



Uncertainty measure and dynamic feature selection based on multigranulation entropy

Xiaoyan Zhang*, Meixia Zhou 

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

ARTICLE INFO

Keywords:

Multigranulation rough set
Multigranulation entropy
Incremental feature selection

ABSTRACT

Feature selection has received increasing attention, particularly with approaches based on multigranulation rough set models and entropy theory. By leveraging multiple levels of granularity, these methods improve robustness and information representation. Nevertheless, existing approaches face critical challenges, including excessively biasing granular division, inadequately quantifying inter-granular interaction importance, and delivering poor computational efficiency amid dynamic feature shifts. To address these issues, this study proposes a multigranulation fuzzy neighborhood rough set model based on K-means clustering, enabling objective and flexible granule construction. Additionally, a multigranulation entropy measure is introduced to evaluate inter-granular interaction importance and it is integrated into a weighted k-nearest neighbor rough set framework for effective feature selection. Furthermore, two incremental algorithms are developed to support dynamic updates of feature subsets. Finally, Extensive experiments on 15 benchmark datasets validate the efficacy and computational efficiency of the proposed methods in both static and dynamic feature selection scenarios.

1. Introduction

Feature selection [1] is a vital process in modern data analysis [2,3], particularly for high-dimensional datasets [4,5]. Such datasets often contain numerous redundant, irrelevant, or noisy features, which can impair the predictive performance [6,7] of machine learning models [8–10] and increase computational costs. Machine learning has been successfully applied across a wide range of practical domains, including price forecasting and medical diagnosis [11,12], but the effectiveness of these models depends heavily on the quality of input features. Consequently, effective feature selection is essential for building accurate, efficient, and robust machine learning systems [13,14].

In the realm of feature selection, metaheuristic algorithms and deep learning-based strategies [15] have been widely adopted for classification [16,17] and intelligent medical diagnosis tasks [18]. Nevertheless, these methods generally fail to explicitly quantify data uncertainty and suffer from poor mathematical interpretability. As a reliable and interpretable mathematical tool, Rough Set Theory (RST) proposed by Pawlak in 1982 [19] exhibits unique superiority in handling uncertain and incomplete data [20,21]. It excavates inherent data knowledge without relying on prior assumptions, making it highly suitable for uncertainty-aware feature screening. So, it can serve as an effective complement to metaheuristic and deep learning based feature selection methods. However, the classical rough set model is limited to discrete data, which constrains its applicability in real-world scenarios. To address this limitation, Neighborhood Rough Set (NRS) is subsequently introduced. Hu et al. initially proposed replacing the equivalence relations in classical rough sets

* Corresponding author.

E-mail addresses: zxy19790915@163.com (X. Zhang), 18852981926@163.com (M. Zhou).

<https://doi.org/10.1016/j.fss.2026.110099>

Received 28 October 2025; Received in revised form 25 June 2026; Accepted 25 August 2026

Available online 29 August 2026

0165-0114/© 2026 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Table 1
The summary of existing methods.

Method	Category	Limitation
Neighborhood Rough Set (NRS)	Basic distance-based NRS [22] Covering-based NRS [23] Weighted NRS [24]	Relies on fixed-radius neighborhoods
Fuzzy Neighborhood Rough Set (FNRS)	Single-granularity FNRS [25]	Limited to single-granularity analysis
Multigranulation Rough Set (MNRS)	Equal-weight MNRS [26] Weighted MNRS [29,30]	Subjective granule partitioning
	Traditional incremental rough set algorithms [32]	High redundant computation cost

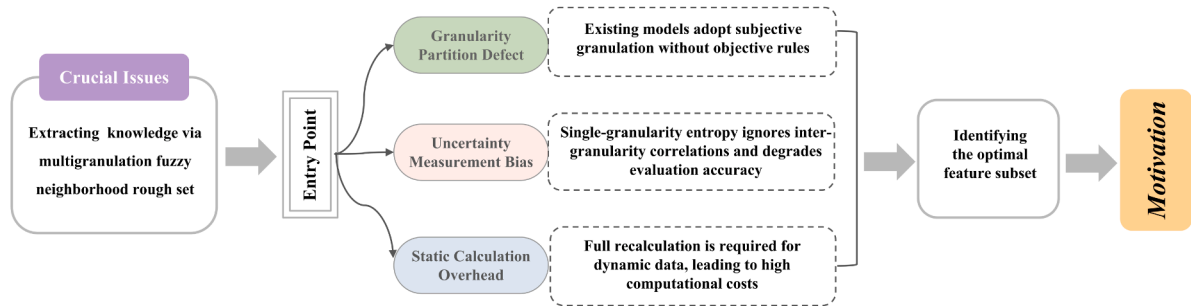


Fig. 1. The motivation of feature selection.

with distance-based neighborhood relations, and simultaneously expanded the application scope of NRS [22]. Building on this, D'eer et al. extended the neighborhood operators from the perspective of covering-based rough sets and investigated the mathematical properties of various neighborhood definitions [23]. Meanwhile, Zhang et al. introduced Weighted Neighborhood Rough Set (WNRS), which characterizes local data structures based on the weights of neighborhood relations and improves the effectiveness of feature evaluation [24]. Although NRS has improved the applicability of RST to some extent, its reliance on a fixed-radius neighborhood limits its ability to handle boundary objects. Specifically, it can only determine in a binary manner (0 or 1) whether an object is included or not, which hinders its effectiveness in capturing the membership degrees of objects located in fuzzy boundary regions. Therefore, Fuzzy Neighborhood Rough Set (FNRS) is established. Chen et al. enhanced the representation of fuzzy indiscernibility relations and improved the accuracy of attribute reduction by introducing a fuzzy discernibility matrix [25]. Limited to a single granular structure, FNRS fails to exploit multi-view data characteristics. As an extended paradigm, Multigranulation Neighborhood Rough Set (MNRS) introduces multiple granular spaces to remedy this defect. Xu et al. formulated a generalized multigranulation rough set framework, which significantly broadens the applicability of rough set theory to multigranular analysis and alleviates the shortcomings of traditional models in representing hierarchical granularity [26]. Nonetheless, most existing multigranulation methods rely on subjective granularity partitioning and lack objective criteria for granule construction. This forms the first challenge tackled in this paper.

Information Entropy (IE), a classical measure of uncertainty and feature–target correlation, has been widely applied in feature selection tasks [27,28]. Various entropy-based metrics have been integrated into MNRS to assess feature importance. Xu et al. designed an improved entropy to quantify uncertainty within the generalized multigranulation fuzzy neighborhood rough set framework [29]. Zhang et al. defined a composite entropy to assess feature importance in this context [30]. Nevertheless, these methods predominantly concentrate on single-granularity entropy, which fails to account for inter-granular interactions and compromises the precision of feature evaluation. This constitutes the second challenge addressed in this research.

In real world, the dynamic nature of data often requires traditional rough set models to perform recalculations after each update, thereby incurring substantial computational costs [31]. In response, incremental feature selection techniques have been investigated, aiming to improve computational efficiency by dynamically adapting to evolving data. Yang et al. devised sample selection to eliminate redundant data and used a feature-based accelerator to improve the efficiency of incremental feature selection [32]. Consequently, adapting feature selection methods to accommodate the dynamic nature of data is the third challenge investigated in this study.

To address the shortcomings of existing methods highlighted in Table 1, this paper presents a novel dynamic feature selection method by integrating the multigranulation fuzzy neighborhood rough set framework with information entropy. As illustrated in Fig. 1, the proposed method is developed from three key motivation perspectives and aims to identify the optimal feature subset. The main contributions of this work are summarized as follows:

- (1) A multigranulation fuzzy neighborhood rough set model based on k-means (K-MFNRS) is proposed. In this model, each knowledge granule is assigned a weight, and the k-means algorithm is applied for adaptive granulation. This approach ensures a more reasonable partitioning of granularities;
- (2) A multigranulation entropy (MGE) measure is formulated to assess the importance of each granulation. This measure analyzes the interactions between granularities, offering a more thorough representation of uncertainty in the information system. It is then integrated into the weighted k-nearest neighbor rough set (WKN-RS) framework to develop a feature selection algorithm (K-FSAMGE) for the optimal search of feature subsets;

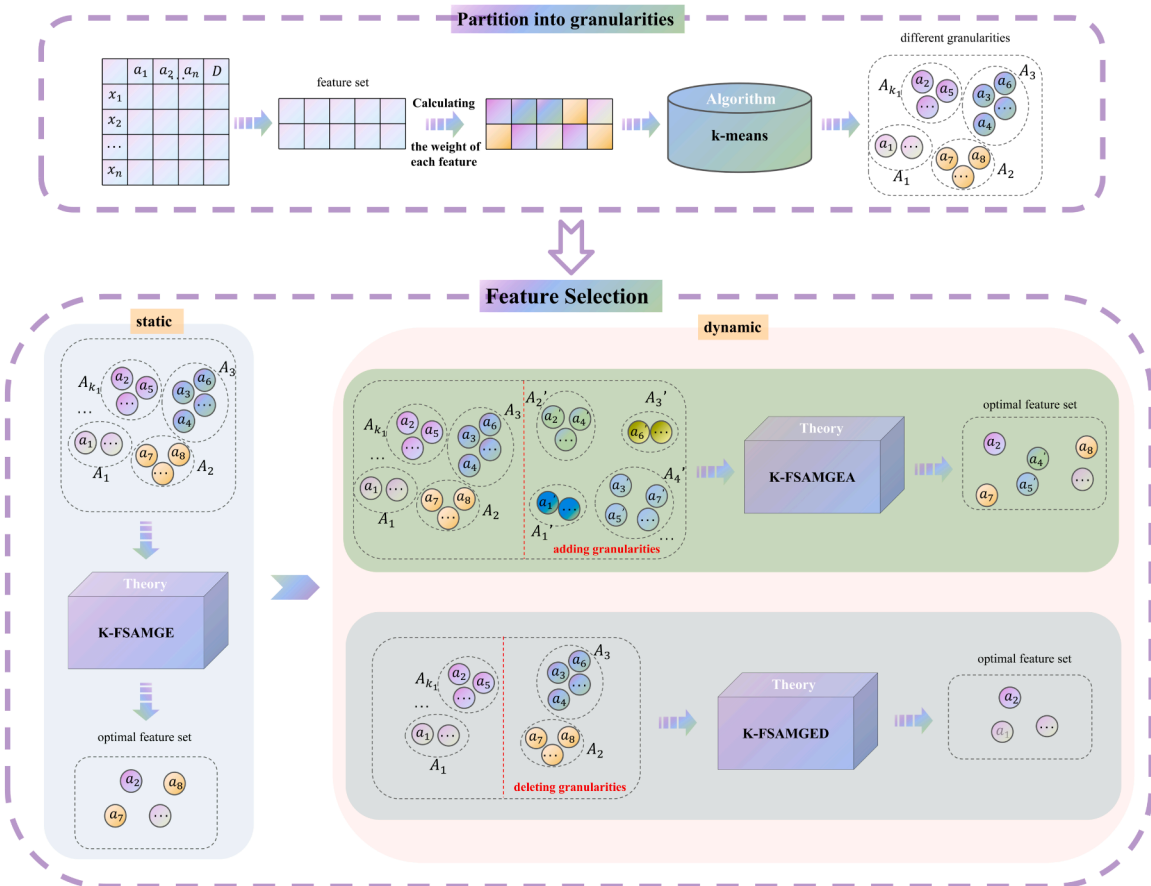


Fig. 2. The overall framework of this study.

Table 2
Nomenclature of abbreviations.

Abbreviation	Full Definition
NRS	Neighborhood Rough Set
FNRS	Fuzzy Neighborhood Rough Set
MNRS	Multigranulation Neighborhood Rough Set
IE	Information Entropy
WKN-RS	Weighted K-nearest Neighbor Rough Set
K-MFNRS	Multigranulation Fuzzy Neighborhood Rough Set based on K-means
SGE	Single-granularity Entropy
MGE	Multigranulation Entropy
K-FSAMGE	Feature Selection Algorithm based on MGE and WKN-RS
K-FSAMGEA	Incremental addition feature selection algorithm
K-FSAMGED	Incremental deletion feature selection algorithm
NIS	Neighborhood Information System
IS	Information System

(3) Two incremental feature selection algorithms (K-FSAMGEA and K-FSAMGED) are established. These algorithms can accelerate computation speed when feature changes.

The structure of the paper is as follows: [Section 2](#) reviews the core concepts, including MNRS, WKN-RS and dynamic updating algorithms. [Section 3](#) introduces the K-MFNRS model and proposes a feature selection algorithm based on MGE. Subsequently, two incremental update methods are designed to handle changes in features. [Section 4](#) demonstrates the effectiveness and robustness of the proposed methods through experiments on 12 datasets. Finally, [Section 5](#) concludes the paper and outlines potential directions for future research. The overall framework of this study is illustrated in [Fig. 2](#). In addition, all abbreviations adopted in this paper are summarized in [Table 2](#).

2. Preliminaries

In this section, we briefly review several fundamental concepts relevant to our study, including multigranulation neighborhood rough set (MNRS), weighted k -nearest rough set (WKN-RS), and dynamic update algorithms.

2.1. Multigranulation neighborhood rough set

In MNRS, the Euclidean distance is typically adopted to characterize the neighborhood relationship between any two objects (x, y) under a subset of condition attributes $N \subseteq AT$. If the distance between two objects is less than or equal to a predefined neighborhood radius δ , then the objects are treated as indistinguishable [30].

The neighborhood class of x with respect to N is defined as

$$[x]_N^\delta = \{y \in U \mid \Delta_N(x, y) \leq \delta\}, \quad (1)$$

where $\Delta_N(x, y)$ denotes the Euclidean distance between x and y based on attributes in N .

Given a $NIS = (U, A \cup \{d\})$, for any $X \subseteq U$ and $N_i \subseteq A$ ($i = 1, 2, \dots, m \leq 2^{|A|}$), the optimistic lower and upper approximations of X with respect to N_i are defined as

$$\underline{\text{OM}}_{\sum_{i=1}^m R_{N_i}^\delta}(X) = \{x \in U \mid \bigvee_{i=1}^m ([x]_{N_i}^\delta \subseteq X)\}, \quad (2)$$

$$\overline{\text{OM}}_{\sum_{i=1}^m R_{N_i}^\delta}(X) = \{x \in U \mid \bigwedge_{i=1}^m ([x]_{N_i}^\delta \cap X \neq \emptyset)\}. \quad (3)$$

Similarly, the pessimistic lower and upper approximations are given by

$$\underline{\text{PM}}_{\sum_{i=1}^m R_{N_i}^\delta}(X) = \{x \in U \mid \bigwedge_{i=1}^m ([x]_{N_i}^\delta \subseteq X)\}, \quad (4)$$

$$\overline{\text{PM}}_{\sum_{i=1}^m R_{N_i}^\delta}(X) = \{x \in U \mid \bigvee_{i=1}^m ([x]_{N_i}^\delta \cap X \neq \emptyset)\}. \quad (5)$$

Here, “ $\wedge$ ” and “ $\vee$ ” denote the logical “and” and “or” operations, respectively.

