

Article

Not peer-reviewed version

---

# Optimizing Attribute Reduction in Multi-granularity Data through a Hybrid Supervised-Unsupervised Model

---

Zeyuan Fan , [Jianjun Chen](#) <sup>\*</sup> , Hongyang Cui , Jingjing Song , Taihua Xu

Posted Date: 26 March 2024

doi: 10.20944/preprints202403.1548.v1

Keywords: rough sets; attribute reduction; multi-granularity; information gain



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Article*

# Optimizing Attribute Reduction in Multi-Granularity Data Through a Hybrid Supervised-Unsupervised Model

Zeyuan Fan, Jianjun Chen \*, Hongyang Cui, Jingjing Song and Taihua Xu

School of Computer Science, Jiangsu University of Science and Technology, Zhenjiang 212100, Jiangsu, China

\* Correspondence: jianjunchen@just.edu.cn

**Abstract:** Attribute reduction is a core technique in the rough set domain and an important step in data preprocessing. Researchers have proposed numerous innovative methods to enhance the capability of attribute reduction, such as the emergence of multi-granularity rough set models, which can effectively process distributed and multi-granularity data. However, these innovative methods still have many shortcomings, such as how to deal with complex constraints and how to perform multi-angle effectiveness evaluations. Based on the multi-granularity model, this study proposes a new method of attribute reduction, namely using multi-granularity neighborhood information gain ratio as the measurement criterion. This method combines both supervised and unsupervised perspectives, and by integrating multi-granularity technology with neighborhood rough set theory, constructs a model that can adapt to multi-level data features. This model can select the optimal granularity level or attribute set according to the requirements. Finally, the proposed method is experimented with on 15 UCI datasets against six other attribute reduction algorithms. The experimental outcomes, upon analysis and comparative review, substantiate the reliability and consistency of the attribute reduction approach introduced in this study.

**Keywords:** rough sets; attribute reduction; multi-granularity; information gain

## 1. Introduction

In this era of information explosion, data is growing exponentially in both dimension and volume, which leads to the attributes of data becoming redundant and vague. How to find valuable information from massive data has become challenging. Rough set theory, introduced by Pawlak [1] in 1982 as a simple and efficient method for data mining, can deal with fuzzy, incomplete, and inaccurate data.

The traditional model of rough sets mainly focuses on describing the uncertainty and fuzziness of data through binary relations. In recent years, multi-granularity rough set models have been proposed to fully mine the multiple granularity levels of target information, extending the traditional single binary relation to multiple binary relations, with the work of Qian et al. [5] being representative. This model has provided a new solution for rough set theory in dealing with distributed data and multi-granularity data. Afterward, researchers continuously improve Qian's multi-granularity rough set model. Some of the improvements combine multi-granularity rough sets with decision-theoretic rough sets to form a multi-granularity decision-theoretic rough set model [6]. In addition, there is research combining multi-granularity rough sets with the three-way decision model, proposing a multi-granularity three-way decision model [7]. Targeting the granulation of attributes and attribute values, Xu proposed an improved multi-granularity rough set model [8]. To expand the applicability of multi-granularity rough sets, Lin et al. integrated the neighborhood relation into the multi-granularity rough set model, proposing the neighborhood multi-granularity rough set. The introduction of this model has made the multi-granularity rough set research branch a hot topic of study [9]. These rough set models can effectively reduce data dimensionality, achieved by attribute reduction.

Attribute reduction can be achieved through supervised or unsupervised constraints, and research on constraints from both supervised and unsupervised perspectives has been extensively explored. [10]. Specifically, some studies propose attribute reduction constraints based on measures from only one perspective, using these constraints to find qualified reductions. For instance, Jiang et al. [11]

and Yuan et al. [12] concentrated on attribute reduction through the lens of supervised information granulation and related supervised metrics, respectively; meanwhile, Yang et al. [13] proposed a concept known as fuzzy complementary entropy for attribute reduction within an unsupervised model. However, whether considering supervised measures or unsupervised measures, single-perspective based measures exhibit inherent constraints. Firstly, measures relying on a single perspective may overlook the multifaceted evaluation of data, leading to the neglect of some important attributes. This is because when only one fixed measure is used for the attribute reduction of data, the importance of each attribute is judged only based on its criterion. However, if other measures are needed for evaluation, then relying only on that criterion may no longer yield accurate results. Secondly, relying only on a single-perspective measure may not fully capture the characteristics of data under complex conditions, resulting in the selection of attributes that are neither accurate nor complete. For instance, if conditional entropy is used as a measure to evaluate attributes [17], the derived reduction may only possess the single feature required for evaluation, without fully considering other types of uncertainty features and learning capabilities.

In solve the limitations of the attribute reduction mentioned above, this paper introduces a new measure that merges both supervised and unsupervised perspectives, leading to a novel rough set model. The model proposed in this paper has the following advantages: (1) It integrates multi-granularity and neighborhood rough sets, making the model more adaptable to data features at different levels; (2) For attribute sets of different granularities, it introduces a fusion strategy, selecting the optimal granularity level or attribute set according to the needs of different tasks and datasets, which can be flexibly adjusted based on specific circumstances.

The rest of this paper is organized as follows. Section 2 reviews related basic concepts. Section 3 provides a detailed introduction to the basic framework and algorithm design of the proposed method. In Section 4, the accuracy of our method is calculated and discussed through experiments. Finally, Section 5 concludes this paper and depicts some future works.

## 2. Preliminaries

### 2.1. Neighborhood Rough Sets

Neighborhood rough sets were proposed by Hu et al. as an improvement over traditional rough sets [18]. The key distinction lies in that neighborhood rough sets are established on the basis of neighborhood relations, as opposed to relations of indiscernibility. Hence, the neighborhood rough set model is capable of processing both discrete and continuous data. [36] Moreover, the partitioning of neighborhoods granulates the sample space, which can reflect the discriminative power of different attributes on the samples.

Within the framework of rough set theory, a decision system is characterized by a tuple, represented by  $DS = (U, AT)$ , where  $U$  denotes a finite collection of samples and  $AT$  encompasses a suite of conditional attributes, including a decision attribute  $d$  [19]. The attribute  $d$  captures the sample labels. For every  $x$  in  $U$  and every  $a$  in  $AT$ ,  $a(x)$  signifies the value of  $x$  for the conditional attribute  $a$ , and  $d(x)$  represents the label of  $x$ . Utilizing  $d$ , one can derive an equivalence relation on  $U$ :

$$IND(d) = \{(x, y) \in U \times U : d(x) = d(y)\} \quad (1)$$

Pursuant to  $IND(d)$ , it leads to a division of  $U/IND(d) = X_1, X_2, \dots, X_q (q \geq 2)$ . Each  $X_k$  within  $U/IND(d)$  is recognized as the  $k$ -th decision category. Notably, the decision category that includes the sample  $x$  can be similarly referred to as  $[x]_d$ .

In rough set methods, binary relations are often used for information granulation, among which neighborhood relations, as one of the most effective binary relations, have received extensive attention. The formation of neighborhood relations is as follows:

$$N_\delta^A = \{(x, y) \in U \times U : \delta_A(x, y) \leq \delta\} \quad (2)$$

where  $r_A$  is a distance function regarding  $A \subseteq AT$ ,  $r \geq 0$  is a radius.

$$r_A = \sum_{a \in A} (a(x) - a(y))^2 \quad (3)$$

Based on  $IND(d)$ , a segmentation of  $U/IND(d) = X_1, X_2, \dots, X_q (q \geq 2)$  can be initiated. For every  $X_k$  within  $U/IND(d)$ , it is identified as the  $k$ -th decision group. In particular, the decision group that encompasses sample  $x$  may also be represented as  $[x]_d$ .

In alignment with Equation (2), the vicinity of a sample  $x$  is established as follows:

$$\delta_A(x) = \{y \in U : r_A(x, y) \leq \delta\} \quad (4)$$

From the perspective of granular computing [21,22], both  $IND(d)$  and  $N_{\delta A}$  are derivations of information granules. The most significant difference between these two types of information granules lies in their intrinsic mechanisms, i.e., the binary relations used. Based on the outcomes of these information granules, the concepts of lower and upper approximations within the context of neighborhood rough sets, as the fundamental units, were also proposed by Cheng et al.

## 2.2. Multi-Granularity Rough Sets

For multi-granularity rough sets [23,24,26], given  $DS = (U, AT)$ , where  $AT = \{A_k | k \in \{1, 2, \dots, m\}\}$  is a set of attributes, and the family of attribute subsets on  $AT$  is represented by  $\{A_1, A_2, \dots, A_m\}$ .

$[x]_{A_i}$  is an equivalence class of  $x$  under  $A_i$ , for any  $X \subseteq U$ , the optimistic multi-granularity lower and upper approximations of  $A_i$  with respect to  $X$  are defined as follows:

$$\sum_{i=1}^m A_i^O(x) = \{x \in U \mid [x]_{A_1} \subseteq X \vee [x]_{A_2} \subseteq X \vee \dots \vee [x]_{A_m} \subseteq X\}, \quad (5)$$

$$\overline{\sum_{i=1}^m A_i^O(x)} = \left( \sum_{i=1}^m A_i^O(x^c) \right)^c \quad (6)$$

If  $\sum_{i=1}^m A_i^O(x) \neq \overline{\sum_{i=1}^m A_i^O(x)}$ , then  $\sum_{i=1}^m A_i^O(x)$  and  $\overline{\sum_{i=1}^m A_i^O(x)}$  are called optimistic multi-granularity rough sets.

