependent on human expertise (domain knowladge). One of the most critical stage in data preprocessing is the reduction of dataset size, which enhances the efficiency and effectiveness of Machine Learning models by decreasing the number of input features. Several algorithms have been developed for such reduction. Principal Component Analysis (PCA) has gained the greatest recognition in the field and is now one of the most widely used approaches [35]. A notable and important method in this field is Rough Set Theory (RS). The innovative concept was introduced by Z. Pawlak [52, 76], who proposed a novel technique for information processing, particularly in cases characterized by uncertainty, imprecision, and vagueness [53]. While seemingly intuitive, addressing these issues in the practical application of this method has led to numerous studies, like [10, 18, 20, 21, 23, 55, 63, 72, 75, 92] and many solutions [1, 22, 28, 36, 43, 47, 50, 60, 69, 73, 77, 80, 83, 85, 88, 89] aimed at minimizing their impact on the Knowledge Discovery in Databases (KDD) process, especially in Decision systems. RS continues to be widely recognized in the scientific community, particularly in Japan, the USA, China, Canada and Poland.

Moreover, the availability of a wide range of programming tools facilitating the efficient application of RS [56] – such as LERS [19], RSES [27], ROSE2 [59], jMAF [8], ROSETTA [37], the R statistical computing package [14, 62], and WEKA [32] – has contributed to its growing popularity among researchers in data analysis, including computer scientists, data scientists, and those interested in the formal aspects of the theory. In the context of Deep Learning (DL) models, which are gaining increasing relevance in the analysis of tabular data, reducing the number of features in the space can lead to faster training times and overall performance improvements. However, despite its potential benefits, the application of RS in conjunction with basic DL models remains still an underexplored research area.

The objective of this study is to investigate the impact of dimensionality reduction based on RS on the classification performance of DL models applied to structured tabular data. The conducted experiments employed an exhaustive algorithm for proper conditional attribute selection. We compared the results of models trained on both original and reduced datasets. The analysis was performed on benchmark dataset. The findings indicate that, in the most cases, feature reduction contributed to improved classification performance. The experiments revealed not only a decrease in training and inference time but also an increase in the ROC AUC score, suggesting better model adaptation to input data. These results support the hypothesis that integrating RS wit DL models presents a promising approach for complex data analysis, warranting further research and development.

2 Deep Learning in tabular datasets processing

Deep Learning has revolutionized modern data analysis by introducing powerful computational models capable of extracting hierarchical patterns from data. These models are the successors of the well-known classical Neural Networks developed as early as the 1980s. However, unlike traditional Machine Learning methods, which often rely on manually engineered features, Deep Neural Networks (DNN) learn abstract, hierarchical representations directly from raw input data. They have achieved State-of-the-art performance across a wide range of domains, including image recognition, Natural Language Processing (NLP), time-series forecasting, etc. Conventional approaches, which required handcrafted features and careful adaptation to rapidly changing input conditions, often failed to deliver satisfactory results. Many complex problems are difficult to formulate using Rule-based algorithms. They have become the domain of DNN models capable of capturing intricate internal regularities and dependencies in data. In this regard, DNN have proven to be unmatched. However, the application of DNNs to multivariate data – such as tabular datasets and time-series – presents unique challenges and opportunities. In their original applications, DNN primarily processed image data, and later expanded to sequential data like au-

dio and more recently to point clouds and other 3D structures. There is no doubt that today computational resources are sufficient to process such data efficiently.

While effective in many domains, DL models face unique challenges when addressing tabular datasets or time-series. These challenges include issues like class imbalance, missing values, limited data availability and the demand for interpretability. Unlike unstructured data (e.g. images or audio), structured data often lack inherent spatial or sequential patterns, making it more difficult for deep architectures to automatically extract meaningful representations. As a result, the development of specialized architectures and preprocessing techniques becomes essential to harness the full potential of DL in these contexts.

2.1 Key DNN Architectures

Deep Learning models for complex data representations leverage architectures designed to capture a wide range of features as temporal dependencies, feature dependencies, and hierarchical representations. Some of the most relevant architectures used in such data processing include:

- Fully Connected Neural Networks (FCNNs): Simple and effective for structured data, can approximate any function given enough hidden units. They are particularly useful in classification and regression tasks where feature interactions are nontrivial. On the other hand, this kind of layers prone to overfitting and requires careful feature engineering. It can be characterized as lacks the ability to capture sequential dependencies. Additionally, FCNNs may struggle when dealing with high-dimensional data without proper regularization.

- Convolutional Neural Networks (CNNs) are efficient at capturing spatial and local patterns in sequences, parameter sharing reduces computational cost. These layers are useful for time-series classification and even anomaly detection. CNNs can extract local dependencies and hierarchical patterns, making them well-suited for structured feature learning. At the same time, CNNs are limited ability to model long-range dependencies. They may require large datasets for effective training. CNNs may also struggle with interpretability in certain applications, making it difficult to determine the relevance of extracted features.

- Recurrent Neural Networks (RNNs) are designed mainly for sequential data, they maintain an internal state that captures temporal dependencies. Especially, RNNs are useful for time-series forecasting and NLP modelling. RNNs allow for dynamic temporal modeling, which is essential for data with complex time dependencies. In contrast to many ML models, they suffer from vanishing gradient problems, challenging to train over long sequences and is computationally expensive. Due to their sequential nature, RNNs can be difficult to parallelize, leading to increased training time.

- Long Short-Term Memory (LSTM) Networks addresses vanishing gradient issues, retains long-term dependencies, widely used in sequential data processing. LSTM units contain gates that allow for better control over information flow, improving performance on long time-series. The main disadvantages of this kind of layers are that they have a high computational cost and requires significant training time. LSTMs may also suffer from high memory consumption, which limits their scalability in resource-constrained environments.

- Gated Recurrent Units (GRUs) can be considered to be more computationally efficient than LSTMs, effective in capturing dependencies in sequential data. They require fewer parameters. GRUs simplify the gating mechanisms found in LSTMs while maintaining comparable performance. However, may not perform as well as LSTM on tasks that require longer memory retention. GRUs also lack the explicit memory cell structure of LSTMs, which can be a disadvantage for modeling very long sequences.

- Hybrid Architectures (e.g. CNN-RNN, Transformer-based models) leverage the strengths of multiple architectures and improve performance on complex tasks. They are useful in many domains, not only in financial forecasting and biomedical signal analysis. Hybrid models allow for both spatial and temporal feature extraction, leading to robust representations

of sequential data. The increase in complexity is the main issue of such complex architectures. Sometimes, they require more computational resources and are difficult to interpret by humans. Training hybrid models may also involve more hyperparameter tuning and computational overhead.

In many applications, ML models such as Gradient Boosted Decision Trees (e.g. LightGBM, CatBoost, XGBoost) still outperform DL approaches in both accuracy and efficiency, especially for small- to medium-sized datasets. However, architectures like TabNet [2] and FTTransformer [78] attempt to bridge this gap by combining interpretability and automatic feature learning.

2.2 Challenges in applying Deep Learning

Despite its success in various fields, DL faces several challenges when applied to tabular and time-series data. Some of them was listed below and more described in [7, 17]

- High Dimensionality: Many real-world datasets contain numerous irrelevant or redundant features, which can lead to increased computational complexity and overfitting. Regularization techniques such as dropout and weight decay help mitigate this issue. Additionally, feature selection methods such as RS can enhance model performance by removing unnecessary attributes.
- Data Scarcity: Unlike image or text datasets, where large labeled corpora are available, tabular datasets often have limited labeled instances, making training DNN challenging. Semi-supervised learning and data augmentation techniques can be useful in addressing this problem. Transfer learning and self-supervised approaches also show promise in leveraging knowledge from related tasks.
- Feature Importance Interpretation: DL models are often criticized for their lack of interpretability. Understanding which features contribute most to model decisions is crucial for applications in finance, healthcare, and other high-stakes domains. Methods such as SHAP values, LIME provide insights into feature importance. Additionally, explainability techniques such as attention mechanisms and surrogate models offer avenues for improving interpretability in DL applications

The incorporation of DL models for complex data analysis presents both challenges and opportunities. While architectures such as FCNNs, CNNs, RNNs, and their hybrid models offer powerful solutions, their effectiveness is highly dependent on the quality of input features. The application of RS for feature selection addresses key limitations by enhancing model efficiency, improving classification accuracy and reducing training complexity. Moreover, incorporating advanced techniques such as attention mechanisms, data augmentation, and interpretability tools further refines the performance of DL models on datasets. The results of this study indicate that leveraging RS in conjunction with DL constitutes a promising approach to analyzing tabular datasets and time-series, warranting further exploration and refinement. Future research should focus on the integration of DL models with hybrid feature selection techniques, as well as the development of novel architectures tailored for data applications.

2.3 Integration of Rough Sets with Deep Learning for feature selection

Feature selection techniques, such as those based on RS, play a crucial role in mitigating the aforementioned challenges. By eliminating redundant and irrelevant features, RS can:

- Improve model generalization by reducing overfitting
- Decrease training and inference time by lowering computational complexity
- Enhance interpretability by highlighting the most significant attributes

Integrating RS with DL models provides a structured approach to handling attribute-value data efficiently. The experiments carried out in this study demonstrate that reducing the number of the features before training DNN leads to better classification performance and higher ROC AUC scores. Using RS, models can focus on the most informative features, leading to improved decision-making processes and robustness against noise.

3 Fundamentals of Rough Set Theory

Rough Set Theory was introduced in the early 1980s by Z. Pawlak. The theory is one of the possible generalizations of classical set theory, alongside other well-known approaches such as fuzzy set theory (which expresses element membership through a characteristic function) and L-fuzzy set theory (which is based on a lattice algebraic structure). Despite the passage of time, RS remains a highly relevant and widely explored research area, particularly among scientists and researchers from China, Japan, and the United States. Its ongoing popularity and significance are reflected in international conferences dedicated to the topic, as well as an extensive body of literature, including prestigious publication series such as Transactions on Rough Sets (Springer).

The contributions of numerous researchers – both in terms of theoretical-mathematical foundations and practical applications – have led to the development of various tools that facilitate the use of RS in diverse domains, particularly within artificial intelligence applications. The continuous publication of research articles and real-world applications demonstrates that, while RS is a well-established approach, it continues to evolve in both theoretical advancements and practical implementations.

This chapter provides a concise introduction to the fundamental concepts of RS. A more detailed and rigorous treatment can be found in the seminal monograph [52].

3.1 Information System

In the context of RS, an information system represents the knowledge of an agent, which is interpreted as the ability to distinguish (classify) objects in the real world. From a mathematical perspective, classification is described using equivalence relations (or families of such relations). The finer the classification (i.e., the greater the distinguishability of objects and the smaller the equivalence classes), the more knowledge the agent possesses about its environment.