The pair $(\underline{\text{OM}}_{\sum_{i=1}^m R_{N_i}^\delta}(X), \overline{\text{OM}}_{\sum_{i=1}^m R_{N_i}^\delta}(X))$ is termed the optimistic multigranulation neighborhood rough set (OMNRS), whereas $(\underline{\text{PM}}_{\sum_{i=1}^m R_{N_i}^\delta}(X), \overline{\text{PM}}_{\sum_{i=1}^m R_{N_i}^\delta}(X))$ is called the pessimistic multigranulation neighborhood rough set (PMNRS).

2.2. Weighted k -nearest neighbor rough set

NRS struggles with uneven sample distributions and noise in real-world datasets. To overcome these limitations, we introduce the weighted k -nearest neighbor rough set (WKN-RS) [33], which assigns greater importance to closer and more stable neighbors.

Consider a sample x with k nearest neighbors y_i . Let $d(x, y_i)$ denote the distance between x and y_i . The similarity between x and y_i is defined as

$$\text{Sim}(x, y_i) = \frac{1}{d(x, y_i)}. \quad (6)$$

Let $\text{std}(x, y_i)$ represent the standard deviation of neighbor y_i within the neighborhood of x . The weight assigned to y_i is computed as

$$w(y_i) = \frac{1}{\text{std}(x, y_i)}. \quad (7)$$

Thus, the weighted quality measure of the k nearest neighbors of sample x is expressed as

$$\text{SIM}_k(x) = \sum_s w(y_i) \cdot \text{Sim}(x, y_i) - \sum_d w(y_i) \cdot \text{Sim}(x, y_i), \quad (8)$$

where the first term represents the aggregated similarity to samples of the same class, and the second term represents the aggregated similarity to samples of different classes.

Given an $IS = (U, A \cup D)$, where $B \subseteq A$ and $\forall X \subseteq U$, the lower and upper approximations of X with respect to B are given by

$$\underline{\text{KN}}_B(X) = \{x \in U \mid \text{SIM}_k(x) > \alpha\}, \quad (9)$$

$$\overline{\text{KN}}_B(X) = \{x \in U \mid \text{SIM}_k(x) \geq -\alpha\}, \quad (10)$$

where α is a small real number close to 0.

The pair $(\underline{\text{KN}}_B(X), \overline{\text{KN}}_B(X))$ constitutes WKN-RS. Furthermore, given an $IS = (U, A \cup D)$, where $B \subseteq A$ and $U/D = \{D_1, D_2, \dots, D_s\}$, the positive region of D with respect to B is defined as

$$\text{KPOS}_B(D) = \bigcup_{i=1}^s \underline{\text{KN}}_B(D_i). \quad (11)$$

The dependency degree of D on B is then calculated by

$$K\gamma_B(D) = \frac{|KPOS_B(D)|}{|U|}. \quad (12)$$

2.3. Dynamic updating algorithms

In dynamic environments, the set of attributes or objects may evolve over time. Dynamic updating algorithms aim to exploit previously computed results to efficiently adjust the relevant rough set measures, thereby avoiding complete recomputation [34].

In classical rough set, when the attribute set changes from S to S' , only the affected positive region needs to be updated, as follows

$$POS_{S'}(D) = POS_S(D) \cup \Delta POS^+(D) \setminus \Delta POS^-(D), \quad (13)$$

where $\Delta POS^+(D)$ and $\Delta POS^-(D)$ denote the sets of objects newly added to or removed from the positive region, respectively, due to the attribute modification.

In NRS, each object $x \in U$ is associated with a neighborhood granule $\mathcal{N}_S(x)$. When the attribute set changes from S to S' , only the neighborhoods of affected objects are updated as

$$\mathcal{N}_{S'}(x) = \mathcal{N}_S(x) \oplus \Delta \mathcal{N}(x), \quad (14)$$

where the operator $\oplus$ adjusts the neighborhood based on the updated distance or similarity measures.

In MNRS, multiple granularities $G_1, G_2, \dots, G_m$ are considered simultaneously, and the multigranulation dependency degree is computed as

$$\gamma_S^{MG}(D) = \Phi(\gamma_S^{G_1}(D), \gamma_S^{G_2}(D), \dots, \gamma_S^{G_m}(D)), \quad (15)$$

where $\Phi(\cdot)$ is an aggregation function. When the attribute set changes from S to S' , only the dependency degrees for the affected granularities are recomputed as

$$\gamma_{S'}^{MG}(D) = \Phi(\gamma_{S'}^{G_1}(D), \dots, \gamma_{S'}^{G_m}(D)). \quad (16)$$

3. Uncertainty measure based on MGE and WKN-RS

In this section, we present multigranulation fuzzy neighborhood rough set based on k-means. To quantify the importance of different granularities, a multigranulation entropy is introduced. By incorporating this entropy into weighted k-nearest neighbor rough set, we design a feature selection algorithm to identify the optimal subset of features. Furthermore, two incremental feature selection algorithms are developed to ensure efficient adaptation in dynamic environments.

3.1. Multigranulation fuzzy neighborhood rough set based on k-means

In practice, knowledge granules differ in their importance. Many existing methods overlook granule quality, which can reduce partitioning accuracy. We improve this by assigning each granule a quality-based weight and using k-means to group them into suitable granularities.

Definition 1. Given a dataset $X \in \mathbb{R}^{n \times m}$ and a target variable $D \in \mathbb{R}^n$, where n denotes the number of samples and m represents the number of features. The feature weight vector is defined as

$$w_j = \begin{cases} \text{RFI}(X, D), & m > 1000, \\ \text{MI}(X, D), & \text{otherwise,} \end{cases} \quad (17)$$

where RFI refers to the Random Forest Importance and MI denotes Mutual Information.

The weight vector is normalized using min-max normalization, which can be computed as

$$\bar{w}_j = \frac{w_j - \min(w)}{\max(w) - \min(w)}, \quad (18)$$

where $\bar{w}_j$ is the normalized weight of the j th feature, and w_j is its original weight obtained via RFI or MI.

After computing the feature weights, the k -means algorithm is applied to cluster features into k_1 distinct granulations based on their weight values. These granulations are denoted as A_t , where $t = 1, 2, \dots, k_1$.

Definition 2. Let $NIS = (U, AT \cup D)$, where $A_t \subseteq AT$ ($t = 1, 2, \dots, k_1$) and $U/D = \{D_1, D_2, \dots, D_s\}$. Given a neighborhood radius δ , the neighborhood class of x with respect to A_t is defined as

$$[x]_{A_t}^\delta = \{y \in U \mid \Delta(x, y) \leq \delta\}. \quad (19)$$

For $\forall D_i \in U/D$, the lower and upper approximations with respect to R_{A_t} are given by

$$\underline{R}_{A_t}^\delta(D_i) = \{x \in D_i \mid [x]_{R_{A_t}}^\delta \subseteq D_i\}, \quad (20)$$

$$\overline{R}_{A_t}^\delta(D_i) = \{x \in D_i \mid [x]_{R_{A_t}}^\delta \cap D_i \neq \emptyset\}. \quad (21)$$

The pair $\left(\bigcup_{i=1}^{k_1} \underline{R}_{A_t}^\delta(D_i), \bigcap_{i=1}^{k_1} \overline{R}_{A_t}^\delta(D_i)\right)$ is termed the multigranulation fuzzy neighborhood rough set based on k -means (K-MFNRS).

3.2. Multigranulation entropy (MGE) in K-MFNRS

Entropy is a fundamental metric in feature selection, measuring the uncertainty or impurity of an information system. Most existing entropy-based approaches, however, compute measures independently at each granularity, without coordinated cross-level analysis. To bridge this gap, we propose MGE, which supports interactive evaluation across multiple granularities.

Definition 3. Given a $NIS = (U, AT \cup D)$, where $A_t \subseteq AT$ ($t = 1, 2, \dots, k_1$) and $U/D = \{D_1, D_2, \dots, D_s\}$, for $\forall D_i \in U/D$, single-granularity entropy (SGE) is specified as

$$SGE(A_t, D) = - \sum_{i=1}^s \frac{|D_i|}{|U|} \log \left(\frac{\left| \underline{R}_{A_t}^\delta(D_i) \right| + \epsilon}{\left| \overline{R}_{A_t}^\delta(D_i) \right| + \epsilon} \right), \quad (22)$$

where $\underline{R}_{A_t}^\delta(D_i)$ and $\overline{R}_{A_t}^\delta(D_i)$ denote the lower and upper approximations of D_i under R_{A_t} , respectively. A tiny positive smoothing constant $\epsilon = 10^{-8}$ is embedded to resolve undefined logarithmic inputs and numerical instability: it avoids division-by-zero when $\left| \underline{R}_{A_t}^\delta(D_i) \right| = 0$ and suppresses sharp value jumps where $\left| \underline{R}_{A_t}^\delta(D_i) \right| = \left| \overline{R}_{A_t}^\delta(D_i) \right|$.

MGE comprises two components: the first characterizes the uncertainty associated with a given granularity level, while the second quantifies its internal information content [35]. Formally, MGE is expressed as

$$MGE(AM) = - \sum_{i=1}^{k_1} P_i \cdot \log P_i + \sum_{i=1}^{k_1} P_i \cdot SGE(A_t, D), \quad (23)$$

where $AM \subseteq \bigcup_{i=1}^{k_1} A_t$ denotes the current set of selected granulated attributes, and $P_i = \frac{|A_t|}{|AM|}$ represents the relative proportion of the i th granularity.

Proposition 1 (Boundedness of MGE). *Given a granularity set $AM \subseteq \bigcup_{i=1}^{k_1} A_t$, the multigranulation entropy is bounded as*

$$0 \leq MGE(AM) \leq \log k_1 + \sum_{i=1}^{k_1} P_i \cdot SGE_{\max}(A_t, D), \quad (24)$$

where $SGE_{\max}(A_t, D)$ denotes the maximum possible single-granularity entropy for the i th granularity.

Proof. The first term of MGE, representing the Shannon entropy of the granularity distribution, ranges from 0 to $\log k_1$. The second term, the weighted sum of single-granularity entropies, is non-negative and bounded above by $\sum_{i=1}^{k_1} P_i \cdot SGE_{\max}(A_t, D)$. Combining these two terms gives the overall lower and upper bounds of MGE. $\square$

Remark 1. Although MGE has a well-defined upper bound, it is not strictly monotonic with respect to the number of granularities. Adding redundant or low-information granules may not reduce MGE. Therefore, the optimal granularity set should be determined by minimizing MGE, rather than assuming that more granularities always decrease it.

Example 1. To illustrate the execution pipeline of K-FSAMGE, we adopt the fuzzy decision system in Table 3. Hyperparameters are fixed as $k_1 = 3$, $k_2 = 2$, $\delta = 0.3$, $\epsilon = 10^{-8}$.

We first compute mutual information feature weights for a_1 to a_6 as

$$[0.30, 0.40, 0.43, 0.35, 0.32, 0.28],$$

and the corresponding normalized weights are

$$[0.133, 0.800, 1.000, 0.467, 0.267, 0.000].$$

K-means with $k_1 = 3$ generates three initial granular subsets:

$$A_1 = \{a_2, a_3\}, \quad A_2 = \{a_1, a_4, a_5\}, \quad A_3 = \{a_6\}.$$

The decision partition is $U/D = \{D_1 = \{x_1, x_2, x_3\}, D_2 = \{x_4, x_5, x_6\}\}$, $|U| = 6$, $|D_1| = |D_2| = 3$.

Table 3
The fuzzy decision system.

U	a_1	a_2	a_3	a_4	a_5	a_6	d
x_1	0.31	0.85	0.90	0.22	0.14	0.08	1
x_2	0.27	0.81	0.93	0.25	0.11	0.10	1
x_3	0.33	0.78	0.88	0.20	0.16	0.06	1
x_4	0.72	0.35	0.24	0.81	0.75	0.68	2
x_5	0.68	0.31	0.20	0.85	0.79	0.72	2
x_6	0.75	0.38	0.27	0.78	0.72	0.65	2

For $A_1 = \{a_2, a_3\}$, the lower and upper approximations are:

$$\underline{R}_{A_1}^\delta(D_1) = \{x_1, x_2, x_3\}, \overline{R}_{A_1}^\delta(D_1) = \{x_1, x_2, x_3\},$$

$$\underline{R}_{A_1}^\delta(D_2) = \{x_4, x_5, x_6\}, \overline{R}_{A_1}^\delta(D_2) = \{x_4, x_5, x_6\}.$$

For $A_2 = \{a_1, a_4, a_5\}$, the lower and upper approximations are:

$$\underline{R}_{A_2}^\delta(D_1) = \{x_1, x_2\}, \overline{R}_{A_2}^\delta(D_1) = \{x_1, x_2, x_3\},$$

$$\underline{R}_{A_2}^\delta(D_2) = \{x_4, x_5\}, \overline{R}_{A_2}^\delta(D_2) = \{x_3, x_4, x_5\}.$$

For $A_3 = \{a_6\}$, the lower and upper approximations are:

$$\underline{R}_{A_3}^\delta(D_1) = \{x_1\}, \overline{R}_{A_3}^\delta(D_1) = \{x_1, x_2, x_3\},$$

$$\underline{R}_{A_3}^\delta(D_2) = \{x_5\}, \overline{R}_{A_3}^\delta(D_2) = \{x_4, x_5, x_6\}.$$

So, we obtain the SGE values are:

$$\text{SGE}(A_1, D) = -\frac{3}{6} \log\left(\frac{3+\varepsilon}{3+\varepsilon}\right) - \frac{3}{6} \log\left(\frac{3+\varepsilon}{3+\varepsilon}\right) = 0,$$

$$\text{SGE}(A_2, D) = -0.5 \log\left(\frac{2+\varepsilon}{3+\varepsilon}\right) - 0.5 \log\left(\frac{2+\varepsilon}{3+\varepsilon}\right) \approx 0.4055,$$

$$\text{SGE}(A_3, D) = -0.5 \log\left(\frac{1+\varepsilon}{3+\varepsilon}\right) - 0.5 \log\left(\frac{1+\varepsilon}{3+\varepsilon}\right) \approx 1.0986.$$

Assuming $k_2 = 2$, we can obtain three candidate granularity sets: $\{A_1, A_2\}$, $\{A_2, A_3\}$, and $\{A_1, A_3\}$.

