





A novel approach to antinoise multi-granularity classification through graph-based feature selection

Xiaoyan Zhang ^{*}, Xuan Shen , Weicheng Zhao

College of Artificial Intelligence, Southwest University, Chongqing, 400715, PR China

ARTICLE INFO

Keywords:

Weighted multi-granulation fuzzy rough set
Feature selection
Graph theory
Fuzzy information system

ABSTRACT

This study proposes a novel anti-noise multi-granularity classification method that incorporates graph-based feature selection to effectively reduce computational complexity while enhancing classification accuracy and robustness. Traditional multi-granularity fuzzy rough set (MFRS) models are often highly sensitive to noisy samples, which hinders the acquisition of reliable knowledge. To address this issue, we design a heuristic feature selection algorithm within the framework of the weighted multi-granulation neighborhood-constrained fuzzy rough set (WMNcFRS) model. The algorithm first utilizes graph theory to evaluate inter-feature correlation and redundancy, enabling efficient partitioning of multi-granularity spaces. It then introduces a granularity-weighted strategy that ranks and prioritizes granularities based on approximation precision, thereby improving approximation capability. Finally, the feature selection strategy is formulated using the fuzzy dependency measure defined in the WMNcFRS model, which effectively suppresses noise interference, reduces the computational costs, and enhances the model's generalization ability. Extensive experiments on fifteen publicly available datasets demonstrate that, compared to six state-of-the-art algorithms, the proposed method exhibits superior robustness and classification performance, validating its effectiveness in practical applications.

1. Introduction

Feature selection [1], a robust data preprocessing tool, has found extensive applications in areas like data mining and image processing. It can extract key features from high-dimensional and redundant data, thereby enhancing the efficiency of models [2,3].

The Fuzzy Rough Set (FRS) model, recognized for its unique approach to rough approximation in managing uncertainty, has gained significant popularity in feature selection. Tsang et al. [4] utilize discernibility matrices based on FRS for feature selection. Hu et al. [5] use an evaluative function of separability to assess attributes and subsequently perform attribute reduction. Al-shami [6] introduced containment neighborhood rough sets and applied them to classification tasks in medical data. In practical applications, data commonly contain noise. The FRS model is highly sensitive to this noise due to its close dependence on the distance between fuzzy relations and sample data. To tackle this issue, An et al. [7] developed the Relative FRS model, which utilizes relative fuzzy membership degrees for attribute reduction. Zhang et al. [8] proposed a feature selection and image edge detection method based on an improved overlap-function fuzzy rough set model. Furthermore, Beynon [9] devised a method for β -reduction in decision information systems, employing β -reduction criteria within variable precision rough sets. Hu et al. [10] utilized particle swarm

^{*} Corresponding author.

E-mail addresses: zhangxiaoyan@swu.edu.cn (X. Zhang), sx18908529566@163.com (X. Shen), 2504975527@qq.com (W. Zhao).

<https://doi.org/10.1016/j.ins.2025.122632>

Received 13 November 2024; Received in revised form 17 July 2025; Accepted 21 August 2025

Table 1

The summary of existing methods.

Methods	Categories	Disadvantages
Fuzzy rough set	Feature selection based on fuzzy rough sets. [4–6]	The model is highly sensitive to noise.
	Fuzzy rough sets that focus on sample relations. [7,8]	The model lacks sufficient robustness and fails to address the inclusiveness between information granules and concepts.
	Approaches considering the relationship between information granules and concepts. [9–11]	The influence of decision information on neighborhood relationships is overlooked.
	Methods adjusting the relationships between samples. [12,13]	
Multi-granulation rough set	Methods assuming all granularities are equally important. [14–17]	The importance of granularity is different in actual scenarios.
	Weighted multi-granulation rough set [18–21]	These methods are not applied to fuzzy sets and involve high computational complexity.
Graph theory	Methods using a priori information about graph structures to construct the feature space. [23–26]	These methods rely heavily on prior knowledge.

optimization to develop a granular fuzzy model, incorporating mixed granules and incomplete data. Meanwhile, Hu et al. [11] investigated multi-view data through anchors and graph subspace clustering, employing information granules for fuzzy modeling. The models mentioned above aim to bolster robustness by prioritizing information granules and conceptual relationships. In contrast, some scholars have improved the robustness of fuzzy rough sets (FRS) by adjusting inter-sample relationships. For example, An et al. [12] explored similarity measurement through probability density and introduced the concept of probability granule distance, Hu et al. [13] quantified object similarity using the k-nearest neighbor samples of each object. However, these studies generally overlook the influence of decision information on neighborhood construction. Therefore, this paper places particular emphasis on investigating neighborhood structures guided by decision information, aiming to improve both the discriminative power and robustness of the model—an aspect that remains largely underexplored in existing research.

Furthermore, models based on a single binary relation fail to meet the diversity requirements of complex decision systems. To address this limitation, Qian et al. [14] extended the classical rough set model into a multi-granulation framework. Building on this idea, Feng et al. [15] proposed two variable precision multi-granulation fuzzy rough set models, introducing the concepts of maximum and minimum membership degrees. To further enhance decision modeling, Sun et al. [16] integrated the multi-granulation perspective into the decision-theoretic rough set model and proposed a variable precision multi-granulation fuzzy decision-theoretic rough set for handling complex decision-making tasks. Jiang et al. [17] have conducted extensive research on covering-based multi-granulation $(\mathcal{I}, \mathcal{T})$ fuzzy rough sets and their variable-precision variants. All of the above models assume equal importance across multiple granularities. However, in practical scenarios, it is crucial to assign distinct weights to different granular sources. To this end, Xu [18] introduced a series of granularity weighting strategies and constructed a generalized multi-granularity rough set model. Building upon this work, Guo et al. [19] proposed a double-weighted generalized multi-granularity decision-theoretic rough set model tailored for multi-source decision systems. In a related study, Ji et al. [20] investigated a multi-granularity intuitionistic fuzzy rough set model with a specific emphasis on granularity weighting. More recently, Yu et al. [21] presented a multi-label feature selection method based on variable-degree multi-granulation decision-theoretic rough sets, Zhang et al. [22] proposed a feature selection algorithm based on generalized multigranulation fuzzy neighborhood rough sets, demonstrating its flexibility and effectiveness in handling uncertainty. Nevertheless, there remains a significant gap in the integration of granularity weighting mechanisms into multi-granulation fuzzy rough set (MFRS) models. In large-scale data environments requiring efficient processing, incorporating granularity weighting offers a promising solution by filtering out low-importance granular sources and reducing computational complexity. Therefore, fusing granularity weighting with MFRS not only addresses this research gap but also opens a new direction for developing robust and scalable feature selection frameworks.

It is noteworthy that with the expansion of data sample size and the increase in feature dimensionality, the aforementioned models and methodologies face the significant challenge of escalating computational complexity. Graphs, being a highly effective and commonly used means for representing data relationships, are increasingly being utilized in feature selection methods owing to their exceptional performance [23,24]. Feature selection algorithms based on graphs can be divided into two critical stages. The first is graph construction, which focuses on generating a graph structure that accurately maps the distribution of the actual sample space, in order to capture the complex relationships between data features. The second is the feature evaluation stage, where through in-depth analysis performed on this structural graph, features that contain important information are retained, while those with high redundancy are removed, thereby optimizing the feature subset [25,26].

Table 1 presents a classification and summary of existing studies based on different methods and application scenarios. A comprehensive analysis of prior work reveals that current feature selection approaches exhibit significant limitations when dealing with noisy data and complex fuzzy information systems. To address the aforementioned issues, this paper proposes a heuristic feature selection method based on weighted multi-granulation neighborhood-constrained fuzzy rough sets (WMNcFRS), from a multi-granularity perspective and guided by the two key entry points illustrated in Fig. 1. Unlike existing methods, our approach addresses the lack of decision-guided neighborhood construction and the assumption of equal granularity importance. It effectively reduces noise by leveraging decision information to refine neighborhood relationships and constructs an undirected graph in the fuzzy information

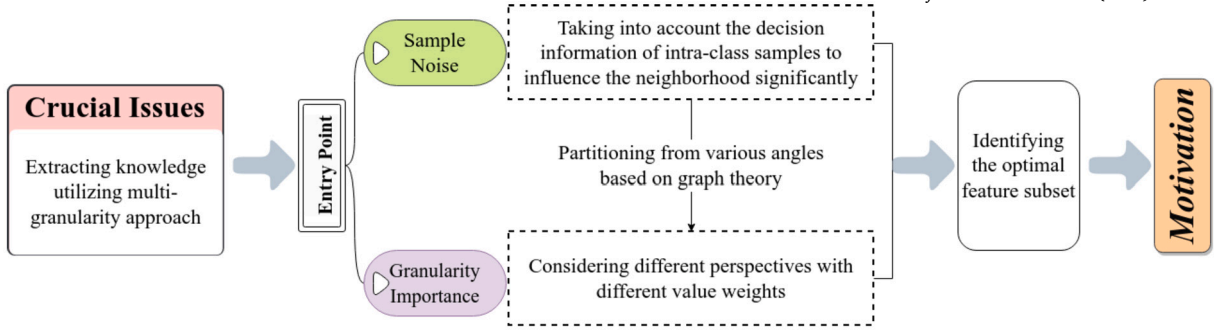


Fig. 1. The motivation framework of feature selection.

system to evaluate feature importance based on both relevance and redundancy. Furthermore, by introducing granularity weighting, a new weighted MFRS model is developed, which significantly reduces computational complexity and improves efficiency. The main innovations and contributions of this paper are summarized as follows:

1. By leveraging information from nearest neighbors within the same class to measure distances between different classes, the interference of noisy samples on attribute approximation ability is mitigated.
2. A graph is constructed in the fuzzy information system based on graph theory. Features are first processed according to correlation and non-redundancy principles, while the use of matrix power series properties significantly reduces computational complexity.
3. Incorporating the granularity weighting approach based on approximation precision, a novel WMNcFRS model is proposed, and a heuristic feature selection algorithm is devised using weighted optimistic multi-granularity fuzzy dependency.

The framework of the paper is outlined as follows. Section 2 presents a comprehensive overview of several related concepts, including neighborhood-constrained fuzzy rough sets, graph theory, and classical multi-granularity rough sets. In Section 3, a graph within the fuzzy information system is constructed, introducing a weighted multi-granularity neighborhood-constrained fuzzy rough set model and devising a corresponding feature extraction algorithm. Section 5 validates the effectiveness and efficiency of the proposed approach through experimentation on UCI datasets. The paper concludes by summarizing the findings and outlining future research directions in Section 6.

2. Preliminaries

In this section, we systematically introduce concepts related to neighborhood-constrained fuzzy rough sets, graph theory, and multi-granularity rough sets.

2.1. Neighborhood-constrained fuzzy rough set

Neighborhood-constrained fuzzy rough sets are capable of capturing local similarities among samples and effectively identifying decision regions with ambiguous boundaries and complex distributions, especially in scenarios involving uncertainty or fuzziness.

Let $FIS = \langle O, B \cup d, G \rangle$ denote a fuzzy information system, where $O = \{o_1, o_2, \dots, o_m\}$ is the set of objects, $B = \{b_1, b_2, \dots, b_n\}$ is the set of conditional attributes, and d is the decision attribute. Each $g(o, a)$ denotes the value of attribute a for object o , and the collection of all such values is represented as $G = \{g(o, a) | o \in O, a \in B \cup d\}$.

