DOI: 10.15849/IJASCA.240730.17

Int. J. Advance Soft Compu. Appl, Vol. 16, No. 2, July 2024
Print ISSN: 2710-1274, Online ISSN: 2074-8523

Optimizing Attribute Reduction in Rough Set Theory using Re-heat Simulated Annealing for Classification and Data Mining

Alwi M Bamhdi[1] Ana Luiza Barros[2], Tahira Jehan Makda[3], Marcial Fernandez[4], Ahmed Patel[5], Laleh Golafshan[6]

[1]Umm Al-Qura University, College of Computing, Al-Qunfudah, Kingdom of Saudi Arabia.
email: ambamhdi@uqu.edu.sa
[2,4,5]Computer Science Program, State University of Ceará (UECE), Fortaleza, Brazil.
email: [2]analuiza.barros@uece.br, [4]marcial.fernandez@uece.br, [5]zapazap@gmail.com
[3]College of Business Management, University of Doha for Science and Technology (UDST), Doha, Qatar.
email: tahira.makda@udst.edu.qa
[6]Department of Computer Engineering, Azad University, Meybod, Iran.
email: laleh.golafshan@gmail.com

Abstract

Data classification is a crucial aspect of knowledge discovery using machine-learning algorithm for supervised learning approach where the goal is to predict the categorical labels of new instances based on past observations. This research presents an innovative classification technique that utilizes Rough Set Attribute Reduction. The proposed method introduces the Re-heat Simulated Annealing (Re-heat SA) algorithm as a meta-heuristic approach. Rough set theory, a mathematical tool dealing with uncertainty and fuzziness in data, is employed to uncover hidden patterns in big data through feature selection. This paper introduces a novel meta-heuristic classification approach that utilizes rough set attribute reduction to achieve optimal accuracy. Re-heat SA effectively optimizes the problem by controlling the dependency degree to identify the minimal reducts required for classification prediction using the Rosetta software. Experimental results demonstrate that Re-heat SA outperforms comparable classification algorithms in discovering classification rules. The results reveal that three datasets achieved 100% accuracy, four datasets achieved accuracy rates ranging from 60% to 99%, and six datasets achieved accuracy rates ranging from 30% to 59%. Additionally, this paper discusses the need for standardization concerning the machine learning pipeline processes as big data and its handling grows exponentially.

Keywords: *Attributes reduction, big data, classification, data mining, meta-heuristic classification optimization, re-heat simulated annealing, rough set theory, standards*

Received 1 May 2023; Accepted 12 July 2024

1 Introduction

Classification is a fundamental task in knowledge discovery and data mining, wherein distinct classes within a dataset are identified using machine learning algorithms for supervised learning. Due to its significance in data differentiation, it has been extensively studied. The primary metric for evaluating classification models is accuracy, along with essential metrics like Precision, Recall, F1 score, and Area under ROC Curve (AUC)[1]. Various algorithms, such as statistical classification, decision trees, rule induction, fuzzy rule induction, and neural networks, are commonly employed for classification. In this research, we specifically focus on rule induction as the chosen method, which utilizes "IF-THEN" statements to generate rules. Previous works, including the CN2 rule-based induction algorithm [1], Supervised Inductive Algorithm (SIA) [2], AprioriC [3], Grammar-based Genetic Programming algorithm (GGP) [4], ACORI [5], RIFT [6], and Rule Induction Using Set-Based Particle Swarm Optimization [7], have employed this method. However, the performance of classification algorithms can be adversely affected by the "Curse of Dimensionality," particularly depending on the dataset used. To achieve satisfactory results, employing dimensionality reduction techniques to reduce the number of attributes in the dataset is essential. Attribute reduction involves identifying correlations among dataset attributes based on their relevance and redundancy. This process eliminates unnecessary and unimportant attributes, mitigating challenges posed by high-dimensional datasets. The attributes retained in a minimal subset are related to the decision attributes, significantly contributing to the performance of the classification task, and improving accuracy and efficiency.

The research presented in this paper investigates the classification accuracy of a dataset reduction generated by the Re-heat Simulated Annealing approach, which utilizes rough set attribute reduction as a single-based algorithm. The objective is to find improved solutions by incorporating a re-heat mechanism into the original simulated annealing algorithm. This research contributes by presenting attribute reduction in rough set theory and utilizes the Re-heat Simulated Annealing algorithm to assess the quality of minimal reducts and classification accuracy and optimization.

Section 2 surveys the background and related works. Section 3 sets out the necessary notations and basic concepts related to rough set theory. Section 4 addresses the strategy for mining rules and evaluating classification accuracy, outlining the methodology and techniques used. Section 5 presents the Re-heat Simulated Annealing algorithm, explaining its components like temperature control and search strategies. Section 6 outlines the experimental design, while Section 7 presents and analyses the experimental results. Section 8 proposes standardization techniques for feature selection and dimensionality reduction. Finally, Section 9 discusses and concludes the key findings, contributions, and future research directions.

2 Related Work

The literature extensively discusses attribute reduction and the usefulness of feature selection in pattern recognition. For instance, Liang et al. [8] proposed a feature selection algorithm called the distance discriminant (FSDD), which overcame the computational costs and suboptimal drawbacks associated with other methods. Guo et al. [9] introduced

[1] https://www.analyticsvidhya.com/blog/2021/07/metrics-to-evaluate-your-classification-model-to-take-the-right-decisions/

a framework based on mutual information (MI) criterion for feature selection in high-dimensional data, specifically addressing image classification and non-Gaussian data. Zhang et al. [10] proposed an improved filtering method for feature selection, utilizing pairwise constraints to evaluate features. These studies are just a few examples, and other approaches such as Genetic Algorithm (GenRSAR) [11], Simulated Annealing (SimRSAR) [11], Ant Colony (AntRSAR and ACORA) [12, 13], Tabu Search (TSAR) [14], and Scatter Search (SSAR) [15] have also been explored.

More recently, Alsahaf et al. [16] introduced a novel framework for feature selection based on boosting or sample re-weighting. Zhang and Yao [17] investigated tri-level attribute reducts, focusing on sequential development and hierarchical deepening of attribute reduction. Additionally, Dokeroglu et al. [18] provided a comprehensive survey on recent metaheuristics for feature selection.

To contribute to this discussion, the main objective of this research is to evaluate the performance of the Re-heat Simulated Annealing (Re-heat SA) algorithm in estimating the accuracy of a classification method based on rule induction. Simulated Annealing (SA), originally proposed by Kirkpatrick et al. [19], is a widely used and effective algorithm for solving optimization problems. It leverages principles from thermodynamics and performs a stochastic search in the neighbourhood space [20]. Our proposed Re-heat SA algorithm introduces a re-heat mechanism to enhance the extraction of knowledge or patterns from large databases.

The first aim of our feature selection approach is to reduce data dimensionality and eliminate noise, while the second aim is to enhance mining performance in terms of predictive accuracy, learning speed, and simplicity and comprehensibility of mining models.

Rough Set Theory (RST) provides a fundamental concept for discovering patterns in inconsistent data [21, 22]. Pawlak [22] introduced RST as a mathematical tool for identifying minimal subsets in data, particularly in the presence of uncertain and incomplete data. One major application of RST is attribute reduction, which is considered a theoretical research problem classified as NP-hard.

Janecek et al. [24] studied the impact of feature selection on classification accuracy using email and drug discovery datasets. Assareh et al. [25] explored the accuracy of different combinations of six classification algorithms on high-dimensional cancer proteomic datasets. Hayward et al. [26] presented a performance evaluation of logistic regression techniques on a clinical database of cancer patients, emphasizing the significance of attribute selection in improving classifier performance. Xia et al. [27] proposed GBNRS, a novel NRS method, which demonstrated superior performance and classification accuracy compared to existing NRS methods. Li and Cui [28] developed a parallel attribute reduction processing algorithm for classification, resulting in improved time efficiency.

The Re-heat SA algorithm optimizes the search space by dynamically resetting the temperature when no improvement in solution quality is observed after a certain number of iterations. This mechanism enhances the algorithm's flexibility in accepting temporarily worse solutions for subsequent exploration of better solutions. The proposed approach is evaluated on well-known UCI datasets [29]. It introduces a novel meta-heuristic classification approach that uses rough set attribute reduction for optimal accuracy. The Re-heat Simulated Annealing (SA) algorithm effectively controls the dependency degree to identify minimal reducts required for classification prediction using Rosetta software.

Experimental results show that Re-heat SA outperforms comparable classification algorithms in discovering classification rules.

3 Rough Set Theory

Rough Set Theory (RST) is a technique that employs set approximation to analyse and extract patterns from the classification space. Pawlak is credited with introducing RST as a mathematical tool for discovering subsets of reducts [21, 22]. RST has found applications in various domains, including knowledge discovery from data, data reduction, data significance evaluation, decision rule generation, and data-driven inference interpretation [31].

In the context of data mining, RST plays a significant role in identifying common patterns in large datasets characterized by uncertainty and incompleteness. Importantly, RST can extract useful information from the data without requiring additional external information. This makes it valuable for analysing datasets with limited or missing data.

Attribute reduction is a key aspect of RST, as it enables the identification and retention of the most informative attributes while discarding others with minimal loss of information [12, 31, 32]. Informative attributes are most predictive of the class attribute or have high relevance to the target variable. RST facilitates the calculation of reducts in an information system, which refers to the minimal subsets of attributes that still preserve the discriminatory power needed for classification.

The ability of RST to find minimal reducts contributes to its effectiveness as a method for attribute reduction. By identifying minimal reducts, RST allows for the generation of more general decision rules, which aids in understanding and interpreting the data. Attribute reduction in RST is considered a crucial aspect in solving problems related to feature selection and dimensionality reduction.

In summary, RST is a valuable technique in data mining and pattern recognition due to its capability to identify common patterns in uncertain and incomplete data. It provides a mathematical framework for attribute reduction, enabling the extraction of informative attributes and the generation of more general decision rules.

3.1 Fundamental Rough Set Concepts

A dataset can be represented as an Information System for formal analysis, denoted by $I = (U, A)$. Here, U represents the universe containing a non-empty set of finite objects, and A is a non-empty finite set of attributes. For every attribute $a \in A$, there exists a mapping $U \rightarrow Va$, where Va represents the domain of attribute a[18].

Let us introduce the concept of an indiscernibility relation over a subset of attributes $P \subseteq A$, which is denoted by $IND(P)$:

Definition 1: For a subset of attributes $P \subseteq A$, the indiscernibility relation $IND(P)$ is associated with the equivalence relation as for any two objects $x, y \in U$, $x\ IND(P)\ y$ if and only if for every attribute $a \in P$, $x(a) = y(a)$.

This definition implies that two objects x and y are indiscernible with respect to the attributes in P if their attribute values are the same for each attribute in P.

The segment of U generated by $IND(P)$ is denoted by U/P and can be calculated as:

Definition 2: For a subset of attributes $P \subseteq A$, the segment of U generated by $IND(P)$ is denoted as U/P and can be calculated as: $U/P = \{ [x]P \mid x \in U \}$ where $[x]P$ represents the equivalence class of x with respect to $IND(P)$.

Now, let us define rough sets PX and $P'X$ based on the information system $I = (U, A)$, where $X \subseteq U$:

Definition 3: Let $X \subseteq U$. The *P-lower* approximation and *P-upper* approximation of set X are defined as the *P-lower* approximation of X, denoted by $LX(P)$, is the set of objects in U/P that are indiscernible from X with respect to the attributes in P: $LX(P) = \{ x \in U/P \mid [x]P \cap X \neq \emptyset \}$

The *P-upper* approximation of X, denoted by $UX(P)$, is the set of objects in U/P that have at least one member in X: $UX(P) = \{ x \in U/P \mid [x]P \cap X \neq \emptyset \text{ or } [x]P \cap X' \neq \emptyset \}$

where $X' = U - X$ (complement of X in U).