For full set $AM = \{A_1, A_2, A_3\}$ with $P_1 = 2/6, P_2 = 3/6, P_3 = 1/6$, the MGE value is:

$$\begin{aligned} \text{MGE}(\{A_1, A_2, A_3\}) &= -\left(\frac{2}{6} \log \frac{2}{6} + \frac{3}{6} \log \frac{3}{6} + \frac{1}{6} \log \frac{1}{6}\right) \\ &\quad + \left(\frac{2}{6} \times 0 + \frac{3}{6} \times 0.4055 + \frac{1}{6} \times 1.0986\right) \\ &\approx 1.3973. \end{aligned}$$

For $\{A_1, A_2\}$ with $P_1 = 2/5, P_2 = 3/5$, the MGE value is:

$$\begin{aligned} \text{MGE}(\{A_1, A_2\}) &= -\left(\frac{2}{5} \log \frac{2}{5} + \frac{3}{5} \log \frac{3}{5}\right) \\ &\quad + \frac{2}{5} \times 0 + \frac{3}{5} \times 0.4055 \\ &\approx 0.9163 \end{aligned}$$

For $\{A_1, A_3\}$ with $P_1 = 2/3, P_3 = 1/3$, the MGE value is:

$$\begin{aligned} \text{MGE}(\{A_1, A_3\}) &= -\left(\frac{2}{3} \log \frac{2}{3} + \frac{1}{3} \log \frac{1}{3}\right) \\ &\quad + \left(\frac{2}{3} \times 0 + \frac{1}{3} \times 1.0986\right) \\ &\approx 1.0027. \end{aligned}$$

For $\{A_2, A_3\}$ with $P_2 = 3/4, P_3 = 1/4$, the MGE value is:

$$\begin{aligned} \text{MGE}(\{A_2, A_3\}) &= -\left(\frac{3}{4} \log \frac{3}{4} + \frac{1}{4} \log \frac{1}{4}\right) \\ &\quad + \left(\frac{3}{4} \times 0.4055 + \frac{1}{4} \times 1.0986\right) \\ &\approx 1.1411. \end{aligned}$$

The combination $\{A_1, A_2\}$ achieves the minimal MGE value, which is selected as the optimal granular set.

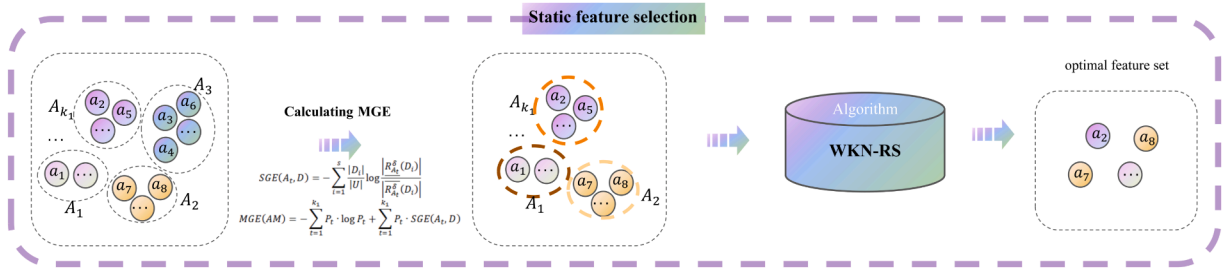


Fig. 3. The flowchart of static feature selection.

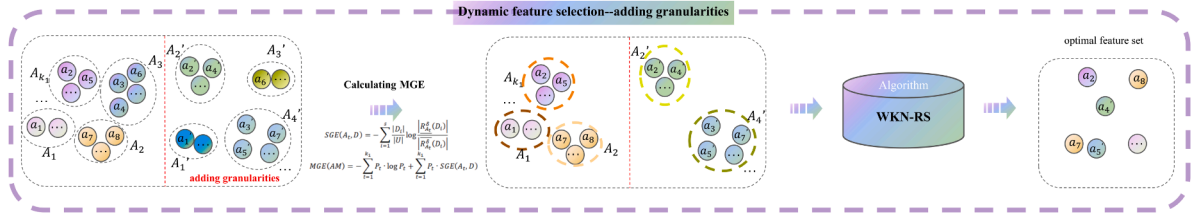


Fig. 4. The flowchart of dynamic feature selection with granularity addition.

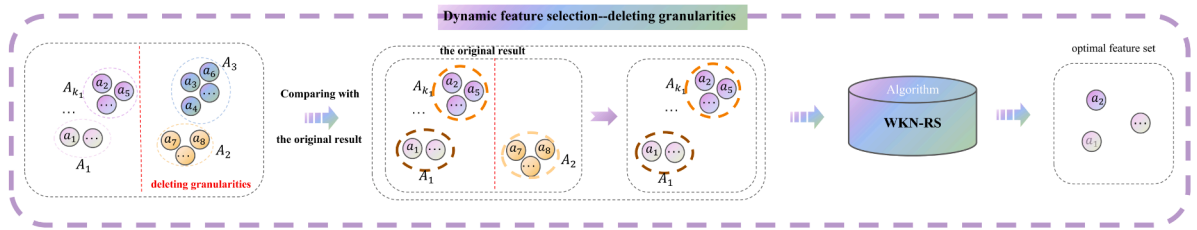


Fig. 5. The flowchart of dynamic feature selection with granularity deletion.

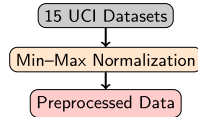


Fig. 6. Data preprocessing framework.

3.3. Feature selection algorithm based on MGE and WKN-RS (K-FSAMGE)

We first eliminate redundant granularities using MGE, and then apply a feature selection algorithm based on WKN-RS to identify the optimal feature subset from the retained granularities.

Definition 4. Let $NIS = (U, AT \cup D)$, where $A_t \subseteq AT$ ($t = 1, 2, \dots, k_1$), k_1 denotes the number of clusters, and k_2 denotes the number of retained granularities. The optimal granularity set is defined as

$$AM^* = \arg \min_{\substack{AM \subseteq \bigcup_{t=1}^{k_1} A_t, \\ |AM| \leq k_2}} MGE(AM). \quad (25)$$

Proposition 2 (Equivalence between MGE minimization and dependency maximization). *Minimizing $MGE(AM)$ is equivalent to maximizing the fuzzy neighborhood dependency degree $K\gamma_{AM}(D)$ in WKN-RS:*

$$AM^* = \arg \max_{AM} K\gamma_{AM}(D). \quad (26)$$

Proof.

- (1) Baseline relation: For any A_t , smaller $SGE(A_t, D)$ corresponds to tighter bounds of lower and upper approximations and stronger discriminative ability to U/D . A smaller SGE directly yields a higher $K\gamma_{A_t}(D)$.
- (2) Weighted aggregation for non-degenerate weights: MGE aggregates per-granularity uncertainty weighted by P_t . Minimizing global $MGE(AM)$ prioritizes retaining granules with low SGE , which elevates the global $K\gamma_{AM}(D)$.

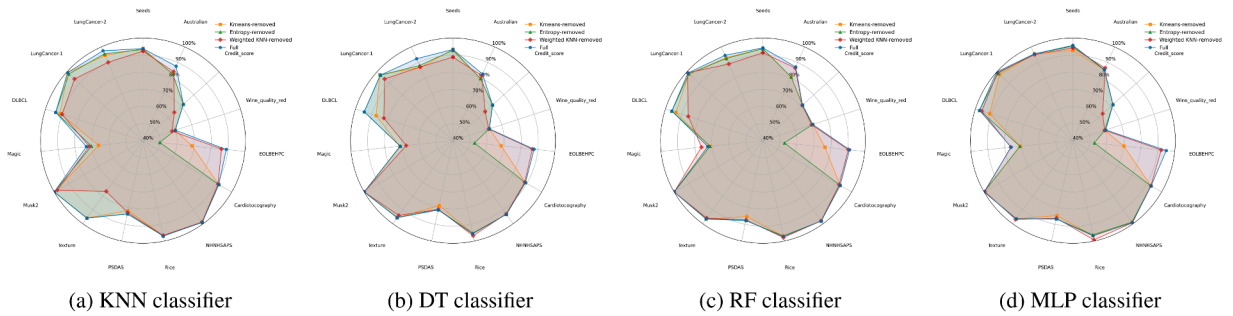


Fig. 7. The classification accuracy of four models under different classifiers.

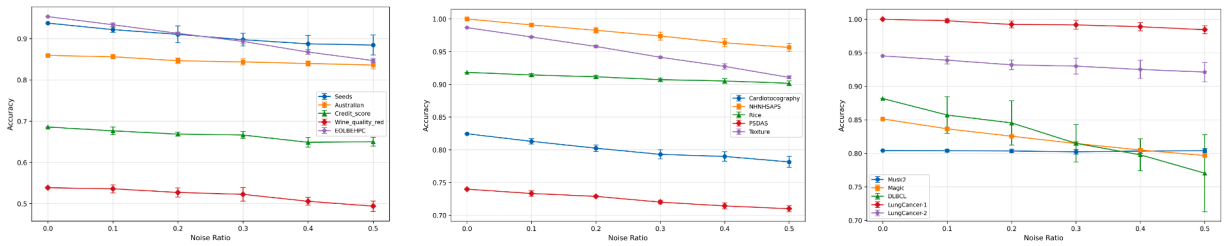


Fig. 8. The classification accuracy of K-FSAMGE under different noise proportions.

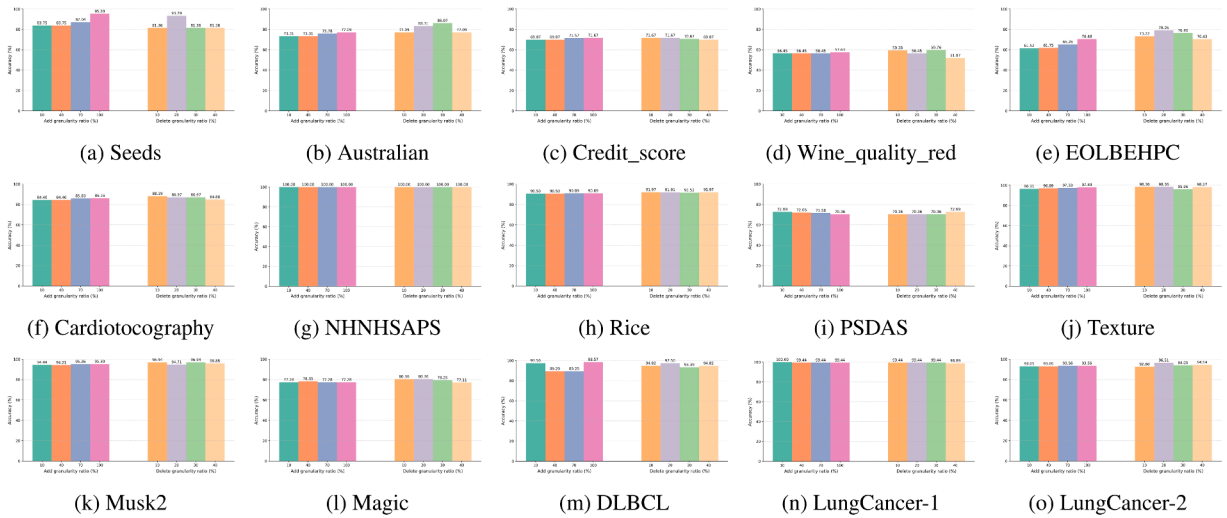


Fig. 9. The classification accuracy of K-FSAMGE when adding or deleting different granularity under KNN Classifier.

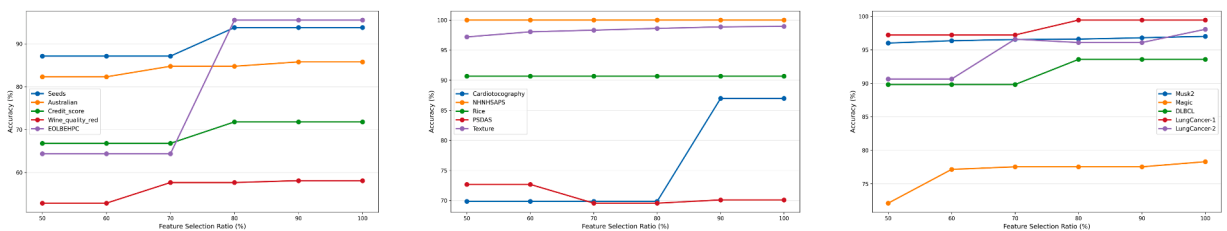


Fig. 10. The classification accuracy of different feature selection ratio.

Table 4
Datasets description.

NO.	Datasets	Samples	Features	Class
1	Seeds	210	7	3
2	Australian	690	14	2
3	Credit_score	1000	84	2
4	Wine quality-red	1599	11	6
5	EOLBEHPC	2111	16	2
6	Cardiotocography	2126	20	3
7	NHNHSAPS	2278	9	2
8	Rice	3810	7	2
9	PSDAS	4424	36	2
10	Texture	5500	40	11
11	Musk2	6598	166	2
12	Magic	19,020	10	2
13	DLBCL	77	7129	2
14	LungCancer-1	181	12,533	2
15	LungCancer-2	203	12,600	5

Table 5
Parameter settings of K-FSAMGE.

Parameter	Description	Range	Step
k_1	Number of clusters	$[2, F_{total})$	1
k_2	Number of retained granules	$[1, k_1]$	1
k_3	Number of nearest neighbors	$[1, 15]$	1
δ	Neighborhood radius	$[0.1, 1]$	0.1
α	Similarity threshold	$[0.1, 1]$	0.1

- (3) Degenerate granularity weights: If weights collapse to a degenerate distribution, MGE degenerates to $SGE(A_{t^*}, D)$. Minimizing this scalar value still maximizes $K\gamma_{A_{t^*}}(D)$, matching the equivalence rule. Near-degenerate weight ratios follow identical monotonic mapping.
- (4) Maximal SGE with poor discriminability: For a granule with $SGE(A_t, D) = SGE_{\max}(A_t, D)$ but negligible class separation power, its large SGE value heavily penalizes the total MGE sum. The minimization objective automatically suppresses high weights for such low-discrimination granules, preventing spurious elevation of global dependency estimates.