Based on the given fuzzy information system, where $\forall x, y \in O$, $b \in B$, and κ represents a positive number, the κ -nearest neighbor distance between x and y within b can be expressed as

$$d_\kappa(x, y) = \left| \frac{1}{|\mathbb{N}_\kappa(x)|} \sum_{x_i \in \mathbb{N}_\kappa(x)} g(x_i, b) - \frac{1}{|\mathbb{N}_\kappa(y)|} \sum_{y_i \in \mathbb{N}_\kappa(y)} g(y_i, b) \right| \quad (1)$$

where $\mathbb{N}_\kappa(x)$ represents the κ nearest samples including x , and these κ samples belong to the same class as x , that is $g(x, d) = g(y, d)$. Let $\mathbb{N}(x)$ represent all samples of the same class as object x , and $|\mathbb{N}(x)|$ is the number of samples in that class. If $|\mathbb{N}(x)| \leq \kappa$, then $\mathbb{N}_\kappa(x) = \mathbb{N}(x)$. Otherwise, the nearest κ neighbors of x within the same class will be chosen based on their order of occurrence.

Given $FIS = \langle O, B \cup d, G \rangle$, $\forall x, y \in O$, and $b \in B$, the neighborhood-constrained fuzzy similarity relation between x and y with respect to attribute b is defined as

$$\mathcal{R}_b^\kappa(x, y) = 1 - d_\kappa(x, y) \quad (2)$$

The neighborhood-constrained fuzzy similarity matrix of a over O is derived according to Eq. (2). Thus, for $\mathcal{A} \subseteq B$, the neighborhood-constrained fuzzy similarity matrix is designated as $\mathcal{R}_\mathcal{A}^\kappa(x, y) = \min_{b \in \mathcal{A}} \mathcal{R}_b^\kappa(x, y)$.

For $\mathcal{A} \subseteq B$, with $\mathcal{R}_\mathcal{A}$ represents the fuzzy similarity relation on \mathcal{A} , the fuzzy lower and upper approximations of $\mathcal{F} \in F(O)$, with $F(O)$ signifies the set of all fuzzy sets on O , are defined as

$$\begin{aligned}\underline{\mathcal{R}}_{\mathcal{A}}^{\mathcal{K}}(\mathcal{F})(x) &= \inf_{y \in O} \max(1 - \mathcal{R}_{\mathcal{A}}^{\mathcal{K}}(x, y), \mathcal{F}(y)), \\ \overline{\mathcal{R}}_{\mathcal{A}}^{\mathcal{K}}(\mathcal{F})(x) &= \sup_{y \in O} \min(\mathcal{R}_{\mathcal{A}}^{\mathcal{K}}(x, y), \mathcal{F}(y)).\end{aligned}\quad (3)$$

Additionally, for $U \subseteq O$, express Eq. (3) in simplified form as

$$\begin{aligned}\underline{\mathcal{R}}_{\mathcal{A}}^{\mathcal{K}}(\mathcal{F})(x) &= \inf_{y \notin U} (1 - \mathcal{R}_{\mathcal{A}}^{\mathcal{K}}(x, y)), \\ \overline{\mathcal{R}}_{\mathcal{A}}^{\mathcal{K}}(\mathcal{F})(x) &= \sup_{y \in U} (\mathcal{R}_{\mathcal{A}}^{\mathcal{K}}(x, y)).\end{aligned}\quad (4)$$

2.2. Graph theory

In [27], the attribute set $B = \{b_1, b_2, \dots, b_n\}$ can be depicted using weighted undirected fully connected graphs $G = (V, E)$ with an equal number of nodes, where the relationships between attributes are characterized by the edges in the graph. Graph G is represented by its adjacency matrix M , where $M(i, j)$, $1 \leq i, j \leq n$, denotes the relationship between features b_i and b_j (the node \vec{v}_i and \vec{v}_j), encapsulating the information among nodes. Each element $M(i, j)$ can be described using a weight function defined as a weighted linear combination of two statistical measures, specified as follows:

$$\psi(\vec{v}_i, \vec{v}_j) = \alpha \mathcal{H}_{ij} + (1 - \alpha) Q_{ij}$$

where $\alpha \in [0, 1]$ is the load indicator, its value is estimated through ten-fold cross-validation in classification tasks. \mathcal{H}_{ij} reflects the correlation information between the two samples, with higher values being better. Conversely, the term $Q_{ij} = 1 - |\xi(b_i, b_j)|$ is used to control non-redundancy and is opposite to the first term, where ξ denotes the spearman correlation coefficient, defined as follows:

$$\xi = 1 - \frac{6 \sum_{i=1}^m r_i^2}{m(m^2 - 1)}$$

m denotes the number of samples, while r indicates the difference in ranks among two objects. It can be seen that $\xi \in [-1, 1]$, with values closer to the extremes indicating stronger correlation.

2.3. Weighted multi-granulation rough sets

Let $FIS = \langle O, B \cup d, G \rangle$ denote a fuzzy information system, $\mathcal{A}_i \subseteq B$, $i = 1, 2, \dots, k (k \leq 2^{|B|})$ represent a granularity composed of some attributes, for $\forall o \in O$, $X \subseteq O$, the support feature function expressed as

$$\mathcal{P}_X^{\mathcal{A}_i}(o) = \begin{cases} 1, & \mathbb{R}_{\mathcal{A}_i}(o) \subseteq X, \\ 0, & \text{others,} \end{cases}$$

where the feature support function illustrates the inclusion relationship between the equivalence class $\mathbb{R}_{\mathcal{A}_i}(o)$ and X , that is, it determines whether the sample o belongs to the equivalence class defined by granularity \mathcal{A}_i within the set X .

Given $FIS = \langle O, B \cup d, G \rangle$, $\mathcal{P}_X^{\mathcal{A}_i}(o)$ as the support feature function, for $\mathcal{A}_i \subseteq B$, $i = 1, 2, \dots, k (k \leq 2^{|B|})$, the weighted optimistic multi-granulation rough set lower and upper approximations of any $X \subseteq O$ on \mathcal{A}_i , respectively, are defined as

$$\begin{aligned}\underline{\mathcal{O}\mathcal{M}}_{\sum_{i=1}^k \mathcal{A}_i}^{\omega}(X) &= \left\{ x \in O \mid \sum_{i=1}^k \varpi_i \mathcal{P}_X^{\mathcal{A}_i}(o) > 0 \right\}, \\ \overline{\mathcal{O}\mathcal{M}}_{\sum_{i=1}^k \mathcal{A}_i}^{\omega}(X) &= \left\{ x \in O \mid \sum_{i=1}^k \varpi_i (1 - \mathcal{P}_{\sim X}^{\mathcal{A}_i}(o)) \geq 1 \right\},\end{aligned}\quad (5)$$

likewise, the weighted pessimistic multi-granulation rough set lower and upper approximations of any $X \subseteq O$ on $\mathcal{A}_i \subseteq B (i = 1, 2, \dots, k \leq 2^{|B|})$, respectively, are defined as

$$\begin{aligned}\underline{\mathcal{P}\mathcal{M}}_{\sum_{i=1}^k \mathcal{A}_i}^{\omega}(X) &= \left\{ x \in O \mid \sum_{i=1}^k \varpi_i \mathcal{P}_X^{\mathcal{A}_i}(o) \geq 1 \right\}, \\ \overline{\mathcal{P}\mathcal{M}}_{\sum_{i=1}^k \mathcal{A}_i}^{\omega}(X) &= \left\{ x \in O \mid \sum_{i=1}^k \varpi_i (1 - \mathcal{P}_{\sim X}^{\mathcal{A}_i}(o)) > 0 \right\},\end{aligned}\quad (6)$$

where ϖ_i stands for the weight of granularity \mathcal{A}_i , with $\sum_{i=0}^k \varpi_i = 1$.

Only when $\underline{\mathcal{P}\mathcal{M}}_{\sum_{i=0}^k \mathcal{A}_i}^{\omega}(X) \neq \overline{\mathcal{P}\mathcal{M}}_{\sum_{i=0}^k \mathcal{A}_i}^{\omega}(X)$, the set X is both pessimistic and rough. This pair $\left(\underline{\mathcal{P}\mathcal{M}}_{\sum_{i=0}^k \mathcal{A}_i}^{\omega}(X), \overline{\mathcal{P}\mathcal{M}}_{\sum_{i=0}^k \mathcal{A}_i}^{\omega}(X) \right)$ is referred to as the weighted pessimistic multi-granulation rough set model. The scenario is equally applicable in optimistic contexts.

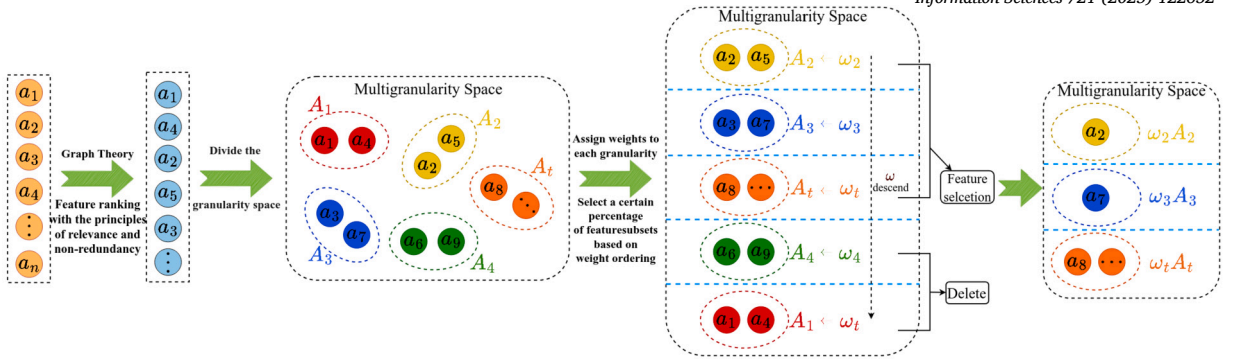


Fig. 2. The architecture of feature selection. Features are ranked using graph theory, divided into multi-granularity spaces, and selected based on granularity weights to construct the final subset.

It is worth noting that in the optimistic multi-granulation model, an object is considered to belong to the target concept if it is supported by at least one granularity. In contrast, the pessimistic model requires all granularities to support the membership decision.

Furthermore, this paper provides a concise overview of an approximate precision-based approach for granularity weighting. Other notable methods include approximate quality weighting and granularity entropy weighting, detailed further in reference [18].

Given $FIS = \langle O, B \cup d, G \rangle$, $\mathcal{A}_i \subseteq B$, $i = 1, 2, \dots, k$ ($k \leq 2^{|B|}$), $O/d = \{d_1, d_2, \dots, d_s\}$, the approximate accuracy associated with granularity \mathcal{A}_i is designated as

$$\varpi_i = \frac{\sum_{j=1}^s \tau_i(d_j)}{\sum_{i=1}^k \sum_{j=1}^s \tau_i(d_j)} \quad (7)$$

where $\tau_i(d_j) = \frac{|\mathbb{R}_{\mathcal{A}_i}(d_j)|}{|\overline{\mathbb{R}_{\mathcal{A}_i}}(d_j)|}$ signifies the approximate precision weight of \mathcal{A}_i on d_j , while $\mathbb{R}_{\mathcal{A}_i}(d_j)$ and $\overline{\mathbb{R}_{\mathcal{A}_i}}(d_j)$ respectively denote the lower and upper approximations of the classical Pawlak rough set on d_j . It is stipulated that when $\sum_{i=1}^k \sum_{j=1}^s \tau_i(d_j) = 0$, then $\varpi_i = 0$.