Given  $DS = (U, AT)$ , where  $AT = \{A_k | k \in \{1, 2, \dots, m\}\}$  is a set of attributes, and the family of attribute subsets on  $AT$  is represented by  $\{A_1, A_2, \dots, A_m\}$ .

$[x]_{A_i}$  is an equivalence class of  $x$  under  $A_i$ , for any  $X \subseteq U$ , the pessimistic multi-granularity [27] lower and upper approximations of  $A_i$  with respect to  $X$  are defined as follows:

$$\sum_{i=1}^m A_i^P(x) = \{x \in U \mid [x]_{A_1} \subseteq X \wedge [x]_{A_2} \subseteq X \wedge \dots \wedge [x]_{A_m} \subseteq X\}, \quad (7)$$

$$\overline{\sum_{i=1}^m A_i^P(x)} = \left( \sum_{i=1}^m A_i^P(x^c) \right)^c \quad (8)$$

If  $\sum_{i=1}^m A_i^P(x) \neq \overline{\sum_{i=1}^m A_i^P(x)}$ , then  $\sum_{i=1}^m A_i^P(x)$  and  $\overline{\sum_{i=1}^m A_i^P(x)}$  are called pessimistic multi-granularity rough sets.

## 2.3. Multi-Granularity Neighborhood Rough Sets

In the literature [25], Lin et al. proposed two types of neighborhood multi-granularity rough sets, which can be applied to deal with incomplete systems containing numerical and categorical attributes [29]. To simplify the problem, when dealing with incomplete systems, only the application of neighborhood multi-granularity rough sets to numerical data is considered.

Given  $DS = (U, AT)$ , where  $AT = \{A_k \mid k \in \{1, 2, \dots, m\}\}$ ,  $U = \{x_i \mid i \in \{1, 2, \dots, n\}\}$ ,  $X \subseteq U$ , in the optimistic neighborhood multi-granularity rough sets, the neighborhood multi-granularity approximation of  $X$  is defined as:

$$\sum_{i=1}^m N_i^O(x) = \{x_i \in U \mid \delta A_1(x_i) \subseteq X \vee \delta A_2(x_i) \subseteq X \vee \dots \vee \delta A_k(x_i) \subseteq X\} \quad (9)$$

$$\overline{\sum_{i=1}^m N_i^O(x)} = \{x_i \in U \mid \delta A_1(x_i) \cap X \neq \emptyset \wedge \delta A_2(x_i) \cap X \neq \emptyset \wedge \dots \wedge \delta A_m(x_i) \cap X \neq \emptyset\} \quad (10)$$

where  $\delta A_k(x_i)$  is the neighborhood granularity of  $x_i$ , based on the granularity structure  $A_k$ .

Given  $DS = (U, AT)$ , where  $AT = \{A_k \mid k \in \{1, 2, \dots, m\}\}$ ,  $U = \{x_i \mid i \in \{1, 2, \dots, n\}\}$ ,  $X \subseteq U$ , in the pessimistic neighborhood multi-granularity rough sets, the neighborhood multi-granularity approximation of  $X$  is defined as:

$$\sum_{i=1}^m N_i^P(x) = \{x_i \in U \mid \delta A_1(x_i) \subseteq X \wedge \delta A_2(x_i) \subseteq X \wedge \dots \wedge \delta A_k(x_i) \subseteq X\} \quad (11)$$

$$\overline{\sum_{i=1}^m N_i^P(x)} = \{x_i \in U \mid \delta A_1(x_i) \cap X \neq \emptyset \vee \delta A_2(x_i) \cap X \neq \emptyset \vee \dots \vee \delta A_k(x_i) \cap X \neq \emptyset\} \quad (12)$$

where  $\delta A_k(x_i)$  is the neighborhood granularity of  $x_i$ , based on the granularity structure  $A_k$ .

#### 2.4. Supervised Attribute Reduction

It is well known that neighborhood rough sets are often used in supervised learning tasks, especially in enhancing generalization performance and reducing classifier complexity.[47] The advantage of attribute reduction lies in its easy adaptation to different practical application requirements, hence a variety of forms of attribute reduction have emerged in recent years. For neighborhood rough sets, information gain and split information value are two metrics that can be used to further explore the forms of attribute reduction.

Given the data  $DS = \langle U, AT, d, \delta \rangle$ , for any  $A \subseteq AT$ , the neighborhood information gain of  $D$  based on  $A$  is defined as:

$$IG_{NRS}(d, A) = H_{NRS} - H_{NRS}(d, A) \quad (13)$$

Here,  $H_{NRS}$  is the entropy of the entire dataset  $D$ , calculated based on the distribution under neighborhood lower or upper approximation.  $H_{NRS}(d, A)$  is the expected value of uncertainty considering attribute  $A$ , defined as:

$$H_{NRS}(d, A) = -\frac{1}{|U|} \sum_{x \in U} |\delta_s(x) \cap [x]_d| \log_2 \frac{|\delta_s(x) \cap [x]_d|}{|\delta_s(x)|} \quad (14)$$

Given the data  $DS = \langle U, AT, d, \delta \rangle$ , for any  $A \subseteq AT$ , the neighborhood split information value of  $d$  based on  $A$  is defined as:

$$SI_{NRS}(d, A) = -\sum_{j=1}^n \frac{|\delta_A(X_j)|}{|U|} \log_2 \frac{|\delta_A(X_j)|}{|U|} \quad (15)$$

Here  $\delta_A(X_j)$  represents the sample set  $X_j$  within the neighborhood formed by attribute  $A$ , and  $n$  is the number of different neighborhoods formed by  $A$ .

The combination of neighborhood information gain and split information value helps to more comprehensively assess the impact of attributes on dataset classification, thereby making more effective decisions in attribute reduction.

### 2.5. Unsupervised Attribute Reduction

It's widely recognized that supervised attribute reduction necessitates the use of sample labels, which are time-consuming and expensive to obtain in many practical tasks. [48] In contrast, unsupervised attribute reduction does not require these labels, hence it has received more attention recently.

In unsupervised attribute reduction, if it is necessary to measure the importance of attributes, one can construct models by introducing pseudo-label strategies and using information gain and split information as metrics.

Given unsupervised data  $IS = \langle U, AT \rangle$  and  $\delta$ , for any  $A \subseteq AT$ , the unsupervised information gain based on  $A$  is defined as:

$$IG_{NRS}(d, A) = H_{NRS} - H_{NRS}(d, A) \quad (16)$$

where  $H_{NRS}(d, A)$  is the expected value of uncertainty considering attribute  $A$ , defined as:

$$H_{NRS}(d, A) = \frac{1}{|A|} \sum_{a \in A} (H_{NRS}(d, A)(d^a)) \quad (17)$$

$d_a$  denotes the pseudo-label decision for samples generated using conditional attribute  $a$ .

Given unsupervised data  $IS = \langle U, AT \rangle$  and  $\delta$ , for any  $A \subseteq AT$ , the unsupervised split information based on  $A$  is defined as:

$$SI_{NRS}(d, A) = \frac{1}{|A|} \sum_{a \in A} (SI_{NRS}(d, A)(d^a)) \quad (18)$$

Here,  $d_a$  is a pseudo-label decision, recorded by using conditional attribute  $a$  for sample pseudo-labels.

These definitions provide a new method for evaluating attribute importance in an unsupervised setting. Information gain reflects the contribution of an attribute to data classification, while split information measures the degree of confusion introduced by an attribute in the division of the dataset. This approach helps in more effective attribute selection and reduction in unsupervised learning.

## 3. Proposed Method

### 3.1. Definition of Multi-Granularity Neighborhood Information Gain Ratio

Considering a dataset  $DS = \langle U, AT, d, \delta \rangle$ , with  $U$  representing the sample set,  $AT$  indicating the attribute set,  $d$  denoting the decision attribute, and  $\delta$  specifying the neighborhood radius.

For any  $A \subseteq AT$ , the multi-granularity neighborhood information gain ratio is defined as:

$$\epsilon_A(d) = \frac{SI_{NRS}(d, A)}{e^{|IG_{NRS}(d, A)|}} \times W_A \quad (19)$$

where  $SI_{NRS}(d, A)$  is the neighborhood split information quantity based on  $A$ ,  $e^{|IG_{NRS}(d, A)|}$  is the information gain for decision attribute  $d$  based on attribute  $A$  with the base of the natural logarithm, and  $W_A$  is the granularity space coefficient of attribute  $A$  in the multi-granularity structure, reflecting its importance in the multi-granularity structure.

For the calculation of the granularity space coefficient, given a set of granularities  $G_1, G_2, \dots, G_n$ , the performance of attribute  $A$  under each granularity can be measured by the quantitative indicator  $P_{G_i}(A)$ . The definition of the granularity space coefficient  $W_A$  is as follows:

$$W_A = \frac{\sum_{i=1}^n \beta_i \cdot P_{G_i}(A)}{\sum_{i=1}^n \beta_i} \quad (20)$$

where  $\beta_i$  is the granularity space allocated to each granularity  $G_i$ , reflecting the importance of different granularities. These granularity spaces are usually determined based on the specific background knowledge or experimental verification of the problem.