Now, let us explore the concepts of positive, negative, and boundary regions based on the equivalence relations P and Q over U:

Definition 4: For the equivalence relations P and Q over U, the following regions can be defined as the *positive region (PosQ(P))* is the set of all objects in U that can be classified into classes of U/Q using the information in attributes P *as*: $PosQ(P) = \{ x \in U \mid \text{for all } y \in U: y \; IND(P) \; x \text{ implies } y \; IND(Q) \; x \}$

The *negative region (NegQ(P))* is the set of all objects in U that cannot be classified into classes of U/Q using the information in attributes P *as*: $NegQ(P) = \{ x \in U \mid \text{for all } y \in U: x \; IND(P) \; y \text{ implies } x \; IND(Q) \; y \}$

The *boundary region (BndQ(P))* is the set of objects in U that belong neither to the positive region nor the negative region as: $BndQ(P) = U - (PosQ(P) \cup NegQ(P))$

The positive region helps us identify objects suitably classified with the information provided by attributes P, while the negative region includes objects not well-classified. The boundary region consists of objects that are ambiguous in their classification.

The notion of dependency in rough sets can be defined as:

Definition 5: For subsets $P, Q \subseteq A$, it is said that Q depends on P with a degree k $(0 \leq k \leq 1)$, denoted $P \Rightarrow k \; Q$, if the following condition holds: for every $x, y \in U: x \; IND(P) \; y$ implies $x \; IND(Q) \; y$ with a probability of at least k.

According to the formulation above, three cases can be assigned to k to determine the dependency relation:

1. If $k = 1$, then Q depends absolutely on P.
2. If $0 < k < 1$, then Q depends partially (to a degree k) on P.
3. If $k = 0$, then Q does not depend on P.

Sets of attributes can be generated through equivalence relations, and in this case, the comparison equivalence relations help achieve attribute reduction. A minimal subset is defined as a subset R of the conditional attribute set C, denoted as $\gamma R(D) = \gamma C(D)$. The set R of all reducts is defined as:

Definition 6: The set R of all reducts is given by: $R = \{ R \subseteq A \mid LX(R) = LX(A) \text{ for all } X \subseteq U \}$

The second point of reducts is about the *minimal cardinality Rmin* of the conditional attribute set:

Definition 7: The minimal cardinality *Rmin* of the conditional attribute set is given by: *Rmin = { R ⊆A | R* is a reduct and there is no subset *R'* of R such that *LX(R') = LX(R) for all X ⊆U }.*

The intersection of all sets in *Rmin* is called the *core*:

Definition 8: The core, denoted as *Core (U, A),* is the intersection of all sets in *Rmin given by*: *Core (U, A) = ∩{ R | R is in Rmin }*

Mostly, a minimal set of attributes called the minimal reducts was not only used to protect segmentation but also able to achieve the best classification for a dataset.

3.2 Attribute Reduction (AR) in Rough Set

Attribute reduction is a crucial aspect of rough set theory. It revolves around the concept of reducts, which are minimal subsets of attributes that preserve the partitioning of the universe and enable effective classifications.

A reducts subset is defined as a minimal set of attributes that retains the discriminatory power necessary for accurate classification [33, 34]. By selecting a reducts subset, redundant or irrelevant attributes are eliminated, reducing the dimensionality of the dataset while preserving the classification performance. This reduction in attributes not only simplifies the dataset but also improves the efficiency and interpretability of the resulting classification models.

The identification and utilization of reducts have significant applications in RST. By extracting reducts, we can uncover the essential attributes that are sufficient to achieve reliable classifications. These minimal attribute sets capture the core information needed for accurate decision-making, allowing for efficient data analysis and interpretation.

The concept of reducts and attribute reduction plays a leading role in various aspects of RST, including feature selection, data reduction, decision rule generation, and data-driven inference interpretation. By identifying minimal attribute subsets, attribute reduction facilitates efficient classification and data analysis by preserving the partitioning of the universe and the ability to perform accurate classifications. It enables researchers and practitioners to identify the most relevant attributes that contribute significantly to the classification process, improving the efficiency and effectiveness of data analysis and decision-making.

3.3 Rough Set Attribute Reduction (RSAR)

Rough Set Attribute Reduction (RSAR) serves as a filter-based tool for discovering concise knowledge from a given domain [31]. In the context of rough set theory (RST), the objective is to explore and understand the relationship between the conditional attributes and the decision attributes in a dataset.

RSAR focuses on finding a minimal subset of attributes from a given dataset, thereby selecting the most informative attributes for the task at hand. This process involves identifying and retaining the attributes that contribute significantly to the knowledge discovery and data mining tasks while eliminating or reducing the impact of unimportant or redundant attributes.

Large datasets often contain many attributes, which can lead to challenges such as increased computational complexity, decreased interpretability, and potentially irrelevant information. By employing attribute reduction using rough set theory, these challenges can be addressed by effectively reducing the dimensionality of the dataset and extracting the most relevant attributes.

The attribute reduction process in rough set theory aims to identify the minimal subset of attributes that can accurately capture the patterns and relationships in the data, ensuring that important knowledge is retained while eliminating unnecessary complexity. This reduction facilitates more efficient and effective data mining and knowledge discovery tasks.

By applying attribute reduction using rough set theory, researchers and practitioners can improve the quality and efficiency of knowledge discovery and data mining processes. The resulting reduced dataset contains a concise representation of the essential attributes, enabling more focused analysis and interpretation.

4 Strategy of Mining Rules & Classification Accuracy

The field of data mining involves various important tasks, including association rule discovery, sequential pattern discovery, classification, clustering, forecasting, deviation or anomaly detection, and regression. Each task requires specific methods and techniques to execute the necessary operations effectively. In the context of this research, the focus is on the classification task, for which several methods have been commonly used. These methods include:

- Decision Tree: Decision tree algorithms construct a tree-like model to represent decisions or classifications based on feature values. They partition the data based on attribute values and create branches that lead to different classes or outcomes.
- Bayesian Classifier: Bayesian classifiers apply probabilistic methods based on Bayes' theorem to classify data. They model the probability distribution of the target class given the attribute values and make predictions based on these probabilities.
- Artificial Neural Network: Artificial neural networks are computational models inspired by the biological neural networks. They consist of interconnected nodes (neurons) and learn from training data to make predictions or classifications.
- Genetic Algorithm: Genetic algorithms are inspired by the process of natural selection and evolution. They use evolutionary principles such as selection, crossover, and mutation to search for optimal solutions in a population-based manner.
- Fuzzy Logic: Fuzzy logic allows for the representation and manipulation of uncertainty or imprecise information in the classification process. It incorporates degrees of membership to different classes, enabling more flexible and nuanced classifications.
- Rough Set Theory: Rough set theory, as discussed earlier, provides a mathematical framework for handling uncertainty and incomplete data. It aims to identify common patterns and relationships in data and plays a significant role in attribute reduction and classification tasks.

These methods, among others, serve as tools for performing the classification task in data mining. Researchers and practitioners choose appropriate methods based on the nature of the data, the problem at hand, and the desired objectives. The selected method should be capable of effectively extracting patterns and making accurate predictions or classifications based on the available features and attribute values.

Supervised classification is a technique that utilizes the class (target) attribute of a dataset to guide the construction of a classification model. The dataset is typically divided into separate sets for training and testing purposes.

During the training phase, the classification model, also known as a classifier, is built using the training set. The classifier learns from the input features and their corresponding class labels to identify patterns and relationships that can be used for classification.

Once the model is constructed, it is evaluated and validated using the testing set. The performance of the classifier is assessed by comparing its predictions on the testing set with the true class labels. The accuracy of the model is calculated based on the number of correct predictions (hits) divided by the total number of instances in the testing set.

To ensure robustness and reliability of the evaluation, the experiments in your research adopted a 10-fold cross-validation scheme. In this scheme, the dataset is divided into 10 subsets or folds. The classification model is trained and tested 10 times, each time using a different fold as the testing set and the remaining folds as the training set. This process allows for a comprehensive evaluation of the model's performance across different subsets of the data.

The use of cross-validation, specifically 10-fold cross-validation, helps to mitigate the potential biases and variances that can arise from using a single training and testing split. It provides a more robust estimate of the classifier's accuracy and generalizability by incorporating multiple iterations of training and testing.

By employing the 10-fold cross-validation scheme, your experiments account for the training, testing, and validation sets, ensuring a thorough evaluation of the classification model's performance and accuracy.

The classification method employed in your research is based on *IF-THEN* rules, where each rule is defined as IF a certain condition is met, *THEN* a specific conclusion or class label is assigned to the instance.

The accuracy of a rule is assessed using the formula specified in Equation (1):

$$\boldsymbol{accuracy}(\boldsymbol{RULE}) = \frac{\boldsymbol{n}_{\boldsymbol{correct}}}{\boldsymbol{n}_{\boldsymbol{covers}}} \qquad (1)$$

where $\boldsymbol{n}_{\boldsymbol{correct}}$ represents the number of tuples (instances) that are correctly classified by the rule, while and $\boldsymbol{n}_{\boldsymbol{covers}}$ denotes the total number of tuples covered by the rule. A tuple is considered covered by the rule when the condition in the IF part of the rule is satisfied.

The accuracy of a rule quantifies its ability to correctly classify instances relative to the total number of instances it covers. It provides a measure of how well the rule performs in terms of correctly assigning class labels to the instances it applies to.

By calculating the accuracy of each rule, you can assess the performance and effectiveness of individual rules in the classification process. This information can help understand the strengths and weaknesses of specific rules and their contributions to the classification model's accuracy.

5 Re-heat Simulated Annealing (Re-heat SA)

In this section, some useful mechanisms in the Re-heat Simulated Annealing (Re-heat SA) approach are discussed. Several mechanisms are employed to effectively explore the solution space and find optimal solutions. These mechanisms include:

- Solution Representation: The Re-heat SA approach requires a suitable representation for the solutions. This representation defines how the attributes or variables are encoded and manipulated. It could be a binary representation, integer encoding, or any other suitable format depending on the problem at hand.
- Neighbourhood Structures: Define the set of moves or transformations that can be applied to a solution to generate neighbouring solutions. These structures determine the search space exploration strategy. In the context of Re-heat SA, neighbourhood structures define the possible changes that can be made to the current solution to generate alternative solutions.
- Solution Quality Measurement: Evaluate the quality of a solution, and a measurement or evaluation function is employed. This function assesses how well a solution satisfies the objectives or criteria of the problem. In the Re-heat SA approach, the solution quality measurement function assesses the performance or fitness of a solution based on the specific problem being addressed.
- Cooling Schedule: It controls the rate at which the temperature decreases during the simulated annealing process. It determines the exploration-exploitation trade-off by balancing between exploration (higher temperature) and exploitation (lower temperature). The cooling schedule plays a crucial role in the convergence and search behaviour of the algorithm.
- Re-heat Mechanism: This mechanism is a distinctive feature of the Re-heat SA approach. It addresses the situation where the algorithm gets stuck in a suboptimal solution or local minimum. When there is no improvement in the solution quality for a certain number of iterations, the Re-heat mechanism is triggered. It resets the temperature to its initial value, allowing the algorithm to explore the search space more extensively and potentially escape from local optima.