3. Our approach

The proposed feature selection method consists of three main steps, as illustrated in Fig. 2. First, a graph is constructed based on the fuzzy information system, and graph-theoretic principles are applied to determine the granularity levels. Second, the granularities are reordered according to their computed weights. Finally, a certain proportion of features is selected based on the granularity ranking, and further refinement is performed to obtain the final feature subset.

3.1. Graph building and feature prioritizing for FIS

To build the graph, it's crucial to first establish the correlation and non-redundancy indices. Therefore, we initially define the dominance set based on neighborhood constraints to derive the ranking cardinality of attributes.

Definition 1. Given a $FIS = \langle O, B \cup d, G \rangle$, for $\forall o_i, o_j \in O$ and $\mathcal{A} \subseteq B$, the dominance relation based on neighborhood constraints is designated as

$$D_{\mathcal{A}}^{\geq} = \left\{ (o_i, o_j) \in O \times O \mid \frac{1}{|\mathbb{N}_{\kappa}(o_i)|} \sum_{o_i \in \mathbb{N}_{\kappa}(o_i)} g(o_i, b) \geq \frac{1}{|\mathbb{N}_{\kappa}(o_j)|} \sum_{o_j \in \mathbb{N}_{\kappa}(o_j)} g(o_j, b) \right\}, \quad (8)$$

where $\mathbb{N}_{\kappa}(o_i)$ denotes the set of κ nearest neighbors of object o_i (including itself), and $|\mathbb{N}_{\kappa}(o_i)|$ indicates the number of elements in this set. The same interpretation applies to object o_j .

Then, the neighborhood-constrained dominating and dominated sets of $o_i \in O$ on \mathcal{A} are designated as

$$D_{\mathcal{A}}^+(o_i) = \{ o_i \in O \mid o_i D_{\mathcal{A}}^{\geq} o_j \},$$

$$D_{\mathcal{A}}^-(o_i) = \{ o_j \in O \mid o_j D_{\mathcal{A}}^{\geq} o_i \},$$

through the defined neighborhood constraints of dominating and dominated sets, we can ascertain the ranking of sample o_i within feature b_j . Consequently, the Spearman rank correlation coefficient can be calculated as follows:

Table 2
A fuzzy information system.

O	b_1	b_2	b_3	b_4	b_5	b_6	d
o_1	0.26	0.28	0.63	0.29	0.37	0.60	1
o_2	0.41	0.42	0.89	0.28	0.55	0.40	1
o_3	0.23	0.24	0.88	0.42	0.46	0.30	1
o_4	0.18	0.18	0.33	0.46	0.09	0.22	1
o_5	0.45	0.42	0.72	0.30	0.18	0.45	1
o_6	0.45	0.43	0.84	0.43	0.18	0.46	2
o_7	0.27	0.29	0.58	0.58	0.91	0.80	2
o_8	0.33	0.33	0.92	0.80	0.73	0.40	2
o_9	0.14	0.15	0.47	0.59	0.46	0.09	2
o_{10}	0.24	0.32	0.61	0.62	0.33	0.41	2

$$\xi_{i,j} = 1 - \frac{6 \sum_{i=1}^m (|D_{b_i}^-(o_i)| - |D_{b_j}^-(o_i)|)^2}{m(m^2 - 1)}$$

For $FIS = \langle O, B \cup d, G \rangle$, the conditional attribute set $B = \{b_1, b_2, \dots, b_n\}$ is depicted as nodes $V = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ in a graph. The edges that connect these nodes represent relationships between attributes, thereby forming the adjacency matrix M of the graph G . Each element $M(i, j)$ in the matrix is defined using a weight function as follows:

$$M(i, j) = \psi(\vec{v}_i, \vec{v}_j) = \alpha \mathcal{H}_{ij} + (1 - \alpha) \mathcal{Q}_{ij}. \quad (9)$$

\mathcal{H}_{ij} denotes the larger value of the max-min normalized standard deviations of the two objects distributions, i.e., $\mathcal{H}_{ij} = \max(s_i, s_j)$, where s_i refers to the standard deviation of objects with respect to attribute b_i . $\mathcal{Q}_{ij} = 1 - |\xi_{b_i, b_j}|$ is a non-redundant metric that is inversely related to the correlation coefficient.

After constructing the adjacency matrix using the weight function, we employ the Matrix Power Series (MPS) method for path selection to obtain an initial ranking of feature scores. Specifically, the implementation begins as follows. Given the identity matrix \mathcal{E} and the eigenvalues $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ of matrix M , the parameter $\varkappa = 0.9/\rho(M)$ serves as a real-valued regularization factor, where $\rho(M) = \max_{\lambda_i \in \{\lambda_1, \lambda_2, \dots, \lambda_n\}} (|\lambda_i|)$ represents the spectral radius. Consequently, the partial score matrix for the attribute set can be calculated as

$$\tilde{\Phi} = (\mathcal{E} - \varkappa M)^{-1} - \mathcal{E},$$

next, with a unit column vector \mathbf{I} where all components are 1, the final score for each feature can be derived by marginalizing the resulting partial score matrix

$$\tilde{\phi}(i) = [\tilde{\Phi} \mathbf{I}]_i.$$

Next, a straightforward example will be presented to demonstrate the graph-based feature scoring calculation described above.

Example 1. Table 2 illustrates a $FIS = \langle O, B \cup d, G \rangle$, where $O = \{o_1, o_2, \dots, o_{10}\}$ represents the set of objects, $B = \{b_1, b_2, \dots, b_6\}$ denotes the set of conditional attributes, and d is the decision attribute with $O/d = \{d_1, d_2\}$. Assuming the neighborhood constraint radius $\kappa = 3$, then we have

$$\begin{aligned} s(b_1) &= 0.0121, s(b_2) = 0.0097, s(b_3) = 0.0392, s(b_4) = 0.0288, s(b_5) = 0.0656, s(b_6) = 0.0381, \\ \mathcal{H}_{12} &= \mathcal{H}_{21} = 0.0421, \mathcal{H}_{13} = \mathcal{H}_{31} = 0.5264, \mathcal{H}_{14} = \mathcal{H}_{41} = 0.3407, \mathcal{H}_{15} = \mathcal{H}_{51} = 1.0000, \\ \mathcal{H}_{16} &= \mathcal{H}_{61} = 0.5077, \mathcal{H}_{23} = \mathcal{H}_{32} = 0.5264, \mathcal{H}_{24} = \mathcal{H}_{42} = 0.3407, \mathcal{H}_{25} = \mathcal{H}_{52} = 1.0000, \\ \mathcal{H}_{26} &= \mathcal{H}_{62} = 0.5077, \mathcal{H}_{34} = \mathcal{H}_{43} = 0.5264, \mathcal{H}_{35} = \mathcal{H}_{53} = 1.0000, \mathcal{H}_{36} = \mathcal{H}_{63} = 0.5264, \\ \mathcal{H}_{45} &= \mathcal{H}_{54} = 1.0000, \mathcal{H}_{46} = \mathcal{H}_{64} = 0.5077, \mathcal{H}_{56} = \mathcal{H}_{65} = 1.0000. \end{aligned}$$

According to Definition 3, the Spearman correlation matrix ξ is given by

$$\begin{pmatrix} 1.0000 & 0.9273 & 0.6000 & -0.3576 & 0.0303 & 0.6121 \\ 0.9273 & 1.0000 & 0.6485 & -0.2485 & -0.0061 & 0.5152 \\ 0.6000 & 0.6485 & 1.0000 & -0.2364 & 0.3818 & 0.0909 \\ -0.3576 & -0.2485 & -0.2364 & 1.0000 & 0.1515 & -0.2727 \\ 0.0303 & -0.0061 & 0.3818 & 0.1515 & 1.0000 & 0.1030 \\ 0.6121 & 0.5152 & 0.0909 & -0.2727 & 0.1030 & 1.0000 \end{pmatrix}.$$

$$\text{When } \alpha = 0.7, \text{ the adjacency matrix } M \text{ is } \begin{pmatrix} 0.0295 & 0.0513 & 0.4885 & 0.4312 & 0.9909 & 0.4717 \\ 0.0513 & 0.0000 & 0.4739 & 0.4639 & 0.9982 & 0.5008 \\ 0.4885 & 0.4739 & 0.3685 & 0.5976 & 0.8855 & 0.6412 \\ 0.4312 & 0.4639 & 0.5976 & 0.2385 & 0.9545 & 0.5736 \\ 0.9909 & 0.9982 & 0.8855 & 0.9545 & 0.7000 & 0.9691 \\ 0.4717 & 0.5008 & 0.6412 & 0.5736 & 0.9691 & 0.3554 \end{pmatrix}.$$

Subsequently, we can calculate the eigenvalues of M and the final feature scores as follows

$$\{\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6\} = \{3.6730, -1.0742, -0.0330, -0.3670, -0.2332, -0.2738\};$$

$$\alpha = 0.9/\rho(M) = 0.24502961;$$

$$\check{\Phi} = (\mathcal{E} - \alpha M)^{-1} - \mathcal{E} = \begin{pmatrix} 0.8151 & 0.8301 & 1.1194 & 1.0704 & 1.6994 & 1.1409 \\ 0.8301 & 0.8276 & 1.1277 & 1.0890 & 1.7179 & 1.1594 \\ 1.1194 & 1.1277 & 1.3547 & 1.3599 & 2.0766 & 1.4481 \\ 1.0704 & 1.0890 & 1.3599 & 1.2333 & 2.0188 & 1.3849 \\ 1.6994 & 1.7179 & 2.0766 & 2.0188 & 2.9390 & 2.1388 \\ 1.1409 & 1.1594 & 1.4481 & 1.3849 & 2.1388 & 1.4144 \end{pmatrix}.$$

$$\check{\Phi}(i) = [\check{\Phi}I]_i = (6.6753, 6.7517, 8.4864, 8.1563, 12.5904, 8.6864).$$

Consequently, the resulting final feature score sequence is $b_5, b_6, b_3, b_4, b_2, b_1$.

3.2. Weighted multi-granulation neighborhood-constrained fuzzy rough sets based on graph partition

In fuzzy information systems, feature scores are first obtained through graph-based construction. These scores are then sorted in descending order to generate a score sequence. Subsequently, the granularity space is partitioned. Unlike previous methods that adopt a fixed granularity scheme, the proposed approach introduces a tunable parameter t to flexibly control the granularity level. Based on this flexible partitioning, a weighted multi-granularity rough set model is further constructed to enhance the feature evaluation process.