Therefore, minimizing $MGE(AM)$ is equivalent to maximizing $K\gamma_{AM}(D)$. $\square$

Definition 5. Given an $IS = (U, AT \cup D)$, where $B \subseteq AT$, the significance of decision attribute D with respect to the attribute subset B is given by

$$\text{Sig}(a, B, D) = K\gamma_{B \cup \{a\}}(D) - K\gamma_B(D), \quad (27)$$

where $K\gamma_B(D)$ denotes the fuzzy neighborhood dependency of D on B .

We propose a feature selection approach, termed K-FSAMGE (Algorithm 1), which integrates MGE with WKN-RS. Using the evaluation criteria described in Definitions 4 and 5, this algorithm identifies the optimal feature subset from the original feature set.

In K-FSAMGE, we first compress and filter all initial granularity sets according to Definition 4. Among all possible granularity combinations, the set containing no more than k_2 granularities and achieving the minimal MGE value is selected. This ensures that the retained structure is highly informative and minimally redundant.

Next, we evaluate the significance of features in the retained granularities based on Definition 5. This definition quantifies the contribution of a feature to the current subset through fuzzy neighborhood dependency. A higher significance value indicates that the feature is more important for improving classification performance. The flowchart of static feature selection is shown in Fig. 3.

3.4. Dynamic feature selection algorithm based on MGE

In the era of big data, processing rapidly evolving information with static methods is highly time-consuming. Therefore, we employ MGE to support incremental and decremental granularity updating, which significantly improves computational efficiency for dynamic data environments.

3.4.1. Updating mechanism of K-FSAMGE when adding granularities (K-FSAMGEA)

To formalize the process of updating the granularity set when new granularities are introduced, we first define the optimal granularity set after addition.

Definition 6. Given a $NIS = (U, AT \cup D)$, let the newly generated granularity set be $AT' = \{A_1, A_2, \dots, A_t\}$, with k'_1 representing the updated number of clusters and k'_2 denoting the updated number of retained granularities. The optimal granularity set after adding

Table 6
Classification accuracy of different clustering algorithms under four classifiers.

Datasets	KNN				DT				RF				MLP			
	k-means		Density-based		k-means		Hierarchical		Density-based		k-means		Density-based		k-means	
	Hierarchical	Density-based	Hierarchical	Density-based	k-means	Hierarchical	Density-based	Hierarchical	Density-based	k-means	Hierarchical	Density-based	k-means	Hierarchical	Density-based	Density-based
Seeds	93.81±5.65	93.81±5.65	93.81±5.65	93.81±5.65	93.33±4.86	93.33±4.86	93.33±4.86	93.33±4.86	93.33±4.86	94.29±4.67	94.29±4.67	94.29±4.67	95.71±3.96	95.71±3.96	95.71±3.96	95.71±3.96
Australian	87.68±3.13	82.75±3.97	85.22±3.98	82.75±3.97	82.61±6.31	82.35±5.33	79.71±4.00	82.35±5.33	79.71±4.00	87.22±2.83	82.81±5.68	80.72±4.05	85.51±4.30	85.94±5.50	85.36±5.04	85.36±5.04
Credit_score	71.80±2.14	71.80±2.14	71.80±2.14	71.80±2.14	71.20±2.60	71.20±2.60	71.20±2.60	71.20±2.60	71.20±2.60	71.20±2.60	71.20±2.60	71.20±2.60	71.60±2.84	71.60±2.84	71.60±2.84	71.60±2.84
Wine_quality_red	59.85±2.90	59.85±2.90	59.85±2.90	59.85±2.90	62.42±3.16	62.42±3.16	62.42±3.16	62.42±3.16	62.42±3.16	70.67±3.81	70.67±3.81	70.67±3.81	60.29±2.13	59.60±2.82	59.60±2.82	59.60±2.82
EOLBEHPC	95.97±1.66	95.97±1.66	95.97±1.66	95.97±1.66	95.66±1.49	95.66±1.49	95.66±1.49	95.66±1.49	95.66±1.49	96.61±1.01	96.61±1.01	96.61±1.01	96.40±1.49	96.40±1.49	96.40±1.49	96.40±1.49
Cardiotocography	89.04±1.96	84.34±1.47	80.95±1.47	84.34±1.47	87.77±1.59	82.68±2.88	88.26±1.77	82.68±2.88	88.26±1.77	91.05±2.14	85.04±2.47	89.85±2.06	95.01±0.53	88.05±2.14	94.69±1.33	94.69±1.33
NHNHSAPS	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
Rice	91.26±1.81	91.26±1.81	90.68±1.53	91.26±1.81	89.03±1.43	88.56±1.49	89.03±1.43	88.56±1.49	89.03±1.43	92.20±0.99	88.57±1.50	92.20±0.99	92.91±1.28	92.91±1.28	92.78±1.54	92.78±1.54
PSDAS	73.33±2.70	66.14±6.57	66.14±6.57	66.14±6.57	71.04±1.87	71.25±2.04	71.25±2.04	71.25±2.04	71.25±2.04	72.48±1.78	71.25±2.06	71.25±2.06	76.36±2.09	71.11±2.39	71.11±2.39	71.11±2.39
Texture	99.04±0.30	99.04±0.30	99.04±0.30	99.04±0.30	93.25±0.98	93.25±0.98	93.25±0.98	93.25±0.98	93.25±0.98	98.09±0.56	98.09±0.56	98.09±0.56	99.27±0.24	98.85±0.35	98.85±0.35	98.85±0.35
Musk2	97.03±0.62	96.27±0.77	96.83±0.75	96.83±0.75	95.66±0.58	94.94±0.57	95.38±1.07	94.94±0.57	95.38±1.07	97.29±0.56	96.72±0.71	97.40±0.63	96.89±0.48	94.73±1.00	96.26±0.95	96.26±0.95
Magic	83.77±0.63	83.77±0.82	83.77±0.63	83.77±0.63	81.40±0.68	80.91±0.42	81.40±0.68	80.91±0.42	81.40±0.68	87.75±0.45	87.36±0.45	87.75±0.45	86.75±0.42	86.42±0.63	86.75±0.42	86.75±0.42
DLBCL	93.75±8.39	93.75±8.39	93.75±8.39	93.75±8.39	94.64±9.18	93.14±10.32	80.71±11.89	93.14±10.32	80.71±11.89	96.07±8.21	95.32±7.81	80.82±11.10	97.50±5.00	93.57±8.53	78.93±12.42	78.93±12.42
LungCancer-1	99.47±1.58	98.89±2.22	98.89±2.22	98.89±2.22	97.25±2.75	98.44±2.48	97.25±2.75	98.44±2.48	97.25±2.75	99.11±1.95	98.78±2.24	98.33±2.55	100.00±0.00	100.00±0.00	98.92±2.16	98.92±2.16
LungCancer-2	97.57±4.42	95.55±3.49	97.57±4.42	95.55±3.49	92.45±5.00	92.45±5.00	86.69±10.21	92.45±5.00	86.69±10.21	94.59±5.14	94.59±5.14	92.81±4.54	95.55±2.69	95.55±2.69	95.55±2.69	95.55±2.69

Table 7

Standard deviation of three entropy metrics under different noise ratios.

Noise(%)	Seeds			Australian			Credit score			Wine_quality_red			EOLBEHPC			Cardiotocography		
	IE	CIE	MGE	IE	CIE	MGE	IE	CIE	MGE	IE	CIE	MGE	IE	CIE	MGE	IE	CIE	MGE
2.5	1.9367	0.3834	0.6423	6.4685	0.3150	0.4593	9.9372	0.0070	0.2491	1.8038	1.4612	1.1452	7.2411	0.4808	1.4060	5.0670	0.7790	0.8129
5.0	1.9367	0.3834	0.6423	6.4685	0.3150	0.4593	9.9562	0.0040	0.1759	1.8038	1.4612	1.1452	7.2411	0.4808	1.4060	6.3444	0.7527	0.8163
7.5	1.9367	0.3834	0.6423	7.3452	0.2492	0.4549	9.9638	0.0020	0.1106	1.8038	1.4612	1.1452	7.9027	0.5200	1.4402	6.3261	0.7543	0.8163
10.0	1.9367	0.3834	0.6423	6.8550	0.4634	0.5587	9.9638	0.0000	0.1255	2.8810	1.4650	1.1477	7.6881	0.5476	1.4245	6.9949	0.7107	0.8113
12.5	1.9367	0.3834	0.6423	6.7256	0.3029	0.4586	9.9658	0.0000	0.0271	2.8942	1.4447	1.1455	8.9608	0.3494	1.4064	6.7359	0.7352	0.8123
15.0	2.7774	0.3766	0.6442	7.3076	0.2784	0.4735	9.9658	0.0000	0.0129	3.0125	1.4288	1.1429	8.5918	0.4810	1.4314	7.7554	0.7079	0.8144
17.5	2.7127	0.3832	0.6467	8.0079	0.1887	0.4515	9.9658	0.0000	0.0560	2.9062	1.4355	1.1448	8.4964	0.4519	1.4202	8.0952	0.7041	0.8269
20.0	2.7578	0.4556	0.6685	7.7276	0.2151	0.4647	9.9658	0.0000	0.0078	4.1640	1.3892	1.1391	9.3866	0.3243	1.4258	8.8702	0.5143	0.7912
22.5	2.7446	0.4762	0.6696	7.9167	0.2192	0.4471	9.9658	0.0000	0.0010	4.0972	1.4167	1.1368	9.1812	0.5047	1.5115	9.1496	0.5125	0.8112
25.0	2.8021	0.3906	0.6521	7.8966	0.1925	0.4663	9.9658	0.0000	0.0021	3.5624	1.5270	1.1558	9.6912	0.4070	1.4760	9.6635	0.4514	0.7959
Std	0.4117	0.0335	0.0103	0.5766	0.0782	0.0306	0.0086	0.0023	0.0813	0.8419	0.0347	0.0048	0.8389	0.0692	0.0324	1.3846	0.1143	0.0097
Noise(%)	NHNHSAPS			Rice			PSDAS			Texture			Musk2			Magic		
	IE	CIE	MGE	IE	CIE	MGE	IE	CIE	MGE	IE	CIE	MGE	IE	CIE	MGE	IE	CIE	MGE
2.5	2.8096	0.2940	0.3649	1.2678	0.4063	0.6083	10.6091	0.3915	0.7591	4.7697	0.8485	1.5760	12.5702	0.0000	0.0053	1.7596	0.7170	0.6108
5.0	2.8096	0.2940	0.3649	1.2678	0.4063	0.6083	11.1238	0.2855	0.7475	6.0370	0.7988	1.5695	12.6816	0.0000	0.0012	1.7596	0.7170	0.6108
7.5	2.8096	0.2940	0.3649	1.2678	0.4063	0.6083	11.5849	0.1683	0.7360	7.3192	0.7531	1.5432	12.6875	0.0000	0.0005	1.7596	0.7170	0.6108
10.0	2.8096	0.2940	0.3649	1.2678	0.4063	0.6083	11.8078	0.0961	0.7203	8.6474	0.6677	1.5267	12.6878	0.0000	0.0005	2.6831	0.7133	0.6093
12.5	3.9075	0.2776	0.3651	1.2678	0.4063	0.6083	11.9586	0.0469	0.6619	9.7750	0.5568	1.5413	12.6878	0.0000	0.0000	2.9768	0.7058	0.6066
15.0	3.9183	0.2801	0.3655	2.3562	0.4048	0.6131	11.9335	0.0563	0.7114	10.8026	0.4308	1.4929	12.6878	0.0000	0.0000	2.8361	0.7126	0.6091
17.5	3.0824	0.2759	0.3621	2.2943	0.4461	0.6278	12.0489	0.0221	0.6588	11.5979	0.2264	1.4803	12.6878	0.0000	0.0000	2.1626	0.8638	0.6318
20.0	3.0810	0.2695	0.3618	2.3484	0.3875	0.6061	12.0758	0.0115	0.5685	12.0473	0.1098	1.3748	12.6878	0.0000	0.0000	3.0673	0.8619	0.6308
22.5	4.1648	0.2559	0.3618	2.2851	0.4449	0.6286	12.0944	0.0071	0.6092	12.2592	0.0477	1.3678	12.6878	0.0000	0.0000	3.6495	0.7187	0.6107
25.0	4.2752	0.2777	0.3662	2.3097	0.4240	0.6059	12.0979	0.0063	0.5707	12.3694	0.0168	1.2675	12.6878	0.0000	0.0000	3.8696	0.7199	0.6120
Std	0.5885	0.0122	0.0016	0.5258	0.0178	0.0082	0.4727	0.1262	0.0681	2.6138	0.3077	0.0979	0.0351	0.0000	0.0016	0.7361	0.0592	0.0086

Table 8

The feature selection time and memory usage between K-FSAMGE and K-FSAMGEA when adding different proportions of granularity.

Datasets	10%				40%				70%				100%			
	K-FSAMGE		K-FSAMGEA		K-FSAMGE		K-FSAMGEA		K-FSAMGE		K-FSAMGEA		K-FSAMGE		K-FSAMGEA	
	Time(s)	Memory(MB)	Time(s)	Memory(MB)	Time(s)	Memory(MB)	Time(s)	Memory(MB)	Time(s)	Memory(MB)	Time(s)	Memory(MB)	Time(s)	Memory(MB)	Time(s)	Memory(MB)
Seeds	0.08	0.16	0.03	0.14	0.10	0.21	0.04	0.02	0.14	0.06	0.06	0.04	0.19	0.09	0.07	0.05
Australian	0.54	0.35	0.43	0.09	0.62	0.36	0.47	0.02	0.93	0.13	0.45	0.01	1.39	0.09	0.55	0.02
Credit_score	3.01	0.91	2.77	0.02	3.32	0.41	2.60	0.04	7.87	0.46	2.62	0.04	13.28	0.98	2.72	0.02
Wine_quality_red	1.01	0.51	0.83	0.14	1.77	0.47	0.88	0.02	2.62	0.30	0.80	0.01	3.01	0.06	1.10	0.02
EOLBEHPC	2.33	3.20	1.76	0.06	4.99	1.03	2.53	0.05	3.25	0.22	2.52	0.13	3.98	2.16	3.79	0.62
Cardiotocography	11.93	4.87	0.87	0.07	12.47	1.12	1.07	0.04	14.09	1.41	5.41	0.04	14.61	0.94	13.54	0.03
NHNHSAPS	0.97	0.37	0.68	0.19	1.09	0.29	0.68	0.01	1.52	0.28	0.69	0.02	1.81	0.49	0.95	0.04
Rice	1.89	0.70	0.95	0.22	2.30	0.28	0.89	0.14	3.38	0.17	1.82	0.01	3.28	0.15	2.24	0.02
PSDAS	376.59	7.56	5.96	1.31	494.94	3.24	12.10	2.29	538.05	5.86	9.95	3.02	479.46	6.64	14.50	3.55
Texture	28.76	3.35	22.06	1.35	42.29	1.29	29.43	0.29	52.25	1.43	40.15	0.45	89.80	2.30	64.44	0.11
Musk2	396.56	0.49	356.85	0.24	592.21	75.33	382.67	46.82	1057.80	31.87	510.94	4.96	809.88	13.37	416.61	0.63
Magic	27.49	2.27	4.15	0.97	32.16	3.87	7.84	1.99	45.36	4.62	29.63	3.20	57.25	4.91	29.63	3.76
DLBCL	240.13	0.18	137.81	0.09	379.19	2.15	138.17	1.50	519.72	3.45	244.13	0.06	518.85	1.07	286.06	1.02
LungCancer_H	3719.04	5.22	663.57	0.62	4838.89	2.86	1666.83	2.84	5659.81	1.66	2600.44	0.84	6786.10	1.46	3521.85	0.39
LungCancer_I	625.56	84.64	569.91	1.33	666.70	58.48	468.49	0.84	813.03	0.92	492.55	0.54	1001.80	4.65	480.72	0.78
Average	362.39	7.65	117.91	0.46	471.54	10.09	180.98	3.79	581.32	3.52	262.81	0.89	652.31	2.62	322.58	0.74

Table 9
The feature selection time and memory usage between K-FSAMGE and K-FSAMGED when deleting different proportions of granularity.