These mechanisms collectively contribute to the effectiveness and efficiency of the Re-heat SA approach. They enable the algorithm to explore the solution space, evaluate solution quality, and address the issue of being trapped in local optima. By utilizing these mechanisms, the Re-heat SA approach enhances the search process and improves the chances of finding better solutions.

5.1 Solution Representation

In the Re-heat SA approach, a solution is represented as a binary array. The size of the array is equal to the number of conditional features, denoted as $|C|$. Each cell of the array, denoted as $y_i \; for \; i = 1, 2, \cdots, |C|$, represents whether the corresponding attribute is included or not in the solution subset.

For example, let us consider an array *y* with a size of $|C|$. If the value of cell y_i is one, it indicates that the *ith* attribute is included in the solution subset. On the other hand, if the value of cell y_i is *zero*, it signifies that the *ith* attribute is not included in the solution subset.

To illustrate this representation, Fig. 1 depicts a subset of a solution. The figure shows an array where certain cells have a value of *one*, indicating that the corresponding attributes are included in the solution. In this example, attributes *1, 2, 4, 7, 9,* and *13* are part of the solution subset, as indicated by the ones in their respective cells.

Conditional Features	**1**	**1**	**0**	**1**	**0**	**0**	**1**	**0**	**1**	**0**	**0**	**0**	**1**
Subset Solution	**1**	**2**	**-**	**4**	**-**	**-**	**7**	**-**	**9**	**-**	**-**	**-**	**13**

Fig. 1. Solution representation

This binary array representation allows for a compact and efficient encoding of the solution, indicating which attributes are selected or excluded. It facilitates the manipulation and exploration of the solution space during the Re-heat SA algorithm, enabling the algorithm to search for optimal subsets of attributes that contribute to accurate classification.

5.2 Encoding Candidate Solutions

In the Re-heat SA approach, the generation of a trial solution plays a crucial role in exploring the search space and potentially finding near-optimal results. In this context, three different neighborhood structures are utilized to generate trial solutions:

1) NS1: In this neighbourhood structure, two attributes are randomly selected from the current solution, and their corresponding cells are changed from one to zero. This operation removes the selected attributes from the solution subset.

2) NS2: This neighbourhood is applied when the current solution is set as the best solution encountered so far. In this case, one attribute is randomly chosen from the current solution, and its corresponding cell is changed from one to zero. This operation removes a single attribute from the solution subset.

3) NS3: This neighborhood is employed when the current solution is set as an accepted worse solution. In this neighbourhood structure, one attribute is randomly selected, and its corresponding cell is modified by changing its value to either zero or one based on its current value. Additionally, another attribute is randomly chosen, and its cell is changed to either zero or one, also based on its current value. This operation adds or removes one attribute from the solution subset and modifies the value of the selected cells.

Each neighborhood structure represents a specific way of modifying the current solution to create a new candidate solution.

By employing these three neighborhood structures, the Re-heat SA approach explores various ways of modifying the current solution, allowing for diverse changes in the attribute subset. This promotes the exploration of the search space and enables the

algorithm to potentially find near-optimal solutions by iteratively generating and evaluating trial solutions.

5.3 Solution Quality Measurement

In the Re-heat SA approach, the quality of a solution is measured using the concept of dependency degree *(γ)* based on rough set theory. The dependency degree indicates the extent to which a solution satisfies the desired objectives or criteria.

When comparing two solutions, such as the current solution *(Sol)* and a trial solution *(Sol*)*, the trial solution is accepted if its dependency degree *(γ)* is greater than the dependency degree of the current solution *(γ)*. In other words, if $\gamma(Sol^*) > \gamma(Sol)$, the trial solution is considered superior in terms of meeting the desired criteria.

In situations where the dependency degrees of both solutions are the same, a tiebreaker rule is applied based on the cardinality of the solutions. The solution with a lower cardinality, i.e., a smaller number of attributes, is preferred or accepted.

This measurement approach allows the Re-heat SA algorithm to compare and evaluate different solutions based on their dependency degrees. Solutions with higher dependency degrees are considered to have better quality in terms of capturing patterns and relationships in the data.

By utilizing this solution quality measurement based on dependency degrees and considering the cardinality tiebreaker rule, the Re-heat SA approach promotes the selection of solutions that exhibit stronger dependencies while favoring solutions with fewer attributes, leading to more concise and informative solutions.

5.4 Cooling Schedule

The cooling schedule in the Re-heat SA approach plays a critical role in determining the trade-off between exploration and exploitation during the optimization process. It influences the convergence behavior and the quality of the final solution.

The choice of a cooling schedule impacts the algorithm's search strategy. A faster cooling schedule leads to quicker convergence to local optima, while a slower cooling schedule allows for a more extensive search and potentially higher-quality solutions.

In this implementation, the initial temperature (T_0) is set to *2* times the cardinality of the conditional attributes $(|C|)$. The initial temperature represents the exploration phase of the algorithm, allowing for a more extensive search initially.

The temperature is updated using the formula specified in Equation (2):

$$T(t+1) = \alpha * T(t) \qquad (2)$$

In this equation, $T(t)$ represents the temperature at iteration t, and α is a constant related to the temperature update. In this work, α is set to *0.93*, which indicates a gradual reduction of temperature over iterations. (Alternatively α can be set to operate in a range between greater than a lower bound value like *0.83* and less than or equal to an upper bound value like *0.93.)*

The choice of α is adapted from Jensen and Shen [12], who have determined the value based on their empirical observations and experimentation.

The algorithm starts with an initial high temperature, allowing for more exploration, and gradually decreases the temperature to focus on exploiting promising regions of the

search space. This approach aims to find high-quality solutions while also considering the computational efficiency of the algorithm.

5.5 Re-heat Mechanism

The Re-heat mechanism in the Re-heat SA approach is a crucial component that addresses the issue of getting stuck in local optima. It involves resetting the current temperature to the initial temperature *(T_0)* and allowing the algorithm to explore the search space more extensively.

The Re-heat mechanism is triggered when there is no improvement in the solution quality for a certain number of iterations. Specifically, if the solution quality does not improve for a predetermined number of non-improving solutions, the mechanism is activated.

When the Re-heat mechanism is triggered, the current temperature is reset to the initial temperature *(T_0)*. Additionally, the best solution obtained thus far is set as the current solution. This reset allows the algorithm to have more flexibility in accepting worse solutions during exploration.

By resetting the temperature and accepting worse solutions, the Re-heat SA algorithm can escape from local optima and continue searching for better solutions. It promotes exploration and increases the chances of finding globally optimal or near-optimal solutions.

In this work, the number of non-improving solutions is set to 3. This means that if the solution quality does not improve for three consecutive iterations, the Re-heat mechanism is activated. The specific value of 3 is determined empirically, based on preliminary experiments, and can be adjusted based on the characteristics of the problem being solved and the algorithm's behavior.

5.6 Proposed Algorithm: Re-heat SA

In the Re-heat SA algorithm, the optimization process is summarized as follows:

1. Initialisation: Generate an initial solution *(Sol)* at random. Encode the solution using a binary array representation, as shown in Figure 1. Set the initial temperature, maximum number of iterations *(NumOfIteration)*, and maximum number of non-improving solutions *(MaxUnImprovement)*. Calculate the initial dependency degree *($\gamma(Sol)$)* and the number of attributes *(cardinality, $|Sol|$)* for the initial solution.
2. Termination Criterion: Check if the termination criterion *(NumOfIteration)* is met. If the maximum number of iterations is reached, the algorithm terminates and returns the best solution obtained so far. Otherwise, the algorithmic process continues as follows:
    i. Generate Trial Solution: Generate a trial solution *(Sol*)* using one of the neighbourhood structures discussed previously. The neighbourhood structure determines how the trial solution is generated by modifying the current solution.
    ii. Calculate Quality: Calculate the dependency degree *(γ)* of the trial solution, *$\gamma(Sol^*)$,* using rough set theory.

iii. Compare Solutions: Compare the quality of the trial solution, $\gamma(Sol^*)$, with the quality of the current solution, $\gamma(Sol)$. If $\gamma(Sol^*)$ is greater than or equal to $\gamma(Sol)$, the trial solution is accepted, and the current solution is updated $(Sol \leftarrow Sol^*)$. Additionally, if the quality of the trial solution is better than the quality of the best solution obtained so far $(\gamma(Sol^*) > \gamma(Sol\_best))$, update the best solution $(Sol\_best \leftarrow Sol^*)$. Furthermore, if the trial solution has the same quality as the current solution but a lower number of attributes $(|Sol^*| < |Sol|)$, it is also accepted.

iv. Iteration Update: Update the iteration counter and go back to step 2 to continue the optimization process until the termination criterion is met.

The Re-heat SA algorithm iteratively generates trial solutions, evaluates their quality using the dependency degree, and updates the current solution and the best solution based on certain acceptance criteria. By exploring different solutions and accepting worse solutions probabilistically, the algorithm aims to escape local optima and search for better solutions with higher dependency degrees and potentially lower cardinalities. Moves that improve the solution, with respect to the objective function, will always be accepted, whilst bad moves are accepted with a certain probability determined by the Boltzmann probability, P, calculated by using Equation (3):

$$P = e^{-\beta / T} \qquad (3)$$

where β is the difference in the objective function evaluation between the current and the trial solutions, and T is the temperature parameter, which periodically decreases during the search process, according to some cooling schedule.

The algorithm continues this iterative process until the termination criterion (maximum number of iterations) is reached. At that point, the algorithm terminates and returns the best solution obtained throughout the optimization process.

Note that the specific details of the neighborhood structures and their generation process are not provided in this summary, but were discussed earlier.

In the Re-heat SA algorithm, the probability of accepting a worse trial solution is determined based on Equation (4):

$$exp\left[\frac{\left(-\left(\gamma(Sol^*)-\gamma(Sol)\right)\right)}{T} \geq RandNum[0,1]\right] \qquad (4)$$

In this equation, T represents the current temperature, $\gamma(Sol^*)$ is the dependency degree of the trial solution, $\gamma(Sol)$ is the dependency degree of the current solution, and $RandNum[0,1]$ is a randomly generated number between 0 and 1.

To accept a worse trial solution, the exponential difference between the dependency degrees of the trial and current solutions divided by the current temperature should be greater than or equal to the randomly generated number. This probabilistic condition ensures that worse solutions have a chance of being accepted during the optimization process, allowing for exploration and potential escape from local optima.

If the condition in Equation (4) is not satisfied for a certain number of iterations, it indicates that there is no improvement in the solution quality. In this case, the re-heat mechanism is invoked. The current temperature is reset to the initial temperature $(T \leftarrow T_0)$, and the best solution obtained so far is set as the current *solution* $(Sol \leftarrow Sol\_best)$. This allows the algorithm to explore the search space again and potentially find better solutions.

The process of generating trial solutions, evaluating their quality, and accepting or rejecting them based on the probabilistic condition continues until the termination criterion is met. The termination criterion can be defined based on the maximum number of iterations or other stopping conditions specific to the problem being solved.

The algorithm repeats this iterative process until the stopping condition is met, resulting in an optimized solution with higher dependency degrees, that balances exploration and exploitation. This is illustrated in Algorithm 1.