Definition 2. Let $FIS = \langle O, B \cup d, G \rangle$ denote a fuzzy information system, $\mathcal{A}_i \subseteq B$, $i = 1, 2, \dots, k (k \leq 2^{|B|})$, where ϖ_i represents the approximation precision weight under \mathcal{A}_i , the weighted optimistic multi-granulation neighborhood-constrained fuzzy lower and upper approximations are designated as

$$\begin{aligned} \underline{\mathcal{OMF}}_{\sum_{i=1}^k \mathcal{A}_i}^{\omega}(\mathcal{F})(x) &= \bigvee_{i=1}^k \varpi_i \underline{\mathcal{R}}_{\mathcal{A}_i}^K(\mathcal{F})(x), \\ \overline{\mathcal{OMF}}_{\sum_{i=1}^k \mathcal{A}_i}^{\omega}(\mathcal{F})(x) &= \bigwedge_{i=1}^k \varpi_i \overline{\mathcal{R}}_{\mathcal{A}_i}^K(\mathcal{F})(x), \end{aligned} \quad (10)$$

where “ \bigvee ” and “ \bigwedge ” signify the disjunction and conjunction operators, respectively, while $\underline{\mathcal{R}}_{\mathcal{A}_i}^K(\mathcal{F})(x)$ and $\overline{\mathcal{R}}_{\mathcal{A}_i}^K(\mathcal{F})(x)$ denote the neighborhood constraint fuzzy lower and upper approximations on \mathcal{A}_i . Therefore, the pair $\left(\underline{\mathcal{OMF}}_{\sum_{i=1}^k \mathcal{A}_i}^{\omega}(\mathcal{F})(x), \overline{\mathcal{OMF}}_{\sum_{i=1}^k \mathcal{A}_i}^{\omega}(\mathcal{F})(x) \right)$ is referred to as the weighted optimistic multi-granular neighborhood-constrained fuzzy rough set (WOMNcFRS).

Property 1. Given a $FIS = \langle O, B \cup d, G \rangle$, $\mathcal{A}_i \subseteq B$, $i = 1, 2, \dots, k \leq 2^{|B|}$. For any $\mathcal{U}, \mathcal{V} \in F(O)$, then

1. $\mathcal{U} \subseteq \mathcal{V} \Rightarrow \underline{\mathcal{OMF}}_{\sum_{i=1}^k \mathcal{A}_i}^{\omega}(\mathcal{U}) \subseteq \underline{\mathcal{OMF}}_{\sum_{i=1}^k \mathcal{A}_i}^{\omega}(\mathcal{V});$
2. $\mathcal{U} \subseteq \mathcal{V} \Rightarrow \overline{\mathcal{OMF}}_{\sum_{i=1}^k \mathcal{A}_i}^{\omega}(\mathcal{U}) \subseteq \overline{\mathcal{OMF}}_{\sum_{i=1}^k \mathcal{A}_i}^{\omega}(\mathcal{V}).$

Proof. 1. Let $\mathcal{U}, \mathcal{V} \in F(O)$ such that $\mathcal{U} \subseteq \mathcal{V}$ and $x \in O$. Then $\underline{\mathcal{OMF}}_{\sum_{i=1}^k \mathcal{A}_i}^{\omega}(\mathcal{U})(x) = \bigvee_{i=1}^k \varpi_i \underline{\mathcal{R}}_{\mathcal{A}_i}^K(\mathcal{U})(x) = \bigvee_{i=1}^k \bigwedge_{y \in O} \varpi_i \max(1 - \mathcal{R}_{\mathcal{A}_i}^K(x, y), \mathcal{U}(y)) \leq \bigvee_{i=1}^k \bigwedge_{y \in O} \varpi_i \max(1 - \mathcal{R}_{\mathcal{A}_i}^K(x, y), \mathcal{V}(y)) = \underline{\mathcal{OMF}}_{\sum_{i=1}^k \mathcal{A}_i}^{\omega}(\mathcal{V})(x).$

2. The proof is analogous to that of 1).

Suppose $O/d = \{d_1, d_2, \dots, d_s\}$, for $\mathcal{A}_i \subseteq B$, $i = 1, 2, \dots, k (k \leq 2^{|B|})$, then the weighted optimistic positive region and fuzzy dependency degree for d concerning \mathcal{A}_i are designated as

$$\begin{aligned} \mathcal{POS}_{\mathcal{A}}^{\omega, K}(d) &= \bigcup_{i=1}^s \underline{\mathcal{OMF}}_{\sum_{i=1}^k \mathcal{A}_i}^{\omega}(d_j), \\ \Gamma_{\mathcal{A}}^{\omega, K}(d) &= \frac{|\mathcal{POS}_{\mathcal{A}}^{\omega, K}(d)|}{|O|}. \end{aligned} \quad (11)$$

Definition 3. Let $FIS = \langle O, B \cup d, G \rangle$ denote a fuzzy information system, $\mathcal{A}_i \subseteq B$, $i = 1, 2, \dots, k (k \leq 2^{|B|})$, where ϖ_i represents the approximation precision weight under \mathcal{A}_i , the weighted pessimistic multi-granulation neighborhood-constrained fuzzy lower and upper approximations are designated as

$$\begin{aligned}\overline{\mathcal{PMF}^\omega_{\sum_{i=1}^k \mathcal{A}_i}}(\mathcal{F})(x) &= \bigwedge_{i=1}^k \overline{\varpi_i \mathcal{R}^\kappa_{\mathcal{A}_i}}(\mathcal{F})(x), \\ \overline{\mathcal{PMF}^\omega_{\sum_{i=1}^k \mathcal{A}_i}}(\mathcal{F})(x) &= \bigvee_{i=1}^k \overline{\varpi_i \mathcal{R}^\kappa_{\mathcal{A}_i}}(\mathcal{F})(x),\end{aligned}\quad (12)$$

the pair $\left(\overline{\mathcal{PMF}^\omega_{\sum_{i=1}^k \mathcal{A}_i}}(\mathcal{F})(x), \overline{\mathcal{PMF}^\omega_{\sum_{i=1}^k \mathcal{A}_i}}(\mathcal{F})(x)\right)$ is referred to as the weighted pessimistic multi-granulation neighborhood-constrained fuzzy rough set (WPMNcFRS).

Suppose $O/d = \{d_1, d_2, \dots, d_s\}$, for $\mathcal{A}_i \subseteq B$, $i = 1, 2, \dots, k (k \leq 2^{|B|})$, then the weighted pessimistic positive region and fuzzy dependency degree for d concerning \mathcal{A}_i are designated as

$$\begin{aligned}\mathcal{POS}_{\mathcal{A}}^{\mathcal{P}, \kappa}(d) &= \bigcup_{i=1}^s \overline{\mathcal{PMF}^\omega_{\sum_{i=1}^k \mathcal{A}_i}}(d_i), \\ \Gamma_{\mathcal{A}}^{\mathcal{P}, \kappa}(d) &= \frac{|\mathcal{POS}_{\mathcal{A}}^{\mathcal{P}, \kappa}(d)|}{|O|}.\end{aligned}\quad (13)$$

Property 2. Let $FIS = \langle O, B \cup d, G \rangle$ denote a fuzzy information system, $\mathcal{A}_1 \subseteq \mathcal{A}_2 \subseteq B$, $O/d = \{d_1, d_2, \dots, d_s\}$. Therefore, the following property holds:

1. $\Gamma_{\mathcal{A}_1}^{\mathcal{O}, \kappa}(d) \leq \Gamma_{\mathcal{A}_2}^{\mathcal{O}, \kappa}(d) \leq \Gamma_B^{\mathcal{O}, \kappa}(d)$;
2. $\Gamma_{\mathcal{A}_1}^{\mathcal{P}, \kappa}(d) \leq \Gamma_{\mathcal{A}_2}^{\mathcal{P}, \kappa}(d) \leq \Gamma_B^{\mathcal{P}, \kappa}(d)$.

Proof. 1. Following the extension of Eq. (2), for any $x, y \in O$, $\mathcal{R}_{\mathcal{A}_1}^\kappa(x, y) = \mathcal{R}_{\mathcal{A}_1}^\kappa(x, y) \wedge \mathcal{R}_{\mathcal{A}_2 - \mathcal{A}_1}^\kappa(x, y) \leq \mathcal{R}_{\mathcal{A}_1}^\kappa(x, y)$. Consequently,

$\mathcal{R}_{\mathcal{A}_1}^\kappa(d_i) \subseteq \mathcal{R}_{\mathcal{A}_2}^\kappa(d_i)$, which implies $\mathcal{POS}_{\mathcal{A}_1}^{\mathcal{O}, \kappa}(d) \subseteq \mathcal{POS}_{\mathcal{A}_2}^{\mathcal{O}, \kappa}(d)$ and $\Gamma_{\mathcal{A}_1}^{\mathcal{O}, \kappa}(d) \leq \Gamma_{\mathcal{A}_2}^{\mathcal{O}, \kappa}(d)$. Likewise, it can be demonstrated that $\Gamma_{\mathcal{A}_2}^{\mathcal{O}, \kappa}(d) \leq \Gamma_B^{\mathcal{O}, \kappa}(d)$.

2. Similarly, it can employ the approach described in (1) for proof.

Example 2. Continuing with Example 1, we have obtained the feature score ranking as $b_5, b_6, b_3, b_4, b_2, b_1$. With the granularity parameter $t = 2$, three granularities are partitioned, denoted as $\mathcal{A}_1 = \{b_5, b_6\}$, $\mathcal{A}_2 = \{b_3, b_4\}$ and $\mathcal{A}_3 = \{b_2, b_1\}$.

The subsequent calculations are based on WOMNcFRS. Firstly, using Eq. (7), we can compute the approximation precision weight for each granularity as $\varpi_1 = 0.124$, $\varpi_2 = 0.210$, $\varpi_3 = 0.039$;

As observed, the ranking sequence based on graph theory can shift after partitioning granularity and reordering according to granularity weights. This indicates that features with higher rankings might also include some redundant attributes, causing a decrease in their respective granularity weights. As a result, the new granularity partitions are $\mathcal{A}_1 = \{b_3, b_4\}$, $\mathcal{A}_2 = \{b_5, b_6\}$ and $\mathcal{A}_3 = \{b_2, b_1\}$.

Next, using the updated granularities, we can calculate the lower approximations of each granularity concerning the decision as follows

$$\begin{aligned}\overline{\mathcal{PMF}^\omega_{\sum_{i=1}^k \mathcal{A}_1}}(d_1) &= 1.073, \overline{\mathcal{PMF}^\omega_{\sum_{i=1}^k \mathcal{A}_1}}(d_2) = 1.027, \overline{\mathcal{PMF}^\omega_{\sum_{i=1}^k \mathcal{A}_2}}(d_1) = 0.493, \overline{\mathcal{PMF}^\omega_{\sum_{i=1}^k \mathcal{A}_2}}(d_2) = 0.747, \\ \overline{\mathcal{PMF}^\omega_{\sum_{i=1}^k \mathcal{A}_3}}(d_1) &= 0.107, \overline{\mathcal{PMF}^\omega_{\sum_{i=1}^k \mathcal{A}_3}}(d_2) = 0.283.\end{aligned}$$

3.3. Significance based on WOMNcFRS

To make effective decisions across various dimensions, [28] adopted a conservative approach. Building on this, we will delve deeper into the WOMNcFRS model, focusing on extracting rules based on weighted optimistic multi-granulation fuzzy dependency degrees (WOMFD), as well as inner and outer importance.