Let U be a non-empty, finite set of objects, referred to as the *universe of discourse*, generated through observations. The set U thus includes all entities, objects, persons, and events mentioned in the previous section, upon which we build our knowledge. For each object, we can specify a set of characteristics, referred to as attributes. The non-empty, finite set of all possible attributes for the objects in the universe is denoted by A, with individual attributes represented by lowercase letters such as a_1. Given these sets, we define:

$$a : U \to V_a \tag{1}$$

where V_a represents the set of possible values for a given attribute. More precisely,

$$A = \bigcup_{a \in A} V_a. \tag{2}$$

A pair $IS = (U, A)$ defines an *information system*. It is most commonly represented in the form of a *decision table*, where rows correspond to examined entities and columns represent attributes. The cells at the intersection of rows and columns contain the specific attribute values assigned to objects in the domain.

3.2 Decision Table

In practical applications of RS, the most common data representation is the *decision table*. It extends the information system (Section 3.1) by incorporating an additional attribute known as the *decision attribute*. The decision attribute is crucial as it enables the precise assignment of specific observations (e.g. measurements, individuals) to predefined classes. However, this assignment may sometimes lead to inconsistencies among objects within the domain. Let

$$\text{TabDec} = (U,\ A \cup \{d\})$$

be a decision table. According to [52], the decision attribute is not strictly considered an element of the attribute set A. The tabular representation allows for distinguishing between *conditional attributes* and the *decision attribute* (conclusion), which does not belong to A. This distinction justifies the use of the union notation $A \cup \{d\}$ in the formula for TabDec.

As previously noted, the data processed in decision tables often exhibit redundancy, incompleteness, or inconsistency. Some attributes may be unnecessary, objects and attribute values may be repeated, and inconsistencies may arise when objects

with identical conditional attribute values are assigned to different decision classes.

3.3 Indiscernibility Relation

The objects forming the universe of discourse U possess attributes that differentiate them from or group them with other objects. In large datasets, such as medical records from hospital patients, it is expected that certain symptoms will consistently characterize specific diseases. However, rather than providing an excessively detailed description of reality, it is often preferable to focus on the discrimination of subtle details.

The *indiscernibility relation* to groups of objects that share similar attribute values. Let

$$IS = (U, A)$$

be an information system. Given this system, we define a subset of attributes $A' \subseteq A$ and formalize the indiscernibility relation as follows:

$$\mathrm{IND}_{IS}(A') = \{(x,y) \in U \times U \mid \forall a \in A' : a(x) = a(y)\}$$

From an algebraic perspective, this relation is an *equivalence relation* since it is symmetric, reflexive, and transitive [52].

Objects that are indiscernible in U share identical values for the attributes under consideration. As an equivalence relation, indiscernibility partitions the universe U into disjoint subsets known as *equivalence classes*. The equivalence class of an object x under $\mathrm{IND}_{IS}(A')$ is denoted as:

$$[x]_{A'}$$

Each equivalence class groups objects exhibiting similarity with respect to attributes in A', meaning $x \sim y$ under $\mathrm{IND}_{IS}(A')$ if and only if their attribute values match. Z. Pawlak linked the concept of indiscernibility to the ability to classify objects, thereby defining knowledge within the framework of RS. The indiscernibility relation partitions the universe into a family of disjoint subsets – *abstraction classes* – which serve as the fundamental elements of an agent's knowledge. Equivalent terms in this context include concepts, categories, and notions. It is important to note that the universe U is not characterized by a single indiscernibility relation but by a family of such relations, each defining equivalence classes according to different sets of attributes. This family of indiscernibility relations, along with the universe generated by the objects, constitutes what Pawlak termed the *knowledge base* of an agent [52].

3.4 Reduct

Data sets often exhibit redundancy due to the presence of attributes or instances that are excessive and unnecessary for further analysis. As noted in [52], a specific form of data reduction is an identification of structures such as equivalence classes (where a single element from the set U is sufficient to represent the entire equivalence class to which it belongs). There is a need to extract from the dataset only those attributes from the dataset that are essential while ensuring that classification based on the reduced structure does not deteriorate. Additionally, the approximation of the set must be preserved. To achieve this, the concept of a *reduct* is introduced.

A *reduct* of an information system IS is defined as a *minimal subset* B of the attribute set A of the system, such that the following property holds regarding the inclusion of attribute sets:

$$B \subseteq A \tag{3}$$

The indiscernibility relation is preserved, i.e.,

$$IND_{IS}(A) = IND_{IS}(B) \tag{4}$$

Finding the minimal reduct set belongs to the class of *NP-hard* problems and is considered a *bottleneck* [53] in this formalism. For an information system IS defined using r attributes, the number of potential reducts is given by:

$$\binom{r}{\lfloor r/2 \rfloor} \tag{5}$$

Reducts allow for significant simplification of knowledge by restricting the set of attributes to only those necessary to maintain the approximation of the set [52]. To express the concept of a reduct more precisely, let R denote the *family of equivalence relations constructed over the space* U (this consideration is made in the context of a relational system,

[1] The term "dispensable relations" is also used.

[2] The term "indispensable relations" is also used.

which constitutes a knowledge base). It is easy to see that for such a defined system, some relations are *removable* [1] while others are *non-removable* [2].

A relation $B \in R$ is **removable** in R if the following holds:

$$IND_{IS}(R) = IND_{IS}(R - \{B\}) \tag{6}$$

Similarly, a relation B is considered **non-removable** in R if:

$$IND_{IS}(R) \neq IND_{IS}(R - \{B\}) \tag{7}$$

In this case, if any relation T is considered non-removable, it can be added that for every non-removable $B \in R$, the family R is **independent**.

These mathematical observations allow us to define, *de facto*, the *reduct* R [3] as a set of certain *independent relations* (possibly all of them) $B' \subseteq R$, where:

$$IND_{IS}(R) = IND_{IS}(B') \tag{8}$$

For a given system IS, there may exist more than one reduct. Most often, the goal is to obtain a minimal reduct (in terms of size), meaning it consists only of the necessary attributes.

3.5 Core

The *core* $CORE(R)$ can be expressed using the previously introduced relationships as the intersection of the considered reducts:

$$CORE(R) = \bigcap_{B \in RED(R)} B \tag{9}$$

It is the set of all *indispensable* relations in R.

3.6 Relative Reduct and Core

The definitions introduced in the previous sections essentially concern indiscernibility in the information system IS. To extend them to the indiscernibility of objects concerning decision values assigned to individual objects within a class, the concepts of *relative reduct* and *relative core* are introduced in [52].

Let P and Q be equivalence relations constructed over the space of consideration generated by the available objects U. Then, the knowledge contained in one classification can describe the knowledge contained in the other.

The *P-positive region of Q* is the set of all objects $x \in U$ that can be correctly classified into category Q using the partition resulting from the knowledge of P[4]:

$$POS_P(Q) = \bigcup_{X \in U/Q} \underline{P}X \tag{10}$$

If we refer to P and Q as *families of equivalence relations* built over a finite space of consideration U, then $R \in P$ is *Q-dispensable* in P if the following condition holds:

$$POS_{IND(P)}(IND(Q)) = POS_{IND(P-\{R\})}(IND(Q)) \tag{11}$$

Otherwise, R is considered *Q-indispensable*. If every R in P is *Q-indispensable*, then P is *Q-independent*.

In turn, this leads to the definition of the *Q-reduct P* (relative reduct) as a *family S*, where $S \subseteq P$, if S is a *Q-independent subfamily* of P and the following condition is satisfied:

$$POS_S(Q) = POS_P(Q) \tag{12}$$

The *Q-core for P*[5] (relative core) is defined as the set of all *Q-indispensable* elementary relations in P.

3.7 Discernibility Matrix

In problems related to knowledge reduction and the determination of reduct sets, the concept of *discernibility matrix* is often encountered. It is defined as follows:

$$c_{ij} = \left\{ a \in A \mid a(x_i) \neq a(x_j) \right\}, \tag{13}$$

where n is the number of objects, and the indices are in the range $i, j = 1..n$. The matrix is *symmetric*, and each cell stores the attributes for given objects x_i and x_j that allow these objects to be distinguished from each other.

The concept of the discernibility matrix can be extended to a *k-relative discernibility matrix*, in which the cells of such a structure contain only

[3] Denoted as $RED(R)$

[4] Classification U/P

[5] Denoted as $CORE_Q(P)$

those attributes that allow distinguishing the object x_i from x_j in the case where they belong to different decision classes (equivalence classes with respect to the k-th equivalence class) [52].

From our perspective, the discernibility matrix is one of the most interesting research objects. The process of constructing the discernibility matrix is time-consuming and serves as the foundation for many algorithms based on RS. Consequently, the input matrix frequently needs to be simplified before the reduct can be derived [87]. The information produced by this matrix facilitates a deeper understanding of the interdependencies between attributes and additionally offers statistical descriptors that can support the assessment of attribute significance, providing a complementary perspective to standard feature-importance measures.

3.8 Discernibility Function

The *discernibility function* is a monotonic *Boolean function* defined as follows:

$$f_{IS}(a_1^*,\ldots,a_m^*) = \bigwedge \left\{ \bigvee c_{ij}^* \;\middle|\; 1 \le j \le i \le n,\ c_{ij} \neq \emptyset \right\}, \tag{14}$$

where $c_{ij} = \left\{ a^* \mid a \in c_{ij} \right\}$. This function[6] is constructed from the *conjunction* of subsequent columns of the discernibility matrix (each cell of the discernibility matrix consists of *disjunctions* of attributes that distinguish these objects, while each row of the discernibility function corresponds to a column in the discernibility matrix). Applying the well-known logical *absorption law*, this function can be significantly simplified. Similarly to the discernibility matrix [52], this concept can be extended to selected *k-columns* of the discernibility matrix (for a given decision class k, forming the *k-relative discernibility function*.

3.9 Feature Stability and Co-occurrence Network Analysis

To assess the significance and robustness of representative conditional attributes, the RS framework based on the *Discernibility matrix* and derived *reducts* was applied.

For a universe of objects U described by conditional attributes $A = \{a_1,\ldots,a_N\}$ and a decision attribute d, the discernibility matrix is defined as:

$$D_{ij} = \{\, a \in A \mid a(x_i) \neq a(x_j) \,\}, \quad \text{for all } d(x_i) \neq d(x_j).$$

Each entry D_{ij} lists the set of attributes that differentiate two instances belonging to the distinct decision classes. Minimal subsets of attributes that preserve the same discernibility power are identified as *reducts*.