The granularities  $G_1, G_2, \dots, G_n$  in the multi-granularity structure are determined according to the data characteristics, problem requirements, etc., and each granularity reflects different levels or details of the data. When calculating the granularity space coefficient, the performance of the attribute under different granularities is considered, in order to more accurately reflect its importance in the multi-granularity structure.

Neighborhood rough set is a method for dealing with uncertain and fuzzy data, which uses neighborhood relations instead of the indiscernible relations in traditional rough sets. In this method, data is decomposed into different granularities, each representing different levels or details of the data. Information gain ratio is a method for measuring the importance of attributes in data classification. It is based on the concept of information entropy and evaluates the classification capability of an attribute by comparing the entropy change of the dataset with and without the attribute.

Therefore,  $\epsilon$  combines these two concepts, i.e., neighborhood information gain at different granularities and the split information value of attributes, to evaluate the importance of attributes in multi-granularity data analysis. This method not only considers the information gain of attributes but also their performance at different granularities, thus providing a more comprehensive method of attribute evaluation.

Given a decision system  $DS$  and a threshold  $\theta \in [0, 1]$ , an attribute  $A$  is considered significant if it satisfies the following conditions:

1.  $\frac{\epsilon_A(d)}{\epsilon_{AT}(d)} \geq \theta$ ;
2. There is no proper subset  $A'$  of attribute  $A$  such that  $\frac{\epsilon_{A'}(d)}{\epsilon_{AT}(d)} < \theta$ .

In this definition, significant attributes are determined based on their contribution to the information gain ratio, aiming to select attributes that are informative yet not redundant for the decision-making process. This method is based on greedy search techniques for attribute reduction and helps identify attributes that significantly impact the decision outcome.

Given a dataset  $DS = \langle U, AT, d \rangle$ , where  $U$  is the set of objects,  $AT$  is the set of conditional attributes, and  $d$  is the decision attribute. For any attribute subset  $A \subseteq AT$  and any  $a \in AT - A$  (i.e., any attribute not in  $A$ ), the significance of attribute  $a$  regarding the multi-granularity neighborhood information gain ratio is defined as follows:

$$Sig_{\epsilon_a(d)} = \epsilon_{A \cup \{a\}}(d) - \epsilon_A(d) \quad (21)$$

The aforementioned significance function suggests that an increase in value enhances the importance of a conditional attribute, making it more probable to be included in the reduction set. For example, if  $Sig_{\epsilon_{a1}(d)} < Sig_{\epsilon_{a2}(d)}$ , where  $a1, a2 \in AT - A$ , then  $\epsilon_{A \cup \{a1\}}(d) < \epsilon_{A \cup \{a2\}}(d)$ . Such a result indicates that choosing  $a2$  to join  $A$  would lead to a higher multi-granularity neighborhood information gain ratio compared to  $a1$ .

### 3.2. Detailed Algorithm

Based on the significance function, Algorithm 1 is designed to find the  $\epsilon$ -reduct.

To streamline the analysis of the computational complexity for Algorithm 1, we initiate by applying  $k$ -means clustering to generate pseudo labels for the samples. With  $T$  denoting the iteration count for  $k$ -means clustering and  $k$  indicating the cluster count, the complexity of creating pseudo labels is  $O(k \cdot T \cdot |U| \cdot |AT|)$ , where  $|U|$  is the total number of samples, and  $|AT|$  signifies the attribute

count. Subsequently, the calculation of  $\epsilon_{A \cup a}(d)$  occurs no more than  $(1 + |AT|) \cdot |AT|/2$  times. In conclusion, the computational complexity of Algorithm 1 equates to  $O\left(\frac{|U|^2 \cdot |AT|^3}{2} + k \cdot T \cdot |U| \cdot |AT|\right)$ .

---

**Algorithm 1** Forward Greedy Searching for  $\epsilon$ -reduct with Neighborhood Rough Set (NRS- $\epsilon$ )

---

**Input:** A decision system  $DS = (U, AT, d)$ , a neighborhood radius  $\delta$ , a significance threshold  $\theta$ .

**Output:** An  $\epsilon$ -reduct  $A$ .

```

1: Initialize  $A = \emptyset$ ;
2: Calculate initial neighborhood rough set characteristics for  $DS$ ;
3: for each attribute  $a \in AT$  do
4:   Generate neighborhood relations  $N_\delta^a$  for  $a$ ;
5: end for
6: repeat
7:   for each  $a \in AT - A$  do
8:     Calculate the neighborhood information gain  $IG_{NRS}(d, A \cup \{a\})$ ;
9:     Calculate the neighborhood split information  $SI_{NRS}(d, A \cup \{a\})$ ;
10:    Compute  $\epsilon_{A \cup \{a\}}(d) = \frac{SI_{NRS}(d, A \cup \{a\})}{e^{IG_{NRS}(d, A \cup \{a\})}} \times W_{A \cup \{a\}}$ ;
11:   end for
12:   Select attribute  $b = \arg \max \{\epsilon_{A \cup \{a\}}(d) : a \in AT - A\}$ ;
13:   Update  $A = A \cup \{b\}$ ;
14: until No attribute can improve the significance threshold  $\theta$  or  $\frac{\epsilon_A(d)}{\epsilon_{AT}(d)} \geq \theta$  return  $A$ 

```

---

## 4. Experimental Analysis

### 4.1. Dataset Description

To evaluate the performance of the proposed measure, 15 UCI datasets are used in this experiment. Table 1 summarizes the statistical information of these datasets.

**Table 1.** Dataset descriptions.

ID	Datasets	Samples	Attributes	Labels
1	Adult Income	48,842	14	2
2	Iris Plants Database	150	4	3
3	Wine	178	13	3
4	Breast Cancer Wisconsin (Original)	699	10	2
5	Climate Model Simulation Crashes	540	20	2
6	Car Evaluation	1728	6	4
7	Human Activity Recognition Using Smartphones	10,299	561	6
8	Statlog (Image Segmentation)	2310	18	7
9	Yeast	1484	8	10
10	Seeds	210	7	3
11	Ultrasonic Flowmeter Diagnostics-Meter D	180	43	4
12	Spambase	4601	57	2
13	Mushroom	8124	22	2
14	Heart Disease	303	75	5
15	Letter Recognition	20,000	16	26

### 4.2. Experimental Configuration

The experiment is performed on a personal computer running Windows 11, featuring an Intel Core i5-12500H processor (2.50 GHz) with 16.00 GB RAM. MATLAB R2023a served as the development environment.

In this experiment, a double means algorithm is adopted to recursively allocate attribute granularity space, while utilizing the  $k$ -means clustering method [38] to generate pseudo labels for samples, and the information gain ratio as the criterion for evaluating attribute reduction. Notably, the selected  $k$ -value needs to match the number of decision categories in the dataset. Moreover, the effect of the neighborhood rough set is significantly influenced by the preset radius size. To demonstrate the effectiveness and applicability of the proposed method, a series of experiments are designed

using 20 different radius values, incremented by 0.02, ranging from 0.02 to 0.40. Through 10-fold cross-validation, the simplified reasoning process is validated. Specifically, for each specific radius, the dataset was divided into ten subsets, nine for training and one for testing. This cross-validation process was repeated 10 times to ensure that each subset had the opportunity to serve as the test set, thereby evaluating the classification performance and ensuring the reliability and stability of the model.

In the experiment, the proposed measure is compared with six advanced attribute reduction algorithms using Regression Trees (CART) [36], K-Nearest Neighbors (KNN, K=3) [41], and Support Vector Machines (SVM) [37]. The performance of the reducer is evaluated in aspects of the stability, accuracy, and timeliness of classification, as well as the stability of reduction. The attribute reduction algorithms included for comparison are:

MapReduce-Based Attribute Reduction Algorithm (MARA) [40]

Robust Attribute Reduction Based On Rough Sets (RARR) [42].

Bipolar Fuzzy Relation System Attribute Reduction Algorithms (BFRS) [43]

Attribute Group (AG) [44];

Separability-Based Evaluation Function (SEF) [45];

Genetic Algorithm-based Attribute Reduction (GAAR) [46].

#### 4.3. Comparison of Classification Accuracy

In this part, the classification accuracy of each algorithm is evaluated using KNN, SVM, and CART for predicting test samples. Regarding attribute reduction algorithms, within a decision system  $DS$ , the definition of classification accuracy post-reduction is as follows:

$$Acc_{red} = \frac{|\{x_i \in red | Pre_{red}(x_i) = d(x_i)\}|}{|U|}, \quad (22)$$

where,  $Pre_{red}(x_i)$  is the predicted label for  $x_i$  using the reduced set  $red$ .