Definition 4. Given a $FIS = \langle O, B \cup d, G \rangle$, for $\mathcal{A} \subseteq B$ and $b \in A$, if $\Gamma_{\mathcal{A}-b}^{\mathcal{O}, \kappa}(d) = \Gamma_{\mathcal{A}}^{\mathcal{O}, \kappa}(d)$, then b is called a redundant attribute of A with respect to d , conversely, b is considered a necessary attribute of A concerning d . With respect to d , we can thus obtain the reduction of B satisfying:

1. $\Gamma_{\mathcal{A}}^{\mathcal{O}, \kappa}(d) = \Gamma_B^{\mathcal{O}, \kappa}(d)$
2. $\forall b \in A, \Gamma_{\mathcal{A}-b}^{\mathcal{O}, \kappa}(d) < \Gamma_{\mathcal{A}}^{\mathcal{O}, \kappa}(d)$

Definition 4 guarantees that in the process of reducing a feature set, the resulting subset maintains equivalent discriminatory capabilities to the original set while encompassing all essential attributes.

Next, the internal and external significance of attribute $A \subseteq B$ concerning d is designated as

$$\mathcal{S}_{in}(b, B, d) = \Gamma_{B-b}^{\mathcal{O}, \kappa}(d) - \Gamma_B^{\mathcal{O}, \kappa}(d),$$

$$\mathcal{S}_{out}(b, A, d) = \Gamma_A^{\mathcal{O}, \kappa}(d) - \Gamma_{A \cup b}^{\mathcal{O}, \kappa}(d).$$

When the value of $\mathcal{S}_{in}(b, B, d)$ is higher, it demonstrates that attribute b is more crucial relative to B . On the other hand, a higher $\mathcal{S}_{out}(b, A, d)$ value signifies the importance of attribute b to A .

The above significance measures are used as evaluation criteria to guide the feature selection process. In the implementation, both internal and external significance are considered, and the feature with the highest combined significance is selected and added to the current feature subset. This process continues iteratively until no further feature yields a significant improvement, resulting in the final optimal feature subset. This approach effectively ensures that the selected features maintain strong dependency with the decision attributes while minimizing redundancy.

4. WOMFD-based feature selection algorithm

In this section, we provide a comprehensive description of the feature selection algorithm based on WOMFD. To enhance feature selection efficiency, we employ a heuristic search strategy. Initially, we select the attribute b_0 that satisfies $\arg\max_{b \in B} \mathcal{S}_{in}(b, B, d)$ as the initial set, and then proceed to select other important features based on the highest \mathcal{S}_{out} value principle, where θ is set to 0.0. The detailed selection process is presented in Algorithm 1. Furthermore, the complexity of the algorithm is examined.

Firstly, regarding the construction of the *FIS* graph and the calculation of feature scores, the time complexity for steps 1 through 10 is $O(n^2 \cdot m)$, where m denotes the number of objects and n represents the number of features. Step 11, which computes the eigenvalues and the real-valued factors of the adjacency matrix, has a time complexity of $O(n^2)$. Then, step 12 involves calculating the partial score matrix, which requires the computation of the inverse matrix, resulting in a time complexity of $O(n^3)$. Lastly, step 13, where the final scores are computed, has a time complexity of $O(n^2)$.

In terms of the further heuristic feature selection process, step 15, encompassing sorting and granularity partitioning, carries a time complexity of $O(n \cdot \log n)$. Step 16, calculating weights, exhibits a time complexity of $O(n)$. Additionally, both step 17 and step 18 possess a time complexity of $O(n \cdot m)$. The complexity of step 19, which involves selecting the initial set b_k , is $O(n)$. The while loop, when executed in its worst-case scenario, iterates n times, thereby contributing to the overall time complexity of steps 21 through 32 amounts to $O(n^2 \cdot m)$. As a result, the overall complexity of the algorithm is $O(n^3 + n^2 \cdot m)$. From this, it can be seen that even with a large data sample, our algorithm significantly improves the reduction efficiency and greatly reduces the time consumption.

5. Experiment

In this section, we assess the efficiency and effectiveness of the feature selection algorithm based on *WOMFD* proposed in this paper through validation on fifteen datasets sourced from the UCI database and Kaggle. Table 4 illustrates the detailed information of the selected datasets. All experiments were performed using the hardware and software configuration outlined in Table 3. We compared the classification accuracy of our two algorithms with six other algorithms, including Hybrid Kernel-based Fuzzy Complementary Entropy Algorithm (HKCMI) [29], Granular-Rectangular Neighborhood Feature Selection Method (GRRS) [30], Matrix-based incremental feature selection method (W-MGMN) [31], Graph-based Local Search Feature Selection (GLSFS) [32], Back Propagation Neural Network (BPNN), and Graph Theory-based Interval Value Feature Selection Method (IGUFS) [27], on four classifiers, namely K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Naive Bayes (NB), and Random Forest (RF), in terms of their classification accuracy.

5.1. Experimental design

At first, the reduction times of the two proposed algorithms relative to a faster algorithm, among others are presented, along with a comparison of the optimal subset sizes achieved by these algorithms across various datasets. Secondly, the classification accuracy of our two algorithms was compared with six contrasting algorithms and the original dataset. Finally, two hypothesis tests were devised to further substantiate the distinctions between WOMFD and the comparative algorithms. Three parameters, α , κ , and t , are considered in the design of WOMFD. Parameter α is used to govern a balance between correlation and non-redundancy, while parameter κ determines the radius of neighborhood constraints, and the level of granularity is regulated by parameter t . According to the definition, changes in different parameters will inevitably have varying effects on classification accuracy. Therefore, for parameter α , we define its range from 0 to 1 with a step size of 0.1. Similarly, parameter κ spans from 1 to 9 with increments of 2. Regarding parameter t , depending on the number of features in the dataset, for datasets with more features, it can be set to 3, 6, and 9, and for datasets with fewer features, it can be set to 1, 2, and 3. Furthermore, for more flexible feature extraction, predefined granularity ratios of 65% and 85% are designated for the WOMFD algorithm selection. Moreover, the reason for not choosing a larger proportion is that our method involves directly selecting and discarding granularities. Choosing a larger proportion of granularities may result in too many redundant features, greatly increasing the time cost for subsequent feature selection. Similarly, not choosing a smaller proportion is due to the uncertainty of the features contained within each granularity. With fewer features, a smaller proportion might result in the loss of essential features. Ultimately, the optimal parameters are determined based on ten-fold cross-validation performed on the training dataset.

Algorithm 1 Feature Selection Algorithm Based on WOMFD.

Input: $FIS = \langle O, B \cup d, G \rangle$, α, κ, t , where $O = \{o_1, o_2, \dots, o_m\}$, $B = \{b_1, b_2, \dots, b_n\}$, $O/d = \{d_1, d_2, \dots, d_s\}$, $\alpha \in [0, 1]$, $\kappa > 0$.

Output: The optimal reduced set C .

```

1: for  $i = 1 \rightarrow n$  do
2:   for  $j = 1 \rightarrow n$  do
3:     for  $k = 1 \rightarrow m$  do
4:       Compute  $D_{b_i}^-(o_k)$  and  $D_{b_j}^-(o_k)$ ;
5:     end for
6:     Compute  $H_{ij} \leftarrow \max(s(b_i), s(b_j))$ ;
7:     Compute  $Q_{ij} \leftarrow |1 - \xi(b_i, b_j)|$ ;
8:      $M(i, j) \leftarrow \alpha H_{ij} + (1 - \alpha) Q_{ij}$ ;
9:   end for
10: end for
11:  $\kappa = 0.9 / \rho(M)$ .
12:  $\tilde{\Phi} = (\mathcal{E} - \kappa M)^{-1} - \mathcal{E}$ .
13:  $\hat{\phi}(i) = [\tilde{\Phi} \mathbf{I}]_i$ .
14: Initialize  $C \leftarrow \emptyset$ .
15: Partitioning granularity  $\mathcal{A}_i$  based on  $t$  after sorting.
16: Compute the approximation precision weight  $\varpi_i$  for each granularity  $\mathcal{A}_i$ .
17: Compute the  $\Gamma_B^{\mathcal{O}, \kappa}(d)$  of feature set  $B$ .
18: Compute the  $\Gamma_{B-b}^{\mathcal{O}, \kappa}(d)$  for  $b \in B$ .
19: Choose attribute  $b_0 = \underset{b \in B}{\operatorname{argmax}} \mathcal{S}_{in}(b, B, d)$ .
20:  $C \leftarrow b_0$ .
21: while  $tag = 1$  do
22:   for each  $b \in B - C$  do
23:     Compute  $\Gamma_{C \cup b}^{\mathcal{O}, \kappa}(d)$  and  $\mathcal{S}_{out}(b, B, d)$ ;
24:   end for
25:   Choose  $b_p = \underset{b \in B - C}{\operatorname{argmax}} \mathcal{S}_{out}(b, B, d)$ ;
26:   if  $\mathcal{S}_{out}(b_p, B, d) \geq \delta$  then
27:      $C \leftarrow C \cup \{b_p\}$ ;
28:      $tag = 1$ ;
29:   else
30:      $tag = 0$ ;
31:   end if
32: end while
33: return  $C$ .
```

Table 3

Configuration of experimental environment.

Name	Model	Parameter
System	Window 10	64 bit
CPU	Intel(R) Core(TM) i5-8250U	1.60 GHz
Platform	Python	3.10
Memory	DDR4	8.00 GB

Table 4

The exhibition of datasets details.

No.	Data sets	Abbreviation	Instances	Features	Class
1	Immunotherapy	Immu	90	8	2
2	Concrete Compressive Strength	CCS	1030	9	14
3	Abalone	Aba	4177	9	3
4	Breast cancer	Breast	699	10	2
5	Shill Bidding	Bid	6321	11	2
6	Dry Bean	DB	13611	16	7
7	Cardiotocography	Card	2126	20	3
8	Wisconsin diagnostic breast cancer	Wdbc	569	31	2
9	Ionosphere	Iono	351	34	2
10	Ultrasonic flowmeter diagnostics	UFD	181	43	4
11	Turkish Music Emotion	TME	400	50	4
12	Connectionist Bench	Sonar	208	60	2
13	Gait Classification	Gait	48	321	16
14	DARWIN	DARW	174	451	2
15	Period Changer	PC	90	1177	2

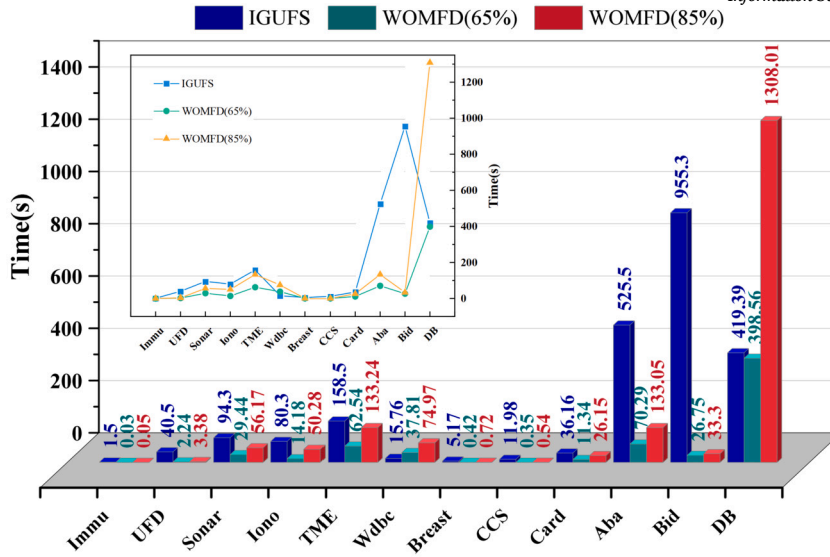


Fig. 3. The comparison of reduction time for WOMFD and IGUFS.