Reduct stability. To evaluate stability across the $K = 5$ cross-validation folds, overlap-based metrics were computed between each pair of reducts. The *Jaccard Index* measures direct subset agreement:

$$J(R_i,R_j) = \frac{|R_i \cap R_j|}{|R_i \cup R_j|},$$

while the *Kuncheva Index* (KI) accounts for chance-level overlap [68]:

$$KI = \frac{rN - k^2}{k(N-k)},$$

where N is the total number of attributes, $k = |R_i| = |R_j|$ is the subset size, and $r = |R_i \cap R_j|$ denotes intersection size. Values of $KI > 0.5$ indicate strong consistency beyond random coincidence. The mean Jaccard and Kuncheva indices across folds quantify global stability of the selection process.

Feature co-occurrence graph. A co-occurrence matrix C was constructed to capture joint participation of attributes in reducts:

$$C_{pq} = \left|\{R \in \mathcal{R} : p \in R \wedge q \in R\}\right|,$$

This symmetric matrix was normalized to form a weighted, undirected graph $G = (V,E,w)$, where nodes represent attributes and edge weights correspond to normalized co-occurrence frequencies.

Network-level [38, 45, 84] descriptors were then computed to reveal interaction patterns:

– *Degree centrality* $C_D(v) = \frac{\sum_u A_{vu}}{N-1}$ – measures direct connectivity. The total number of attributes is denoted by $N = |V|$. A_{vu} represents the weight of the edge between nodes. v and u, and the summation $\sum_u A_{vu}$ corresponds to the weighted degree of node v.

[6] The set of all prime implicants forming the discernibility function determines the identification of all reducts in the system in which it is defined.

– *Betweenness centrality* $C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$ – identifies bridging attributes. In this formula, σ_{st} denotes the total number of shortest paths between nodes s and t, while $\sigma_{st}(v)$ can be described as the number of those paths that pass through node v. Shortest paths are computed using edge distances defined as the inverse of co-occurrence weights.

– *Eigenvector centrality* $Ax = \lambda x$ – captures global influence via connected important nodes. In this measure, x is the eigenvector associated with the largest eigenvalue λ of the adjacency matrix A, and the v-th component of x quantifies the relative influence of node v in the network.

– *Weighted clustering coefficient* – quantifies the tendency of attributes to form tightly connected groups, accounting for co-occurrence strength. The coefficient was computed using a weighted formulation implemented in NetworkX [24].

– *Graph density* $= \frac{2|E|}{N(N-1)}$ – reflects overall interdependence.

– *Modularity*

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j),$$

indicating the strength of community structure (e.g., behavioral, financial, demographic clusters). In this Modularity formula, k_i and k_j denote the degrees of nodes i and j, respectively, $m = |E|$ is the total number of edges in the graph, and $\delta(c_i, c_j)$ is the Kronecker delta equal to 1 if nodes i and j belong to the same community, and 0 otherwise.

Feature frequency and interpretation. The frequency of each attribute across folds was computed as:

$$f(a) = \frac{1}{K} \sum_{i=1}^{K} \mathbf{1}_{a \in R_i}.$$

In this formula, a denotes a single attribute, K is the total number of folds, R_i represents the reduct obtained in the i-th fold, and $\mathbf{1}_{a \in R_i}$ is the indicator function equal to 1 if attribute a is included in R_i, and 0 otherwise.

The stability and network measures constitute the basis for the interpretation of selected attributes in reducts; this interpretation is presented in the subsequent sections. In general, attributes with both high $f(a)$ and high centrality values can be interpreted as *core features*, exhibiting stable inclusion and strong relational influence.

Consequently, high Jaccard and Kuncheva values indicated reproducible reducts; high eigenvector and degree centralities identified central predictors (e.g. payment history variables); whereas high betweenness scores corresponded to bridging attributes (e.g. credit limit) linking distinct conceptual groups. Moderate density and positive modularity further confirmed non-random, structured dependencies among conditional features.

4 Proposed method of feature selection using RS

This chapter outlines the foundations of our method. In the context of dimensionality reduction, there are various approaches to identifying the most representative and effective features. Some of these methods incorporate concepts from classical RS. One such approach is the exhaustive method which allows for the identification of all possible reducts. While this approach has several advantages including providing a complete set of potential reducts, it is computationally expensive. The first subsection presents several examples of such methods, while the second focuses on our proposed approach.

4.1 Attribute Reduction Algorithms

Knowledge reduction plays a crucial role in data exploration, particularly in terms of usability and efficiency in data analysis. Reduction is performed by identifying minimal subsets of attributes, known as reducts, which preserve the classification structure derived from indiscernibility relations (approximations) and additionally do not degrade classification accuracy after reduction. Many tools implementing these methods also include additional functionalities, such as the identification of core attributes. By applying these approaches, the volume of processed data can be significantly reduced while maintaining the quality of information.

Common Attribute Reduction Algorithms:

– Exhaustive Algorithm [74] – computes all mini-

mal reducts by searching the entire space of possible solutions

- Genetic Algorithm [42, 71] – inspired by natural evolution and optimization techniques
- Dynamic Reduct Algorithm [29] – identifies reducts with the highest stability, meaning they remain reducts across subtables of the decision table
- Lattice-Based Reduction Algorithm [13, 86] – utilizes lattice theory for attribute reduction
- Heuristic Search-Based Reduction Algorithm [52, 91] – uses heuristics to efficiently explore the search space.
- Johnson's Algorithm [34, 46]– a greedy algorithm that selects one frequent reduct (but does not guarantee the optimal solution)
- Holte's 1R Algorithm [26] – generates one reduct per attribute in the reduct set.
- QuickReduct Algorithm [33, 70]– a dependency function-based reduction method. It searches for a single most significant attribute subset without exhaustive search, starting from an empty subset.

It is worth noting that over the past few years, new approaches to generating sets of reducts have been developed. Despite the passage of time, there is still ongoing interest in classical methods, such as the exhaustive algorithm for reduct generation [44].

4.2 Our methodology

In this work, we focus on the first approach, namely the exhaustive algorithm (implemented in RSES) which searches the whole space of possible solutions to find all minimal reducts. This feature is incorporated into the proposed concept. The primary goal of our research was to determine whether reduction, in accordance with the RS approach, would generally improve the performance of simple ML/DL models in terms of model Accuracy, Balanced Accuracy, Training time, and Testing time, while maintaining an acceptable level of fit to existing data. The second goal is to evaluate stability of generated reducts. In this case, we used well-known framework for a complex network analysis (NetworkX [24]). In addition, we determine attribute significance coefficients and use statistical information obtained from the generated discriminability matrix for comparison purposes.

The Figure 1 outlines the steps in our approach. This method involves constructing a discernibility matrix, generating all subsets of attributes, and identifying those that preserve the partition of objects resulting from the indiscernibility relation. Although the algorithm guarantees an optimal solution, its computational complexity grows exponentially, limiting its applicability to large datasets. However, it enables a precise analysis and elimination of redundant attributes, which is crucial for data exploration and dimensionality reduction. In our methodology, we compute a set of irreducible attributes for the input dataset. Then, we limit the input dataset to these features. In the subsequent phase, we perform training and testing of selected ML models (e.g. Logistic Regression) with special focus on standard DNN architectures such as TabNet and a three-layer multilayer perceptron (MLP). The data are preprocessed and adapted to the appropriate formats required by these layers. During Training and Testing phase, we evaluate the models using various classification metrics such as Accuracy, Precision, Recall, as well as Training and Inference times for each model. Additionally, leveraging the knowledge from the indiscernibility matrix about the frequency of individual features, we determine the significance of the attributes for each training set using several importance measures commonly used in Information Theory, such as Information Gain and Entropy. We also consider other techniques based on models like SHAP, XGBoost or even Random Forest. Subsequently, for comparison purposes, we compare the obtained information with additional insights derived from the indiscernibility matrix. Throughout the research, we conduct a comprehensive evaluation of the results produced by the models. We make an effort to include a variety of layer types, including some less commonly used ones offered by the Keras library [9] and PyTorch [51]. We evaluate the performance of several ML/DL models when trained and tested on both full datasets (including all original features) and reduced datasets.

To minimize the impact of overfitting on the trained models, additional protective mechanisms

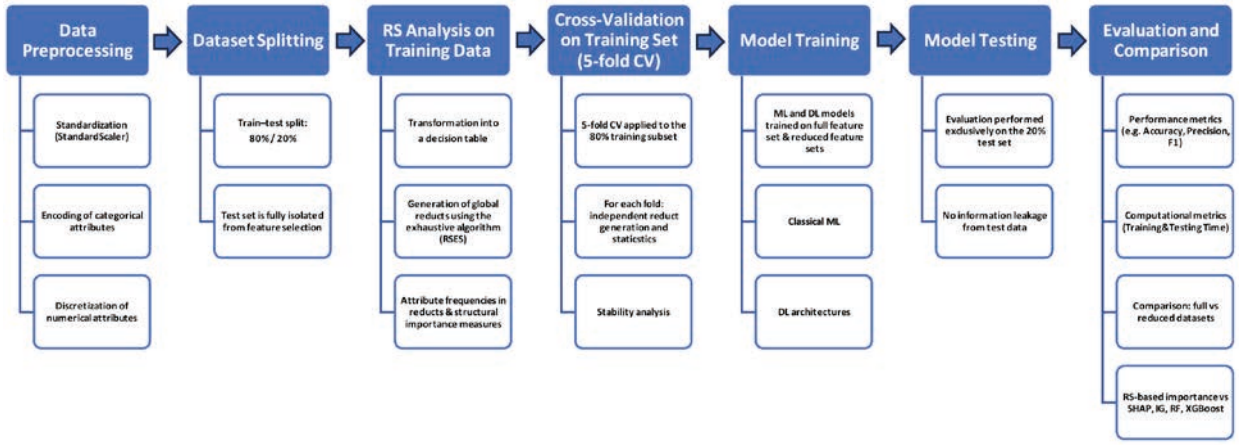

Figure 1. Workflow of the proposed feature selection and evaluation framework. The input data are first preprocessed by standardization, encoding, and discretization. The dataset is then split into training (80%) and test (20%) subsets. All RS–based operations, including the construction of the decision table, DM and exhaustive generation of reducts are performed exclusively on the training data. Global and local (fold-specific) reducts are computed within a K-fold cross-validation framework to assess feature stability. Reduced feature subsets are subsequently used to train selected ML and DL models. Final evaluation is conducted on the held-out test set using standard classification and computational performance metrics.

are implemented, such as Early stopping and appropriate regularization of selected layers in some models. The overall analysis of the results shows that attribute reduction improves the performance of ML/DL models. In particular, using the full dataset was less effective in scenarios where at least one attribute – identified as more significant according to the applied measures – is already included in the individual reducts.