Datasets	10%				20%				30%				40%			
	K-FSAMGE		K-FSAMGED		K-FSAMGE		K-FSAMGED		K-FSAMGE		K-FSAMGED		K-FSAMGE		K-FSAMGED	
	Time(s)	Memory(MB)	Time(s)	Memory(MB)	Time(s)	Memory(MB)	Time(s)	Memory(MB)	Time(s)	Memory(MB)	Time(s)	Memory(MB)	Time(s)	Memory(MB)	Time(s)	Memory(MB)
seeds	0.07	0.20	0.03	0.18	0.03	0.11	0.02	0.04	0.08	0.09	0.03	0.03	0.08	0.05	0.03	0.01
australian	1.61	0.40	0.75	0.20	1.48	0.09	0.79	0.05	0.94	0.11	0.38	0.02	1.54	0.09	0.70	0.02
credit score	13.37	2.54	12.98	0.90	11.02	1.01	10.51	0.93	4.34	0.39	4.18	0.23	2.04	0.05	1.86	0.03
wine quality-red	2.82	0.35	2.65	0.17	0.78	0.16	0.72	0.08	0.88	0.12	0.83	0.07	0.68	0.09	0.60	0.02
EOLBHEPC	2.53	1.81	2.36	0.67	3.53	0.45	1.04	0.06	1.72	3.09	0.46	0.04	1.72	0.63	0.30	0.07
card	3.12	0.56	1.13	0.04	4.16	0.55	1.36	0.07	4.08	0.34	1.39	0.03	2.27	0.23	0.22	0.05
NHNHSAPS	0.78	1.27	0.28	0.02	0.76	0.17	0.40	0.02	0.77	0.24	0.49	0.01	0.54	0.04	0.27	0.02
rice	9.41	0.66	1.06	0.61	5.49	0.07	1.31	0.01	2.81	0.09	0.94	0.01	2.69	0.07	0.65	0.03
PSDAS	15.35	2.45	6.76	1.21	12.44	4.11	4.95	1.08	11.79	3.70	6.25	1.62	9.02	2.90	1.82	1.00
texture	71.72	1.70	63.23	0.21	72.29	2.00	70.05	0.19	32.95	2.96	32.59	2.07	47.78	2.83	47.69	2.64
Musk	694.87	3.44	681.07	1.13	293.44	2.55	201.89	0.99	696.85	1.46	604.38	0.08	1146.18	5.38	1047.66	0.35
magic	48.79	2.17	46.63	1.65	53.66	2.84	50.61	2.83	20.94	3.28	19.45	0.17	25.73	2.84	24.98	1.21
DLBCL	88.84	0.33	85.19	0.04	88.89	0.55	88.01	0.02	4.16	0.42	1.91	0.02	1.92	0.04	1.87	0.01
LungCancerH	6244.03	98.96	228.13	0.21	6217.38	47.25	276.74	0.96	5964.52	59.05	279.15	24.38	6800.47	7.07	587.00	3.44
LungCancerI	1485.34	132.72	1177.53	63.28	123.35	2.80	99.54	0.11	137.26	0.12	101.35	0.03	971.31	0.30	815.51	0.03
Average	578.84	16.64	153.99	4.70	459.25	4.31	53.86	0.50	458.94	5.03	70.25	1.92	600.93	1.51	168.74	0.60

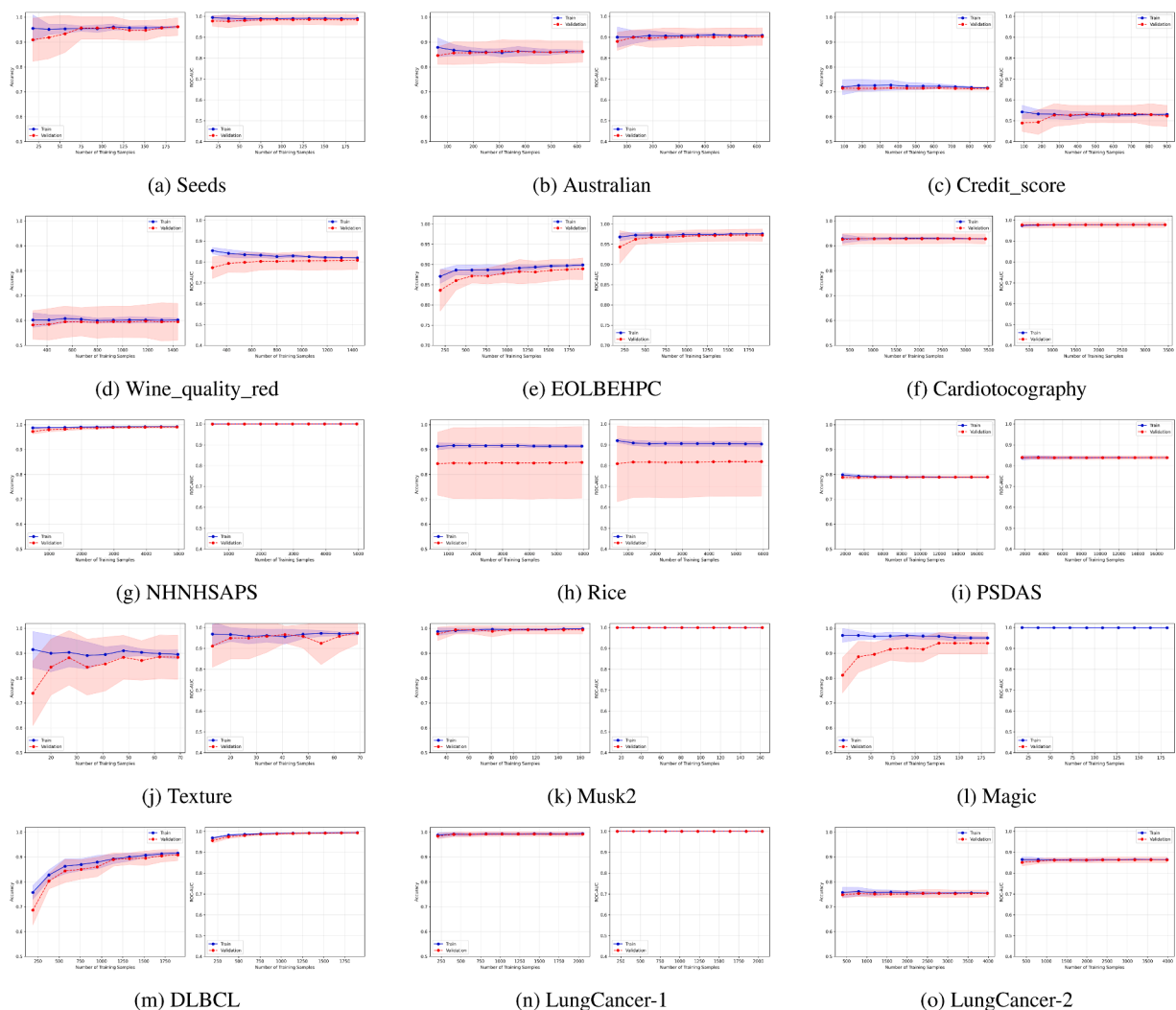


Fig. 11. Learning curves for accuracy and ROC-AUC.

Table 10
Running time of three component modules.

Datasets	Features	K-means	MGE	Weighted KNN
Seeds	7	0.4004	0.0362	0.0973
Rice	7	0.4875	0.0000	2.4496
NHHNSAPS	9	0.6894	12.6729	0.7114
Magic	10	0.3332	24.4537	17.0997
Wine_quality_red	11	0.3832	0.0000	2.5938
Australian	14	0.4260	1.6294	1.2215
EOLBEHPC	16	0.3253	0.8350	1.6871
Cardiotocography	20	0.3399	13.5564	1.4977
PSDAS	36	0.3534	393.5391	15.2264
Texture	40	0.7277	0.0000	201.1279
Credit_score	84	0.6148	0.0000	11.4105
Musk2	166	0.6016	316.6666	934.2559
DLBCL	7129	0.5940	1.0636	130.3878
LungCancer-1	12,533	0.8241	9.8747	341.7895
LungCancer-2	12,600	0.6771	8990.3667	1229.1319

Table 11

The classification accuracy of different algorithms based on KNN (%).

Datasets	RAW	FNGCE	IGUFS	KND-UFS	WKN-RS	CHCQX	PSO-MI	DCGA	K-FSAMGE
Seeds	93.33±4.36	93.33±4.36	86.19±6.88	92.38±3.16	93.33±4.36	91.90±6.04	92.86±5.32	91.90±6.04	93.81±5.65
Australian	82.75±5.00	84.78±4.59	83.04±5.61	84.78±4.41	82.75±3.97	83.19±4.06	84.49±4.35	86.81±3.86	87.68±3.13
Credit_score	64.70±3.87	64.70±3.35	64.60±3.50	65.10±5.45	71.80±2.14	63.60±5.16	65.30±3.38	65.30±3.52	71.80±2.14
Wine quality-red	57.91±2.56	57.91±2.63	54.85±3.38	54.78±2.77	59.85±2.90	56.41±2.89	55.85±2.51	56.72±4.44	59.85±2.90
EOLBEHPC	78.83±3.42	77.22±3.49	60.35±2.48	82.28±3.81	95.97±1.66	80.10±3.42	79.02±3.26	91.95±2.10	95.97±1.66
Cardiotocography	81.18±2.81	81.23±3.01	72.43±3.25	80.24±3.28	49.87±6.12	85.56±3.46	85.04±1.76	82.64±3.02	89.04±1.96
NNHSAPS	96.58±1.35	81.56±1.21	80.16±1.59	84.02±1.39	100.00±0.00	98.46±0.63	100.00±0.00	100.00±0.00	100.00±0.00
Rice	91.05±1.16	90.92±1.16	91.36±1.51	91.47±1.18	91.26±1.81	91.05±1.59	91.26±1.81	91.57±1.43	91.26±1.81
PSDAS	65.69±1.37	66.21±1.79	55.97±2.50	64.69±1.92	70.32±2.19	68.38±1.63	72.20±2.29	71.09±1.71	73.33±2.70
Texture	98.82±0.37	98.82±0.40	97.75±0.64	97.51±0.71	99.04±0.30	98.85±0.29	98.95±0.48	98.75±0.43	99.04±0.30
Musk2	96.57±0.87	96.59±0.88	95.60±0.58	93.89±0.50	96.83±0.75	96.30±0.91	95.83±0.97	96.59±0.88	97.03±0.62
Magic	83.14±0.71	81.43±1.12	75.70±0.96	79.93±1.01	83.77±0.63	83.05±0.51	81.01±0.67	83.77±0.82	83.77±0.63
DLBCL	89.82±12.30	89.82±12.30	89.82±12.30	80.54±12.86	93.75±8.39	N/A	88.57±11.82	81.07±17.84	93.75±8.39
LungCancer-1	93.39±4.10	N/A	N/A	84.01±10.61	98.33±2.55	N/A	93.39±4.10	99.44±1.67	99.47±1.58
LungCancer-2	90.62±4.68	N/A	N/A	67.93±4.83	95.55±3.49	N/A	89.64±5.10	90.62±5.96	97.57±4.42
Average	84.29±3.26	81.89±3.10	77.52±3.48	80.24±3.86	85.49±2.75	83.07±2.55	84.89±3.19	85.88±3.58	88.89±2.53

Table 12

The classification accuracy of different algorithms based on DT (%).