Table 5

The count of selected features by various algorithms.

Datasets	RAW	HKCMI	GRRS	W-MGMN	GLSFS	BPNN	IGUFS	WOMFD(65%)	WOMFD(85%)
Immu	7	2.6	3	5	4	3	3	3	4
UFD	43	2	2	3	27	19	19	6	10
Sonar	60	1	2	13	39	27	21	40	49
Iono	34	4	5	12	22	15	14	17	27
TME	50	1.4	2	17	32	23	22	24	39
Wdbc	31	2	-	16	20	20	13	20	26
Breast	10	3	-	8	6	5	4	3	5
CCS	8	2	7	5	5	4	3	2	3
Card	20	2	6	9	13	9	9	5	5
Average	29.2	2.2	1	23.2	12	13.9	12	13.3	18.7
Aba	8	2	3	8	5	4	3	3	6
Bid	11	2	-	5	7	4	4	3	4
Gait	321	1	1	16	208	232	144	79	110
DARW	451	-	-	18	293	380	202	290	383
PC	1177	-	1	25	765	792	529	764	998
DB	16	-	2	7	10	13	7	8	13

5.2. Experimental analysis

If an algorithm takes more than 24 hours to process a dataset, the experiment will be terminated due to excessive time consumption. If a comparison algorithm fails to handle the dataset, the experiment will also be stopped. The time consumption of our algorithm WOMFD across 12 low-dimensional datasets is depicted in Fig. 3. The reduction time of the algorithm is taken as the maximum runtime among all parameters in this paper. For the dataset DB with the largest sample size of 13611 and 16 attributes, WOMFD incurs a time consumption of 1308.01 seconds at a granularity ratio of 85%, equivalent to approximately 22 minutes. And for the high-dimensional dataset, PC with 1177 features, it has an approximate simplification time of 25106.65 seconds, which is about seven hours. Nevertheless, there are still three comparison algorithms on this dataset that exceed a runtime of 24 hours. Except for the longer time consumption on datasets WDBC and DB compared to the IGUFS algorithm, our algorithm performs two to four times faster than IGUFS in all other cases. The extended time on these two datasets is due to significant time spent in preprocessing and clustering during the initial stages, thereby increasing the overall time cost.

Table 5 presents the number of features selected by our proposed algorithm and the comparison algorithms across 15 datasets, illustrating the varying effectiveness of feature reduction. Compared with other methods, IGUFS and GLSFS offer greater flexibility in controlling the proportion of selected features, enabling customization based on specific needs. However, our algorithm, WOMFD, introduces a novel perspective by incorporating multigranulation, allowing flexible granularity proportions to be selected according to practical requirements. This makes WOMFD more adaptable to real-world applications. For instance, in datasets with a large number of attributes (e.g., Sonar and TME), reducing the selection ratio proportionally can be beneficial. Conversely, datasets with fewer attributes may benefit from increasing the selection ratio appropriately.

Table 6

The classification of various algorithms based on KNN (%).

Datasets	RAW	HKCMI	GRRS	W-MGMN	GLSFS	BPNN	IGUFS	WOMFD(65%)	WOMFD(85%)
Immu	76.67±11.6	77.78±7.03	78.75±1.57	70.97±1.48	76.67±1.59	72.22±7.03	75.56±10.30	86.67±1.19	84.44±2.02
UFD	78.45±9.44	46.67±13.43	72.22±0.86	91.67±0.26	79.04±0.46	71.67±5.67	90.00±4.16	81.17±0.52	82.37±0.27
Sonar	84.19±7.65	56.86±17.33	63.29±0.64	77.83±0.31	82.31±0.81	80.77±4.28	83.18±4.25	88.02±0.33	87.57±0.41
Iono	85.17±4.60	78.03±8.20	84.57±0.25	86.29±0.36	86.31±0.24	87.74±4.12	91.16±2.93	87.74±0.44	85.44±0.43
TME	63.75±8.08	30.25±3.39	37.11±0.69	55.41±0.83	58.00±0.41	46.75±2.81	62.50±2.96	69.50±0.61	70.00±0.91
Wdbc	97.01±1.58	90.70±3.67	-	93.50±0.14	96.14±0.08	94.73±2.48	95.08±3.02	96.13±0.03	97.37±0.03
Breast	96.64±2.26	94.58±2.18	-	62.44±0.35	95.32±0.05	95.47±2.10	94.74±2.91	96.19±0.02	96.78±0.03
CCS	35.73±3.06	41.26±5.48	24.10±0.18	41.79±0.27	38.25±0.12	36.50±3.79	23.59±4.12	37.96±0.20	45.15±0.16
Card	80.99±2.51	48.82±1.95	84.75±0.04	73.88±0.02	75.25±0.10	86.31±2.97	80.71±2.03	87.58±0.05	85.56±0.06
Aba	52.98±2.23	46.23±1.28	47.70±0.11	52.95±0.07	52.81±0.04	49.37±1.11	50.42±1.72	50.47±0.05	51.74±0.05
Bid	98.48±0.42	89.32±0.93	-	97.85±0.00	87.38±0.01	89.86±0.73	98.83±0.21	98.01±0.00	98.88±0.00
Gait	43.50±4.20	6.50±1.80	13.00±1.96	29.50±2.82	43.50±5.00	41.50±4.70	40.00±3.80	74.50±5.92	70.50±6.92
DARW	72.32±9.21	-	-	68.07±1.49	71.73±1.28	71.21±8.54	71.24±6.32	75.23±0.84	74.05±0.87
PC	63.33±12.22	-	59.58±0.75	52.78±2.18	63.33±1.49	58.89±1.99	63.33±0.75	67.78±1.10	67.78±1.59
DB	91.56±0.67	-	57.26±0.01	69.60±0.01	76.22±0.01	91.62±0.56	88.26±0.81	92.15±0.00	91.97±0.00
Average	77.62±5.64	62.77±6.96	56.58±0.64	68.30±0.71	72.15±0.78	74.68±3.92	77.39±4.08	81.22±0.38	81.63±0.48

Table 7

The classification of various algorithms based on SVM (%).

Datasets	RAW	HKCMI	GRRS	W-MGMN	GLSFS	BPNN	IGUFS	WOMFD(65%)	WOMFD(85%)
Immu	77.78±16.48	78.89±8.89	78.75±1.57	78.75±1.57	78.89±2.58	78.89±8.89	80.00±9.03	84.44±1.04	81.11±2.48
UFD	47.54±11.53	58.89±6.43	40.00±0.42	41.11±0.44	46.43±1.63	21.22±4.51	65.56±4.84	53.65±1.72	51.43±1.65
Sonar	84.19±5.25	71.21±8.37	66.31±1.28	78.33±0.59	78.90±0.46	82.22±2.83	81.78±6.09	84.62±0.46	86.10±0.76
Iono	93.71±3.79	85.74±7.74	87.43±0.26	93.43±0.07	93.44±0.13	91.16±2.93	93.44±3.34	94.87±0.08	94.86±0.13
TME	78.00±3.67	30.00±6.52	39.63±0.48	54.37±0.36	73.75±0.27	52.00±2.57	70.75±1.27	79.00±0.22	81.75±0.44
Wdbc	97.89±2.05	91.92±2.62	-	91.55±0.05	96.84±0.04	95.44±2.38	97.19±1.40	96.48±0.03	97.54±0.03
Breast	96.19±1.88	93.85±1.12	-	64.95±0.21	95.76±0.04	95.77±2.93	94.88±2.44	96.35±0.09	97.22±0.02
CCS	41.26±3.29	44.47±4.61	41.21±0.31	41.21±0.31	41.26±0.11	41.26±3.29	41.26±3.29	45.24±0.16	46.31±0.19
Card	86.36±2.29	53.43±2.30	80.24±0.06	59.90±0.08	74.36±0.10	89.04±1.25	81.84±2.28	90.26±0.04	89.04±0.03
Aba	55.18±2.25	54.32±2.15	51.29±0.01	53.57±0.03	55.42±0.07	51.33±1.50	54.66±2.12	54.68±0.06	55.52±0.07
Bid	98.05±0.55	89.32±0.93	-	98.20±0.00	89.32±0.02	90.21±0.33	98.77±0.43	98.20±0.00	98.91±0.00
Gait	24.50±3.02	0.00±0.00	0.00±0.00	0.00±0.00	20.50±1.62	18.50±1.20	10.50±1.12	54.50±4.42	48.00±3.16
DARW	93.73±5.91	-	-	71.60±0.85	87.94±0.57	93.17±6.08	96.05±5.02	97.12±0.08	94.25±0.34
PC	71.11±13.33	-	69.58±1.54	69.58±1.54	71.11±1.78	71.11±9.16	71.11±1.78	71.11±1.78	71.11±1.78
DB	91.39±0.57	-	63.49±0.01	70.02±0.02	63.27±0.01	91.43±0.55	89.55±1.05	92.32±0.00	92.47±0.00
Average	78.10±5.58	67.60±5.40	56.18±0.54	64.44±0.41	69.95±0.63	71.89±3.51	78.52±3.78	80.55±0.43	80.60±0.64

The classification accuracy of the original data and nine algorithms—including our two proposed methods—under KNN, SVM, NB, and RF classifiers are shown in Tables 6, 7, 8, and 9, respectively. In each table, the algorithm achieving the highest accuracy for each dataset is highlighted in bold. Analysis of these tables shows that our proposed method consistently delivers superior performance in most cases. Under the KNN classifier, WOMFD(85%) fails to achieve the highest accuracy only on the UFD, Iono, Card, and Aba datasets. Although WOMFD(65%) outperforms other algorithms on only 8 datasets, its average classification accuracy still exceeds the best average accuracy of other methods by 3.82%. For SVM, both variants of WOMFD continue to outperform competitors on most datasets, with WOMFD(85%) maintaining the highest accuracy across three large-scale datasets. In the case of the NB classifier, WOMFD(85%) achieves top accuracy on 5 datasets, while WOMFD(65%) leads in 10 cases. Despite this, WOMFD still records the highest overall average accuracy under NB. For RF, WOMFD(65%) yields slightly lower accuracy on only 5 datasets, while WOMFD(85%) underperforms on just 3, achieving an impressive average accuracy of 87.19%. It is worth noting that under all four classifiers, our proposed WOMFD algorithm consistently achieves the highest classification accuracy on large-scale datasets, demonstrating its outstanding capability in handling complex and high-dimensional data.

Across all four classifiers and datasets, WOMFD consistently achieves the highest classification performance in the majority of cases. In comparison, IGUFS attains the best results on 8 datasets, while other algorithms lead in only 2 or 3 cases. The results of this comprehensive analysis indicate that WOMFD significantly enhances classification performance while accurately selecting the most informative features, leading to more robust and efficient feature selection outcomes.

5.3. Statistical test

To further verify the effectiveness of the proposed algorithms, this study adopts the Friedman test and the Wilcoxon signed-rank test to statistically evaluate the classification performance of nine algorithms across four types of classifiers, namely KNN, SVM, NB, and RF. Since some experiments on specific datasets required more than 24 hours to complete or failed to generate valid results, all subsequent statistical tests were conducted only on datasets where all algorithms successfully produced outputs.