Lastly developed bireduct concept [31] may be also applied in the redundancy reduction. Bireduct extends the classical notion of reduct by ensuring not only the preservation of the dataset classification ability but also the elimination of unnecessary dependence on specific objects. A bireduct is a subset of attributes which, when considered together with a corresponding subset of objects, maintain the discernibility between objects while reducing redundancy in both dimensions. This means that bireducts strike a balance between attribute reduction and object selection, making them particularly useful in scenarios where the dataset contains noise or irrelevant information [31]. The process of computing bireducts is inherently more complex than that of finding classical reducts, as it requires simultaneous exploration of both attribute and object subsets. Due to the dual constraints on attributes and objects, the computational burden increases significantly. By contrast, the exhaustive algorithm for computing classical reducts is considerably more efficient and scalable, as it focuses solely only on subsets of attributes. This lower computational complexity makes the reduct-based approach a more practical option in many real-world applications, especially when object-level noise is minimal or can be addressed in earlier stages of data preprocessing. Moreover, the minimal reducts benefit from a strong theoretical foundation and offer high interpretability, which further supports their widespread use. Despite its greater complexity, the bireduct approach offers important advantages in similar tasks where both irrelevant attributes and outlier objects need to be filtered out. The classical exhaustive algorithm for computing reducts remains a fundamental benchmark due to its lower computational cost and solid theoretical grounding.

5 Experiments

This chapter presents the key findings from experiments evaluating the effectiveness of an approach based on RS in enhancing the performance of selected DL models for tabular data processing. The created models were implemented in Python. Reducts generations were conducted using RSES analystical tool. The study investigates the impact

of RS on feature selection and model optimization aiming to improve both overall computational efficiency and Accuracy metrics for each case. The results provide insights into the practical applicability of this methodology in various scenarios.

5.1 Data preparation

5.1.1 Dataset

The dataset used in this part of the study is the **Taiwan Credit Card Default dataset**, a widely recognized benchmark for evaluating predictive models in credit risk analysis and financial decision-making. Originally released by the Taiwan National University and made publicly available on the *UCI Machine Learning Repository*, this dataset has been used extensively in research on credit scoring, financial risk prediction, and customer behavior modeling.

It contains demographic and financial features of credit card holders, along with a binary target variable indicating whether a client defaulted on payment in the following month. This dataset provides an excellent opportunity to evaluate Feature selection methods and classification models in a high-dimensional, imbalanced, and real-world financial context.

5.1.2 Dataset Attributes

The dataset consists of 24 attributes – including both numerical and categorical variables. A brief description is provided below:

- **ID**: Unique identifier for each client.
- **LIMIT_BAL**: Amount of the given credit (in NT dollars).
- **SEX**: Gender (1 = male, 2 = female).
- **EDUCATION**: Level of education (1 = graduate school, 2 = university, 3 = high school, 4 = others).
- **MARRIAGE**: Marital status (1 = married, 2 = single, 3 = others).
- **AGE**: Age of the client (years).
- **PAY_0 – PAY_6**: Repayment status for the past six months (-1 = pay duly, 1 = payment delay by one month, etc.).
- **BILL_AMT1 – BILL_AMT6**: Amount of bill statement (in NT dollars) for each of the past six months.
- **PAY_AMT1 – PAY_AMT6**: Amount of previous payments (in NT dollars).
- **default payment next month**: Target variable (1 = default, 0 = no default).

5.1.3 Sample Data

Table 1 presents a representative sample of the dataset.

This dataset was used to evaluate correlations between financial, demographic, and behavioral variables and the probability of credit default.

5.1.4 Dataset Characteristics

Table 2 summarizes key statistical characteristics of the dataset.

Table 2. Statistical characteristics of the Taiwan Credit dataset

Characteristic	Value
Samples	30,000
Features	23 (excluding ID)
Positive Class (Default)	22.1%
Negative Class (No Default)	77.9%
Age Range (years)	21–79
Missing Values	0%
Feature Types	6 continuous, 8 categorical, 9 ordinal

The dataset exhibits a class imbalance typical of financial default data, which makes it an excellent test case for evaluating model robustness and fairness.

5.1.5 Preprocessing Pipeline

Table 3 summarizes the preprocessing steps applied to the dataset, along with their rationale.

All preprocessing steps were performed using the same pipeline for each feature selection approach.

The dataset is publicly available via the *UCI Machine Learning Repository*. Full preprocessing and modeling code is available upon request

Table 1. Representative sample of the Taiwan Credit Card Default dataset.

ID	LIMIT_BAL	SEX	EDU	MARR	AGE	PAY_0	BILL_AMT1	PAY_AMT1	Default
1	20000	2	2	1	24	2	3913	0	1
2	120000	2	2	2	26	-1	2682	1000	1
3	90000	2	2	2	34	0	29239	1500	0
4	50000	1	2	1	37	0	46990	2000	0

Table 3. Preprocessing steps and rationale for the Taiwan Credit Card Default dataset. To ensure the reproducibility of the results, each part of the dataset has been saved as a separate file. Some preprocessing phases were conducted using RSES programming tool based on RS.

Step	Technique	Rationale
Discretization	Global method [3] (implemented in RSES)	Computationally slower than the local approach it often produces fewer cut points for LIMIT_BAL, AGE, and payment history.
Normalization	StandardScaler Standarization	Ensured comparable feature ranges for neural and distance-based models.
Feature Selection (our approach)	Rough Set–based reducts	Reduced redundancy and improved interpretability.

from the corresponding author (see Data Availability Statement).

Table 4 summarizes the empirical feature importance ranking for the *Taiwan Credit Card Default* dataset. The results combine measures derived from RS, Information Theory, and Ensemble-based models (like Random Forest, XGBoost).

Repayment history attributes (`PAY_0`–`PAY_6`) show the strongest discriminative power across all metrics. In particular, `PAY_0` (the most recent repayment status) achieves the highest SHAP value (0.54) Information Gain (0.78) and XGBoost importance (0.43) among others features, confirming its dominant influence on the prediction of default behavior.

Financial features such as `LIMIT_BAL`, `BILL_AMT1-2`, and `PAY_AMT1-2` also exhibit meaningful but secondary contributions, capturing credit capacity and repayment magnitude. Demographic features (e.g., `AGE`) have minimal Information Gain and low ensemble importances, indicating a limited predictive role.

Overall, the ranking confirms that payment delay indicators provide the most informative description of client creditworthiness, while RS reduct frequencies emphasize their frequent occurrence among minimal attribute subsets.

The training data was transformed into a decision table, which is required by the exhaustive algorithm to generate the set of reducts. In particular, precise calculations were made regarding the frequency of occurrence of specific attributes in the rows of an indiscernibility matrix (DM), and the overall frequency of these attributes in individual reducts. These frequencies were then presented on in the appropriate forms in the tables. During training and testing, the models were analyzed by observing various metrics specific to the classification task. Additionally, the importance of attributes for each training set was determined using various feature importance measures employed in Information Theory, as well as measures based on XGBoost and Random Forest. Subsequently, for comparison purposes, the obtained information was compared with additional data contained in the Discernibility matrix.

5.2 Selected results

In this experiment, the attribute importance analysis was performed using a fixed 80% subset of the *Taiwan Credit Card Default* dataset (outer_train). The same input subset was employed across all evaluated methods, ensuring full comparability of the resulting importance measures. Table 4 presents all computed importance scores for attributes. The applied techniques can be divided into two methodological categories. The first group comprises measures derived from RS theory, namely *DM Obj*, *DM Dec*, and the attribute fre-

Table 4. Attribute importance measures for the *Taiwan Credit Card Default* dataset. DM Obj, DM Dec, Reduct Frequency (Red. Freq.), Information Gain, and Gain Ratio were computed on discretized attributes, with Information Gain approximated using mutual information. Model-based importance measures were obtained from models trained on the original continuous attributes. Entropy was reported as a descriptive characteristic of attribute value distributions.

Attribute	DM Obj	DM Dec	Red. Freq.	Information Gain	Gain Ratio	Random Forest	XGBoost	Permutation	ReliefF	SHAP	Entropy
PAY_0	69.13	77.80	16	0.077	0.056	0.104	0.425	0.062	0.102	0.536	0.089
LIMIT_BAL	94.41	93.99	16	0.017	0.007	0.061	0.026	0.001	0.085	0.215	0.336
BILL_AMT1	90.42	90.43	16	0.001	0.001	0.059	0.021	0.001	0.037	0.141	0.998
PAY_AMT2	85.43	85.62	1	0.012	0.006	0.048	0.028	0.002	0.000	0.109	0.675
PAY_AMT1	84.73	84.88	6	0.014	0.007	0.050	0.020	-0.001	0.003	0.092	0.679
PAY_AMT3	85.73	85.46	7	0.011	0.005	0.046	0.017	0.001	0.002	0.080	0.648
PAY_2	65.16	71.63	5	0.049	0.041	0.041	0.151	0.002	0.043	0.092	0.071
BILL_AMT3	99.07	99.02	15	0.001	0.001	0.052	0.012	-0.001	0.037	0.040	0.972
BILL_AMT6	77.18	77.53	1	0.002	0.001	0.051	0.012	-0.001	0.035	0.041	0.930
BILL_AMT4	19.13	18.66	1	0.001	0.001	0.050	0.014	0.002	0.038	0.042	0.964
PAY_3	64.91	69.78	5	0.037	0.031	0.026	0.054	0.001	0.031	0.084	0.069
PAY_6	63.30	67.09	6	0.026	0.022	0.021	0.032	0.000	0.020	0.059	0.064
PAY_4	63.05	67.49	6	0.032	0.028	0.024	0.042	0.000	0.018	0.050	0.065
PAY_5	61.66	65.88	3	0.003	0.003	0.021	0.031	-0.002	0.014	0.049	0.061
AGE	96.30	96.35	16	0.002	0.001	0.066	0.012	0.002	0.069	0.023	0.332
EDUCATION	62.85	62.70	16	0.001	0.001	0.020	0.011	0.001	0.019	0.023	0.050
MARRIAGE	50.95	51.19	16	0.000	0.001	0.014	0.008	0.001	0.005	0.029	0.010
BILL_AMT5	18.74	18.43	1	0.001	0.001	0.049	0.013	0.002	0.037	0.036	0.946