Algorithm 1: The Pseudo Code for Re-heat SA Process

```
Variables / Operators:
Sol       // Initial solution
Sol*      // Trial solution
Solbest   // Best solution
γ         // Cost function
||        // The number of attributes of Sol
BEGIN
   for NumOfIteration do
      T = T * α
      generate Sol*
      if (γ(Sol*) ≥ γ(Sol) && |Sol*| ≤ |Sol|) then
         Sol = Sol*
         Solbest = Sol*
         UnImprovement = 0
      else
         β = γ(Sol*) - γ(Sol)
         generate RandomNum in [0, 1]
         if (RandomNum ≤ e^(-β / T)) then
            Sol = Sol*
            UnImprovement = 0
         else
            UnImprovement++
         end if
      end if
      if (UnImprovement == MaxUnImprovement) then
         T = T0
         UnImprovement = 0
         Sol = Solbest
      end if
   end for
   return Solbest, γ(Solbest), |Solbest|
END
```

6 Experimental Design

The overall experimental design for the solution proposed in this work encompasses several steps, as illustrated in Fig. 2.

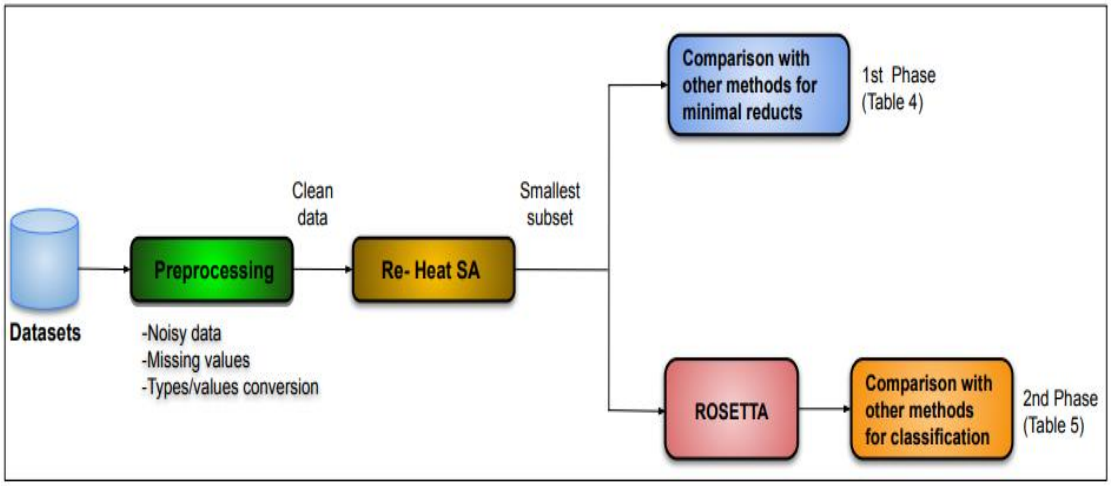

Fig. 2. Experiments stratagem

Each one of the steps is defined as follows:

- Preprocessing: The first step involves preprocessing the dataset to clean it from vague, imprecise, and inconsistent data. This process aims to enhance the quality of the data by addressing issues such as missing values, outliers, and noise.
- Attribute Reduction: The cleaned dataset is then used for attribute reduction, which involves finding the minimal reducts. Attribute reduction is performed using the Re-heat SA algorithm, as described earlier. The algorithm selects the most informative attributes by optimizing the search in the attribute space.
- Classification: After attribute reduction, the reduced dataset is used for classification. The classification task involves building a classifier using the "IF-THEN" rules generated from the reduced attribute set. The classifier aims to accurately assign class labels to the instances in the dataset.
- Accuracy Estimation: To estimate the classification accuracy, the Rosetta software [37, 38] is utilized. Rosetta is a tool commonly used for evaluating the performance of classification models. It assesses the accuracy of the classifier based on the correctly classified instances and the total number of instances.
- Comparison to Existing Approaches: The experimental results obtained from the proposed Re-heat SA algorithm are compared to other approaches available in the literature. These approaches may include different methods for attribute reduction and classification. The comparison is conducted to assess the effectiveness and performance of the proposed method relative to existing techniques.

By following this experimental design, the proposed Re-heat SA algorithm's performance in terms of attribute reduction and classification accuracy can be evaluated.

The preprocessing step ensures the dataset's quality before applying the algorithm, and the comparison with existing approaches provides insights into the algorithm's competitiveness and potential advantages.

In the evaluation of the experiments, a total of 13 well-known datasets from the UCI Machine Learning Repository [29] are utilized. These datasets are selected to test the performance of the proposed Re-heat SA algorithm. Table 1 provides an overview of the

main characteristics of these datasets, including the number of dataset instances, attributes, and classes.

Table 1. Characteristics of the selected datasets

Datasets	No. of Attributes	No. of Classes
M-of-N	13	1000
Exactly	13	1000
Exactly2	13	1000
Heart	13	294
Vote	16	300
Credit	20	1000
Mushroom	22	8124
LED	24	2000
Letters	25	26
Derm	34	366
Derm2	34	358
WQ	38	521
Lung	56	32

By using these 13 datasets, the proposed algorithm was tested and evaluated on a diverse range of data characteristics, allowing for a comprehensive assessment of its performance in attribute reduction and classification tasks.

In the datasets shown in Table 1, nine of them (M-of-N, Exactly, Exactly2, Heart, Vote, Credit, LED, Letters, and WQ) do not have missing values, and the class attribute is in the last column. Therefore, these datasets do not require any preprocessing steps.

Four datasets, however, required preprocessing due to missing values and other data issues:

- Mushroom dataset: This dataset is the largest among the chosen datasets. Preprocessing steps are performed, including data cleaning, and converting some characters to the integer type. These steps are necessary to make the data suitable for the calculation of rough sets.
- Derm dataset: Preprocessing is needed for the Derm dataset due to missing values and noisy data. Additionally, discretization is applied to the age attribute.
- Derm2 dataset: The Derm2 dataset is created by removing objects with missing values from the Derm dataset. After removing these objects, the Derm2 dataset is formed with 358 objects. The preprocessing steps for Derm2 are the same as the Derm dataset, except for handling missing values.
- Lung dataset: The Lung dataset has the largest number of attributes among all the chosen datasets. It differs from the others in terms of the class attribute, which is in the first column. Additionally, the Lung dataset contains missing values at objects 15, 19, 21, and 26.

For each dataset, the Re-heat SA algorithm is executed multiple times with different initial solutions. Specifically, the algorithm is run 20 times for each dataset, except for Heart, Vote, and Derm2 datasets, which are executed 30, 30, and 10 times, respectively. These numbers were determined based on previous works [12, 14]. More recently, the authors from [36] in their experimental work used a similar approach.

The results obtained from the Re-heat SA algorithm are analyzed in two phases: minimal reducts and rules generation for accuracy. This allows for the evaluation of the algorithm's performance in finding minimal reducts and generating rules for classification. The accuracy of the generated rules is also assessed to determine the classification performance of the algorithm on each dataset.

By conducting these experiments and analyzing the obtained results, the effectiveness and performance of the Re-heat SA algorithm can be evaluated for each dataset.

7 Results

The experiments were conducted in two phases, and the outcomes for each phase are presented below.

7.1 Results of the First Phase (Minimal Reducts)

In this study, we introduced a method called Re-heat SA. In the first phase of the experiments, the focus was on finding minimal reducts using the Re-heat SA algorithm. The parameters of the Re-heat SA algorithm were set as shown in Table 2.

Table 2: Re-heat SA Parameter Settings

Parameter	Definition	Value
T_0	Initial temperature	2*|C|
T(t+1)	Decreasing temperature rate	T * 0.93
Max-un-improvement	Maximum number of unimproved solutions	3
NumOfIteration	Maximum number of iterations	250

The initial temperature (T_0) was set as 2 times the number of classes. The temperature was decreased according to the formula T(t+1) = T * 0.93, where T represents the current temperature.

A maximum number of 3 unimproved solutions was allowed before invoking the re-heat mechanism. The maximum number of iterations was set to 250.

These parameter settings were chosen based on previous works [12, 14] and were used consistently across all datasets.

By running the Re-heat SA algorithm with these parameter settings, the minimal reducts for each dataset were obtained. The results of the first phase, concerning the minimal reducts for each dataset, provide insights into the effectiveness of the algorithm in identifying minimal subsets of attributes that are relevant for classification tasks. The

results of the second phase, concerning the classification task, focus on rules generation and accuracy.

In the first phase of the experiments, the minimal reducts generated by Re-heat SA for the M-of-N dataset are presented in Table 3. The "No. of Attributes of Reducts" column indicates the number of attributes in each minimal reducts, which is 6 for all runs. The "Minimal Reducts" column lists the attributes included in the minimal reducts for each run. The "Run Time(s)" column shows the execution time in seconds for each run.

Across all 20 runs, Re-heat SA consistently identified reducts comprising 6 attributes, specifically attributes 3, 5, 9, 11, 1, and 7, from the original set of 13 attributes. The specific attributes included in the minimal reducts varied across different runs.

These results demonstrate the ability of the Re-heat SA algorithm to consistently find minimal reducts with a fixed number of attributes for the M-of-N dataset. The algorithm explores different combinations of attributes to identify the most informative subset that preserves the classification accuracy.

Table 3: Observed reducts generated by Re-heat SA for minimal subset reducts in M-of-N dataset

M-of-N	No. of Attributes of Reducts	Minimal Reducts	Run Time(s)
1	6	3, 5, 9, 11, 1, 7	130
2	6	3, 5, 11, 7, 1, 9	117
3	6	9, 11, 3, 1, 5, 7	146
4	6	1, 3, 9, 11, 7, 5	175
5	6	1, 3, 11, 9, 5, 7	157
6	6	1, 5, 7, 3, 9, 11	179
7	6	1, 5, 7, 11, 3, 9	127
8	6	3, 5, 9, 7, 1, 11	151
9	6	7, 11, 5, 9, 3, 1	151
10	6	3, 5, 7, 11, 9, 1	115
11	6	5, 11, 1, 3, 7, 9	180
12	6	11, 9, 7, 3, 5, 1	182
13	6	5, 11, 7, 9, 3, 1	195
14	6	1, 3, 5, 11, 9, 7	136

15	6	3, 5, 9, 7, 11, 1	100
16	6	3, 11, 9, 7, 1, 5	103
17	6	1, 5, 11, 7, 3, 9	110
18	6	3, 5, 11, 8, 1, 7	125
19	6	1, 3, 7, 9, 11, 5	189
20	6	1, 9, 11, 7, 3, 5	121

Table 4 presents a comparison of subset reducts results for different datasets using various algorithms, including Re-heat SA. The table includes the datasets, followed by the results obtained by different algorithms such as TSAR, SimRSAR, AntRSAR, GenRSAR, ACOAR, SSAR, and Re-heat SA.

In the comparison of various methods for finding minimal reducts, we considered two types of approaches: single-based solution approaches and population-based approaches.

The selected methods for comparison were as follows:

- TSAR: Tabu search [14]
- SimRSAR: Simulated annealing [11]
- AntRSAR: Ant algorithm [12]
- GenRSAR: Genetic algorithm [11]
- ACOAR: Ant colony optimization [13]
- SSAR: Scatter search [15]

The best minimal reducts (minimum number of attributes) obtained by each method are highlighted in bold. The superscripts in parentheses represent the number of runs in which the methods achieved the minimal reducts. When no superscripts are present, it indicates that the method consistently obtained these minimal reducts across all runs.

Upon analysing the results presented in Table 4, we observe that Re-heat SA demonstrates comparable performance with TSAR and SimRSAR in the first comparison. However, there is a notable exception for the WQ dataset, where Re-heat SA performs worse, resulting in minimal reducts equal to 15, while TSAR and SimRSAR achieve 12 and 13 minimal reducts, respectively.