Datasets	RAW	FNGCE	IGUFS	KND-UFS	WKN-RS	CHCQX	PSO-MI	DCGA	K-FSAMGE
Seeds	89.05±6.75	89.05±6.75	88.57±9.81	90.00±5.41	92.57±5.12	90.10±7.08	87.81±6.78	90.57±5.15	93.33±4.86
Australian	81.74±2.84	82.32±4.48	83.77±4.66	81.74±5.43	79.71±4.00	84.78±3.79	84.20±4.32	82.14±4.92	82.61±6.31
Credit_score	65.40±5.75	65.20±4.45	61.80±4.71	60.50±3.41	71.20±2.60	62.20±3.25	63.74±3.56	59.66±3.89	71.20±2.60
Wine quality-red	62.04±3.61	63.98±2.59	61.41±3.33	61.04±2.62	62.42±3.16	61.41±3.36	61.22±3.21	62.54±3.17	62.42±3.16
EOLBEHPC	93.74±1.74	93.74±2.25	67.63±2.16	94.80±1.43	95.56±1.58	95.20±1.57	94.03±1.52	94.96±1.41	95.66±1.49
Cardiotocography	85.84±2.26	85.42±2.49	74.46±2.41	82.55±2.71	52.68±2.32	86.41±1.67	87.72±1.59	86.32±2.26	87.77±1.59
NNHSAPS	100.00±0.00	77.54±2.68	72.67±2.62	84.02±1.39	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
Rice	88.92±1.62	88.90±1.81	88.92±1.17	88.87±1.93	89.03±1.43	88.87±1.86	88.75±1.22	88.77±1.10	89.03±1.43
PSDAS	68.95±1.96	68.90±2.12	49.74±1.01	64.82±2.63	71.04±1.87	68.15±1.88	68.05±1.67	67.70±1.85	71.04±1.87
Texture	92.98±1.45	92.75±1.30	91.09±0.93	90.35±1.47	93.25±0.98	92.35±0.88	92.58±0.92	92.73±1.41	93.25±0.98
Musk2	94.85±1.00	96.70±0.79	96.04±0.74	94.82±0.75	95.36±1.06	95.99±0.58	96.32±0.91	96.92±0.61	95.66±0.58
Magic	80.44±1.06	79.04±0.73	73.56±0.86	76.54±1.12	81.18±0.75	80.59±0.92	79.30±0.81	80.86±0.45	81.40±0.68
DLBCL	87.14±11.62	84.18±13.02	89.64±5.22	76.82±15.36	94.64±9.18	N/A	85.36±12.03	84.96±8.91	94.64±9.18
LungCancer-1	94.53±4.78	N/A	N/A	81.93±10.94	97.25±2.75	N/A	95.67±6.16	94.02±4.51	97.25±2.75
LungCancer-2	86.14±7.67	N/A	N/A	61.45±6.02	88.12±7.80	N/A	89.67±5.60	88.72±7.91	92.45±5.00
Average	85.45±3.81	82.27±3.73	79.49±3.51	81.48±4.28	84.47±2.84	84.59±2.93	85.03±2.89	84.92±3.24	87.87±2.39

granularities is defined as

$$(AM')^* = AM^* \cup \arg \min_{\substack{AM' \subseteq \bigcup_{i=1}^{k_1'} A_i, \\ |AM'| \leq k_2}} \text{MGE}(AM'). \quad (28)$$

Proposition 3 (Monotonicity of dependency). *Given the original optimal granularity set AM^* and the MGE-screened new granular subset AM' , the updated optimal set is $(AM')^* = AM^* \cup AM'$. The fuzzy neighborhood positive region and dependency degree satisfy the following monotonic relation:*

$$KPOS_{AM^*}(D) \subseteq KPOS_{(AM')^*}(D), \quad (29)$$

$$K\gamma_{AM^*}(D) \leq K\gamma_{(AM')^*}(D). \quad (30)$$

Proof. The newly screened granularities supplement valid neighborhood information and refine the multigranulation structure. All samples classified as positive in AM^* remain positive after incorporating additional qualified granules. Accordingly, the positive region is non-shrinking, which further guarantees the non-decreasing property of the fuzzy neighborhood dependency degree. Therefore, granularity addition maintains strict monotonicity of rough set approximation. $\square$

Proposition 4 (Approximation Consistency). *Since newly added granularities are strictly screened via MGE minimization, only granules with low uncertainty and strong discriminability are retained. The updated knowledge rough approximations satisfy:*

$$\underline{KN}_{AM^*}(D) \subseteq \underline{KN}_{(AM')^*}(D), \quad (31)$$

$$\overline{KN}_{(AM')^*}(D) \subseteq \overline{KN}_{AM^*}(D). \quad (32)$$

Table 13

The classification accuracy of different algorithms based on RF (%).

Datasets	RAW	FNGCE	IGUFS	KND-UFS	WKN-RS	CHCQX	PSO-MI	DCGA	K-FSAMGE
Seeds	91.90±3.72	91.90±3.72	91.43±6.67	91.90±4.29	93.81±4.29	93.81±4.29	92.38±5.83	94.10±4.77	94.29±4.67
Australian	86.38±4.31	86.81±4.32	83.62±4.63	85.65±3.97	80.72±4.05	84.64±4.06	84.38±4.34	86.20±4.18	87.22±2.83
Credit_score	71.00±3.49	70.90±3.45	70.60±3.01	72.00±3.58	71.20±2.60	71.10±3.51	71.26±3.96	70.12±3.70	71.20±2.60
Wine quality-red	70.05±3.20	69.30±2.89	66.98±3.30	66.73±2.87	70.67±3.81	68.73±3.80	67.86±3.75	69.12±3.53	70.67±3.81
EOLBEHPC	95.41±1.41	94.71±1.43	72.26±2.30	95.04±1.53	96.49±1.17	94.72±1.97	95.11±1.36	95.02±1.72	96.61±1.01
Cardiotocography	89.13±2.33	89.18±2.13	81.04±1.57	84.81±2.50	52.68±2.32	88.36±2.04	89.42±2.16	89.15±2.93	91.05±2.14
NNHNSAPS	100.00±0.00	84.29±1.12	72.74±2.39	84.02±1.39	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
Rice	92.13±1.74	92.05±1.74	92.44±1.43	92.18±1.70	92.20±0.99	92.07±1.39	91.92±1.17	92.24±1.51	92.20±0.99
PSDAS	77.89±1.81	77.76±1.65	57.02±1.82	75.31±1.78	71.12±1.94	75.94±1.94	75.37±1.48	74.82±1.53	72.48±1.78
Texture	97.95±0.65	97.53±0.45	96.84±0.56	96.02±0.52	98.09±0.56	97.67±0.66	98.05±0.53	97.85±0.55	98.09±0.56
Musk2	96.71±0.50	97.73±0.48	97.24±0.66	96.12±0.82	96.98±0.65	97.69±0.51	97.56±0.66	97.64±0.48	97.29±0.56
Magic	87.08±0.77	85.87±0.80	80.23±1.22	83.16±0.71	87.68±0.61	87.37±0.64	85.55±0.62	87.33±0.44	87.75±0.45
DLBCL	89.64±9.48	86.79±10.76	88.21±10.48	77.25±14.18	96.07±8.21	N/A	89.46±7.27	81.36±10.29	96.07±8.21
LungCancer-1	98.36±2.50	N/A	N/A	87.60±9.88	98.33±2.55	N/A	98.89±2.22	98.89±2.22	99.11±1.95
LungCancer-2	91.64±5.34	N/A	N/A	69.33±6.62	92.64±5.01	N/A	91.14±5.69	91.11±4.93	94.59±5.14
Average	89.55±2.62	86.52±2.62	82.91±2.92	84.14±3.45	86.40±2.55	87.65±2.89	88.09±2.90	87.85±3.05	89.95±2.49

Table 14

The classification accuracy of different algorithms based on MLP (%).

Datasets	RAW	FNGCE	IGUFS	KND-UFS	WKN-RS	CHCQX	PSO-MI	DCGA	K-FSAMGE
Seeds	94.76±4.97	94.76±4.97	88.10±6.48	94.29±3.56	95.24±3.69	93.81±5.24	93.33±5.30	93.81±5.24	95.71±3.96
Australian	84.78±5.28	84.78±5.28	85.51±5.61	85.36±4.96	85.36±5.04	85.22±4.97	84.93±4.99	84.93±3.90	85.51±4.30
Credit_score	63.50±3.72	63.40±4.52	64.20±3.63	65.90±4.93	71.60±2.84	64.00±3.85	67.80±3.57	71.10±3.75	71.60±2.84
Wine quality-red	59.60±2.59	60.98±2.66	56.53±3.11	57.91±3.44	59.60±2.82	59.66±3.03	58.41±2.82	59.22±2.70	60.29±2.13
EOLBEHPC	96.16±2.36	96.16±1.78	61.92±3.40	94.31±8.15	96.40±1.49	96.40±2.34	96.30±1.64	95.36±1.50	96.40±1.49
Cardiotocography	94.31±2.08	94.31±2.08	72.48±3.53	90.17±2.02	52.68±2.32	89.37±1.71	93.98±2.03	94.50±2.01	95.01±0.53
NNHNSAPS	99.61±0.54	84.20±0.87	83.89±1.33	84.02±1.39	100.00±0.00	99.78±0.29	100.00±0.00	100.00±0.00	100.00±0.00
Rice	92.78±1.51	92.91±1.51	92.62±1.48	92.91±1.59	92.78±1.54	92.91±1.51	92.78±1.54	92.83±1.63	92.91±1.28
PSDAS	74.21±1.29	74.48±1.66	58.95±1.84	73.26±1.37	71.16±2.40	75.95±1.13	75.95±1.76	75.23±1.53	76.36±2.09
Texture	99.36±0.22	99.16±0.49	97.42±0.47	97.96±0.40	98.85±0.35	98.91±0.20	99.31±0.23	99.25±0.26	99.27±0.24
Musk2	99.52±0.27	99.58±0.21	96.89±0.42	94.09±0.97	96.26±0.95	98.61±0.53	97.92±0.66	98.15±0.55	96.89±0.48
Magic	87.23±0.74	84.81±0.73	79.00±0.90	83.14±0.71	86.75±0.42	87.26±0.73	83.28±0.59	86.42±0.63	86.75±0.42
DLBCL	96.07±6.02	96.07±6.02	96.07±6.02	85.89±11.78	97.50±5.00	N/A	96.07±6.02	97.50±5.00	97.50±5.00
LungCancer-1	99.44±1.67	N/A	N/A	88.45±8.27	98.92±2.16	N/A	99.44±1.67	99.44±1.67	100.00±0.00
LungCancer-2	93.57±4.91	N/A	N/A	72.40±6.40	95.55±4.71	N/A	94.55±3.51	95.55±3.43	95.55±2.69
Average	89.71±2.56	86.83±2.49	80.64±2.99	84.00±4.06	87.14±2.33	87.42±2.62	88.54±2.57	89.17±2.52	90.38±1.89

Proof. By the definition of MGE, a smaller entropy value corresponds to lower system uncertainty and more accurate $\underline{KN}$, $\overline{KN}$ approximation bounds. MGE-guided screening eliminates low-quality granules that cause structural disturbance. The incremental update only optimizes and refines the original approximation space rather than altering it arbitrarily, thus preserving approximation consistency.

According to Definition 6 and the above propositions, we introduce a feature addition selection algorithm based on MGE and WKN-RS, termed K-FSAMGEA (Algorithm 2). Different from naive set union operations, the proposed MGE-constrained incremental update theoretically guarantees monotonic dependency growth and stable approximation consistency. K-FSAMGEA initially applies MGE to screen newly generated granular features, subsequently merges the selected granularities with the existing granularity set, and ultimately identifies the optimal feature subset through WKN-RS. The flowchart of dynamic feature selection with granularity addition is illustrated in Fig. 4. $\square$

3.4.2. Updating mechanism of K-FSAMGE when deleting granularities (K-FSAMGED)

Similarly, to handle the removal of granularities, we introduce the following definition to formalize the updated optimal granularity.

Definition 7. Given a $NIS = (U, AT \cup D)$, suppose that a subset of granularities $AT' \subseteq AT$ is to be removed. The optimal granularity set after deleting granularities can then be expressed as

$$(AM')^* = AM^* - AT'. \quad (33)$$

Proposition 5 (Information consistency). Let $(AM')^* = AM^* - AT'$ denote the updated granular set after deletion. The positive region and dependency degree satisfy:

$$KPOS_{(AM')^*}(D) \subseteq KPOS_{AM^*}(D), \quad (34)$$

Table 15

The feature selection numbers of different algorithms.

Datasets	RAW	FNGCE	IGUFS	KND-UFS	WKN-RS	CHCQX	PSO-MI	DCGA	K-FSAMGE
Seeds	7	7	3	5	3	4	3	4	3
Australian	14	12	6	7	3	4	4	6	5
Credit_score	84	82	37	27	3	43	41	7	3
Wine quality-red	11	9	4	5	5	7	6	5	5
EOLBEHPC	16	14	7	6	3	8	15	6	3
Cardiotocography	20	18	9	10	1	11	7	17	4
NNHNSAPS	9	7	4	1	1	5	1	1	1
Rice	7	5	3	5	2	3	2	3	2
PSDAS	36	34	16	18	2	16	18	10	5
Texture	40	38	18	20	15	25	22	30	15
Musk2	166	164	74	41	25	89	75	89	24
Magic	10	8	7	5	6	8	5	5	5
DLBCL	7129	7127	3208	5	3	N/A	3480	930	3
LungCancer-1	12,533	N/A	N/A	7	4	N/A	6325	425	3
LungCancer-2	12,600	N/A	N/A	29	9	N/A	6349	4414	8

Table 16

The running time of different algorithms (s).

Datasets	FNGCE	IGUFS	KND-UFS	WKN-RS	CHCQX	PSO-MI	DCGA	K-FSAMGE
Seeds	0.48	2.18	0.23	0.22	54.61	32.33	19.66	0.53
AustralianV	9.79	11.75	1.66	1.36	98.78	38.41	19.20	3.28
Credit_scoreV	812.82	161.27	131.39	21.56	742.60	33.95	17.17	12.03
Wine_quality_red	91.19	30.75	13.65	3.88	180.77	50.79	23.72	2.98
EOLBEHPC	260.72	127.14	35.92	3.55	240.35	131.36	58.93	2.85
Cardiotocography	273.66	85.00	28.92	7.37	222.18	77.81	35.04	15.39
NNHNSAPS	61.67	45.64	22.99	0.95	87.74	40.05	74.05	12.91
Rice	118.51	58.55	49.33	2.43	330.89	62.39	114.12	2.94
PSDAS	2840.84	568.82	174.05	25.92	662.89	175.95	130.05	409.12
Texture	19460.05	1034.40	209.13	260.18	2241.84	179.87	126.70	201.86
Musk2	88368.61	8050.61	23211.89	2871.79	8269.99	1378.26	159.89	1251.52
Magic	4255.93	3091.75	1009.39	61.11	6284.64	560.93	277.08	41.89
DLBCL	31468.26	57287.77	36.07	100.51	N/A	96.71	135.41	132.04
LungCancer-1	N/A	N/A	307.32	414.22	N/A	223.60	104.84	352.49
LungCancer-2	N/A	N/A	844.60	1384.09	N/A	268.86	298.12	10220.18
Average	11386.35	5427.36	1738.44	343.94	1618.11	223.42	106.27	844.13

Table 17

Average ranking results of the Friedman test.

Classifier	FNGCE	IGUFS	KND-UFS	WKN-RS	CHCQX	PSO-MI	DCGA	K-FSAMGE
KNN	5.25	7.08	5.75	3.75	5.25	4.42	3.25	1.25
DT	4.33	6.25	6.67	3.83	4.08	4.83	4.17	1.83
RF	4.50	6.83	5.67	4.58	4.25	4.25	3.83	2.08
MLP	4.41	7.08	5.83	4.67	3.92	4.50	4.00	1.58

Table 18

The p-value of the Friedman test.