quency in reducts. These measures quantify the discriminative potential of individual attributes in terms of their ability to distinguish objects, both with and without consideration of the decision attribute on discretized input dataset. Their goal is to assess structural discernibility within the data, rather than predictive contribution. The second group includes information-theoretic measures (Information Gain (IG), Gain Ratio (GR)), ensemble-based approaches (Random Forest, XGBoost), permutation importance, ReliefF, and SHAP values. IG and GR were computed on discretized attributes using a quantile-based binning strategy, while all the model-based importance measures were derived from models trained on original continuous features. Entropy was reported as a descriptive characteristic of attribute distributions.These methods quantify the predictive contribution of each attribute with respect to the target classification task. Due to the fundamentally different criteria evaluated by these methods, the resulting rankings are not expected to be identical. Additionally, the inherent class imbalance in the dataset further amplifies differences between methods that are more or less sensitive to uneven class distributions. The results indicate that *PAY_0* is the most influential attribute across the majority of methods. It achieves the highest SHAP value (0.536) and is consistently ranked highly by XGBoost, Random Forest, and the RS-based measures, confirming its central role as an indicator of current delinquency. Other delay-related features (*PAY_2–PAY_6*) also show systematically high importance across different methodologies. Attributes related to credit limit and billing amounts (*LIMIT_BAL*, *BILL_AMT1–BILL_AMT6*) play a significant role as well. The *BILL_AMT** features obtain high SHAP values (e.g., *BILL_AMT1* = 0.141), indicating strong non-linear effects that are effectively captured by tree-based models. In contrast, these attributes receive very low scores from classical information-theoretic measures, confirming that such measures fail to capture their complex interactions. Payment amount variables (*PAY_AMT1–PAY_AMT6*) receive moderate to high SHAP values, though their importance varies substantially across other methods. This variability reflects non-linear dependencies and feature interactions that are relevant to predictive models but less evident in structural measures. RS-based measures assign high discernibility to several attributes. For example, *AGE* receives high values in *DM Obj* and *DM Dec* (approximately 96), while its predictive contribution – as measured by SHAP (0.023) – is negligible. This demonstrates that structural discernibility does not necessarily correlate with predictive usefulness. Conversely, demographic attributes such as *EDUCATION* and *MARRIAGE* consistently receive low importance across all methods, confirming their minimal relevance.

In summary, attributes related to delinquency history and current payment status constitute the key predictive determinants of default behaviour.

The observed differences between methods arise from their distinct evaluation principles and sensitivity to class imbalance. These findings highlight the importance of using a heterogeneous set of attribute importance techniques, combining both structural and predictive perspectives.

The next table 5 presents stability metrics computed for the global reduct and the reducts obtained from a 5-fold cross-validation procedure. The results show that the reducts exhibit moderate-to-high stability, indicating that the same subset of attributes tends to reappear across different training partitions. Minor variations observed between folds suggest the presence of redundancy and interactions among several attributes, which is common for financial risk datasets.

The results indicate that the reducts obtained through cross-validation exhibit **very high structural stability**. The average pairwise Jaccard similarity of 0.907 shows that the majority of features are consistently selected across folds, with only small differences involving a few peripheral attributes. An even stronger agreement with the global reduct is observed, as reflected by the Jaccard score of 0.949, confirming that the global reduct represents the attribute structure identified during 5-fold cross-validation. The Kuncheva Index of 0.833, which adjusts for chance agreement, further confirms the high repeatability of the reduction process. The Rogers–Tanimoto Distance, with a value close to zero, indicates that disagreements between fold-wise and global reducts are exceedingly rare. Likewise, the Dice Coefficient, with an average value approaching 1, demonstrates that the reducts differ only marginally, showing nearly complete agreement in the selected attributes. Together, these metrics demonstrate that the attribute reduction process is highly stable and reproducible for the Taiwan Credit dataset. The average reduct size of 15.7 ± 0.9 further confirms that the exhaustive algorithm produces consistently sized reducts, with minimal sensitivity to variations in data partitioning.

The table 6 presents aggregated measures of attribute importance derived from the analysis of global and fold-specific reducts. Topological indicators, including degree and eigenvector centrality were computed for each attribute within a feature co-occurrence graph. It also reports the frequencies with which individual attributes were selected. Attributes with high frequency in both global and fold-specific reducts and high centrality can be interpreted as *core predictors*, consistently selected and central in the co-occurrence structure. Attributes with moderate centrality or variable fold-wise frequency may serve as *linking* or context-dependent features, while those with low frequency and low centrality are *peripheral* or rarely selected. This analysis highlights that features such as `LIMIT_BAL`, `AGE`, `SEX`, `EDUCATION`, `MARRIAGE`, `PAY_0`, and most historical payment variables (`PAY_AMT1`–`PAY_AMT6`, `BILL_AMT1`–`BILL_AMT6`) are stable across folds and occupy central positions, suggesting their robustness in the feature selection process.

The Tables 7 and 8 provide a comprehensive assessment of multiple families of classification ML models, including linear methods, Tree-based ensembles, Gradient boosting algorithms, and modern DL architectures designed for tabular data. Overall, the results indicate that the task exhibits moderate predictive separability, with model performance constrained by nonlinear feature interactions and the inherent class imbalance typical for credit-risk datasets. Models based on gradient boosting (particularly HistGradientBoosting and LightGBM) and selected neural architectures such as FT-Transformer and TabNet achieve the highest overall Accuracy, typically in the range of 0.81–0.82 on the full feature set. However, these gains in Accuracy do not translate into superior performance on metrics that account for class imbalance. Their Balanced Accuracy remains comparatively low (0.63–0.66), and Recall values for the positive class are limited (0.34–0.38). This pattern suggests that these models tend to overfit the predominant class, thereby underdetecting high-risk clients. Such behavior, while boosting Accuracy, undermines their practical utility in real credit-scoring scenarios where correct identification of defaulting applicants is critical. In contrast, tree-based ensembles such as ExtraTrees and RandomForest, although reaching slightly lower Accuracy values (0.77–0.80), demonstrate superior balance across metrics. They achieve the highest Balanced Accuracy (0.70–0.71) and more favorable Precision–Recall trade-offs, with Precision around 0.50–0.56 and Recall around 0.50–0.53, yielding F1-scores in the range of 0.52–0.54. These re-

Table 5. Stability analysis of the global reduct and fold-wise reducts derived from 5-fold cross-validation on the Taiwan Credit dataset. The global reduct was extracted from the 80% training portion of the input data. While minimal reducts emphasize parsimony, longer reducts preserve a broader set of condition attributes contributing to decision discernibility. Consequently, they provide a more reliable basis for stability analysis, interpretability assessment, and downstream machine learning evaluation.

Metric	**Mean $\pm$ SD**	**Interpretation**
Average reduct size	15.7 ± 0.9 attributes	Large and consistent reducts with minimal variation
Average Jaccard Index (Global vs Fold)	0.949 ± 0.041	Fold reducts are almost identical to the global reduct. Minor deviations occur.
Pairwise Jaccard (Fold vs Fold)	0.907 ± 0.056	High similarity between fold reducts. Stable feature selection. is high even after adjusting for chance
Kuncheva Index (KI)	0.833 ± 0.105	Strong repeatability. The probability of selecting the same features across folds
Rogers—Tanimoto Distance	0.036 ± 0.029	Low value indicates very high similarity between fold-wise and global reducts. Most features are consistently selected across folds.
Dice Coefficient	0.973 ± 0.022	Very high similarity. Almost all features are consistently selected.

Table 6. Feature frequency and network centrality for the Taiwan Credit dataset based on global and fold-wise reducts. High frequency and centrality indicate core and stable predictors.

Attribute	**Frequency (All Reducts)**	**Frequency (Folds)**	**Degree Centrality**	**Eigenvector Centrality**	**Interpretation**
AGE	1.00	1.00	0.591	0.303	Core predictor, stable across folds
BILL_AMT1	0.682	0.000	0.591	0.303	Moderately important, variable across folds
BILL_AMT3	0.682	0.000	0.372	0.204	Moderately important, variable across folds
BILL_AMT4	0.273	0.833	0.188	0.096	Linking attribute, occasional inclusion
BILL_AMT5	0.273	0.833	0.186	0.096	Linking attribute, occasional inclusion
BILL_AMT6	0.227	0.667	0.160	0.082	Peripheral predictor, less stable
EDUCATION	1.00	1.00	0.591	0.303	Core predictor, stable across folds
LIMIT_BAL	1.00	1.00	0.591	0.303	Core predictor, stable across folds
MARRIAGE	1.00	1.00	0.591	0.303	Core predictor, stable across folds
PAY_0	1.00	1.00	0.591	0.303	Core predictor, stable across folds
PAY_2	0.273	0.167	0.160	0.089	Moderately important, occasional inclusion
PAY_3	0.227	0.000	0.128	0.072	Peripheral predictor, rare in folds
PAY_4	0.273	0.000	0.152	0.085	Peripheral predictor, rare in folds
PAY_5	0.136	0.000	0.069	0.040	Peripheral predictor, rarely selected
PAY_6	0.273	0.000	0.152	0.085	Peripheral predictor, rare in folds
PAY_AMT1	0.545	1.00	0.335	0.176	Core predictor, stable in folds
PAY_AMT2	0.318	1.00	0.219	0.112	Core predictor in folds, moderate overall
PAY_AMT3	0.591	1.00	0.370	0.194	Core predictor in folds, moderate overall
PAY_AMT4	0.591	1.00	0.370	0.194	Core predictor in folds, moderate overall
PAY_AMT5	0.773	1.00	0.474	0.244	Core predictor, stable across folds
PAY_AMT6	0.955	0.833	0.563	0.290	Core predictor, stable across folds
SEX	1.00	1.00	0.591	0.303	Core predictor, stable across folds

Table 7. Selected results for Taiwan Credit dataset with different attribute subsets