Tables 2 and Figure 1 present the specific classification accuracy outcomes for each algorithm across 15 datasets. From these observations, several insights can be readily inferred:

1. For most datasets, the classification accuracy associated with NRS- $\epsilon$  is superior to other comparison algorithms, regardless of whether the KNN, SVM, or CART classifier is used. For example, in the "Car Evaluation (ID: 6)" dataset, when using the CART classifier, the classification accuracies of NRS- $\epsilon$ , MARA, RARR, BFRS, AG, SEF, and GAAR are 0.5039, 0.4529, 0.4157, 0.494, 0.4886, 0.4909, 0.4719 respectively; when using the KNN classifier, the classification accuracies of NRS- $\epsilon$ , MARA, RARR, BFRS, AG, SEF, and GAAR are 0.6977, 0.6584, 0.535, 0.6747, 0.6675, 0.6586, 0.6579 respectively; when using SVM, the classification accuracies of NRS- $\epsilon$ , MARA, RARR, BFRS, AG, SEF, and GAAR are 0.5455, 0.4307, 0.368, 0.4737, 0.4698, 0.4718, 0.4923 respectively. Therefore, NRS- $\epsilon$  derived simplifications can provide effective classification performance.
2. Examining the average classification accuracy per algorithm reveals that the accuracy associated with NRS- $\epsilon$  is on par with, if not exceeds, that of MARA, RARR, BFRS, AG, SEF, and GAAR. When using the CART classifier, the average classification accuracy of NRS- $\epsilon$  is 0.8012, up to 29.28% higher than other algorithms; when using the KNN classifier, the average classification accuracy of NRS- $\epsilon$  is 0.8169, up to 34.48% higher than other algorithms; when using SVM, the average classification accuracy of NRS- $\epsilon$  is 0.80116, up to 36.38% higher than other algorithms.

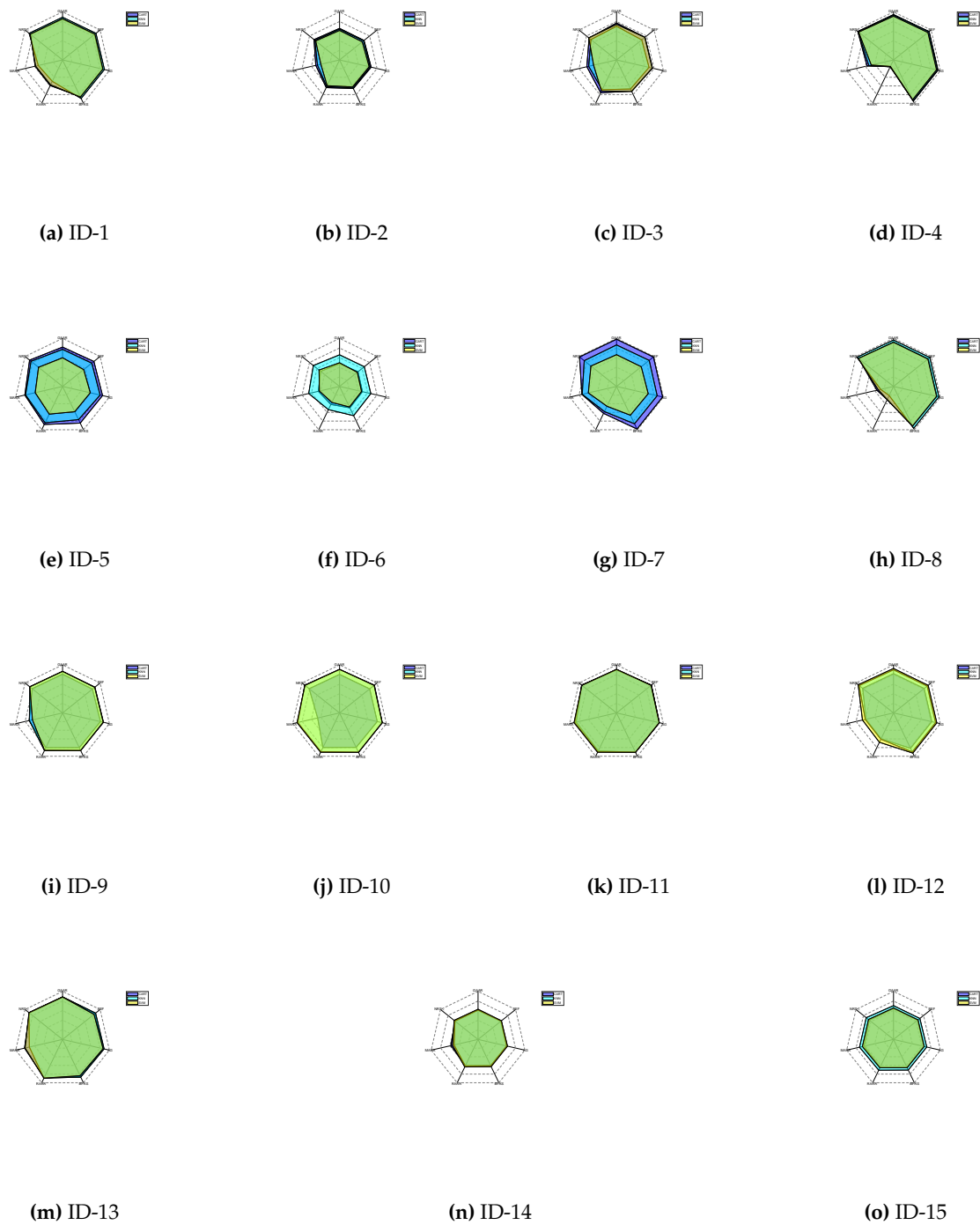


Figure 1. Classification accuracies of three classifiers.

Table 2. The comparisons of the classification accuracies.

ID	CART						
	NRS- $\epsilon$	MARA	RARR	BFRS	AG	SEF	GAAR
1	0.8514	0.5849	0.5911	0.8469	0.8511	0.8494	0.8452
2	0.6795	0.5078	0.6392	0.6604	0.6749	0.6516	0.6573

3	0.7312	0.6342	0.7584	0.7213	0.7671	0.7568	<b>0.7687</b>
4	<b>0.9507</b>	0.5655	0.1499	0.9503	0.9502	0.9466	0.9445
5	<b>0.8783</b>	0.8057	0.8779	0.8468	0.8532	0.834	0.8184
6	<b>0.5039</b>	0.4529	0.4157	0.494	0.4886	0.4909	0.4719
7	<b>0.9848</b>	0.7403	0.6217	0.9831	0.9845	0.9825	0.9820
8	<b>0.9271</b>	0.3540	0.3046	0.9187	0.9262	0.9245	0.9259
9	<b>0.8101</b>	0.7125	0.8015	0.8014	0.8097	0.7944	0.8075
10	<b>0.8091</b>	0.4674	0.7979	0.7947	0.8047	0.8023	0.8022
11	<b>0.9281</b>	0.8612	0.8769	0.9072	0.9140	0.9276	0.8937
12	<b>0.8206</b>	0.5887	0.5967	0.8161	0.8161	0.8161	0.8169
13	<b>0.8834</b>	0.8115	0.8828	0.8652	0.8665	0.8670	0.8794
14	<b>0.6158</b>	0.5784	0.6048	0.6067	0.6095	0.6053	0.6138
15	<b>0.6434</b>	0.6308	0.6422	0.6430	0.6414	0.6421	0.6425
Average	<b>0.8012</b>	0.6197	0.6374	0.7903	0.7972	0.7927	0.7913
rate		29.27% ↑	25.68% ↑	1.37% ↑	0.49% ↑	1.06% ↑	1.24% ↑

KNN							
ID	NRS- $\epsilon$	MARA	RARR	BFRS	AG	SEF	GAAR
1	0.8898	0.5177	0.4930	0.8898	<b>0.8935</b>	0.8891	0.8891
2	<b>0.6547</b>	0.4723	0.6140	0.6465	0.6529	0.6387	0.6410
3	0.6802	0.5942	0.6880	0.6703	0.6930	0.6702	<b>0.7005</b>
4	<b>0.9445</b>	0.5269	0.1590	0.9436	0.9414	0.9419	0.9344
5	<b>0.8392</b>	0.7890	0.8390	0.7659	0.8015	0.7768	0.7701
6	<b>0.6977</b>	0.6584	0.5350	0.6747	0.6675	0.6586	0.6579
7	0.8597	0.7315	0.5796	0.8620	0.8671	<b>0.8691</b>	0.8632
8	<b>0.9743</b>	0.2733	0.2094	0.9671	0.9684	0.9659	0.9658
9	<b>0.8730</b>	0.7087	0.8685	0.8671	0.8729	0.8605	0.8650
10	<b>0.7655</b>	0.3996	0.7533	0.7527	0.7634	0.7577	0.7603
11	<b>0.9267</b>	0.8927	0.8943	0.9081	0.9121	0.9266	0.9042
12	<b>0.9132</b>	0.6035	0.6066	0.8998	0.8981	0.8959	0.8972
13	<b>0.8948</b>	0.7087	0.8945	0.8700	0.8871	0.8791	0.8865
14	<b>0.6148</b>	0.5105	0.6122	0.6079	0.6115	0.6145	0.6104
15	<b>0.7260</b>	0.7254	0.7096	0.7064	0.7094	0.7009	0.7032
Average	<b>0.8169</b>	0.6075	0.6304	0.8021	0.8093	0.8030	0.8033
rate		34.47% ↑	29.59% ↑	1.84 % ↑	0.94% ↑	1.73% ↑	1.70% ↑