Table 4: Comparison of subset reducts results (* reflects best results)

Datasets	TSAR	SimRSAR	AntRSAR	GenRSAR	ACOAR	SSAR	**Re-heat SA**
M-of-N	**6***	**6***	**6***	$6^{(6)}7^{(12)}$	**6***	**6***	**6***
Exactly	**6***	**6***	**6***	$6^{(10)}7^{(10)}$	**6***	**6***	**6***
Exactly2	**10***	**10***	**10***	$10^{(9)}11^{(11)}$	**10***	**10***	**10***
Heart	**6**	$6^{(29)}$ $7^{(1)}$	$6^{(18)}$ $7^{(2)}$	$6^{(18)}7^{(2)}$	**6***	**6***	$6^{(16)}7^{(14)}$
Vote	**8**	$8^{(15)}$ $9^{(15)}$	8	$8^{(2)}9^{(18)}$	**8***	**8***	8
Credit	$8^{(13)}$ $9^{(5)}$ $10^{(2)}$	$8^{(18)}$ $9^{(1)}$ $11^{(1)}$	$8^{(12)}$ $9^{(4)}$ $10^{(4)}$	$10^{(6)}11^{(14)}$	$8^{(16)}$ $9^{(4)}$	$8^{(9)}$ $9^{(8)}$ $10^{(3)}$	$9^{(2)}10^{(7)}11^{(11)}$
Mushroom	$4^{(17)}$ $5^{(3)}$	**4***	**4***	$5^{(1)}6^{(5)}7^{(14)}$	**4***	$4^{(12)}$ $5^{(8)}$	$4^{(8)}5^{(12)}$
LED	**5***	**5***	$5^{(12)}$ $6^{(4)}$ $7^{(3)}$	$6^{(1)}7^{(3)}8^{(16)}$	**5***	**5***	**5***
Letters	$8^{(17)}$ $9^{(3)}$	**8***	**8***	$8^{(8)}9^{(12)}$	**8***	$8^{(5)}$ $9^{(15)}$	$8^{(6)}9^{(10)}10^{(4)}$
Derm	$6^{(14)}$ $7^{(6)}$	$6^{(12)}$ $7^{(8)}$	$6^{(17)}$ $7^{(3)}$	$10^{(6)}11^{(14)}$	**6***	**6***	$6^{(10)}7^{(10)}$
Derm2	$8^{(2)}$ $9^{(14)}$ $10^{(4)}$	$8^{(3)}$ $9^{(7)}$	$8^{(3)}$ $9^{(17)}$	$10^{(4)}11^{(16)}$	***8$^{(4)}$**$9^{(16)}$	$8^{(2)}$ $9^{(18)}$	$9^{(6)}$ $10^{(4)}$
WQ	$12^{(1)}$ $13^{(13)}$ $14^{(6)}$	$13^{(16)}$ $14^{(4)}$	$12^{(2)}$ $13^{(7)}$ $4^{(11)}$	16	***12$^{(4)}$**$13^{(12)}14^{(4)}$	$13^{(4)}$ $14^{(16)}$	$15^{(5)}14^{(15)}$
Lung	$4^{(6)}$ $5^{(13)}$ $6^{(1)}$	$4^{(7)}$ $5^{(12)}$ $6^{(1)}$	**4***	$6^{(8)}7^{(12)}$	**4***	**4***	$5^{(12)}6^{(8)}$

Legend: *For each dataset in Table 4, the best result is marked with an asterisk (*). The numbers in parentheses represent the rank or frequency of occurrence of the corresponding reducts across multiple runs or algorithms.*

Regarding the second comparison, Re-heat SA remains competitive with AntRSAR and outperforms GenRSAR on all tested datasets. However, when compared to ACOAR and SSAR, in most cases, these two approaches perform better than Re-heat SA. Nevertheless, there is a tie (equal finish) on 5 datasets, namely M-of-N, Exactly, Exactly2, Vote, and LED.

As a result, Re-heat SA can be considered a viable alternative approach for finding minimal reducts. The incorporation of a re-heat mechanism, where the temperature is set back to a higher value during the search, facilitates exploration. By accepting solutions that may initially worsen the current solution, the algorithm can escape local optima and delay convergence. The dynamically changing neighbourhood structure based on the current solution's quality aids in exploring the search space and finding better solutions.

The performance of the proposed algorithm is influenced by parameter settings, such as the MaxUnImprovement, which determines the occurrence of the re-heat mechanism when there is no improvement in solution quality after a certain number of iterations. These parameter values may vary from one dataset to another, depending on dataset features, leading to differing results. Thus, proper tuning of parameters holds the potential to enhance the quality of results obtained by Re-heat SA, and this remains a subject for future work and investigation.

In the first comparison with single-based solution approaches (TSAR and SimRSAR), Re-heat SA demonstrates comparable performance, achieving minimal reducts like those

obtained by TSAR and SimRSAR on most datasets. However, there is a slight deviation on the WQ dataset, where Re-heat SA performs marginally worse.

Conversely, in the second comparison with population-based approaches (AntRSAR, GenRSAR, ACOAR, and SSAR), Re-heat SA performs better than GenRSAR on all tested datasets. However, it is outperformed by ACOAR and SSAR on most datasets, indicating that these methods have a competitive advantage in finding minimal reducts. Nonetheless, there are instances where Re-heat SA achieves equal performance (a tie) with ACOAR and SSAR on datasets such as M-of-N, Exactly, Exactly2, Vote, and LED.

The re-heat mechanism employed in Re-heat SA offers an alternative approach to finding minimal reducts. By resetting the temperature to its initial value and accepting worse solutions, the algorithm introduces exploration into the search process. This exploration allows the algorithm to move away from local optima and delay convergence, thereby increasing the chances of finding a better solution. Fine-tuning the parameters of Re-heat SA, including MaxUnImprovement, T_0, cooling schedule, and NumOfIteration, can enhance the quality of the obtained results. Finding the optimal parameter values for each dataset can lead to improved performance and better exploration of the search space.

Our observations indicate that Re-heat SA performs better on certain datasets (M-of-N, Exactly, Exactly2, Vote, and LED) and worse on others (Credit, Derm2, WQ, and Lung). This suggests that the algorithm's performance is influenced by the specific dataset properties and the interplay between the parameter settings and the dataset characteristics.

The probability of accepting a worse solution is determined by the temperature, with higher temperatures corresponding to a higher probability. This flexibility enables the algorithm to explore different regions of the search space and avoid getting trapped in local optima. The combination of the re-heat mechanism, temperature control, and adaptive neighbourhood structures in Re-heat SA provides a powerful approach for exploring and finding minimal reducts, offering an alternative solution to tackle the optimization challenges in attribute reduction. By adapting the neighbourhood structure based on the quality of the current solution, Re-heat SA can dynamically adjust its exploration strategy, improving its efficiency and effectiveness.

We believe that the parameter settings can significantly impact the performance of the Re-heat SA algorithm. The value chosen for MaxUnImprovement, which determines when the re-heat mechanism is triggered, can vary depending on the characteristics of the dataset. This adaptive behaviour allows the algorithm to effectively explore the search space and jump between different regions, increasing the likelihood of finding an optimal or near-optimal solution. However, different datasets may require various levels of exploration to find the best settings for specific problem domains before finding an optimal solution.

The results demonstrate the performance of Re-heat SA in terms of finding minimal subset reducts for different datasets. It can be observed that Re-heat SA achieves competitive results, often obtaining the best or close-to-best reducts compared to other algorithms for both single-based and population-based solution approaches. The specific results vary across datasets, indicating the effectiveness of Re-heat SA in adapting to different problem domains. However, its performance may vary depending on the specific dataset, and there are cases where other methods excel. Further analysis and comparisons could be conducted to gain deeper insights into the strengths and weaknesses of each method. Parameter tuning is an important aspect in optimizing the performance of any algorithm, and it is worth exploring different parameter configurations to find the most suitable settings for specific datasets and application domains. These aspects can be a

direction for future research and experimentation to further optimize the Re-heat SA algorithm.

Overall, the results highlight the capability of Re-heat SA to identify informative attribute subsets that could improve the classification accuracy in various datasets.

7.2 Results of the Second Phase (Rules Generation and Classification Accuracy)

This paper addresses the principal issue between classification accuracy and minimal reducts in data mining.

Table 5 provides a comparison of the number of rules generated and the classification accuracy for different algorithms on various datasets. The algorithms compared include Reheat Simulated Annealing Algorithm, Genetic Algorithm, Johnson's Algorithm, and Holte's 1R.

Table 5: Average number of rules and the classification accuracy for each algorithm on each dataset

	Reheat Simulated Annealing Algorithm		**Genetic Algorithm**		**Johnson's Algorithm**		**Holte's 1R**	
Datasets	***Average # of rules***	***Accuracy %***	***Average # of rules***	***Accuracy %***	***Average # of rules***	***Accuracy %***	***Average # of rules***	***Accuracy %***
M-of-N	62	**100**	64	**100**	64	**100**	26	63
Exactly	607	**100**	64	**100**	64	**100**	26	69
Exactly2	64	62	607	62	607	62	26	76
Heart	262	8	6407	25	261	7	67	64
Vote	136	66	1125	73	125	72	48	88
Credit	883	15	129259	28	882	14	83	63
Mushroom	82	**99**	6401	**100**	90	**100**	117	90
LED	10	**100**	9234	**100**	10	**100**	48	64
Letters	23	46	4388	46	24	46	51	46
Derm	231	33	57231	70	306	11	135	49
Derm2	253	40	50052	78	212	33	128	60
WQ	136	**69**	87267	40	461	4	94	52
Lung	23	35	3467	72	20	49	156	72

Re-heat SA results are compared to three algorithms available in Rosetta software, Genetic Algorithm (GA), Jonson's Algorithm, and Holte's 1R. According to Fig. 2, GA had a significant increase in the average number of generated rules.

It is worth mentioning some general observations and insights on the relationship between classification accuracy, the number of generated rules, and the performance of different algorithms. In addition, in data mining, the trade-off between classification accuracy and the number of generated rules is often a key consideration. Generating more rules may lead to a more comprehensive representation of the data and higher accuracy. However, it can also increase the complexity of the model and make it harder to interpret. The performance of different algorithms can vary depending on the dataset and problem at hand. It is essential to assess the trade-off between accuracy and the number of rules generated, and consider the interpretability and complexity of the resulting model.

Therefore, when comparing Re-heat SA to other algorithms like Genetic Algorithm (GA), Johnson's Algorithm, and Holte's 1R, it is important to consider both the average number of generated rules and the accuracy achieved by each algorithm. GA may have a higher number of generated rules, indicating a more comprehensive exploration of the solution space. However, it is crucial to analyze whether this increase in rules translates into a significant improvement in accuracy or if it introduces unnecessary complexity.

Figures 3 and 4 address the average number of rules generated by different methods respectively. While both figures focus on the average number of rules, they are separated due to differences in scale. In Fig. 3, the Genetic Algorithm (GA) generates rules close to 14,000 for the Credit dataset, while the other two approaches (shown in Fig. 4) do not exceed 1,000 rules for all datasets.