Model	Attributes	Accuracy	Balanced Acc.	Precision	Recall	F1	ROC-AUC	Train Time (s)	Infer. Time (s)
CatBoost	All	0.76	0.71	0.48	0.61	0.53	0.78	5.80	0.02
	red_1	0.76	0.70	0.46	0.61	0.52	0.77	6.02	0.02
	red_2	0.76	0.71	0.47	0.61	0.53	0.77	6.25	0.02
	red_3	0.76	0.71	0.47	0.61	0.53	0.77	6.19	0.02
	red_4	0.77	0.71	0.48	0.61	0.53	0.77	6.27	0.02
	red_5	0.76	0.71	0.47	0.61	0.53	0.77	6.39	0.02
ExtraTrees	All	0.78	0.71	0.50	0.58	0.54	0.78	5.64	0.65
	red_1	0.77	0.70	0.48	0.56	0.52	0.77	4.32	0.58
	red_2	0.78	0.70	0.50	0.56	0.53	0.77	4.74	0.63
	red_3	0.77	0.70	0.49	0.57	0.53	0.77	4.42	0.56
	red_4	0.78	0.70	0.49	0.56	0.53	0.77	4.34	0.54
	red_5	0.77	0.69	0.48	0.56	0.51	0.77	4.30	0.54
	red_6	0.78	0.70	0.50	0.56	0.53	0.77	4.31	0.56
	red_7	0.78	0.70	0.51	0.54	0.52	0.77	4.08	0.55
HistGB	All	0.82	0.66	0.67	0.37	0.48	0.78	0.60	0.01
	red_1	0.82	0.65	0.67	0.36	0.47	0.77	0.58	0.01
	red_2	0.82	0.65	0.67	0.36	0.47	0.77	0.64	0.03
	red_3	0.82	0.65	0.65	0.36	0.46	0.77	0.83	0.02
	red_5	0.82	0.66	0.67	0.36	0.47	0.77	0.53	0.01
	red_11	0.82	0.65	0.67	0.36	0.47	0.77	0.91	0.01
	red_13	0.82	0.65	0.68	0.36	0.47	0.77	0.52	0.01
	red_14	0.82	0.66	0.67	0.37	0.47	0.77	0.66	0.01
	red_16	0.82	0.65	0.67	0.36	0.47	0.77	0.58	0.01
LightGBM	All	0.77	0.70	0.48	0.57	0.52	0.76	5.79	0.02
	red_1	0.76	0.70	0.47	0.58	0.52	0.76	5.46	0.02
	red_2	0.77	0.70	0.48	0.59	0.53	0.76	4.26	0.02
	red_3	0.76	0.69	0.47	0.57	0.51	0.76	4.62	0.02
	red_4	0.77	0.70	0.48	0.58	0.52	0.76	5.00	0.02
	red_7	0.77	0.70	0.48	0.58	0.52	0.76	5.30	0.02
LogisticRegression	All	0.68	0.66	0.37	0.62	0.46	0.71	0.06	0.00
	red_1	0.68	0.66	0.37	0.63	0.46	0.70	0.03	0.00
	red_2	0.69	0.66	0.37	0.61	0.46	0.70	0.03	0.00
	red_3	0.69	0.66	0.37	0.61	0.46	0.70	0.02	0.00
	red_4	0.68	0.66	0.37	0.61	0.46	0.71	0.03	0.00
MLP (128,64,32)	All	0.81	0.65	0.64	0.36	0.46	0.76	3.66	0.02
	red_1	0.82	0.66	0.66	0.37	0.47	0.76	4.32	0.02
	red_3	0.82	0.65	0.68	0.35	0.46	0.76	4.29	0.02
	red_6	0.82	0.66	0.66	0.37	0.47	0.76	4.89	0.02
	red_9	0.82	0.67	0.64	0.39	0.49	0.75	2.89	0.02
RandomForest	All	0.80	0.70	0.54	0.54	0.54	0.77	21.16	0.52
	red_6	0.80	0.70	0.54	0.54	0.54	0.77	13.39	0.53
	red_7	0.80	0.70	0.54	0.52	0.53	0.77	14.52	0.54
	red_12	0.80	0.70	0.54	0.52	0.53	0.77	14.46	0.54

Table 8. Selected results for Taiwan Credit dataset with different attribute subsets

Model	Attributes	Accuracy	Balanced Acc.	Precision	Recall	F1	ROC-AUC	Train Time (s)	Infer. Time (s)
XGBoost	All	0.77	0.69	0.49	0.55	0.52	0.76	5.60	0.03
	red_3	0.77	0.69	0.48	0.55	0.51	0.76	3.54	0.02
	red_6	0.77	0.70	0.49	0.56	0.52	0.75	3.73	0.02
	red_7	0.77	0.69	0.49	0.55	0.52	0.75	4.88	0.02
	red_11	0.77	0.69	0.49	0.55	0.52	0.76	3.69	0.02
	red_13	0.76	0.69	0.47	0.55	0.51	0.76	3.71	0.02
	red_15	0.77	0.69	0.48	0.55	0.51	0.76	4.04	0.02
FT-Transformer	All	0.82	0.66	0.66	0.37	0.48	0.78	31.04	0.03
	red_1	0.82	0.65	0.67	0.35	0.46	0.77	14.56	0.01
	red_2	0.82	0.65	0.66	0.36	0.46	0.77	15.56	0.02
	red_3	0.82	0.65	0.67	0.34	0.45	0.76	15.53	0.02
	red_6	0.82	0.66	0.66	0.37	0.47	0.77	16.61	0.02
	red_16	0.82	0.65	0.67	0.35	0.46	0.77	26.71	0.03
TabNet	All	0.82	0.65	0.67	0.34	0.45	0.75	125.33	0.26
	red_1	0.82	0.65	0.67	0.34	0.45	0.76	133.53	0.26
	red_3	0.82	0.65	0.67	0.35	0.46	0.74	87.11	0.26
	red_4	0.81	0.63	0.66	0.30	0.42	0.75	124.37	0.26
	red_10	0.82	0.64	0.66	0.33	0.44	0.75	102.15	0.26
	red_11	0.82	0.65	0.66	0.34	0.45	0.75	127.37	0.26
	red_12	0.82	0.65	0.68	0.34	0.45	0.75	114.58	0.26

sults highlight their robustness in handling imbalanced distributions and their better alignment with the operational objectives of credit-risk assessment, where both types of misclassifications (false alarms and missed detections) carry significant cost implications. Classical ML Logistic regression performs noticeably worse in terms of overall Accuracy (0.68–0.69), confirming its limited ability to model nonlinear dependencies in the dataset. Nevertheless, the model records relatively high Recall (0.61–0.63), paired with a low Precision of approximately 0.37. This indicates that, while it successfully identifies a substantial fraction of positive cases, it produces a large number of false positives. The trade-off reinforces the need for more expressive models when predictive Precision is essential to downstream decision-making.

A key outcome of the our study is the impact of feature-set reduction on predictive performance. Across nearly all evaluated models, the reduced feature sets (reduct variants) yield Accuracy, Balanced Accuracy, and ROC-AUC values nearly identical to those obtained with the full feature set. For instance, in RandomForest the Accuracy difference between the full dataset (≈ 0.79) and reduced variants rarely exceeds 0.01, while Balanced Accuracy remains stable at 0.70–0.71. This indicates that the reduct–based feature-selection method successfully retains the essential predictive structure of the data. At the same time, feature reduction leads to significant computational benefits, with training times reduced by 20–40% in several models (e.g., RandomForest from 21.16 s to approximately 13–15 s). These findings demonstrate the practical value of reduct-based dimensionality reduction for creating more efficient and interpretable models without sacrificing predictive power.

The achieved results show that no single model dominates across all evaluation criteria. Gradient boosting and DL models provide the highest Accuracy but insufficient sensitivity to the minority class, limiting their strategic usefulness in risk detection. Tree-based ensembles, while slightly less accurate, offer the most balanced performance and are therefore better suited for applications where both detection and stability across metrics are crucial. The minimal degradation observed under feature reduction further emphasizes that compact, computationally efficient models can be constructed without compromising classification quality. Consequently, the optimal model choice should reflect the operational costs associated with different error types, the available computational resources, and the interpretability requirements of the target credit-risk environment.

6 Conclusions

In this study, we proposed a RS–based feature reduction framework aimed at improving both the efficiency and effectiveness of DL models. The approach involves constructing a decision table from discretized training data, generating a discernibility matrix, and exhaustively identifying all minimal RS reducts. These reducts are subsequently analyzed in terms of their frequency, stability and structural relationships among attributes. The resulting reducts are then used to restrict the input feature space of ML and DL models, whose performance is systematically compared against full-feature counterparts with respect to predictive accuracy and computational efficiency. A comprehensive evaluation framework was employed to assess classification quality and robustness. Predictive performance was measured using standard classification metrics, including: Accuracy, Balanced Accuracy, Precision, Recall, F1-score, and ROC-AUC. In parallel, the stability of feature selection was quantified using overlap-based measures such as the Jaccard and Kuncheva indices, enabling an assessment of the repeatability of reduct-based feature subsets across cross-validation folds. The proposed approach was experimentally validated on the Taiwan Credit Card Default dataset. The results demonstrate that restricting the input space to RS-based feature subsets leads to a substantial reduction in Training and Inference time, while maintaining – and in several cases improving – classification performance across a wide range of ML and DL models. Moreover, the high degree of overlap between reducts obtained in different folds confirms the structural stability and robustness of the selected features. The proposed methodology is applicable to classification problems involving structured tabular data. For example, in domains where computational efficiency, robustness of feature selection, and interpretability are important. Such domains include financial risk assessment, credit scor-

ing, and decision-support systems based on data-driven models or in distributed datasets and privacy-sensitive environments, as highlighted in the works of M. Przybyła-Kasperek [57, 58]. In particular, the proposed framework allows for scalable integration of multiple local datasets without recalculating reducts for each dataset individually while leveraging DL to adapt reducts flexibly to local data characteristics. This makes the approach well-suited for federated learning and other distributed decision-making systems. The method aligns with and extends existing RS–based frameworks by integrating exhaustive reduct analysis with modern DL architectures. It should be noted that, in the case of a very large number of conditional attributes, exhaustive reduct generation may increase computational complexity. In addition, the quality of the obtained reducts depends on the discretization strategy applied to continuous features. These limitations motivate further research directions, including the use of heuristic or hybrid reduction algorithms, alternative discretization schemes and scalability analysis on more heterogeneous datasets.

6.1 Further research

While this study demonstrates the feasibility of integrating RS–based feature reduction with DL models for tabular data, several new directions remain open for further investigation. One promising direction is the expansion of the analysis to include other discretization techniques. This approach would enhance the understanding of how the discretization of continuous features affects the performance of DL models, particularly in the context of tabular data. Additionally, it is crucial to examine other available models beyond Keras and PyTorch. Numerous alternative frameworks offer integrated feature selection mechanisms, which may provide a more optimized solution for handling high-dimensional data. A comparison of the effectiveness of these models with traditional DL architectures could yield important insights into the trade-offs between complexity, interpretability, and performance.

Acknowledgement

The research has been supported by the program: Excellence Initiative – Research University for the AGH University of Krakow, Subvention 2025 grant, project: Development of AI models and methods, and cyber-physical systems applications (Rozwijanie modeli i metod sztucznej inteligencji oraz zastosowania w sterowaniu systemów cyberfizycznych).

References

[1] Das A. K., Sengupta S., and Bhattacharyya S. A group incremental feature selection for classification using rough set theory based genetic algorithm. Applied Soft Computing, 65:400–411, 2018.

[2] S.O Arik and T. Pfister. TabNet: Attentive Interpretable Tabular Learning, 2020.

[3] J. Bazan, M. Szczuka, and J. Wróblewski. A New Version of Rough Set Exploration System. In Lecture Notes in Artificial Intelligence, volume 2475 of Lecture Notes in Computer Science, pages 397–404. Springer-Verlag, Berlin, Heidelberg, 2002.