SVM							
ID	NRS- $\epsilon$	MARA	RARR	BFRS	AG	SEF	GAAR
1	<b>0.8616</b>	0.5751	0.5741	0.8573	0.8612	0.8613	0.8580
2	<b>0.6343</b>	0.3989	0.5935	0.6231	0.6314	0.6047	0.6142
3	0.7388	0.4890	0.7211	0.7270	<b>0.7614</b>	0.7538	0.7400
4	<b>0.9280</b>	0.4741	0.1499	0.9169	0.9108	0.9150	0.9106
5	<b>0.6392</b>	0.5834	0.6390	0.5834	0.5912	0.5618	0.5962
6	<b>0.5455</b>	0.4307	0.3680	0.4737	0.4698	0.4718	0.4923
7	<b>0.6809</b>	0.6021	0.4575	0.6636	0.6756	0.6573	0.6603
8	<b>0.9354</b>	0.3234	0.3016	0.9129	0.9193	0.9146	0.9187
9	<b>0.8718</b>	0.6420	0.8636	0.8672	0.8672	0.8549	0.8651
ID	NRS- $\epsilon$	MARA	RARR	BFRS	AG	SEF	GAAR
10	<b>0.7724</b>	0.4580	0.7565	0.7461	0.7543	0.7453	0.7535
11	<b>0.9209</b>	0.9075	0.9078	0.9092	0.9105	0.9201	0.9075

12	<b>0.9402</b>	0.6625	0.6774	0.9267	0.9240	0.9218	0.9249
13	<b>0.894</b>	0.8001	0.8938	0.8371	0.8645	0.8367	0.8827
14	<b>0.6372</b>	0.5436	0.6357	0.6281	0.6286	0.6228	0.6281
15	<b>0.6662</b>	0.6641	0.6544	0.6533	0.6549	0.6522	0.6544
Average	<b>0.7778</b>	0.5703	0.6129	0.75502	0.7621	0.7530	0.7604
rate		36.38% ↑	26.90% ↑	3.01% ↑	2.05% ↑	3.29% ↑	2.27% ↑

4.4. Comparison of Classification Stability

Similar to the evaluation of classification accuracy, this section explores the classification stability obtained by analyzing the classification results of seven different algorithms, including experiments with CART, KNN, and SVM classifiers. In a decision system  $DS = \langle U, AT, d, \delta \rangle$ , assume the set  $U$  is equally divided into  $z$  mutually exclusive groups of the same size (using 10-fold cross-validation, so  $z = 10$ ), that is  $U_1, \dots, U_\tau, \dots, U_z$  ( $1 \leq \tau \leq z$ ). Then, the classification stability based on redundancy reduction  $red_\tau$  (obtained by removing  $U_\tau$  from the set  $U$ ) can be represented as:

$$Stab_{class} = \frac{2}{z \cdot (z - 1)} \sum_{\tau=1}^{z-1} \sum_{\tau'=\tau+1}^z Exa(red_\tau, red_{\tau'})$$

(23)

where  $Exa(red_\tau, red_{\tau'})$  measures the consistency between two classification results, which can be defined according to Table 3.

Table 3. Joint distribution of classification results.

	$Prered_\tau(x) = d(x)$	$Prered_\tau(x) \neq d(x)$
$Prered_{\tau'}(x) = d(x)$	$\psi_1$	$\psi_2$
$Prered_{\tau'}(x) \neq d(x)$	$\psi_3$	$\psi_4$

In Table 3,  $Prered_\tau(x)$  represents the predicted label of  $x$  obtained by  $red_\tau$ . The symbols  $\psi_1, \psi_2, \psi_3$ , and  $\psi_4$  respectively represent the number of samples that satisfy the corresponding conditions in Table 4. Based on this,  $Exa(red_\tau, red_{\tau'})$  is defined as

$$Exa(red_\tau, red_{\tau'}) = \frac{\psi_1 + \psi_4}{\psi_1 + \psi_2 + \psi_3 + \psi_4}.$$

(24)

Table 4. The comparisons of classification stabilities.

ID	CART						
	NRS- $\epsilon$	MARA	RARR	BFRS	AG	SEF	GAAR
1	0.8583	<b>0.8663</b>	0.8369	0.8580	0.8566	0.8538	0.8558
2	<b>0.7364</b>	0.6794	0.7122	0.6981	0.7244	0.7130	0.7276
3	0.7222	0.7453	<b>0.7761</b>	0.7546	0.7553	0.7431	0.7500
4	<b>0.9485</b>	0.6265	0.7733	0.9423	0.9454	0.9480	0.9391
5	<b>0.8837</b>	0.8279	0.8834	0.8370	0.8473	0.8432	0.8407
6	0.6469	0.6557	0.6584	<b>0.669</b>	0.6473	0.6397	0.6632
7	<b>0.9812</b>	0.9334	0.8934	0.9797	0.9806	0.9809	0.9792
8	0.9259	0.6674	<b>0.9501</b>	0.9141	0.9041	0.9088	0.9159
9	<b>0.9017</b>	0.7772	0.9014	0.8727	0.8880	0.8906	0.8948
10	<b>0.8367</b>	0.7041	0.8151	0.8359	0.8082	0.8269	0.8283
11	<b>0.9246</b>	0.8853	0.8994	0.9204	0.9054	0.9014	0.9032
12	<b>0.7724</b>	0.5217	0.6804	0.7532	0.7492	0.7601	0.7565
13	<b>0.9145</b>	0.9144	0.9037	0.8680	0.8515	0.8331	0.8940

14	<b>0.6465</b>	0.5331	0.6395	0.6402	0.6372	0.6441	0.6348
15	<b>0.6420</b>	0.6316	0.6376	0.6383	0.6339	0.6412	0.6393
Average	<b>0.8228</b>	0.7313	0.7974	0.8121	0.8090	0.8085	0.8148
rate		12.51% ↑	3.18% ↑	1.31% ↑	1.71% ↑	1.76% ↑	0.97% ↑
KNN							
ID	NRS- $\epsilon$	MARA	RARR	BFRS	AG	SEF	GAAR
1	<b>0.9460</b>	0.8825	0.8377	0.9427	0.9378	0.9428	0.9387
2	<b>0.8357</b>	0.6380	0.8349	0.8155	0.8145	0.8253	0.8246
3	0.7977	0.6518	<b>0.8102</b>	0.7494	0.7526	0.7656	0.7867
4	<b>0.9924</b>	0.5949	0.9908	0.9737	0.9689	0.9771	0.9660
5	<b>0.9452</b>	0.7945	0.9443	0.8743	0.8709	0.8532	0.8996
6	<b>0.8031</b>	0.7612	0.7018	0.7555	0.7539	0.7597	0.7825
7	0.9062	<b>0.9357</b>	0.9044	0.9024	0.8903	0.9065	0.9002
8	<b>0.9786</b>	0.6906	0.8601	0.9663	0.9592	0.9678	0.9627
9	<b>0.9341</b>	0.6991	0.9284	0.9102	0.9233	0.9296	0.9245
10	<b>0.8854</b>	0.6707	0.8761	0.8749	0.8541	0.8823	0.8707
11	<b>0.9706</b>	0.9705	0.9657	0.9358	0.9309	0.9396	0.9457
12	<b>0.9299</b>	0.5301	0.7138	0.8746	0.8745	0.8935	0.8808
13	0.9222	0.6058	<b>0.9362</b>	0.8632	0.8705	0.8662	0.8935
14	<b>0.7942</b>	0.4714	0.7889	0.7682	0.7727	0.7934	0.7813
15	0.8164	<b>0.8574</b>	0.8273	0.8067	0.8063	0.825	0.8141
Average	<b>0.8972</b>	0.7170	0.8614	0.8676	0.8653	0.8752	0.8781
rate		25.14% ↑	4.16% ↑	3.41% ↑	3.68% ↑	2.51% ↑	2.17% ↑
SVM							
ID	NRS- $\epsilon$	MARA	RARR	BFRS	AG	SEF	GAAR
1	<b>0.9739</b>	0.8679	0.8877	0.9733	0.9671	0.9698	0.9675
2	<b>0.8918</b>	0.6581	0.8774	0.8771	0.8748	0.8852	0.8783
3	0.8754	0.6134	<b>0.9050</b>	0.8575	0.8662	0.8668	0.8477
4	<b>0.9807</b>	0.7282	0.7456	0.9749	0.9678	0.9797	0.9641
5	<b>0.7782</b>	0.7557	0.7779	0.7476	0.7543	0.7576	0.7576
6	<b>0.7926</b>	0.6862	0.7394	0.7609	0.749	0.7519	0.7729
7	0.9247	0.9992	<b>1.0001</b>	0.9311	0.9015	0.9376	0.9287
8	<b>0.9709</b>	0.6467	0.9567	0.9501	0.9267	0.9478	0.9459
9	<b>0.9670</b>	0.7753	0.9655	0.9536	0.9566	0.9646	0.9639
ID	NRS- $\epsilon$	MARA	RARR	BFRS	AG	SEF	GAAR
10	<b>1.0000</b>	<b>1.0000</b>	<b>1.0000</b>	<b>1.0000</b>	<b>1.0000</b>	<b>1.0000</b>	<b>1.0000</b>
11	0.9975	<b>1.0000</b>	0.9995	0.9846	0.9945	0.9974	<b>1.0000</b>
12	<b>0.9694</b>	0.5583	0.8165	0.9141	0.9117	0.9334	0.9225
13	0.9338	0.9144	<b>0.9587</b>	0.8968	0.8978	0.8711	0.9314
14	0.9245	<b>1.0000</b>	0.9240	0.9221	0.9208	0.9323	0.9262
15	<b>0.9619</b>	0.9611	0.9455	0.9382	0.9260	0.9362	0.9421
Average	<b>0.9295</b>	0.8110	0.9000	0.9121	0.9077	0.9154	0.9166
rate		14.61% ↑	3.30% ↑	1.90% ↑	2.41% ↑	1.54% ↑	1.41% ↑

The classification stability index reflects the degree of deviation of prediction labels when data perturbation occurs. Higher values of classification stability mean more stable prediction labels, indicating better quality of the corresponding reduction. Improvements in classification stability mean

increased stability of prediction label results and reduced interference with training samples. After analyzing the 15 datasets using these three classifiers, Table 4 and Figure 2 present the findings of each algorithm in terms of classification stability. It should be noted that the classification stability index reflects the degree of change in prediction labels when data is perturbed. Higher classification stability values indicate more stable prediction labels, meaning the related redundancy reduction has a higher quality.