Classifier	χ^2	F	P
KNN	44.11	12.16	2.31×10^{-10}
DT	31.47	6.59	4.18×10^{-6}
RF	26.44	5.05	9.43×10^{-5}
MLP	35.17	7.92	3.27×10^{-7}

Table 19

The p-value of the Wilcoxon test.

Classifiers	FNGCE	IGUFS	KND-UFS	WKN-RS	CHCQX	PSO-MI	DCGA
KNN	<0.05	<0.05	<0.05	0.06	<0.05	<0.05	<0.05
DT	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05
RF	0.06	<0.05	<0.05	<0.05	0.12	0.14	0.14
MLP	0.06	<0.05	<0.05	<0.05	0.13	<0.05	<0.05

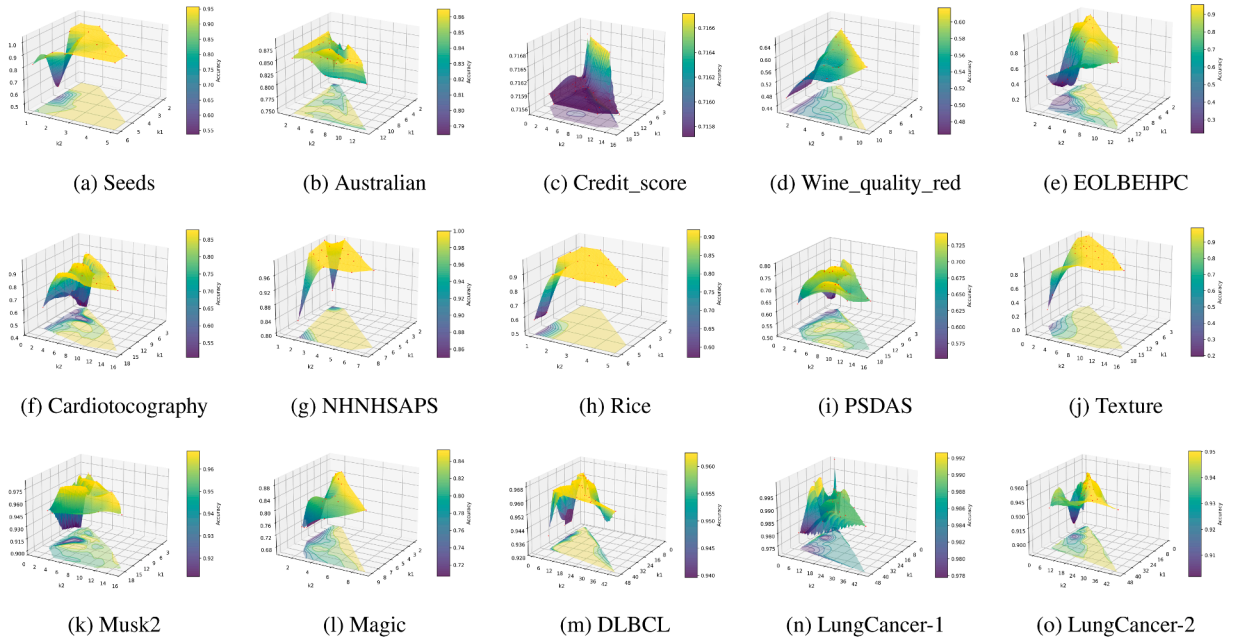


Fig. 12. The classification accuracy of different parameters k_1 and k_2 under KNN classifier.

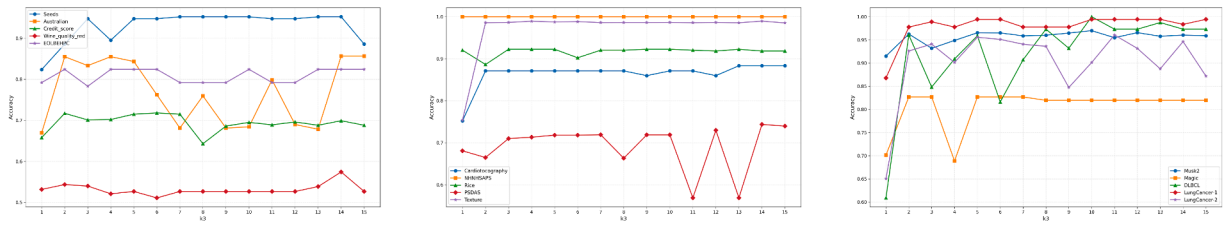


Fig. 13. The classification accuracy of different parameters k_3 under KNN classifier.

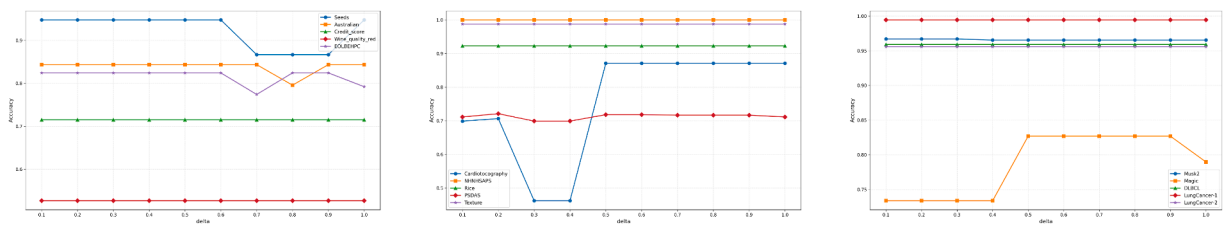


Fig. 14. The classification accuracy of different parameters δ under KNN classifier.

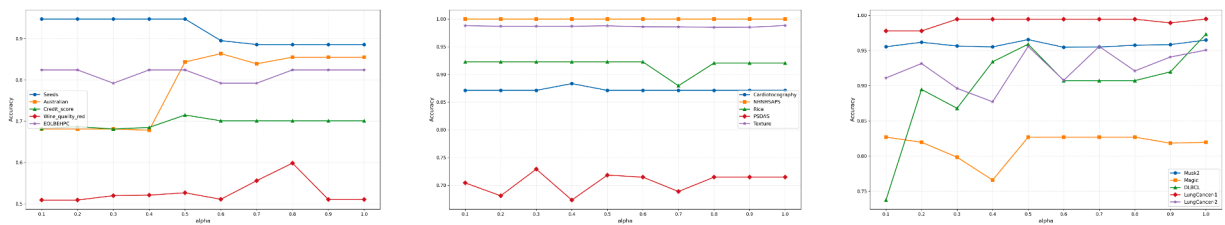


Fig. 15. The classification accuracy of different parameters α under KNN classifier.

Algorithm 1 Feature selection algorithm based on multigranulation entropy.

Input: A decision system $DIS = (U, A \cup \{d\})$, granules $\bigcup_{t=1}^{k_1} A_t$, threshold parameters $\delta \in [0.1, 1]$, $\alpha \in [0.1, 1]$, $k_1 \in [2, |A|]$, $k_2 \in [1, k_1]$, $k_3 \in [1, 15]$

Output: Selected feature subset S

```

1:  $AM \leftarrow \sum_{t=1}^{k_1} A_t$ 
2: if  $|AM| > k_2$  then
3:    $m \leftarrow MGE(AM)$ 
4:   while  $|AM| > k_2$  do
5:      $m_{best} \leftarrow m$ 
6:      $g_{best} \leftarrow \text{None}$ 
7:     for each  $g$  in  $AM$  do
8:        $AM' \leftarrow AM \setminus \{g\}$ 
9:       if  $AM' = \emptyset$  then continue
10:      end if
11:       $m' \leftarrow MGE(AM')$ 
12:      if  $m' < m_{best}$  then
13:         $m_{best} \leftarrow m'$ 
14:         $g_{best} \leftarrow g$ 
15:      end if
16:    end for
17:    if  $g_{best} = \text{None}$  then break
18:    end if
19:    Remove  $g_{best}$  from  $AM$ 
20:     $m \leftarrow m_{best}$ 
21:  end while
22: end if
23:  $F \leftarrow$  all features in  $AM$ 
24:  $S \leftarrow \emptyset$ 
25: while  $F \neq \emptyset$  do
26:    $Sig_{best} \leftarrow -\infty$ 
27:   for each  $a$  in  $F$  do
28:     compute  $SIM(x)$ 
29:     compute  $KPOS_S(U)$ 
30:     compute  $K\gamma_{S \cup \{a\}}(D)$  and  $K\gamma_S(D)$ 
31:     compute  $Sig(a, S, D)$ 
32:     if  $Sig(a, S, D) > Sig_{best}$  then
33:        $Sig_{best} \leftarrow Sig(a, S, D)$ 
34:        $f_{best} \leftarrow a$ 
35:     end if
36:   end for
37:   if  $Sig_{best} > 0$  then
38:      $S \leftarrow S \cup \{f_{best}\}$ 
39:     Remove  $f_{best}$  from  $F$ 
40:   else
41:     break
42:   end if
43: end while
44: return  $S$ 

```

$$K\gamma_{(AM')^*}(D) \leq K\gamma_{AM^*}(D). \quad (35)$$

In particular, if $AT' \cap AM^* = \emptyset$, then $K\gamma_{(AM')^*}(D) = K\gamma_{AM^*}(D)$, indicating that deleting redundant non-optimal granules causes no information loss.

Proof. Deleting granularities reduces the granular space, which may slightly shrink the positive region and leads to bounded non-increasing dependency. Nevertheless, the deletion operation strictly follows the original rough set approximation structure without introducing structural deviation. If the removed granules are not involved in the optimal set AM^* , the discriminative neighborhood structure remains completely unchanged, so no effective information is lost and approximation consistency is maintained.

Following Definition 7 and the proven theoretical properties, we propose a feature deletion selection algorithm by means of MGE and WKN-RS, named K-FSAMGED (Algorithm 3). K-FSAMGED updates the optimal granularity set by determining whether the

Algorithm 2 Feature selection algorithm based on multigranulation entropy with granularity addition.

Input: A decision system $DIS = (U, A \cup \{d\})$; previous selected granules AM^* ; new granules $AM' = \bigcup_{i=1}^{k_1} A_i$; thresholds $\delta \in [0.1, 1]$, $\alpha \in [0.1, 1]$, $k_1 \in [2, |A|]$, $k_2 \in [1, k_1]$, $k_3 \in [1, 15]$

Output: Selected feature subset S

```

1: if  $|AM'| > k_2$  then
2:    $m \leftarrow MGE(AM')$ 
3:   while  $|AM'| > k_2$  do
4:      $m_{best} \leftarrow m$ 
5:      $g_{best} \leftarrow \text{None}$ 
6:     for each  $g$  in  $AM'$  do
7:        $AM_{temp} \leftarrow AM' \setminus \{g\}$ 
8:       if  $AM_{temp} = \emptyset$  then continue
9:       end if
10:       $m' \leftarrow MGE(AM_{temp})$ 
11:      if  $m' < m_{best}$  then
12:         $m_{best} \leftarrow m'$ 
13:         $g_{best} \leftarrow g$ 
14:      end if
15:    end for
16:    if  $g_{best} = \text{None}$  then break
17:    end if
18:    Remove  $g_{best}$  from  $AM'$ 
19:     $m \leftarrow m_{best}$ 
20:  end while
21: end if
22:  $AM'^* \leftarrow AM^* \cup AM'$ 
23:  $F \leftarrow$  all features in  $AM'^*$ 
24:  $S \leftarrow \emptyset$ 
25: while  $F \neq \emptyset$  do
26:    $Sig_{best} \leftarrow -\infty$ 
27:   for each  $a \in F$  do
28:     Compute  $SIM(x)$ 
29:     Compute  $KPOS_S(U)$ 
30:     Compute  $K\gamma_{S \cup \{a\}}(D)$  and  $K\gamma_S(D)$ 
31:     compute  $Sig(a, S, D)$ 
32:     if  $Sig(a, S, D) > Sig_{best}$  then
33:        $Sig_{best} \leftarrow Sig(a, S, D)$ 
34:        $f_{best} \leftarrow a$ 
35:     end if
36:   end for
37:   if  $Sig_{best} > 0$  then
38:      $S \leftarrow S \cup \{f_{best}\}$ 
39:     Remove  $f_{best}$  from  $F$ 
40:   else
41:     break
42:   end if
43: end while
44: return  $S$ 

```

granules to be removed are present in the original optimal set, and eventually yields the optimal feature subset via WKN-RS. The flowchart of dynamic feature selection with granularity deletion is shown in Fig. 5. $\square$

It should be noted that the proposed K-FSAMGEA and K-FSAMGED only implement granularity-level incremental and decremental updates at present, and the current framework does not support real-time streaming sample arrival. In practical online learning scenarios, data distribution may shift continuously, and new samples arrive sequentially, which brings challenges to model stability. For future extension, we plan to adapt to instance streaming by forgetting factor strategies to load newly arrived samples into the granularity calculation module in batches and recalculate neighborhood similarity and MGE without retraining the whole model from scratch. Importantly, the core MGE evaluation and WKN-RS feature selection logic can be reused in the extended online framework, which ensures good scalability of our multigranulation entropy backbone.

Algorithm 3 Feature selection algorithm based on multigranulation entropy with granularity deletion.

Input: A decision system $DIS = (U, A \cup \{d\})$; previous selected granules AM^* ; granules to delete AM' ; thresholds $\delta \in [0.1, 1]$, $\alpha \in [0.1, 1]$, $k_1 \in [2, |A|]$, $k_2 \in [1, k_1]$, $k_3 \in [1, 15]$

Output: Selected feature subset S

```

1: if  $AM' \in AM^*$  then
2:    $AM'^* \leftarrow AM^* \setminus AM'$ 
3: else
4:    $AM'^* \leftarrow AM^*$ 
5: end if
6:  $F \leftarrow$  all features in  $AM'^*$ 
7:  $S \leftarrow \emptyset$ 
8: while  $F \neq \emptyset$  do
9:    $Sig_{best} \leftarrow -\infty$ 
10:  for each  $a \in F$  do
11:    Compute  $SIM(x)$ 
12:    Compute  $KPOS_S(U)$ 
13:    Compute  $K\gamma_{S \cup \{a\}}(D)$  and  $K\gamma_S(D)$ 
14:    Compute significance  $Sig(a, S, D)$ 
15:    if  $Sig(a, S, D) > Sig_{best}$  then
16:       $Sig_{best} \leftarrow Sig(a, S, D)$ 
17:       $f_{best} \leftarrow a$ 
18:    end if
19:  end for
20:  if  $Sig_{best} > 0$  then
21:     $S \leftarrow S \cup \{f_{best}\}$ 
22:    Remove  $f_{best}$  from  $F$ 
23:  else
24:    break
25:  end if
26: end while
27: return  $S$ 

```

3.5. Time complexity analysis

To quantitatively evaluate the computational cost and scalability of the proposed model, we derive the theoretical time complexity of static K-FSAMGE, incremental K-FSAMGEA and decremental K-FSAMGED. Granularity updates are triggered by the addition or removal of raw features. Let n be the number of samples, I the iteration number of k-means clustering, and k_3 the neighbor number adopted in WKN-RS.