[4] J.G. Bazan, S. Bazan-Socha, U. Bentkowska, W. Gałka, M. Mrukowicz, and L. Zaręba. Comparison of aggregation classes in ensemble classifiers for high dimensional datasets. In 2022 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pages 1–10, 2022.

[5] J. Bilski, B. Kowalczyk, L. Dymova, and M. Xiao. Accelerating Neural Network Training with FS-GQR: A Scalable and High-Performance Alternative to Adam. Journal of Artificial Intelligence and Soft Computing Research, 15(2):95–113, 2025.

[6] Ch.M Bishop. Pattern Recognition and Machine Learning. Springer, 2006.

[7] V. Borisov, T. Leemann, K. Seßler, J. Haug, M. Pawelczyk, and G. Kasneci. Deep Neural Networks and Tabular Data: A Survey. IEEE Transactions on Neural Networks and Learning Systems, 35(6):7499–7519, 2024.

[8] J. Błaszczyński, S. Greco, B. Matarazzo, R. Słowiński, and M. Szeląg. jMAF - Dominance-Based Rough Set data Analysis Framework. In Intelligent Systems Reference Library, volume 42, pages 185–209. Springer, 2013.

[9] F. Chollet. Keras. `https://keras.io`, 2015.

[10] S. Dutta and A. Skowron. Information System in the Light of Interactive Granular Computing. In Mengjun Hu, Chris Cornelis, Yan Zhang, Pawan Lingras, Dominik Ślęzak, and JingTao Yao, editors, Rough Sets, pages 223–237, Cham, 2024. Springer Nature Switzerland.

[11] W.H. Elashmawi, A. Sheta, B.S. Alqadi, D.S. AbdElminaam, and D.M. Alsekait. A New Version of the Golden Eagle Optimizer Algorithm And Its Application For Solving A Trio-Objective Skillful Team Formation Problem In A Social Network. Journal of Artificial Intelligence and Soft Computing Research, 15(4):357–384, 2025.

[12] A.E. Ezugwu, A.M. Ikotun, O. O. Oyelade, L. Abualigah, J. O. Agushaka, Ch. I. Eke, and A. A. Akinyelu. A Comprehensive Survey of Clustering Algorithms: State-of-the-Art Machine Learning Applications, Taxonomy, Challenges, and Future Research Prospects. Engineering Applications of Artificial Intelligence, 110:104743, 2022.

[13] B. Ganter and R. Wille. Formal Concept Analysis: Mathematical Foundations. Springer, 1999.

[14] M. Garbulowski, K. Diamanti, K. Smolińska, et al. R.ROSETTA: an interpretable machine learning framework. BMC Bioinformatics, 22(1):110, 2021.

[15] W. Gałka, J. Bazan, U. Bentkowska, K. Szwed, M. Mrukowicz, P. Drygaś, L. Zaręba, M. Szpyrka, P. Suszalski, and S. Obara. Aggregation-Based Ensemble Classifier Versus Neural Networks Models for Recognizing Phishing Attacks. IEEE Access, 13:48469–48487, 2025.

[16] I. Goodfellow, Y. Bengio, and A. Courville. Deep Learning. MIT Press, 2016.

[17] Y. Gorishniy, I. Rubachev, V. Khrulkov, and A. Babenko. Revisiting deep learning models for tabular data. In Proceedings of the 35th International Conference on Neural Information Processing Systems, NIPS '21, Red Hook, NY, USA, 2021. Curran Associates Inc.

[18] M. Grzegorowski, A. Janusz, Ł. Marcinowski, A. Skowron, D. Ślęzak, and G. Śliwa. On Explainability of Cluster Prototypes with Rough sets: A Case Study in the FMCG Market. International Journal of Applied Mathematics and Computer Science, 35(1):19–31, 2025.

[19] J.W. Grzymala-Busse. LERS-A data mining system. In O. Maimon and L. Rokach, editors, Data Mining and Knowledge Discovery Handbook. Springer, Boston, MA, 2005.

[20] G. Góra and A. Skowron. RIONIDA: A novel algorithm for imbalanced data combining instance-based learning and rule induction. Information Sciences, 708:122015, 2025.

[21] Sakai H., Nakata M., Ślęzak D., and Watada J. Rule generation in Rough set Non-deterministic Information analysis (RNIA) and some applications of the obtained rules. Applied Soft Computing, 172:112842, 2025.

[22] T. Hachaj and J. Wąs. An insightful data-driven crowd simulation model based on rough sets. Information Sciences, 692:121670, 2025.

[23] T. Hachaj and J. Wąs. Rough neighborhood graph: A method for proximity modeling and data clustering. Applied Soft Computing, 171:112789, 03 2025.

[24] A.A. Hagberg, D.A. Schult, and P.J. Swart. Exploring Network Structure, Dynamics, and Function using NetworkX. Python in Science Conference, 2008.

[25] J. Han, J. Pei, and M. Kamber. Data Mining: Concepts and Techniques. Elsevier, 4 edition, 2021.

[26] R.C. Holte. Very Simple Classification Rules Perform Well on Most Commonly Used Datasets. Machine Learning, 11:63–90, 1993.

[27] Bazan J. and Szczuka M. The Rough Set Exploration System. Transactions on Rough Sets III, 3400:37–56, 2005.

[28] Jing J. Research on decision-making evaluation system based on rough set theory. In Proceedings of the 2023 8th International Conference on Intelligent Information Processing, ICIIP '23, page 174–178, New York, NY, USA, 2023. Association for Computing Machinery.

[29] Bazan J. G. A Comparison of Dynamic and Non-Dynamic Rough Set Methods for Extracting Laws from Decision Tables. In L. Polkowski and A. Skowron, editors, Rough Sets in Knowledge Discovery 1: Methodology and Applications, pages 321–365. Physica-Verlag, Heidelberg, 1998.

[30] G. James, D. Witten, T. Hastie, and R. Tibshirani. An Introduction to Statistical Learning: with Applications in R. Springer, 2 edition, 2021.

[31] A. Janusz, D. Ślęzak, S. Stawicki, and K. Stencel. A practical study of methods for deriving insightful attribute importance rankings using decision bireducts. Information Sciences, 645:119354, 2023.

[32] R. Jensen. Fuzzy-Rough Data Mining. In Rough Sets and Current Trends in Computing (RSCTC 2010), pages 31–35, Berlin, Heidelberg, 2011. Springer.

[33] R. Jensen and Q. Shen. Semantics-Preserving Dimensionality Reduction: Rough and Fuzzy-Rough-Based Approaches. IEEE Transactions on Knowledge and Data Engineering, 16(12):1457–1471, 2004.

[34] D.S. Johnson. Approximation Algorithms for Combinatorial Problems. Journal of Computer and System Sciences, 9(3):256–278, 1974.

[35] I.T. Jolliffe and J. Cadima. Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2065), 2016.

[36] J. Jomal, M. Bibin, J.J. Sunil, and Jobish V. Integrating Rough Sets and Multidimensional Fuzzy Sets for Approximation Techniques: A Novel Approach. IEEE Access, 12:154796–154810, 2024.

[37] J. Komorowski, A. Øhrn, and A. Skowron. The ROSETTA Rough Set Software System. In W. Klösgen and J. Zytkow, editors, Handbook of Data Mining and Knowledge Discovery, chapter D.2.3. Oxford University Press, 2002.

[38] V. Latora, V. Nicosia, and G. Russo. Complex Networks: Principles, Methods and Applications. Cambridge University Press, 2017.

[39] Z. Liu, M. Zhao, Sh. Zhao, J. Luo, and F. Luo. Opinion Evolution and Guidance Model Based on Social Networks and Information Networks. Journal of Artificial Intelligence and Soft Computing Research, 16(2):105–123, 2026.

[40] Z. Lv, H. Song, P. Basanta-Val, A. Steed, and M. Jo. Next-Generation Big Data Analytics: State of the Art, Challenges, and Future Research Topics. IEEE Transactions on Industrial Informatics, 13(4):1891–1899, 2017.

[41] E. Masciari, A. Umair, and M.H. Ullah. A systematic Literature Review on AI-Based Recommendation Systems and Their Ethical Considerations. IEEE Access, 12:121223–121241, 2024.

[42] S. Mitra, S.K. Pal, and P. Mitra. Data mining in soft computing framework: a survey. IEEE Transactions on Neural Networks, 13(1):3–14, 2002.

[43] M. Ju. Moshkov, M. Piliszczuk, and B. Zielosko. Partial Covers, Reducts and Decision Rules in Rough Sets: Theory and Applications. Springer Publishing Company, Incorporated, 1st edition, 2009.

[44] S. Naouali and O. El Othmani. Rough Set Theory and Soft Computing Methods for Building Explainable and Interpretable AI/ML Models. Applied Sciences, 15(9):5148, 2025.

[45] M. Newman. Networks. Oxford University Press, 2nd edition, 07 2018.

[46] H.S. Nguyen and A. Skowron. Quantization of Real-Valued Attributes in Decision Tables – A Rough Set Approach. International Journal of Approximate Reasoning, 13(2):119–148, 1995.

[47] R.K. Nowicki, R. Seliga, D. Żelasko, and Y. Hayashi. Performance analysis of rough set–based hybrid classification systems in the case of missing values. Journal of Artificial Intelligence and Soft Computing Research, 11(4):307–318, 2021.

[48] P. Okeleke, D. Ajiga, S. Folorunsho, and Ch. Ezeigweneme. Predictive Analytics for Market Trends using AI: A Study in Consumer Behavior. International Journal of Engineering Research Updates, 7(1), August 2024.

[49] R. Olszowski, M. Zabdyr-Jamróz, S. Baran, P. Pięta, and W. Ahmed. A Social Network Analysis of Tweets Related to Mandatory COVID-19 Vaccination in Poland. Vaccines (Basel), 10(5):750, May 2022.

[50] S.K Pal, D. Bhoumik, and D. Bhunia Chakraborty. Granulated deep learning and z-numbers in motion detection and object recognition. Neural Computing and Applications, 32(7):16533–16548, 2020.

[51] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al. Pytorch: An Imperative Style, High-Performance Deep Learning Library. Advances in Neural Information Processing Systems, 32, 2019.

[52] Z. Pawlak. Rough Sets: Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Norwell, USA, 1991.

[53] Z. Pawlak and A. Skowron. Rudiments of rough sets. Information Sciences, 177(1):3–27, 2007. Z. Pawlak life and work (1926–2006).

[54] Z. Pei, Z. Meng, T. Diao, P. Miao, Y. Meng, and Ch. Tan. A Visually Explainable Dynamic Similarity Network for Few-Shot Classification. Journal of Artificial Intelligence and Soft Computing Research, 16(3):237–256, 2026.