1. Across many datasets, the NRS- $\epsilon$  algorithm exhibits leading performance compared to other algorithms in terms of classification stability, playing a leading role. For example, in the "Iris Plants Database (ID: 2)" dataset, significant differences in classification accuracy were observed under different classifiers for NRS- $\epsilon$  and other algorithms: when using the CART classifier, the accuracy of NRS- $\epsilon$  reached 0.7364, while MARA, RARR, BFRS, AG, SEF, and GAAR had accuracies of 0.6794, 0.7122, 0.6981, 0.7244, 0.7130, and 0.7276 respectively; when using the KNN classifier, the accuracy of NRS- $\epsilon$  was 0.8357, with MARA, RARR, BFRS, AG, SEF, and GAAR algorithms having accuracies of 0.6380, 0.8349, 0.8155, 0.8145, 0.8253, and 0.8246 respectively; when using the SVM classifier, the accuracies of NRS- $\epsilon$ , MARA, RARR, BFRS, AG, SEF, and GAAR were 0.8918, 0.6581, 0.8774, 0.8771, 0.8748, 0.8852, 0.8783 respectively.
2. Regarding average classification accuracy, the stability of NRS- $\epsilon$  markedly surpasses that of competing algorithms. Specifically, when using the CART classifier, the classification stability of NRS- $\epsilon$  was 0.8228, up to 12.51% higher than other methods; when using the KNN classifier, its classification stability was 0.8972, up to 25.14% higher than other methods; and through the SVM classifier, the classification stability of NRS- $\epsilon$  was 0.9295, up to 14.61% higher than other methods.

4.5. Comparisons of Elapsed Time

In this section, the time required for attribute reduction by different algorithms is compared. The results are shown in the Table 5.

Table 5. The elapsed time of all seven algorithms.

ID	NRS- $\epsilon$	MARA	RARR	BFRS	AG	SEF	GAAR
1	<b>429.7783</b>	0.8807	6.6227	57.0501	7.2915	6.5293	26.3625
2	<b>2.4335</b>	0.6513	0.8982	0.9413	0.1933	0.2173	0.3533
3	<b>21.0802</b>	0.3232	0.7559	1.9937	0.2217	0.1986	0.6452
4	<b>7.0519</b>	1.3462	4.2944	7.0359	1.4769	1.1096	1.9095
5	<b>14.0703</b>	0.2167	5.3504	11.7332	1.9639	1.8331	3.4963
6	122.1212	6.9838	<b>421.1056</b>	154.8219	31.4599	33.3661	54.0532
7	<b>233.6055</b>	81.7964	21.1919	80.4803	20.2694	18.389	35.7563
8	36.3329	0.5928	28.5184	<b>66.8708</b>	11.0547	9.1725	15.1738
ID	NRS- $\epsilon$	MARA	RARR	BFRS	AG	SEF	GAAR
9	70.3101	5.7716	<b>532.7057</b>	286.4753	49.7015	38.066	73.976
10	7.9717	0.6485	1.8745	<b>9.1748</b>	1.5362	1.1173	2.0991
11	<b>2.3576</b>	0.5797	0.6115	1.0526	0.2311	0.2115	0.3641
12	18.2902	0.5314	6.0787	<b>43.3079</b>	8.0851	5.6702	11.0026
13	167.9126	15.3875	<b>1605.0476</b>	1466.822	254.7792	157.5132	327.6303
14	<b>8.5942</b>	0.6286	0.7663	3.7765	0.5923	0.5887	0.9772
15	345.9941	10.34	<b>1946.5721</b>	1403.8344	207.5309	167.8453	310.1552
Average rate	99.1936	8.4452	<b>305.4930</b>	239.6914	39.7592	29.4552	57.5970
		1001% $\uparrow$	97.23% $\downarrow$	27.45% $\uparrow$	503.1 % $\uparrow$	34.98% $\uparrow$	48.86 % $\downarrow$

An increase in the value of dimensionality reduction stability correlates with an extended length of reduction. Through an in-depth analysis of the table below, the findings listed below can be easily derived. The reduction length of NRS- $\epsilon$  is longer, suggesting that there's a need to enhance the algorithm's time efficiency throughout the simplification process.

When analyzing the average processing time performance of algorithms, from the perspective of average time consumed, it is noteworthy that the value of NRS- $\epsilon$  is reduced by 97.23% and 48.86% compared to RARR and GAAR, respectively. Taking the dataset "Car Evaluation (ID: 6)" as an example, the time consumed by NRS- $\epsilon$ , MARA, RARR, BFRS, AG, SEF, and GAAR are 122.1212 seconds, 6.9838 seconds, 421.1056 seconds, 154.8219 seconds, 31.4599 seconds, 33.3661 seconds, and 54.0532 seconds, respectively. Hence, under certain conditions, the time NRS- $\epsilon$  takes for attribute reduction is less compared to RARR and BFRS.

Based on the discussion, it is evident that while our novel algorithm exhibits better time efficiency compared to RARR and BFRS on certain datasets, the speed of NRS- $\epsilon$  requires further enhancement.

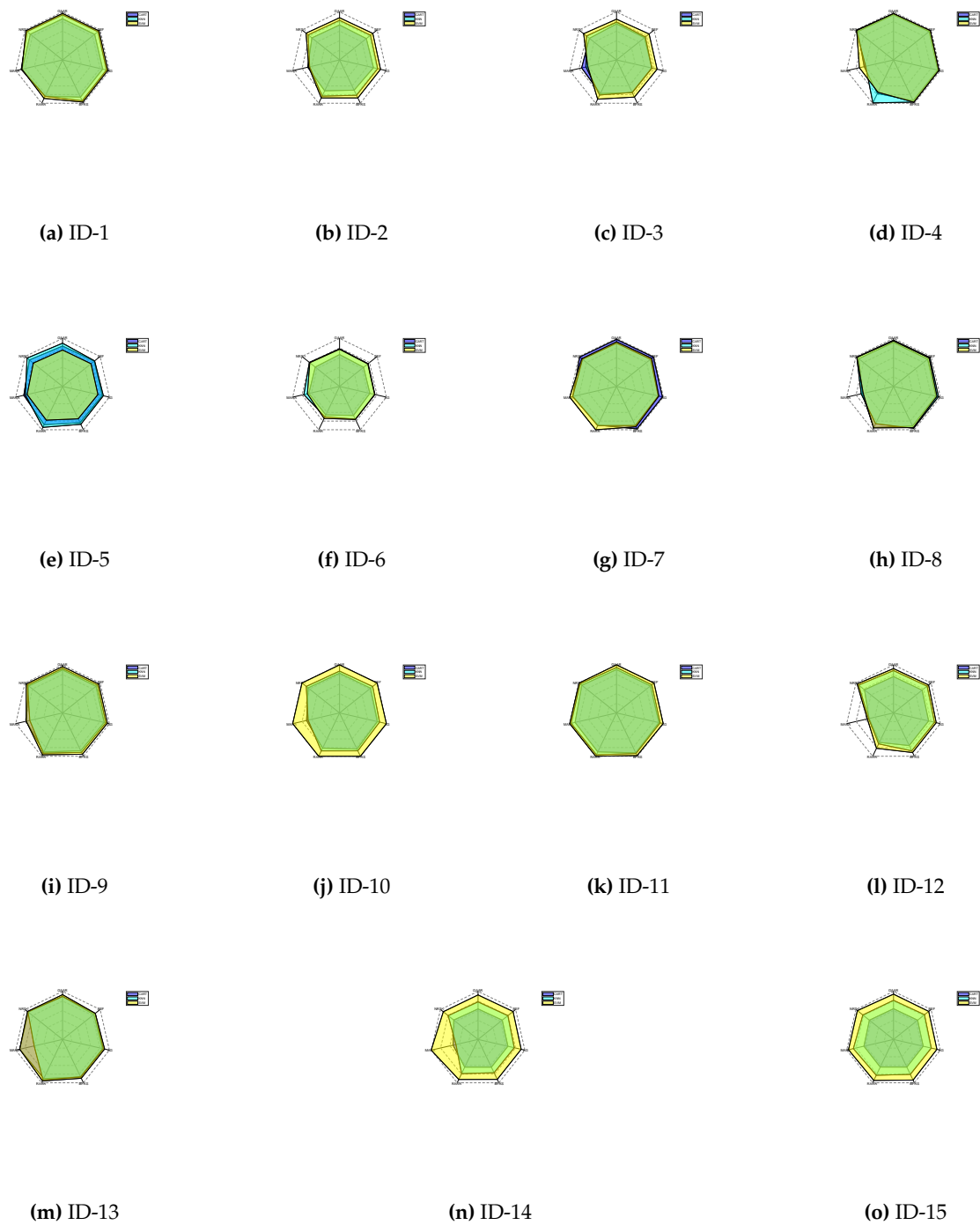
#### 4.6. Comparison of Attribute Dimensionality Reduction Stability