Table 8

The classification of various algorithms based on NB (%).

Datasets	RAW	HKCMI	GRRS	W-MGMN	GLSFS	BPNN	IGUFS	WOMFD(65%)	WOMFD(85%)
Immu	80.00±9.69	78.89±8.39	84.31±2.01	76.53±2.30	77.78±2.22	70.00±8.31	81.11±9.67	84.44±1.53	81.11±1.49
UFD	63.6±10.59	70.56±4.16	62.22±0.60	74.44±0.69	61.93±1.58	76.67±9.40	87.22±6.71	73.54±1.18	77.98±1.36
Sonar	67.29±10.20	64.52±8.56	69.07±1.66	66.76±1.34	68.79±0.76	73.61±5.63	71.63±4.67	70.74±0.85	68.74±1.31
Iono	88.02±6.11	74.10±7.80	70.86±0.23	77.14±0.62	88.60±0.28	88.60±4.70	87.46±3.43	90.31±0.15	87.76±0.16
TME	75.75±5.13	33.50±3.42	40.60±0.67	61.12±0.67	65.25±0.46	66.75±2.81	78.25±3.76	75.00±0.37	76.00±0.21
Wdbc	93.14±3.38	90.34±3.61	-	92.07±0.16	92.97±0.14	92.79±3.38	94.37±3.32	93.67±0.12	93.32±0.14
Breast	95.91±1.81	93.70±2.80	-	87.40±0.17	94.15±0.03	95.32±2.23	94.73±1.75	96.49±0.04	96.64±0.04
CCS	31.75±4.66	43.01±5.44	25.36±0.34	36.83±0.23	29.22±0.26	42.14±5.36	21.26±4.75	43.88±0.18	38.93±0.08
Card	53.25±2.32	53.95±1.81	65.88±0.12	58.21±0.09	52.92±0.07	62.75±2.50	53.15±2.12	73.75±0.06	75.31±0.02
Aba	52.05±1.86	51.86±1.61	52.11±0.04	51.56±0.03	52.43±0.05	49.37±1.11	51.33±1.39	53.39±0.04	53.36±0.08
Bid	97.07±0.51	88.55±0.65	-	68.71±0.01	97.39±0.00	89.86±0.73	97.18±0.71	98.04±0.00	98.04±0.00
Gait	0.00±0.00	6.50±1.00	14.50±3.32	16.5±15.17	0.00±0.00	0.00±0.00	2.00±0.36	16.00±4.44	17.00±1.56
DARW	88.04±6.20	-	-	73.37±0.30	85.13±0.51	86.86±8.31	89.74±7.37	89.74±0.55	88.59±0.36
PC	36.67±13.19	-	56.67±14.44	39.17±3.68	36.67±1.74	56.67±12.86	34.44±1.84	57.78±4.40	44.44±3.21
DB	89.71±0.65	-	63.67±0.01	68.71±0.01	80.89±0.01	89.66±0.73	86.57±0.85	90.69±0.00	90.11±0.00
Average	72.08±5.99	66.95±5.11	55.02±2.13	63.23±1.70	65.61±0.54	74.29±4.92	74.35±4.46	77.98±0.50	77.31±0.53

Table 9

The classification of various algorithms based on RF (%).

Datasets	RAW	HKCMI	GRRS	W-MGMN	GLSFS	BPNN	IGUFS	WOMFD(65%)	WOMFD(85%)
Immu	86.67±10.89	74.44±8.31	85.42±2.47	85.42±1.49	76.67±2.33	71.11±9.56	80.00±8.31	87.78±2.09	88.89±1.48
UFD	97.78±2.72	66.11±9.20	76.67±1.04	95.56±0.17	95.58±0.17	80.00±5.67	88.89±6.09	98.33±0.06	98.89±0.05
Sonar	82.74±8.02	57.84±13.24	67.17±0.53	77.40±1.02	79.86±0.16	86.49±4.60	86.04±3.58	87.98±0.19	85.62±0.31
Iono	92.87±4.10	85.75±5.72	86.57±0.46	91.14±0.12	92.87±0.10	92.87±2.40	94.02±3.18	94.02±0.07	94.02±0.17
TME	79.00±6.44	34.50±5.28	43.37±0.68	67.43±0.81	76.25±0.64	74.25±5.79	78.75±3.71	82.50±0.34	83.25±0.33
Wdbc	95.79±3.44	90.51±2.95	-	94.34±0.11	95.26±0.14	94.91±2.76	95.78±2.74	95.95±0.06	96.66±0.11
Breast	96.64±1.45	93.85±2.44	-	97.50±0.03	96.34±0.03	95.77±2.77	95.03±2.35	96.05±0.03	96.78±0.02
CCS	46.41±5.47	46.21±4.62	23.72±0.21	49.71±5.95	49.71±0.25	48.25±3.96	23.30±4.43	39.61±0.18	49.71±6.85
Card	88.43±2.72	49.95±3.79	85.83±0.03	82.63±0.05	81.18±0.13	90.12±2.30	86.41±2.15	90.88±0.04	90.92±0.05
Aba	54.27±1.23	50.05±1.47	49.31±0.06	54.26±0.06	54.11±0.04	51.33±1.50	53.77±2.30	53.75±0.05	55.21±0.02
Bid	99.43±0.32	88.85±0.35	-	76.33±0.01	99.65±0.00	90.21±0.33	98.80±0.19	98.04±0.00	98.91±0.00
Gait	91.50±1.90	6.00±0.84	17.00±1.56	72.00±3.16	92.00±2.56	90.00±1.80	88.00±1.76	92.00±1.76	92.00±3.36
DARW	93.73±5.30	-	-	78.59±1.25	88.56±0.59	94.28±4.39	95.42±4.28	98.30±0.13	97.75±0.14
PC	68.89±10.89	-	59.86±2.19	69.72±0.72	66.67±1.48	66.67±0.74	70.00±0.51	74.44±1.49	73.33±2.02
DB	92.40±0.68	-	86.99±0.01	76.33±0.01	89.87±0.01	92.41±0.64	89.97±0.84	92.48±0.00	92.66±0.01
Average	85.15±5.03	66.57±6.17	61.99±0.84	77.90±1.00	82.31±0.58	81.53±4.42	80.91±4.06	85.90±0.34	87.19±1.04

Table 10

Average ranking of classification accuracy among different classifiers across various algorithms.

	HKCMI	GRRS	W-MGMN	GLSFS	BPNN	IGUFS	WOMFD(65%)	WOMFD(85%)
KNN	6.93	6.53	5.20	4.13	4.87	4.00	2.13	2.13
SVM	6.27	7.00	5.93	4.20	4.67	3.27	2.07	1.53
NB	6.27	5.93	5.73	5.07	4.20	4.20	1.87	2.53
RF	7.33	6.93	4.80	4.00	4.67	4.07	2.40	1.73

Table 10 presents the average classification accuracy rankings for the two proposed algorithms and six comparative methods across the four classifiers mentioned above. For algorithms that failed to generate results, the lowest possible rank was assigned for the corresponding dataset. Based on these average rankings, the Friedman statistics and corresponding P-values are provided in the Table 11. With the significance level set at 0.05, all P-values fall below the threshold, allowing the rejection of the null hypothesis that all algorithms perform equally. This indicates significant differences in classification performance among the algorithms.

Furthermore, to examine pairwise differences between algorithms, the Wilcoxon signed-rank test was conducted to compare the proposed method with six competing algorithms. The significance level was also set at 0.05. The Wilcoxon test results are shown in the Table 12. Except for a few comparisons with the IGUFS algorithm under specific classifiers where the null hypothesis could not be rejected, most of the remaining comparisons show statistically significant differences. This indicates that the proposed algorithm outperforms most of the competing methods in terms of classification performance. The failure to reject the null hypothesis in some cases may be attributed to insufficient sample size.

Table 11
Friedman test results for various classifiers.

	χ^2	F	P Value
KNN	53.88	14.76	2.48×10^{-9}
SVM	45.61	10.75	1.04×10^{-7}
NB	40.48	8.78	1.02×10^{-6}
RF	64.80	22.57	1.65×10^{-11}

Table 12
P-Values of Wilcoxon tests on various classifiers.

Algorithms	KNN		SVM		NB		RF	
	65%	85%	65%	85%	65%	85%	65%	85%
HKCMI	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05
GRRS	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05
W-MGMN	0.08	<0.05	<0.05	<0.05	<0.05	<0.05	0.07	<0.05
GLSFS	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05	0.16	<0.05
BPNN	<0.05	<0.05	<0.05	<0.05	0.07	0.13	0.07	<0.05
IGUFS	0.06	0.13	0.10	0.10	0.20	0.40	<0.05	<0.05

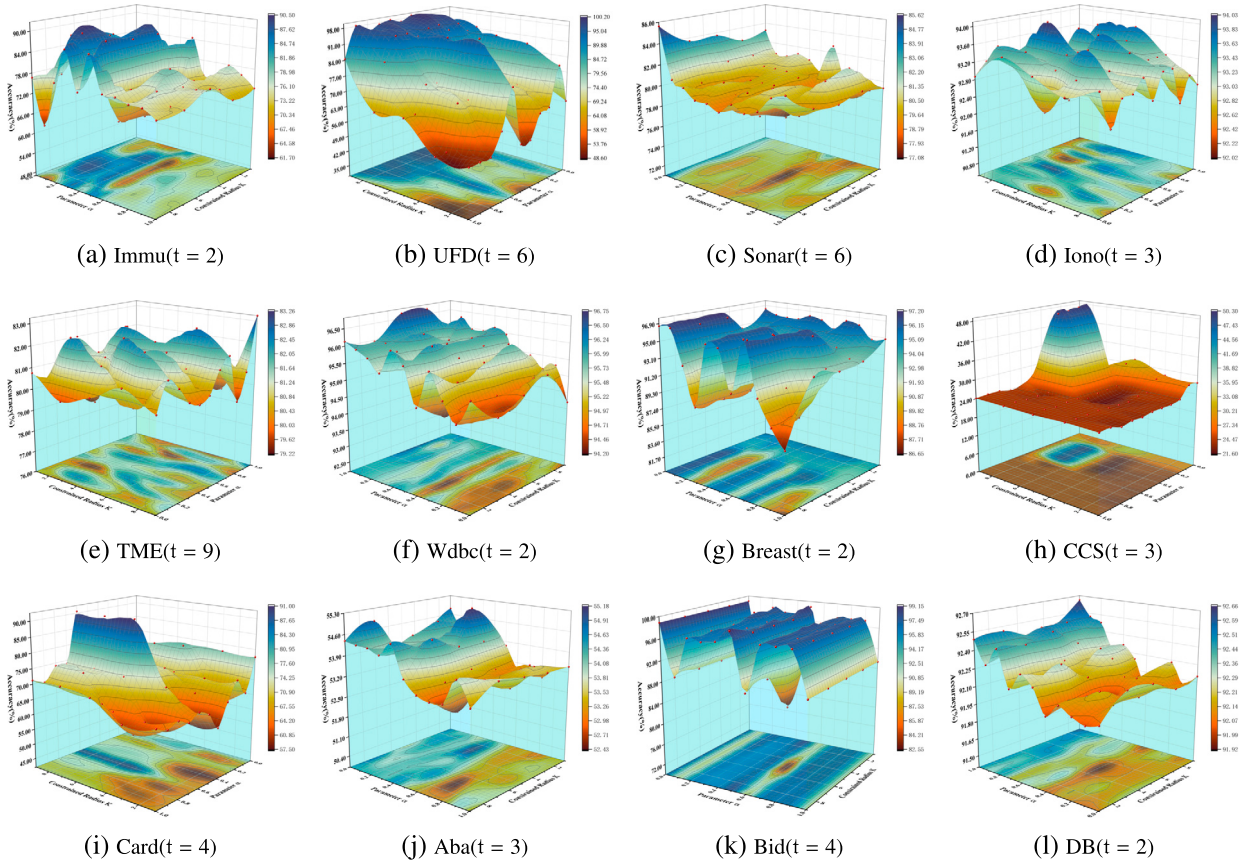


Fig. 4. The classification accuracy across different neighborhood constraint radii κ and parameter α under the optimal granularity t of RF classifier.