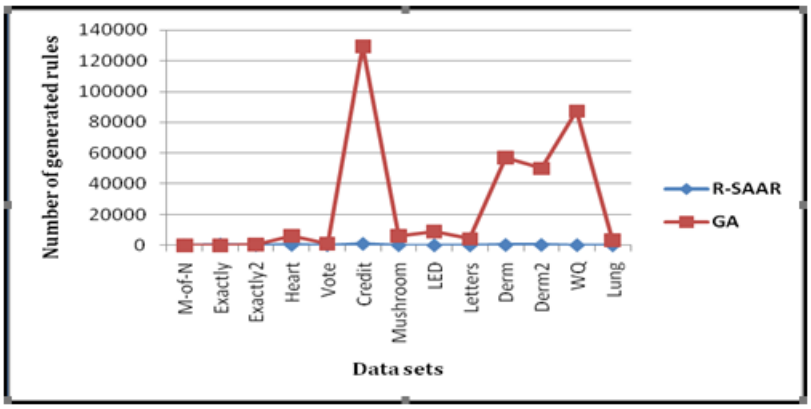

Fig. 3. Comparison of average number of rules in Re-heat SA and GA

Fig. 4 also reveals that Re-heat SA, when compared to the Janson and Holter methods, demonstrates promising results for the average number of rules. Additionally, Table 5 provides further insights into the comparison of these methods, specifically for datasets such as Credit, Vote, Heart, Exactly, and Led. In these cases, Re-heat SA generates more rules or the same number of rules in relation to the other methods.

The discrepancy in the number of rules generated has implications for the practical utility and interpretability of the generated rule sets. Having a significantly higher number of rules may lead to complex models that are harder to comprehend and apply in real-world scenarios. On the other hand, generating fewer rules that still effectively capture the essence of the data is desirable for better interpretability and ease of use in decision-making processes.

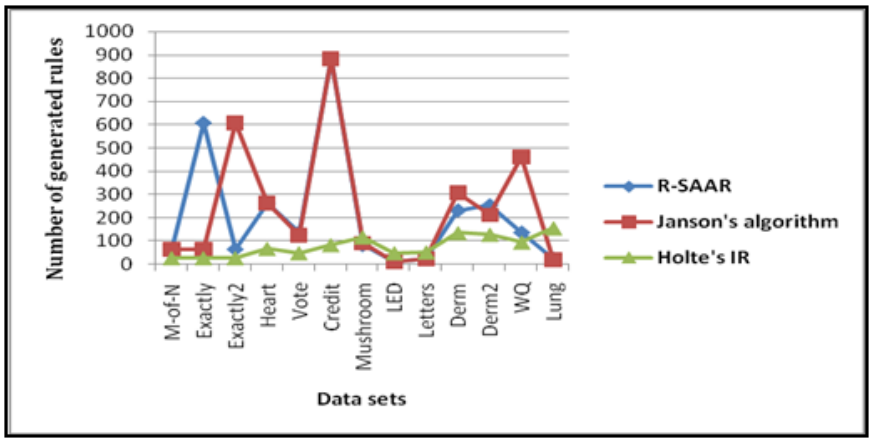

Fig. 4. Comparison of average number of rules among Re-heat SA, Janson & Holte's 1R

Therefore, the results from both Fig. 4 and Table 5 indicate that Re-heat SA performs competitively in terms of the average number of rules generated compared to other methods, especially for certain datasets. Nevertheless, further analysis is necessary to understand the specific patterns and reasons behind the variations in the number of rules across different datasets. It is possible that the characteristics of the datasets, the complexity of the problem, and the algorithm's parameters may all contribute to these variations. As such, investigating the interplay between these factors could be an avenue for future research to optimize the rule generation process and enhance the overall performance of Re-heat SA.

Fig. 5 presents the percentage of accuracy of Re-heat SA compared to the other three selected methods. Since these algorithms use random strategies and have different parameters, it is challenging to determine the most appropriate one among them with certainty. However, these algorithms are chosen as alternative solutions to find nearby optimal solutions, rather than exact optimal solutions.

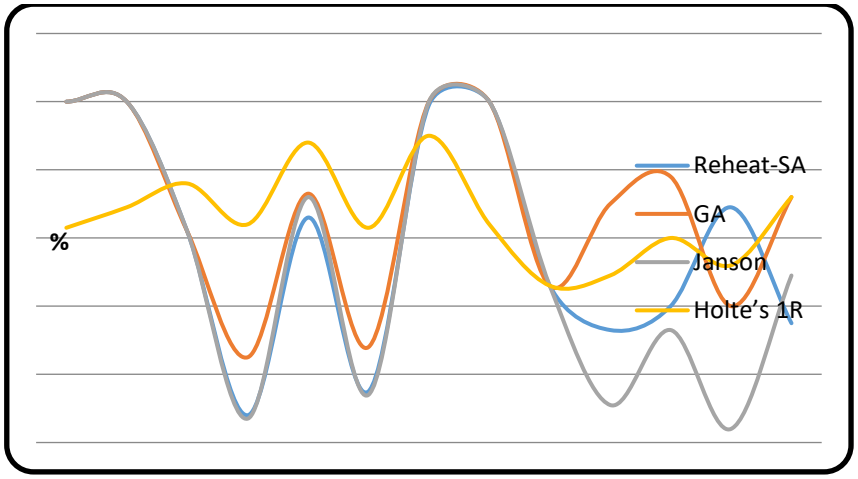

Fig. 5. Comparison of accuracies among Re-heat SA, Janson, GA and Holte's 1R

In general, Re-heat SA demonstrates reliable performance in solving attribute reduction in rough set theory based on the following factors:

- The number of reducts generated with a smaller cardinality, indicating a more concise representation of the dataset.
- Shorter runtime, suggesting faster computational efficiency compared to other methods.
- The quality of the generated reducts, indicating that the selected attributes are relevant and informative for the classification task.
- Promising results in terms of classification accuracy, which implies that Re-heat SA is effective in accurately classifying the data.

These observations highlight the potential strengths of Re-heat SA in attribute reduction and classification tasks within the domain of rough set theory. However, it is important to note that the algorithm's performance may still be influenced by various factors such as dataset characteristics, parameter settings, and the specific problem at hand. Further analysis and experimentation may be necessary to validate and further explore the capabilities of Re-heat SA.

The experimental results indicate that Re-heat SA outperforms other classification schemes, as evidenced by the classification accuracy measured using Rosetta software. Among the tested datasets, three datasets achieved a classification accuracy of 100%, indicating that Re-heat SA was able to accurately classify all instances in those datasets. Four datasets achieved classification accuracy between 60% and 99%, indicating a substantial level of accuracy. Additionally, six datasets achieved classification accuracy between 30% and 59%, suggesting that Re-heat SA was still able to achieve reasonable accuracy on those datasets.

These results demonstrate the effectiveness of Re-heat SA in achieving high classification accuracy across a range of datasets. It indicates that Re-heat SA can accurately classify instances and can be a valuable tool in solving classification problems. However, it is important to note that the actual classification accuracy achieved by Re-heat SA may vary depending on the specific dataset and problem at hand. Further evaluation and comparison with other classification schemes would provide a more comprehensive understanding of its performance.

7.3 Calculated running time

The calculated running time for Re-heat SA shows that in some cases it is longer compared to other methods, such as AntRSAR, GenRSAR, and SimRSAR. For example, in the LED dataset, Re-heat SA takes longer compared to AntRSAR, but both methods achieve the same number of attributes in the smallest reducts. Similar observations can be made for the WQ dataset, where Re-heat SA takes longer than SimRSAR and GenRSAR, but less time compared to AntRSAR.

In the case of the Mushroom dataset, Re-heat SA requires more computational time to discover the smallest reducts compared to AntRSAR, GenRSAR, and SimRSAR. This can be attributed to the larger size of the Mushroom dataset, which has 8124 objects. Similarly, the WQ dataset also requires longer time in Re-heat SA compared to SimRSAR and GenRSAR, but less time compared to AntRSAR.

It is important to note that the number of objects in each dataset plays a significant role in decreasing the running time for Re-heat SA. This is evident in the Mushroom, LED, and WQ datasets, which have more objects.

Despite the longer execution times for some datasets compared to other methods, Re-heat SA is able to discover results with a smaller cardinality, indicating its effectiveness in finding minimal reducts. Therefore, the proposed approach of Re-heat SA outperforms some of the other approaches, as demonstrated in Fig. 6.

Based on the findings of Jensen and Shen [12], SimRSAR and AntRSAR were reported to outperform the other methods in terms of running time, with the order of techniques being SimRSAR ≤ AntRSAR ≤ GenRSAR. However, when comparing these methods to Re-heat SA, the order of techniques in terms of running time may change to Re-heat SA ≤ SimRSAR ≤ AntRSAR ≤ GenRSAR for 10 datasets.

Additionally, the results in Table 3 demonstrate the advantage of the Re-heat SA method in terms of discovering minimal reducts with minimal cardinality compared to other methods. This indicates that Re-heat SA can find more compact and concise solutions, which can be beneficial in interpretability and efficiency.

Therefore, while SimRSAR and AntRSAR may have performed better in terms of running time according to previous studies, the Re-heat SA method shows its superiority in terms of discovering minimal reducts with minimal cardinality, as evident from the results in Table 3.

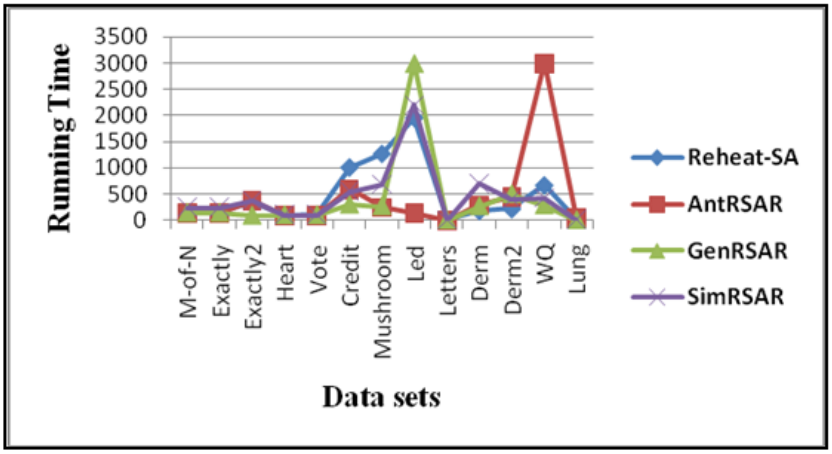

Fig. 6. Comparison of running time among Re-heat SA, AntRSAR, GenRSAR and SimRSAR

In this way, based on the presented results, it can be observed that:

- Reheat Simulated Annealing Algorithm tends to generate a smaller number of rules compared to the other algorithms in most cases.
- The accuracy of Reheat Simulated Annealing Algorithm varies across the datasets, with some datasets achieving 100% accuracy while others have lower accuracy.
- Genetic Algorithm and Johnson's Algorithm generate a higher number of rules compared to Reheat Simulated Annealing Algorithm, but their accuracy varies across the datasets.
- Holte's 1R tends to generate a smaller number of rules compared to the other algorithms, but its accuracy also varies across the datasets.

Overall, Reheat Simulated Annealing Algorithm shows promising results in terms of generating a smaller number of rules and achieving high classification accuracy on certain datasets. However, the performance of the algorithms varies across the different datasets, indicating that the choice of algorithm may depend on the specific characteristics of the dataset.

Finally, Re-heat SA demonstrates reliable performance in solving attribute reduction problems. It offers several advantages such as shorter execution time and the ability to generate reducts with lower cardinality. While it may not achieve the exact optimal results, the solutions obtained by Re-heat SA are near optimal. As a stochastic search algorithm, it explores the search space and provides results that are close to the optimal solution. This makes it a valuable approach for attribute reduction tasks, offering a balance between computational efficiency and solution quality.