[55] J.F. Peters, A. Skowron, and J. Stepaniuk. Rough Sets: Foundations and Perspectives, pages 877–889. Springer US, New York, NY, 2023.

[56] P. Pięta, T. Szmuc, and K. Kluza. Comparative overview of rough set toolkit systems for data analysis. MATEC Web of Conferences, 252:03019, 2019. III International Conference of Computational Methods in Engineering Science (CMES'18), Kazimierz Dolny, Poland, Nov 22–24, 2018.

[57] M. Przybyła-Kasperek and K. Kusztal. Rules' Quality Generated by the Classification Method for Independent Data Sources Using Pawlak Conflict Analysis Model. In Jiří Mikyška, Clélia de Mulatier, Maciej Paszynski, Valeria V.

Krzhizhanovskaya, Jack J. Dongarra, and Peter M.A. Sloot, editors, Computational Science – ICCS 2023, pages 390–405, Cham, 2023. Springer Nature Switzerland.

[58] M. Przybyła-Kasperek and K. Opoku. Decision rules for dispersed data using a federated learning approach. Procedia Computer Science, 225:4305–4313, 2023. 27th International Conference on Knowledge Based and Intelligent Information and Engineering Sytems (KES 2023).

[59] B. Prędki, R. Słowiński, J. Stefanowski, R. Susmaga, and S. Wilk. ROSE - software implementation of the rough set theory. In Proceedings of the 2nd Joint European Conference on the Theory and Practice of Software (ETAPS'98), Lecture Notes in Computer Science, pages 605–608, Berlin, Heidelberg, 1998. Springer.

[60] Y. Qian, X. Liang, Q. Wang, J. Liang, B. Liu, A. Skowron, Y. Yao, J. Ma, and Ch. Dang. Local rough set: A solution to rough data analysis in big data. International Journal of Approximate Reasoning, 97:38–63, 2018.

[61] Z. Qiao, A.A. Heidari, X. Zhao, and H. Chen. An Evolutionary Neural Architecture search method Accelerated by Multi-Fidelity Evaluation and Genetic Decision Controller. Journal of Artificial Intelligence and Soft Computing Research, 15(4):413–446, 2025.

[62] L.S. Riza, A. Janusz, Ch. Bergmeir, Ch. 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, 287:68–89, 2014.

[63] H. Sakai, M. Nakata, D. Ślęzak, and J. Watada. Consideration of Detecting Data and Functional Dependency in Tabular Data with Missing Values by the Obtained Rules. In M. Hu, Ch. Cornelis, Y. Zhang, P. Lingras, D. Ślęzak, and J. Yao, editors, Rough Sets, pages 120–133, Cham, 2024. Springer Nature Switzerland.

[64] R. Saravanan and Pothula Sujatha. A State of Art Techniques on Machine Learning Algorithms: A Perspective of Supervised Learning Approaches in Data Classification. In 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), pages 945–949, 2018.

[65] S. Sarker, P. Sarker, Stone. G., R. Gorman, A. Tavakkoli, G. Bebis, and J. Sattarvand. A comprehensive overview of deep learning techniques for 3D point cloud classification and semantic segmentation. Machine Vision and Applications, 35(4):67, 2024.

[66] D. Saumendra and J. Nayak. Customer Segmentation via Data Mining Techniques: State-of-the-Art Review, pages 489–507. Springer, Singapore, 01 2022.

[67] J. Sawicki, M. Ganzha, and M. Paprzycki. The State of the Art of Natural Language Processing – A Systematic Automated Review of NLP Literature Using NLP Techniques. Data Intelligence, 5:1–47, 07 2023.

[68] R. Sen, A.K. Mandal, and B. Chakraborty. A Critical Study on Stability Measures of Feature Selection with a Novel Extension of Lustgarten Index. Machine Learning and Knowledge Extraction, 3(4):771–787, 2021.

[69] N. Shakhovska, A. Shebeko, and Y. Prykarpatskyy. A Novel Explainable AI Model for Medical Data Analysis. Journal of Artificial Intelligence and Soft Computing Research, 14(2):121–137, 2024.

[70] Q. Shen and R. Jensen. Selecting Informative Features with Fuzzy-Rough Sets and Its Application for Complex Systems Monitoring. Pattern Recognition, 37(7):1351–1363, 2004.

[71] W. Siedlecki and J. Sklansky. A note on genetic algorithms for large-scale feature selection. Pattern Recognition Letters, 10(5):335–347, 1989.

[72] A. Skowron. Informational Granules in Interactive Granular Computing. Computer Sciences & Mathematics Forum, 8(1), 2023.

[73] A. Skowron, A. Jankowski, and S. Dutta. Interactive granular computing. Granular Computing, 1(2):95–113, 2016.

[74] A. Skowron and C. Rauszer. The Discernibility Matrices and Functions in Information Systems, pages 331–362. Springer Netherlands, Dordrecht, 1992.

[75] A. Skowron and J. Stepaniuk. Toward rough set based insightful reasoning in intelligent systems. Information Sciences, 709:122078, 2025.

[76] A. Skowron and D. Ślęzak. Rough sets Turn 40: From Information Systems to Intelligent Systems. In 2022 17th Conference on Computer Science and Intelligence Systems (FedCSIS), pages 23–34, 2022.

[77] V. S. Subha and R. Selvakumar. An Innovative Electric Vehicle Selection with Multi-Criteria Decision-making Approach in Indian brands on a Neutrosophic Hypersoft Rough Set by using Reduct and Core. In 2024 Fourth International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), pages 1–7, Jan 2024.

[78] M. Suzuki. FT-Transformer: Transformer for Tabular Data. https://www.kaggle.com/code/masatakasuzuki/ft-transformer-transformer-for-tabular-data, 2023. Kaggle Notebook, accessed 2026-02-01.

[79] A. Szczur, J.G. Bazan, U. Bentkowska, P. Kruczek, and S. Bazan-Socha. Identifying and Predicting Changes in Behavioral Patterns for Temporal Data in Treatment of Neonatal Respiratory Failure. Applied Sciences, 15(22), 2025.

[80] J. Tan, Q. Hou, X. Liu, and Y. Xiong. Research on energy consumption prediction based on fast attribute reduction of weighted neighborhood rough set with moving horizon. In Proceedings of the 2023 7th International Conference on Machine Learning and Soft Computing, ICMLSC '23, page 18–25, New York, NY, USA, 2023. Association for Computing Machinery.

[81] Sanjeev Verma, Rohit Sharma, Subhamay Deb, and Debojit Maitra. Artificial Intelligence in Marketing: Systematic Review and Future Research Direction. International Journal of Information Management Data Insights, 1(1):100002, 2021.

[82] D. Vetrithangam, Naresh Kumar Pegada, R. Himabindu, and A. Ramesh Kumar. A State of Art Review on Image Analysis Techniques, Datasets, and Applications. AIP Conference Proceedings, 3072(1):040002, 03 2024.

[83] Q. Wang, Y. Qian, X. Liang, Q. Guo, and J. Liang. Local neighborhood rough set. Knowledge-Based Systems, 153:53–64, 2018.

[84] S. Wasserman and K. Faust. Social Network Analysis: Methods and Applications. Structural Analysis in the Social Sciences. Cambridge University Press, 1994.

[85] S. Xia, C. Wang, G. Wang, X. Gao, W. Ding, J. Yu, Y. Zhai, and Z. Chen. GBRS: A Unified Granular-Ball Learning Model of Pawlak Rough Set and Neighborhood Rough Set. IEEE Transactions on Neural Networks and Learning Systems, 36(1):1719–1733, Jan 2025.

[86] Y. Yao. A Comparative Study of Formal Concept Analysis and Rough Set Theory in Data Analysis. In Shusaku Tsumoto, Roman Słowiński, Jan Komorowski, and Jerzy W. Grzymała-Busse, editors, Rough Sets and Current Trends in Computing, pages 59–68, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.

[87] Y. Y. Yao and Yiyu Zhao. Discernibility matrix simplification for constructing attribute reducts. Information Sciences, 179(5):867–882, 2009.

[88] K. Yuan, W. Xu, and D. Miao. A Local Rough Set method for feature selection by Variable Precision composite measure. Applied Soft Computing, 155:111450, 2024.

[89] Q. Zhang, Q. Xie, and G. Wang. A survey on rough set theory and its applications. CAAI Transactions on Intelligence Technology, 1(4):323–333, 2016.

[90] W. Zhang and Z. Yang. Methods and Applications of Data Management and Analytics. Applied Sciences, 14(24), 2024.

[91] W. Ziarko. Variable Precision Rough Set Model. Journal of Computer and System Sciences, 46(1):39–59, 1993.

[92] R. Świniarski and A. Skowron. Rough set method in feature selection and recognition. Pattern Recogn. Lett. 24, 833-849. Pattern Recognition Letters, 24:833–849, 03 2003.

Piotr Pięta is affiliated with the AGH University of Science and Technology in Kraków, Poland. He received his M.Sc. degree in Computer Science in 2018. He is a member of the "Smart methods in software engineering and data analysis" research group at the Department of Computer Science and Artificial Intelligence, AGH University of Science and Technology. His research activity focuses on Rough Set Theory and its applications in big data analysis and data mining, including Fuzzy-Rough Sets, Machine Learning, Collective Intelligence, Artificial Intelligence, and social network analysis.
https://orcid.org/0000-0001-8490-7117

Tomasz Szmuc received MSc (1972) in Electrical and Control Engineering from the AGH University of Science and Technology (AGH). Since the beginning, he has been employed at AGH University of Science and Technology, where he received Ph.D. (1979) and Sc.D. (1989) degrees (both in Computer Science), and Professor title (1999). His research may be assigned to two main areas: formal methods supporting software development and hybrid approaches in big data analysis. The first area focuses on applications of Petri nets, process algebras, and temporal logics in the development. The combined use of Fuzzy-Rough Sets and neural networks forms the basis of his second research branch.
https://orcid.org/0000-0003-4922-5369

Rafał Mrówka was employed as an Assistant at AGH University of Science and Technology in 2007. In 2008 he defended his doctoral dissertation regarding formal methods in software engineering and began working as a lecturer. During this period, he conducted research and teaching activities requiring practical knowledge of Software Engineering. He co-creators the Alvis formal modelling language based on process algebra. He published several articles dealing with the subject of UML translation into Petri nets and the use of bisimulation in the process of verifying the requirements of real-time systems. Recent research has focused on using artificial intelligence in practical software engineering. In particular, he is interested in automated processing and generating spatial models of objects using artificial neural networks.

https://orcid.org/0000-0001-5143-3488