In this section, the attribute dimensionality reduction stability related to 15 datasets is presented. Table 6 shows that the dimensionality reduction stability of NRS- $\epsilon$  is slightly lower than GAAR and SEF but still maintains a leading position. Compared to MARA, RARR, BFRS, and AG, the average dimensionality reduction stability value of NRS- $\epsilon$  has increased by 100.2%, 49.89%, 27.19%, and 14.15%, respectively, while it only decreased by 19.323% and 6.677% compared to GAAR and SEF.

**Table 6.** The stabilities of all seven algorithms.

ID	NRS- $\epsilon$	MARA	RARR	BFRS	AG	SEF	GAAR
1	<b>0.6587</b>	0.1275	0.4888	0.1535	0.1232	0.2059	0.6033
2	0.5761	0.6308	<b>0.9277</b>	0.2903	0.2941	0.3647	0.5004
3	0.8504	0.2051	<b>0.9506</b>	0.6211	0.4254	0.6466	0.9209
4	0.9271	0.4013	0.9224	0.7869	0.8033	<b>0.9356</b>	0.898
5	0.8246	<b>1.0000</b>	0.9045	0.7659	0.7281	0.8687	0.8007
6	<b>0.9006</b>	0.1498	0.6007	0.5958	0.5818	0.7154	0.6374
7	0.5577	0.4933	<b>0.7277</b>	0.2996	0.1764	0.3813	0.5346
8	0.9153	0.3103	<b>1.0000</b>	0.8698	0.6712	0.8967	0.7945
9	0.8959	<b>1.0000</b>	<b>1.0000</b>	0.7955	0.7183	0.8788	0.8257
10	0.8246	0.4051	<b>0.8588</b>	0.8033	0.7261	0.8280	0.815
11	<b>0.6606</b>	0.0546	0.3773	0.5284	0.3484	0.3265	0.5815
12	0.9146	0.2001	<b>0.9917</b>	0.8328	0.8259	0.9281	0.8708
13	0.9054	0.3037	<b>1.0000</b>	0.7155	0.5962	0.7672	0.7502
14	0.7816	0.1602	<b>0.8678</b>	0.6017	0.6038	0.7438	0.7165
15	<b>0.8608</b>	0.6001	0.4011	0.7882	0.6549	0.7723	0.7442
Average	<b>0.8034</b>	0.4014	0.8012	0.6299	0.5518	0.6840	0.7329
rate		100.2% $\uparrow$	49.89% $\downarrow$	27.19% $\uparrow$	14.15% $\uparrow$	19.323% $\downarrow$	6.677% $\downarrow$

Although NRS- $\epsilon$  does not fall short of GAAR and SEF's results in terms of dimensionality reduction stability on many datasets, in some cases, its results in attribute dimensionality reduction are superior to the six advanced algorithms. For example, for the "Letter Recognition (ID: 15)" dataset, the dimensionality reduction stability of NRS- $\epsilon$ , MARA, RARR, BFRS, AG, SEF, and GAAR were 0.8608, 0.6001, 0.4011, 0.7882, 0.6549, 0.7723, and 0.7442, respectively. Compared to other algorithms, the results of NRS- $\epsilon$  improved by 43.47%, 115.2%, 9.211%, 31.44%, 11.46%, and 15.67%, respectively. Thus, it's important to recognize that employing NRS- $\epsilon$  favors the selection of attributes better aligned with variations in samples.



**Figure 2.** Classification stabilities of three classifiers.

**5. Conclusion and Future Expectations**

To address the challenges of high-dimensional data, a new attribute reduction search strategy is proposed in this paper. Unlike traditional methods, on the one hand, by integrating multi-granularity into a new model, the new model becomes more flexible and adaptable to data features at different levels. On the other hand, the new model combines supervised and unsupervised perspectives, overcoming the limitations of single-attribute measurement methods.

Through experiments on 15 UCI datasets, it is not difficult to find that our proposed strategy possesses excellent classification performance as well as strong stability during the dimensionality reduction process. However, our strategy still has some shortcomings, which are specifically manifested in the following aspects:

1. Although the use of integrated measurement methods has improved the quality of attribute selection, it has also correspondingly increased the time cost of the selection process. For this reason, future work can explore introducing more acceleration technologies to optimize efficiency and reduce the required time.
2. Given the high universality of the search framework proposed in this paper, future research can attempt to use different rough set-based fundamental measurements to replace the current information gain ratio, to assess the performance of different measurement methods in classification performance.

**Author Contributions:** Conceptualization, J.C.; methodology, J.C.; software, Z.F.; validation, J.S.; formal analysis, H.C.; investigation, T.X.; resources, J.S.; data curation, T.X.; writing—original draft preparation, Z.F.; writing—review and editing, T.X.; visualization, Z.F.; supervision, J.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Natural Science Foundation of China (Grant No.62006099), Industry-school Cooperative Education Program of the Ministry of Education (Grant No. 202101363034).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare there are no conflicts of interest.

## References

1. Z.Pawlak. "Rough sets." International journal of computer & information sciences 11 (1982): 341-356.<https://doi.org/10.1145/219717.219791>
2. M. Dash, H. Liu. "Consistency-based search in feature selection." Artificial intelligence 151, no. 1-2 (2003): 155-176. [https://doi.org/10.1016/S0004-3702\(03\)00079-1](https://doi.org/10.1016/S0004-3702(03)00079-1)
3. M. Dowlatshahi, V. Derhami, H. Nezamabadi-pour. "Ensemble of filter-based rankers to guide an epsilon-greedy swarm optimizer for high-dimensional feature subset selection." Information 8, no. 4 (2017): 152. <https://doi.org/10.3390/info8040152>
4. H. Chen, T. Li, C. Luo, SJ. Horng, G. Wang. "A decision-theoretic rough set approach for dynamic data mining." IEEE Transactions on fuzzy Systems 23, no. 6 (2015): 1958-1970. <https://doi.org/10.1109/TFUZZ.2014.2387877>
5. Y. Qian, J. Liang, Z. Wei, C. Dang. "Information granularity in fuzzy binary GrC model." IEEE Transactions on Fuzzy Systems 19, no. 2 (2010): 253-264. <https://doi.org/10.1109/TFUZZ.2010.2095461>
6. Y. Qian, F. Li, J. Liang, B. Liu, C. Dang. "Space structure and clustering of categorical data." IEEE transactions on neural networks and learning systems 27, no. 10 (2015): 2047-2059. <https://doi.org/10.1109/TNNLS.2015.2451151>
7. J. Qian, C. Liu, D. Miao, X. Yue. "Sequential three-way decisions via multi-granularity." Information Sciences 507 (2020): 606-629. <https://doi.org/10.1016/j.ins.2019.03.052>
8. S. Wan, F. Wang, J. Dong. "A preference degree for intuitionistic fuzzy values and application to multi-attribute group decision making." Information Sciences 370 (2016): 127-146. <https://doi.org/10.1016/j.ins.2016.07.013>
9. Z. Qiang, L. Jian, Y. Fan, S. Qing, Y. ZhiHai. "Subjective weight determination method of evaluation index based on intuitionistic fuzzy set theory." In 2022 34th Chinese Control and Decision Conference (CCDC), pp. 2858-2861. IEEE, 2022. <https://doi.org/10.1109/CCDC55256.2022.10034296>
10. K. Liu, X. Yang, H. Fujita, D. Liu, X. Yang, Y. Qian. "An efficient selector for multi-granularity attribute reduction." Information Sciences 505 (2019): 457-472. <https://doi.org/10.1016/j.ins.2019.07.051>