8. Proposing Standardization for Feature Selection and Dimensionality Reduction

In modern machine learning, the quality of data directly affects the effectiveness of learning through iterative processes. Although we now have access to enormous databases, not all information within these datasets is equally relevant for the learning process [39]. Datasets often contain numerous features that describe examples through specific rules, resulting in a high-dimensional feature space. However, dealing with high dimensionality can pose challenges for machine learning tasks. One significant issue is the complexity of learning models. As the number of features increases, so does the model complexity, leading to higher computational intensity and processing time demands. Moreover, model complexity can negatively affect accuracy and efficiency.

In practice, only a subset of features is usually crucial for distinguishing among examples and enabling knowledge discovery, while the rest may be insignificant or even detrimental to the model, leading to the "curse of dimensionality”. Feature selection plays a vital role in addressing this challenge by identifying the most relevant and non-redundant features for data separability.

In this research, we propose using Rough Set Theory (RST) as the basis for attribute reduction, employing the Re-heat Simulated Annealing algorithm to assess the quality of minimal reducts. RST is a mathematical tool that uncovers subsets of reducts, especially valuable when dealing with uncertain and incomplete data typically encountered in big data scenarios. On the other hand, Re-heat SA aims to optimize the search space to find the optimal solution.

Throughout our work, we have highlighted that knowledge discovery for data classification is a critical aspect, and leveraging rough sets to pre-process data within meta-heuristic and annealing approaches can enhance performance.

Given the insights from our discussion, we strongly advocate for the formalization and standardization of solutions that address the problems and challenges. Such standardization efforts need to be adopted as de facto or international norms, providing consistency, validation, and verifiability in machine learning pipelines across industries.

Our proposed solution can serve as a standard for feature selection and dimensionality reduction, particularly problems of a deep learning nature [40]. It is inevitable that on the one hand dimensionality reduction stands as a fundamental and essential technique in data mining and optimization, particularly in the present big data era, which typically demands

feature selection and feature extraction tasks, each with its advantages and limitations [41]. On the other hand, as the volume of data continues to grow exponentially, the task of extracting and selecting features with robust representational capacity from limited sample data becomes a crucial area of research for the future [42]. Furthermore, we believe that we are pioneers in recognizing the significance and purpose of such standardization.

The standardization process should encompass knowledge discovery and the application of intelligent data analysis strategies and techniques for tackling the challenges discussed above. By establishing industry-wide standards, we can streamline and optimize machine learning processes, enabling reliable and efficient analysis of big data. This effort will foster advancement and adoption of smart data analysis strategies in various fields.

9 Discussion and Conclusion

With the exponential growth of data in real-world applications, the need for dimensionality reduction techniques has become more critical than ever. Attribute reduction and feature selection are prominent methods for handling such large-scale data.

In this paper, we have demonstrated that integrating minimal reducts theory with if-then rules offers a superior approach to tackling the dimensionality problem and classification issue. The results clearly indicate that the proposed method, Re-heat SA, outperforms other approaches in terms of generating high-quality reducts, achieving faster running times, reducing cardinality, and improving classification accuracy. Re-heat SA achieved 100% classification accuracy for three datasets, 60-99% accuracy for four datasets, and 30-59% accuracy for six datasets. These results reinforce the advantages and performance of the proposed method. However, there is much room for further research to enhance knowledge discovery.

In future work, we suggest evaluating Re-heat SA using larger and more challenging datasets with a greater number of attributes. Additionally, comparing the generated reducts with other classification algorithms, such as neural networks, beyond those based on Rosetta and rules would be beneficial for comprehensive analysis. We believe that proper parameter tuning plays a crucial role in the algorithm's performance. Parameters like the maximum number of non-improving solutions, which governs the re-heating process, might vary across datasets due to their unique features. By carefully tuning these parameters, the Re-heat SA algorithm can further improve result quality. As a future research direction, we propose the use of search strategies, such as genetic algorithms, for dynamically tuning Re-heat SA's parameters instead of relying on fixed values.

The dimensionality reduction method proposed in this paper follows a filter-based approach, where attribute reduction and classification tasks are treated separately. An alternative avenue is to develop a wrapper-based solution that integrates both tasks, where the cost function of the algorithm incorporates both classifier accuracy and the dependency degree derived from rough set theory. There is a tremendous opportunity for new research and development work in these subject and topic areas.

The field of knowledge discovery, data mining, and machine learning continues to evolve rapidly, driven by advancements in technology, increasing availability of big data, and emerging challenges in various domains.

Firstly, further investigations need to explore and refine the proposed Re-heat SA algorithm for attribute reduction and classification. This includes exploring its performance on larger and more diverse datasets, evaluating its scalability, and

benchmarking it against other state-of-the-art algorithms. Additionally, parameter-tuning strategies, such as advanced genetic algorithms or other optimization techniques must be utilised to enhance the algorithm's effectiveness.

Secondly, the integration of different feature selection methods with classification algorithms warrants further exploration. Combining filter-based and wrapper-based approaches, as well as exploring embedded methods, can lead to more efficient and accurate solutions for dimensionality reduction.

Thirdly, considering the ever-increasing complexity and heterogeneity of real-world data, the development of advanced techniques that handle high-dimensional and mixed-type data is crucial. Exploring feature selection and dimensionality reduction techniques tailored for specific data types, such as text, images, time series, or network data, can open new avenues for research and development.

Fourthly, in addition to algorithmic advancements, research can also focus on developing frameworks and tools for standardizing and automating the feature selection and dimensionality reduction processes. This includes the formulation of guidelines, best practices, and evaluation metrics to facilitate the comparison and reproducibility of different methods. Standardized workflows and software implementations can further accelerate the adoption of these techniques across industries and domains.

Fifthly, as the field of machine learning continues to grow, it is essential to address ethical considerations and interpretability challenges associated with feature selection and dimensionality reduction methods. Research can explore techniques for ensuring fairness, transparency, and interpretability in the feature selection process to mitigate biases and enable trustworthy decision-making in transparent ways.

Sixthly and lastly, exploring the applicability of feature selection and dimensionality reduction techniques in emerging areas such as federated learning, edge computing, or privacy-preserving settings presents exciting research opportunities. By developing innovative approaches that preserve data privacy, while enabling effective feature selection, can revolutionize the way machine learning in sensitive domains is applied. In addition, based on the technical analysis and experience gained from this work, we strongly advocate for the initiation of standardization efforts to establish a norm for feature selection and dimensionality reduction problems, especially in the context of big data. This standardization should provide consistency, validation, and verifiability to machine learning processes applied industry-wide, encompassing the entire knowledge discovery pipeline, and employing smart intelligent data analysis strategies and techniques.

Finally, in conclusion, the subject and topic areas discussed in this paper offer a vast landscape for new research and development work. The field of feature selection and dimensionality reduction is ripe with opportunities to advance algorithms, explore new data types, standardize processes, address ethical concerns, and explore emerging applications. By embracing these opportunities, researchers can make significant contributions to the field and unlock the potential of machine learning in various domains.

References

[1] P. Clark and T. Niblett, “The CN2 Induction Algorithm,” *Machine Learning* **3**, 261–283, 1989. https://doi.org/10.1023/A:1022641700528

[2] G. S. Venturini, “A Supervised Inductive Algorithm with Genetic Search for Learning Attributes based Concepts,” in: Brazdil, P.B. (eds) *Machine Learning:*

ECML-93, Lecture Notes in Computer Science, vol 667. Springer, Berlin, Heidelberg, 1993. https://doi.org/10.1007/3-540-56602-3_142

[3] V. Jovanoski and N. Lavrac, "Classification Rule Learning with APRIORI-C," in: Brazdil, P., Jorge, A. (eds) *Progress in Artificial Intelligence: EPIA 2001*, Lecture Notes in Computer Science, vol 2258, pp. 44-51, Springer, Berlin, Heidelberg, 2001. https://doi.org/10.1007/3-540-45329-6_8

[4] P. G. Espejo, C. Romero, S. Ventura, and C. Hervas, "Induction of classification rules with grammar-based genetic programming," in Conference on Machine Intelligence, pp. 596-601, 2005. https://sci2s.ugr.es/keel/pdf/keel/congreso/MachineLearning-Pedro05.pdf

[5] S. Asadi and J. Shahrabi, "ACORI: a novel ACO algorithm for rule induction," *Knowledge-Based-Systems*, Vol. 97, pp. 175–187, 2016. https://doi.org/10.1016/j.knosys.2016.01.005

[6] M. Z. Asghar, A. Khan, K. Furqan, and F. M. Kundi, "RIFT: A Rule Induction Framework for Twitter Sentiment Analysis," *Arab J Sci Eng*, Vol. 43, pp. 857–877, 2018. https://doi.org/10.1007/s13369-017-2770-1.

[7] J-P. van Zyl and A. P. Engelbrecht, "Rule Induction Using Set-Based Particle Swarm Optimisation", in *IEEE Congress on Evolutionary Computation (CEC)*, pp. 1-8, 2022. https://ieeexplore.ieee.org/document/9870360

[8] J. Liang, S. Yang, A. Winstanley, and C. Adam "Invariant optimal feature selection: a distance discriminant and feature ranking based solution," *Pattern Recognition* 41 (5), pp. 1429–1439, 2008. https://doi.org/10.1016/j.patcog.2007.10.018

[9] B. Guo, R.I. Damper, S.R. Gunn, and J.D.B. Nelson, "A fast separability-based feature- selection method for high-dimensional remotely sensed image classification," *Pattern Recognition* 41 (5), pp. 1670–1679, 2008. https://doi.org/10.1016/j.patcog.2007.11.007

[10] D. Zhang, S. Chen, and Z. Zhou, "Constraint score: a new filter method for feature selection with pairwise constraints," *Pattern Recognition,* 41 (5), pp. 1440–1451, 2008. https://doi.org/10.1016/j.patcog.2007.10.009

[11] R. Jensen and Q. Shen, "Semantics-preserving Dimensionality Reduction Rough and Fuzzy rough-based Approaches," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 12, pp. 1457-1471, 2004. https://ieeexplore.ieee.org/document/1350758

[12] R. Jensen and Q. Shen, "Finding rough set reducts with ant colony optimization," in *Proceeding of UK Workshop on Computational Intelligence*, vol. 1(2), pp. 15-22, 2003.