5.4. Parameter analysis

This section the impact of varying the neighborhood constraint radius κ and correlation index α on the classification performance of our algorithm WOMFD under the RF classifier is investigated, with Fig. 4 illustrating the classification accuracy across different parameter for various datasets at their respective optimal granularities t . The graph demonstrates that varied performance is observed among different datasets under various parameter combinations, underscoring the importance of selecting the best parameters to optimize classification efficiency. For instance, the small-sample datasets Immu, UFD, and Sonar tend to prefer smaller values for

α , specifically 0.2, 0.4, and 0, respectively. Conversely, they require larger values for the constraint radius κ , specifically 7 and 9. For the Card and DB datasets, achieving optimal classification performance requires larger values for both κ and α . Moreover, the classification results are influenced by the granularity levels applied to different datasets. The selection of granularity is determined based on the number of features in the dataset. For datasets with a larger number of features, the parameter t can be set to 3, 6, and 9, whereas for datasets with fewer features, values of 1, 2, and 3 are recommended.

This study annotates the optimal granularity specifications for each dataset in parentheses within the figure, revealing the diverse granularity preferences across datasets. Datasets with a larger number of features require coarser granularity to achieve better results, whereas those with fewer features need finer granularity. In conclusion, the choice of different parameters can lead to diverse outcomes, underscoring the importance of selecting appropriate parameter values based on specific requirements.

6. Conclusions

This paper proposes a feature selection method for fuzzy information systems that integrates graph theory with granularity weighting in the Weighted Optimistic Multi-Granulation Neighborhood-Constrained Fuzzy Rough Set (WOMNcFRS) model. The approach first ranks features using graph-theoretic measures, which guide the partitioning of the granularity space. Then, a granularity-weighted approximation precision strategy is introduced, leading to the construction of a weighted multi-granulation fuzzy rough set model. Based on this, a heuristic feature selection algorithm is refined by incorporating two different feature selection ratios. The proposed algorithm fully leverages decision information to guide neighborhood construction, significantly enhancing noise resistance. Moreover, by combining graph-based evaluation with granularity weighting, the computational complexity is effectively reduced. Extensive experiments conducted on 12 publicly available datasets, along with comparisons against six alternative algorithms, demonstrate the superior classification performance of the proposed method.

Despite these promising results, several limitations remain. First, the proposed method is built upon the optimistic WOMNcFRS framework, which may retain redundant features due to its inclusive approximation criteria. Exploring a pessimistic or hybrid granulation strategy could help mitigate this limitation and improve selection precision. Second, our approach is designed for static datasets and lacks adaptability to real-time or evolving environments. Future work could investigate incremental or streaming feature selection mechanisms under multi-granular fuzzy settings.

Overall, this study contributes a unified and robust framework for fuzzy feature selection, offering enhanced granularity discrimination and noise resilience, while opening avenues for further exploration in dynamic, hybrid, and adaptive fuzzy systems.

CRedit authorship contribution statement

Xiaoyan Zhang: Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Xuan Shen:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Data curation. **Weicheng Zhao:** Writing – review & editing, Visualization, Validation, Software, Methodology, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank the Associate Editor and the reviewers for their insightful comments and suggestions.

This work was supported by the National Natural Science Foundation of China (Grant No. 12371465) and the Chongqing Natural Science Foundation (Grant No. CSTB2023NSCQ-MSX1063).

Data availability

No data was used for the research described in the article.

References

- [1] Y. Xue, Y. Tang, X. Xu, J. Liang, F. Neri, Multi-objective feature selection with missing data in classification, *IEEE Trans. Emerg. Top. Comput. Intell.* 6 (2) (2021) 355–364.
- [2] Q. Zhou, Q. Wang, Q. Gao, M. Yang, X. Gao, Unsupervised discriminative feature selection via contrastive graph learning, *IEEE Trans. Image Process.* 33 (2024) 972–986.
- [3] Z. Feng, X. Zhang, Supervised incremental feature selection using regularization vector for dynamic multi-scale interval valued datasets, *Pattern Recognit.* (2025) 111985.
- [4] E.C. Tsang, D. Chen, D.S. Yeung, X.-Z. Wang, J.W. Lee, Attributes reduction using fuzzy rough sets, *IEEE Trans. Fuzzy Syst.* 16 (5) (2008) 1130–1141.
- [5] M. Hu, E.C. Tsang, Y. Guo, W. Xu, Fast and robust attribute reduction based on the separability in fuzzy decision systems, *IEEE Trans. Cybern.* 52 (6) (2021) 5559–5572.
- [6] T.M. Al-shami, An improvement of rough sets' accuracy measure using containment neighborhoods with a medical application, *Inf. Sci.* 569 (2021) 110–124.

- [7] S. An, E. Zhao, C. Wang, G. Guo, S. Zhao, P. Li, Relative fuzzy rough approximations for feature selection and classification, *IEEE Trans. Cybern.* 53 (4) (2021) 2200–2210.
- [8] X. Zhang, M. Li, S. Shao, J. Wang, (i, o)-fuzzy rough sets based on overlap functions with their applications to feature selection and image edge extraction, *IEEE Trans. Fuzzy Syst.* 32 (4) (2023) 1796–1809.
- [9] M. Beynon, Reducts within the variable precision rough sets model: a further investigation, *Eur. J. Oper. Res.* 134 (3) (2001) 592–605.
- [10] X. Hu, Y. Shen, W. Pedrycz, Y. Li, G. Wu, Granular fuzzy rule-based modeling with incomplete data representation, *IEEE Trans. Cybern.* 52 (7) (2021) 6420–6433.
- [11] X. Hu, X. Liu, W. Pedrycz, Q. Liao, Y. Shen, Y. Li, S. Wang, Multi-view fuzzy classification with subspace clustering and information granules, *IEEE Trans. Knowl. Data Eng.* 35 (11) (2022) 11642–11655.
- [12] S. An, Q. Hu, C. Wang, Probability granular distance-based fuzzy rough set model, *Appl. Soft Comput.* 102 (2021) 107064.
- [13] M. Hu, Y. Guo, D. Chen, E.C. Tsang, Q. Zhang, Attribute reduction based on neighborhood constrained fuzzy rough sets, *Knowl.-Based Syst.* 274 (2023) 110632.
- [14] Y. Qian, J. Liang, Y. Yao, C. Dang, Mgrs: a multi-granulation rough set, *Inf. Sci.* 180 (6) (2010) 949–970.
- [15] T. Feng, J.-S. Mi, Variable precision multigranulation decision-theoretic fuzzy rough sets, *Knowl.-Based Syst.* 91 (2016) 93–101.
- [16] B. Sun, W. Ma, X. Xiao, Three-way group decision making based on multigranulation fuzzy decision-theoretic rough set over two universes, *Int. J. Approx. Reason.* 81 (2017) 87–102.
- [17] H. Jiang, J. Zhan, D. Chen, Covering-based variable precision (I, \mathcal{T}) -fuzzy rough sets with applications to multiattribute decision-making, *IEEE Trans. Fuzzy Syst.* 27 (8) (2018) 1558–1572.
- [18] W.-H. Xu, *Ordered Information Systems and Rough Sets*, Science Press, Beijing, 2013.
- [19] Y. Guo, E.C. Tsang, W. Xu, A weighted multi-granulation decision-theoretic approach to multi-source decision systems, in: *International Conference on Machine Learning and Cybernetics*, vol. 1, IEEE, 2017, pp. 202–210.
- [20] X. Ji, P. Zhao, S. Yao, Rough set model and decision research in intuitionistic fuzzy information system based on weighted multi-granulation, *Pattern Recognit. Artif. Intell.* 30 (11) (2017) 971–982.
- [21] Y. Yu, M. Wan, J. Qian, D. Miao, Z. Zhang, P. Zhao, Feature selection for multi-label learning based on variable-degree multi-granulation decision-theoretic rough sets, *Int. J. Approx. Reason.* 169 (2024) 109181.
- [22] X. Zhang, W. Zhao, Uncertainty measures and feature selection based on composite entropy for generalized multigranulation fuzzy neighborhood rough set, *Fuzzy Sets Syst.* 486 (2024) 108971.
- [23] X. Xie, Z. Cao, F. Sun, Joint learning of graph and latent representation for unsupervised feature selection, *Appl. Intell.* 53 (21) (2023) 25282–25295.
- [24] C. Tang, X. Zheng, W. Zhang, X. Liu, X. Zhu, E. Zhu, Unsupervised feature selection via multiple graph fusion and feature weight learning, *Sci. China Inf. Sci.* 66 (5) (2023) 152101.
- [25] R. Zhang, Y. Zhang, X. Li, Unsupervised feature selection via adaptive graph learning and constraint, *IEEE Trans. Neural Netw. Learn. Syst.* 33 (3) (2020) 1355–1362.
- [26] L. Xiang, H. Chen, T. Yin, S.-J. Horng, T. Li, Unsupervised feature selection based on bipartite graph and low-redundant regularization, *Knowl.-Based Syst.* 302 (2024) 112379.
- [27] W. Xu, M. Huang, Z. Jiang, Y. Qian, Graph-based unsupervised feature selection for interval-valued information system, *IEEE Trans. Neural Netw. Learn. Syst.* (2023).
- [28] X. You, J. Li, H. Wang, Relative reduction of neighborhood-covering pessimistic multigranulation rough set based on evidence theory, *Information* 10 (11) (2019) 334.
- [29] Z. Yuan, H. Chen, X. Yang, T. Li, K. Liu, Fuzzy complementary entropy using hybrid-kernel function and its unsupervised attribute reduction, *Knowl.-Based Syst.* 231 (2021) 107398.
- [30] S. Xia, S. Wu, X. Chen, G. Wang, X. Gao, Q. Zhang, E. Glem, Z. Chen, Grrs: accurate and efficient neighborhood rough set for feature selection, *IEEE Trans. Knowl. Data Eng.* 35 (9) (2022) 9281–9294.
- [31] W. Xu, Q. Bu, Matrix-based incremental feature selection method using weight-partitioned multigranulation rough set, *Inf. Sci.* 681 (2024) 121219.
- [32] X. Zhang, X. Shen, Graph-driven feature selection via granular-rectangular neighborhood rough sets for interval-valued data sets, *Appl. Soft Comput.* 170 (2025) 112716.