11. Z. Jiang, K. Liu, X. Yang, H. Yu, H. Fujita, Y. Qian. "Accelerator for supervised neighborhood based attribute reduction." *International Journal of Approximate Reasoning* 119 (2020): 122-150. <https://doi.org/10.1016/j.ijar.2019.12.013>
12. Z. Yuan, H. Chen, T. Li, Z. Yu, B. Sang, C. Luo. "Unsupervised attribute reduction for mixed data based on fuzzy rough sets." *Information Sciences* 572 (2021): 67-87. <https://doi.org/10.1016/j.ins.2021.04.083>
13. X. Yang, Y. Yao. "Ensemble selector for attribute reduction." *Applied Soft Computing* 70 (2018): 1-11. <https://doi.org/10.1016/j.asoc.2018.05.013>
14. Y. Chen, P. Wang, X. Yang, H. Yu. "Bee: towards a robust attribute reduction." *International Journal of Machine Learning and Cybernetics* 13, no. 12 (2022): 3927-3962. <https://doi.org/10.1007/s13042-022-01633-4>
15. J. Li, X. Yang, X. Song, J. Li, P. Wang, D. Yu. "Neighborhood attribute reduction: a multi-criterion approach." *International journal of machine learning and cybernetics* 10 (2019): 731-742. <https://doi.org/10.1007/s13042-017-0758-5>
16. J. Dai, W. Wang, H. Tian, L. Liu. "Attribute selection based on a new conditional entropy for incomplete decision systems." *Knowledge-Based Systems* 39 (2013): 207-213. <https://doi.org/10.1016/j.knosys.2012.10.018>
17. Y. Qian, J. Liang, W. Pedrycz, C. Dang. "Positive approximation: an accelerator for attribute reduction in rough set theory." *Artificial intelligence* 174, no. 9-10 (2010): 597-618. <https://doi.org/10.1016/j.artint.2010.04.018>
18. Q. Hu, W. Pedrycz, D. Yu, J. Lang. "Selecting discrete and continuous features based on neighborhood decision error minimization." *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 40, no. 1 (2009): 137-150. <https://doi.org/10.1109/TSMCB.2009.2024166>
19. S. Xu, H. Ju, L. Shang, W. Pedrycz, X. Yang, C. Li. "Label distribution learning: a local collaborative mechanism." *International Journal of Approximate Reasoning* 121 (2020): 59-84. <https://doi.org/10.1016/j.ijar.2020.02.003>
20. Y. Chen, P. Wang, X. Yang, J. Mi, D. Liu. "Granular ball guided selector for attribute reduction." *Knowledge-Based Systems* 229 (2021): 107326. <https://doi.org/10.1016/j.knosys.2021.107326>
21. X. Yang, S. Xu, H. Dou, X. Song, H. Yu, J. Yang. "Multigranulation rough set: a multiset based strategy." *International Journal of Computational Intelligence Systems* 10, no. 1 (2017): 277-292. <https://doi.org/10.2991/ijcis.2017.10.1.19>
22. X. Yang, S. Liang, H. Yu, S. Gao, Y. Qian. "Pseudo-label neighborhood rough set: measures and attribute reductions." *International journal of approximate reasoning* 105 (2019): 112-129. <https://doi.org/10.1016/j.ijar.2018.11.010>
23. Y. Qian, S. Li, J. Liang, Z. Shi, F. Wang. "Pessimistic rough set based decisions: a multigranulation fusion strategy." *Information Sciences* 264 (2014): 196-210. <https://doi.org/10.1016/j.ins.2013.12.014>
24. Y. Qian, J. Liang, Y. Yao, C. Dang. "MGRS: A multi-granulation rough set." *Information sciences* 180, no. 6 (2010): 949-970. <https://doi.org/10.1016/j.ins.2009.11.023>
25. G. Lin, Y. Qian, J. Li. "NMGRS: Neighborhood-based multigranulation rough sets." *International Journal of Approximate Reasoning* 53, no. 7 (2012): 1080-1093. <https://doi.org/10.1016/j.ijar.2012.05.004>
26. Y. Qian, J. Liang, Y. Yao, C. Dang. "MGRS: A multi-granulation rough set." *Information sciences* 180, no. 6 (2010): 949-970. <https://doi.org/10.1016/j.ins.2009.11.023>
27. Y. Qian, H. Zhang, Y. Sang, J. Liang. "Multigranulation decision-theoretic rough sets." *International journal of approximate reasoning* 55, no. 1 (2014): 225-237. <https://doi.org/10.1016/j.ijar.2013.03.004>
28. Z. Xu, D. Yuan, W. Song, W. Cai. "New attribute reduction based on rough set." In *2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery*, vol. 5, pp. 271-275. IEEE, 2008. <https://doi.org/10.1109/FSKD.2008.657>
29. M. Song, J. Chen, J. Song, T. Xu, Y. Fan. "Forward Greedy Searching to  $\kappa$ -Reduct Based on Granular Ball." *Symmetry* 15, no. 5 (2023): 996. <https://doi.org/10.3390/sym15050996>
30. E. Xu, X. Gao, W. Tan. "Attributes Reduction Based On Rough Set." In *2006 International Conference on Machine Learning and Cybernetics*, pp. 1438-1442. IEEE, 2006. <https://doi.org/10.1109/ICMLC.2006.258755>
31. J. Liang, Y. Zhang, Y. Qu. "A heuristic algorithm of attribute reduction in rough set." In *2005 International Conference on Machine Learning and Cybernetics*, vol. 5, pp. 3140-3142. IEEE, 2005. <https://doi.org/10.1109/ICMLC.2005.1527482>

32. X. Xu, Y. Niu, Y. Niu. "Research on attribute reduction algorithm based on Rough Set Theory and genetic algorithms." In 2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), pp. 524-527. IEEE, 2011. <https://doi.org/10.1109/AIMSEC.2011.6010295>
33. J. Liang, X. Zhao, D. Li, F. Cao, C. Dang. "Determining the number of clusters using information entropy for mixed data." Pattern Recognition 45, no. 6 (2012): 2251-2265. <https://doi.org/10.1016/j.patcog.2011.12.017>
34. J. Dai, H. Hu, W. Wu, Y. Qian, D. Huang. "Maximal-discernibility-pair-based approach to attribute reduction in fuzzy rough sets." IEEE Transactions on Fuzzy Systems 26, no. 4 (2017): 2174-2187. <https://doi.org/10.1109/TFUZZ.2017.2768044>
35. Y. Yang, D. Chen, H. Wang. "Active sample selection based incremental algorithm for attribute reduction with rough sets." IEEE Transactions on Fuzzy Systems 25, no. 4 (2016): 825-838. <https://doi.org/10.1109/TFUZZ.2016.2581186>
36. J. Wang, Y. Liu, J. Chen, X. Yang. "An Ensemble Framework to Forest Optimization Based Reduct Searching." Symmetry 14, no. 6 (2022): 1277. <https://doi.org/10.3390/sym14061277>
37. C. Chang, C. Lin. "LIBSVM: a library for support vector machines." ACM transactions on intelligent systems and technology (TIST) 2, no. 3 (2011): 1-27. <https://doi.org/10.1145/1961189.1961199>
38. P. Wang, H. Shi, X. Yang, J. Mi. "Three-way k-means: integrating k-means and three-way decision." International journal of machine learning and cybernetics 10 (2019): 2767-2777. <https://doi.org/10.1007/s13042-018-0901-y>
39. Y. Pan, W. Xu, Q. Ran. "An incremental approach to feature selection using the weighted dominance-based neighborhood rough sets." International Journal of Machine Learning and Cybernetics 14, no. 4 (2023): 1217-1233. <https://doi.org/10.1007/s13042-022-01695-4>
40. L. Yin, J. Li, Z. Jiang, J. Ding, X. Xu. "An efficient attribute reduction algorithm using MapReduce." Journal of Information Science 47, no. 1 (2021): 101-117. <https://doi.org/10.1177/0165551519874617>
41. K. Fukunaga, P. Narendra. "A branch and bound algorithm for computing k-nearest neighbors." IEEE transactions on computers 100, no. 7 (1975): 750-753. <https://doi.org/10.1109/T-C.1975.224297>
42. D. Lianjie, C. Degang, W. Ningling, L. Zhanhui. "Key energy-consumption feature selection of thermal power systems based on robust attribute reduction with rough sets." Information Sciences 532 (2020): 61-71. <https://doi.org/10.1016/j.ins.2020.03.085>
43. G. Ali, M. Akram, J. Alcantud. "Attributes reductions of bipolar fuzzy relation decision systems." Neural computing and applications 32, no. 14 (2020): 10051-10071. <https://doi.org/10.1007/s00521-019-04536-8>
44. Y. Chen, K. Liu, J. Song, H. Fujita, X. Yang, Y. Qian. "Attribute group for attribute reduction." Information Sciences 535 (2020): 64-80. <https://doi.org/10.1016/j.ins.2020.05.010>
45. M. Hu, E. Tsang, Y. Guo, W. Xu. "Fast and robust attribute reduction based on the separability in fuzzy decision systems." IEEE transactions on cybernetics 52, no. 6 (2021): 5559-5572. <https://doi.org/10.1109/TCYB.2020.3040803>
46. F. Iqbal, J. Hashmi, B. Fung, R. Batool, A. Khattak, S. Aleem, P. Hung. "A hybrid framework for sentiment analysis using genetic algorithm based feature reduction." IEEE Access 7 (2019): 14637-14652. <https://doi.org/10.1109/ACCESS.2019.2892852>
47. T. Xing, J. Chen, T. Xu, Y. Fan. "Fusing Supervised and Unsupervised Measures for Attribute Reduction." Intelligent Automation & Soft Computing 37, no. 1 (2023). <https://doi.org/10.32604/iasc.2023.037874>
48. Z. Yin, Y. Fan, P. Wang, J. Chen. "Parallel Selector for Feature Reduction." Mathematics 11, no. 9 (2023): 2084. <https://doi.org/10.3390/math11092084>

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.