[13] L. Ke, Z. Feng, and Z. Ren, "An Efficient Ant Colony Optimization Approach to Attribute Reduction in Rough Set Theory," *Pattern Recognition Letters,* 29 (9), pp. 1351–1357, 2008. https://doi.org/10.1016/j.patrec.2008.02.006

[14] A. R. Hedar, J. Wang, and M. Fukushima, "Tabu Search for Attribute Reduction in Rough Set Theory," *Soft Computing,* 12(9), pp. 909-918, 2008. https://doi.org/10.1007/s00500-007-0260-1

[15] J. Wang, A. -R. Hedar, G. Zheng and S. Wang, "Scatter Search for Rough Set Attribute Reduction," *2009 International Joint Conference on Computational Sciences and Optimization*, Vol. 1, pp. 531-535, 2009. https://ieeexplore.ieee.org/document/5193753

[16] A. Alsahaf, N. Petkov, V. Shenoy, and G. Azzopardi, "A framework for feature selection through boosting," *Expert Systems with Applications,* Vol. 187, 115895, 2022. https://doi.org/10.1016/j.eswa.2021.115895

[17] X. Zhang and Y. Yao, “Tri-level attribute reduction in rough set theory,” *Expert Systems with Applications*, Vol. 190, 2022. https://doi.org/10.1016/j.eswa.2021.116187
[18] T. Dokeroglu, A. Deniz, and H. E. Kiziloz, “A comprehensive survey on recent metaheuristics for feature selection,” *Neurocomputing*, Vol. 494, pp. 269–296, 2022. https://doi.org/10.1016/j.neucom.2022.04.083
[19] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, “Optimization by Simulated Annealing,” *Science*, Vol 220, Issue 4598, pp. 671-680, 1983. https://www.science.org/doi/10.1126/science.220.4598.671
[20] S. Consoli, “Combinatorial Optimization and Metaheuristics,” *Operational Research Report*, Department of Mathematical Sciences Technical Reports, TR/01/06, 2006. http://bura.brunel.ac.uk/handle/2438/9631
[21] Z. Pawlak, "Rough sets,” *International Journal of Computer and Information Sciences*. Vol. 11, pp. 341-356, 1982. https://doi.org/10.1007/BF01001956
[22] Z. Pawlak, “Rough Sets: Theoretical Aspects of Reasoning about Data,” Book Published by Kluwer Academic Publishers, Dordrecht, 1991.
[23] X. Wang, J. Yang, X. Teng, W. Xia, and R. Jensen, “Feature Selection Based on Rough Sets and Particle Swarm Optimization,” *Pattern Recognition Letters*, 28(4), 459–471. 2007. https://doi.org/10.1016/j.patrec.2006.09.003
[24] A. G. K. Janecek, W.N. Gansterer, M.A. Demel, and G.F. Ecker, “On the Relationship Between Feature Selection and Classification Accuracy,” *Journal of Machine Learning Research: Workshop and Conference Proceedings Track 4*, pp. 90-105, 2008. https://www.semanticscholar.org/paper/On-the-Relationship-Between-Feature-Selection-and-Janecek-Gansterer/c21817de5dc8164022a2f077a516a1164fb8f5be
[25] A. Assareh, M. H. Moradi, and L. G. Volkert, “A hybrid random subspace classifier fusion approach for protein mass spectra classification,” in: Marchiori, E., Moore, J.H. (eds) *6th European Conference: Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics*, EvoBIO, Lecture Notes in Computer Science, Springer Berlin Heidelberg, vol 4973. pp. 1-11, 2008. https://link.springer.com/chapter/10.1007/978-3-540-78757-0_1
[26] J. Hayward, S. Alvarez, C. Ruiz, M. Sullivan, J. Tseng, and G. Whalen, “Knowledge discovery in clinical performance of cancer patients,” in *2008 IEEE International Conference on Bioinformatics and Biomedicine*, pp. 51-58, 2008. https://ieeexplore.ieee.org/document/4684872
[27] S. Xia, H. Zhang, W. Li, G. Wang, E. Giem and Z. Chen, "GBNRS: A Novel Rough Set Algorithm for Fast Adaptive Attribute Reduction in Classification," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 3, pp. 1231-1242, 2022. https://ieeexplore.ieee.org/document/9099413
[28] D. Li and Z. Cui, “A Parallel Attribute Reduction Method Based on Classification,” *Complexity*, vol. 2021, Article ID 9989471, 8 pages, 2021. https://doi.org/10.1155/2021/9989471
[29] UCI, “Data Sets,” Welcome to the UC Irvine Machine Learning Repository, Center for Machine Learning and Intelligent Systems. https://archive.ics.uci.edu/
[30] Z. Pawlak and A. Skowron, “Rough sets and conflict analysis,” in: Lu, J., Zhang, G., Ruan, D. (eds) *E-Service Intelligence: Methodologies, Technologies and Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 35-74, 2007. https://doi.org/10.1007/978-3-540-37017-8_2
[31] R. Jensen and Q. Shen, "A rough set-aided system for sorting WWW bookmarks," In *Asia-Pacific conference on web intelligence*, Lecture Notes in Computer

Science, vol 2198.pp. 95-105, Springer, Berlin, Heidelberg, 2001. https://doi.org/10.1007/3-540-45490-X_10

[32] S. K. Pal and A. Skowron, "Rough Fuzzy Hybridization: a New Trend in Decision-Making," Springer-Verlag Berlin and Heidelberg, ISBN-13: ,9814021005-978 1999.

[33] R. B. Perez, A. Nowe, P. Vrancx, Y. Gomez D, and Y. Caballero, "Using ACO and Rough Set Theory to Feature Selection," in: *Proceedings of the 6th WSEAS Int. Conf. on Evolutionary Computing,* 2(5), pp. 251-256, 2005. https://www.academia.edu/13488495/Using_ACO_and_Rough_Set_Theory_to_Feature_Selection

[34] Y. Bengio and Y. Grandvalet, "No Unbiased Estimator of The Variance of K-fold cross-validation," *The Journal of Machine Learning Research,* vol. 5, pp. 1089-1105, 2004. https://www.jmlr.org/papers/volume5/grandvalet04a/grandvalet04a.pdf

[35] H. Almazini, Ku-Mahamud, Ku. Ruhana, and H. F. Almazini, "Enhanced feature clustering method based on ant colony optimization for feature selection," *International Journal of Advances in Intelligent Informatics,* vol. 9(1), pp. 79-95, 2023. https://doi.org/10.26555/ijain.v9i1.987

[36] H. Dhahri, I. Rahmany, A. Mahmood, E. Al-Maghayreh, and W. Elkilani, "Tabu Search and Machine-Learning Classification of Benign and Malignant Proliferative Breast Lesions", *BioMed Research International*, vol. 2020, Article ID 4671349, 10 pages, 2020. https://doi.org/10.1155/2020/4671349

[37] J. Komorowski, P. Synak, and A. Skowron. "The ROSETTA Software System," Online Released Version 3.2, 2011. https://www.rosettacommons.org/software

[38] M. Garbulowski, K. Diamanti, K. Smolińska, et al., R. ROSETTA: an interpretable machine learning framework," *BMC Bioinformatics*, 22, 110, 18 pages, 2021. https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-021-04049-z

[39] N. Talpur, S. J. Abdulkadir, H. Alhussian, M. H. Hasan, N. Aziz, and A. Bamhdi, "Deep Neuro-Fuzzy System application trends, challenges, and future perspectives: a systematic survey," *Artificial Intelligence Review*, 13:1-49, 2023. https://doi.org/10.1007/s10462-022-10188-3 [40]

[40] N. Talpur, S. J. Abdulkadir, H. Alhussian, M. H. Hasan, N. Aziz, and A. Bamhdi, "A comprehensive review of deep neuro-fuzzy system architectures and their optimization methods," *Neural Computing & Applications,* 34: 1837–1875, 2022. https://doi.org/10.1007/s00521-021-06807-9

[41] M. Li, H. Wang, L. Yang, Y. Liang, Z. Shang, and H. Wan, "Fast hybrid dimensionality reduction method for classification based on feature selection and grouped feature extraction," *Expert Systems with Applications*, Vol. 150, 113277, 2020. https://doi.org/10.1016/j.eswa.2020.113277

[42] W. Jia, M. Sun, J. Lian, and S. Hou, "Feature dimensionality reduction: a review," *Complex & Intelligent Systems*, Vol. 8, pp. 2663–2693, 2022. https://doi.org/10.1007/s40747-021-00637-x

Notes on contributors

Alwi M Bamhdi is the Associate Professor in the Department of Computer Sciences, College of Computing (Al Qunfudhah), Umm Al-Qura University, Saudi Arabia. He received his MSc (2010) and PhD (2014) in Computer Science respectively from Heriot-Watt University, Scotland, UK. His research interests include Computer Networking, Mobile Ad Hoc Networks, Wireless Sensor Networks, Internet of Things, Cloud Computing, Artificial Intelligence and Deep Learning, Information & Cyber Security, Digital Investigations & Forensics, Blockchain Technology, Computer Vision, Software Engineering and Performance Evaluation. He has published many papers in his specialized areas of MANETS and in other areas of Computer Science and Information Technology within his research and teaching activities.

Ana Luiza Bessa Barros received her MSc degree in Computer Science from the Federal University of Minas Gerais (UFMG) in 1996 and her PhD degree in Teleinformatics Engineering from the Federal University of Ceará (UFC) in 2013. She is professor and researcher at the State University of Ceará (UECE), Fortaleza, Brazil. Her interests are in Computational Intelligence, including Machine Learning, Artificial and Deep Neural Networks, Pattern Recognition and Evolutionary Computation. She is a member of the State Sectorial Chamber of the Productive Chain of ICT. She has worked on several research and development projects with universities and companies and coordinates the Lab LAURA (Learn, Understand, Reason, Act) at UECE.

Tahira Jehan Makda received her MTech degree in Tourism and Hospitality from the Cape Peninsula University of Technology in 2008. She is currently a Technical Instructor at the University of Doha for Science and Technology (UDST) in Qatar, at the College of Business Management, in the Digital Marketing Department. Although her academic career is centered on Tourism and Events, she has steered her focus to business management in Digital Marketing and Analytics, which she is currently teaching. Her interests are vested in analytics, marketing, tourism, tourist guiding, events, and entrepreneurship. She has published 2 journal papers in event risk management and presented a poster and paper presentations at conferences in Dubai, Finland, and Oman.

Marcial Fernandez is a graduate of the Electronic Engineer Faculty at the Federal University of Rio de Janeiro in 1988 and received his M.Sc. and D.Sc. in Electric Engineering from the same university in 1998 and 2002 respectively. He completed his postdoctoral research in Future Internet at the Technische Universitat Berlin TU-Berlin in 2010. Prof. Fernandez is currently an Associate Professor at Universidade Estadual do Ceara UECE). His academic teaching and research are in the subject areas of Computer Communication Networks, Computer Architecture, Distributed Systems, Network Security, Parallel and Concurrent Programming. His research interests are in Software Defined Networks (SDN), Network Management, Network Provisioning, Content Distribution Networks, Datacentre Infrastructures, Internet of Things (IoT), Cloud

Computing and Network Security. He has special expertise and skills in the application and use of Fuzzy Logic and its policies. He has published over 71 papers in scientific journals, conference proceedings and book chapters plus one book.

Ahmed Patel received his MSc & PhD degrees in Computer Science from Trinity College Dublin (TCD), University of Dublin in 1978 & 1984 respectively, specializing in the design, implementation & performance analysis of packet switched networks. He is Research Professor at Universidade Estadual do Ceará, Fortaleza, Brazil with key research interest in Advanced Computer Networking, Internet of Things, Cloud Computing, Big Data, Predictive Analysis, Use of Advanced Computing Techniques, Impact of e-social Networking, Closing the digital divide ICT gap and ICT Project Management. He has published well-over 276 technical & scientific papers & co-authored three books, two on computer network security & the third on group communications. He co-edited one book on distributed search systems for the Internet and co-edited & co-authored another book entitled: "Securing Information & Communication Systems: Principles, Technologies & Applications". He is a member of the Editorial Advisory Board of International Journals and has participated in Irish, Malaysian, and European funded research projects.

Laleh Golafshan received her B.Sc. in Computer Engineering from Islamic Azad University (IAU), Fars, Iran in 2005, Iran and M.Sc. in Information Technology (Computer Science) from Universiti Kebangsaan (National University) Malaysia (UKM), Malaysia in 2011. Subsequently she was an instructor at the Departments of Computer Engineering and at their Science and Research IT Branch in IAU-Fars, Iran. Her research interests are data mining, optimization algorithms and artificial intelligence. Currently, she is employed as a Software Engineer as an IT Expert in an oil and gas and petrochemical factory in Iran.