For K-FSAMGE: m_{sta} denotes the total number of initial raw features, $k_{1,sta}$ stands for all initial granules generated from the initial features, and m'_{sta} is the count of candidate features covered by the static optimal granule set. K-means clustering over all initial granules costs $O(Ik_{1,sta}m_{sta})$. Iterative MGE pruning over the full global granule pool yields $O(k_{1,sta}^2 n^2 m_{sta})$, where each MGE calculation consumes $O(n^2 m_{sta})$ for fuzzy neighborhood construction. WKN-RS on static optimal granules takes $O(m_{sta}'^2 nk_3)$. So, The total complexity of static K-FSAMGE is

$$T_{K-FSAMGE} = O(Ik_{1,sta}m_{sta}) + O(k_{1,sta}^2 n^2 m_{sta}) + O(m_{sta}'^2 nk_3).$$

For K-FSAMGEA: m_{new} represents the quantity of newly added raw features, and $k_{1,new}$ refers to the new granules generated from these newly added features. The merged candidate feature set input to WKN-RS consists of both preserved historical candidate features and screened new candidate features, with a total scale of $m'_{sta} + m'_{new}$. The overall complexity of K-FSAMGEA is

$$\begin{aligned} T_{K-FSAMGEA} = & O(Ik_{1,new}m_{new}) \\ & + O(k_{1,new}^2 n^2 m_{new}) \\ & + O((m'_{sta} + m'_{new})^2 nk_3). \end{aligned}$$

Unlike re-executing K-FSAMGE from scratch on expanded features, K-FSAMGEA avoids redundant k-means clustering and MGE computation on historical granules, greatly reducing redundant overhead.

For K-FSAMGED: Let m'_{del} denote the number of candidate features corresponding to the shrunk granule set after deletion. The whole computational load comes from rerunning WKN-RS:

$$T_{K-FSAMGED} = O(m_{del}'^2 nk_3).$$

Compared with rerunning K-FSAMGE after feature removal, K-FSAMGED skips both k-means clustering and MGE computation entirely, which delivers evident efficiency advantages.

4. Experimental decision and analysis

In this chapter, a series of experiments are conducted to validate the effectiveness and superiority of the proposed algorithm. All experiments are implemented in Python using PyCharm under a Windows 10 environment on an 11th Gen Intel® Core™ i5-1155G7 CPU (2.50 GHz) with 16 GB RAM. The experimental procedure is carried out as follows:

- (1) *Data Preprocessing*: We employ 15 datasets (Table 4) from the UCI Machine Learning Repository (<https://archive.ics.uci.edu/>) to assess the comparative performance of the algorithms. The entire data preprocessing pipeline is illustrated graphically in Fig. 6. For each dataset, max–min normalization is applied to rescale all features to the range [0,1], as formulated in the following equation:

$$b(x_j) = \frac{b(x_j) - \min b(x)}{\max b(x) - \min b(x)}. \quad (36)$$

where $\max b(x)$ and $\min b(x)$ represent the maximum and minimum values of attribute b , respectively.

- (2) *Ablation study*: We design an ablation study to examine how each component contributes to the K-FSAMGE framework. All configurations are executed on the same datasets under identical parameter settings. We then compare their classification accuracies to reveal how each module affects the overall performance.
- (3) *Robustness verification*: To evaluate the stability of the proposed algorithm, we conduct two types of robustness experiments. The first examines the algorithm's noise resistance by introducing random noise of different intensities. The second tests its adaptability by adding and removing granularities from the original granule set.
- (4) *Performance analysis*: A series of supplementary experiments are conducted to comprehensively evaluate the feature selection capability, generalization stability, and computational efficiency of the proposed K-FSAMGE algorithm.
- (5) *Algorithm validation*: To verify the overall effectiveness of the proposed algorithm, K-FSAMGE is compared with the original dataset results and seven representative feature selection methods. The comparison focused on classification accuracy, the number of selected features and the time of feature selection, assessing whether K-FSAMGE can exhibit outstanding comprehensive performance in classification accuracy, feature reduction and computational efficiency.
- (6) *Statistical testing*: Two hypothesis testing experiments are designed to further analyze the differences in classification performance between our algorithm and the comparison methods.
- (7) *Parameter analysis*: We further investigate the impact of five key parameters on classification performance, including the number of clusters (k_1), the number of retained granules (k_2), the number of nearest neighbors (k_3), the neighborhood radius (δ), and the similarity threshold (α). k_1 varies from 2 to the total number of features with a step size of 1. k_2 is selected from the range $[1, k_1]$ with an increment of 1. k_3 spans from 1 to 15, increasing by 1. δ and α vary between 0.1 and 1 in steps of 0.1. The detailed parameter settings of K-FSAMGE are listed in Table 5.

4.1. Ablation study

To gain deeper insights into the independent contributions and interactions of the key components within the K-FSAMGE algorithm, we conduct a systematic ablation study.

4.1.1. Contribution of internal modules

Four ablation models are designed for comparative analysis: (a) *K-means-removed*: the k-means clustering module is removed while all other components are retained; (b) *Entropy-removed*: the MGE component is removed while all other components are retained; (c) *Weight KNN-removed*: the Weight KNN module is removed while all other components are retained; (d) *Full model*: all the above modules are integrated and maintained.

All experiments use the same classifier, parameter settings, and datasets to ensure fairness. Classification accuracy is adopted as the primary evaluation metric. As presented in Fig. 7, under the KNN, DT, RF and MLP classifiers, the full model achieves the highest accuracy on most datasets and demonstrates more stable performance than the other three variants. These results indicate that the k-means clustering module organizes the feature space more effectively and enhances the discriminative ability of granularity partitioning. Meanwhile, the MGE module refines the granularity weights and reduces redundant information. The weighted KNN module further improves the discriminative accuracy. Working together, the three modules produce a more representative and stable feature subset, leading to improved classification performance across most datasets.

4.1.2. Comparison of different clustering algorithms

To further verify the effectiveness of the adopted clustering strategy, we conduct comparative experiments among k-means clustering, hierarchical clustering and density-based clustering. In this part, only the clustering module is replaced, while all other experimental settings and parameter configurations remain identical. We evaluate the classification accuracy under KNN, DT, RF and MLP classifiers, and the corresponding results are summarized in Table 6. It can be observed that the model equipped with k-means clustering achieves overall higher or comparable classification accuracy compared with the other two clustering approaches across

all classifiers. The results demonstrate that k-means is more suitable for the proposed framework, and it can better excavate intrinsic data structures to facilitate subsequent classification tasks.

4.2. Robustness evaluation

We examine the robustness of the proposed K-FSAMGE model from two perspectives: noise analysis and dynamic feature selection. Classification accuracy, value of entropy, running time, and memory usage serve as primary indicators to reflect performance stability in uncertain and dynamic scenarios.

4.2.1. Noise analysis

To comprehensively explore the anti-interference ability of K-FSAMGE, this section investigates the model performance under different noise levels from classification accuracy and value of entropy.

We first evaluate the model's noise resistance with a class-aware Gaussian noise injection strategy. We add class-scaled Gaussian noise to random samples of each category at noise ratios of 10%, 20%, 30%, 40% and 50%. The noise scale of each class is determined by the average feature standard deviation to match the original data distribution. After noise injection on selected features, classification accuracy is measured and presented in Fig. 8. Experimental results indicate that accuracy decreases marginally with increasing noise ratio, accompanied by slight performance degradation. It demonstrates that K-FSAMGE achieves stable classification and strong robustness to class-aware noise.

We then explore the internal structural stability of K-FSAMGE under noise by comparing the proposed MGE with two traditional entropy metrics, IE [36] and CIE [37]. Class-aware Gaussian noise is injected into the feature space at ten ratios from 2.5% to 25%. We collect the values of the three entropy metrics and calculate their standard deviations to reflect fluctuation degrees, as listed in Table 7. On most datasets, MGE has a significantly smaller standard deviation than IE and CIE. This confirms that MGE maintains stable numerical trends under varying noise, and further validates the superior noise robustness of the granularity measurement mechanism of K-FSAMGE.

4.2.2. Dynamic feature selection

Datasets are prepared for feature addition and deletion. For addition, each preprocessed feature set is split into an initial set B and an additional set B' . Clustering B generates the initial granularity set AT . The K-means clustering is conducted with 10 initializations and a convergence tolerance of 10^{-4} for centroid updates. Features from B' are randomly selected at 10%, 40%, 70%, and 100%, clustered into new sets AT' , and progressively merged into AT , forming the dynamic addition dataset. For deletion, the clustered dataset serves as the initial granularity set AT . Granules are removed at 10%, 20%, 30%, and 40% to construct the dynamic deletion dataset. We record running time, memory usage, and classification accuracy using a KNN classifier to compare the performance of different update strategies. Tables 8 and 9 demonstrate that dynamic feature selection algorithms reduce both computation time and memory consumption, while Fig. 9 illustrates that these algorithms maintain high accuracy. These results indicate that K-FSAMGE adapts efficiently to feature changes and remains stable under dynamic scenarios.

4.3. Performance analysis

To comprehensively evaluate the feature selection capability, generalization stability, and computational efficiency of K-FSAMGE, a series of supplementary experiments are conducted.

Fig. 10 illustrates the classification accuracy of K-FSAMGE under varying feature selection ratios ranging from 50% to 100%. For all datasets, as the feature selection ratio gradually increases, classification accuracy increases rapidly at the initial stage and gradually plateaus after reaching its peak. This indicates that K-FSAMGE effectively identifies highly discriminative features early and maintains stable performance even when the feature subset undergoes minor changes. Moreover, an examination of the selected feature subsets further reveals clear statistical and domain relevance to the corresponding classification objectives. This demonstrates that K-FSAMGE performs targeted, data-driven feature selection rather than indiscriminate dimensionality reduction.

In addition, the fitting behavior and generalization ability of the model are assessed using learning curves for accuracy and ROC-AUC, as shown in Fig. 11. Across all datasets, the training and test curves increase consistently as the training set grows, with a small performance gap. Meanwhile, the accuracy and ROC-AUC curves exhibit highly consistent trends. The observed behavior suggests stable model fitting without evident underfitting or severe overfitting. Therefore, K-FSAMGE exhibits strong generalization capability across different datasets.

Finally, the computational scalability of the algorithm is investigated by decomposing the total runtime into three main components: k-means clustering, MGE computation, and feature selection based on weighted KNN. Table 10 summarizes the runtime variation with respect to the number of features. The results reveal that MGE computation and feature selection based on weighted KNN dominate the total computational cost, while k-means clustering contributes relatively little overhead. As the feature dimensionality increases, the runtime grows steadily and remains manageable. Overall, these findings indicate that K-FSAMGE is scalable and suitable for efficient application to high-dimensional classification tasks.

4.4. Effectiveness validation

To comprehensively assess the overall performance of K-FSAMGE in feature selection and classification tasks, we compare it with seven mainstream feature selection algorithms, as outlined below:

- (1) *Feature Selection Algorithm Based on Fuzzy Neighborhood Generalized Composite Entropy (FSFNGCE)* [30]: The paper quantifies feature importance by jointly evaluating uncertainty and structural consistency within fuzzy neighborhood spaces at multiple granularities, enabling the selection of robust and informative subsets.
- (2) *Graph-Based Unsupervised Feature Selection (IGUFS)* [38]: The method builds a similarity graph to preserve the intrinsic geometric structure of data, retaining features that best capture local neighborhood relationships without label supervision.
- (3) *Unsupervised Feature Selection Algorithm Based on k -Nearest Neighborhood Dependency (KND-UFS)* [39]: The approach assesses feature relevance through local structural dependency analysis among data instances, achieving effective dimensionality reduction without class labels.
- (4) *Weighted k -nearest Neighbor Feature Selection (WKN-RS)* [33]: The algorithm incorporates distance-weighted neighborhood information to strengthen discriminative capability and improve classification outcomes.
- (5) *Feature Selection Based on Genetic Algorithm (CHCQX)* [40]: The technique combines elitism and accelerated convergence strategies to efficiently obtain compact and high-quality feature subsets from high-dimensional data, reducing redundancy while boosting classification accuracy.
- (6) *Feature Selection Based on Particle Swarm Optimization (PSO-MI)* [41]: The strategy exploits the global search ability of PSO to explore the feature space and derive optimal subsets, enhancing classification accuracy and reducing dimensionality.
- (7) *Divide-and-Conquer based Co-evolutionary Genetic Algorithm (DCGA)* [42]: The method decomposes high-dimensional feature spaces via multiple filter metrics and employs co-evolutionary search to efficiently identify compact, high-quality feature subsets.

We evaluate classification accuracy after feature selection under the KNN, DT, RF and MLP classifiers. Ten-fold cross-validation is applied throughout the experiments. The classification accuracies of each algorithm across the four classifiers are presented in Tables 11–14. The “RAW” column indicates the accuracy of the original dataset before feature selection. The label “N/A” indicates that the running time of the corresponding algorithm exceeds 24 h and no valid experimental results are obtained. The highest accuracy in each row is highlighted in bold. The value to the left of the “ $\pm$ ” symbol denotes the mean, while the value to the right represents the variance. The results show that K-FSAMGE consistently improves accuracy over the original datasets. It also outperforms the seven comparison algorithms on most datasets. This performance advantage originates from the multi-granularity evaluation mechanism of K-FSAMGE. The MGE measure assesses feature importance under multi-layered granular structures to filter out redundant and noisy attributes unassociated with classification objectives. Integrated with weighted KNN rough set assessment, the proposed approach preserves only discriminative features highly relevant to decision labels rather than conducting blind dimensionality reduction. Such targeted feature screening largely mitigates disturbances from irrelevant attributes and yields steady classification performance. Additionally, the number of features selected by each method is summarized in Table 15, showing that K-FSAMGE selects significantly fewer features than the other algorithms. The running time of all algorithms is recorded in Table 16. Among all eight methods, our proposed algorithm ranks fourth in terms of computational efficiency, belonging to the upper-middle level. Compared with traditional rough set-based feature selection methods including FNGCE, IGUFS and KND-UFS, K-FSAMGE achieves lower time complexity. Considering its superior classification performance and prominent dimensionality reduction effect, such time consumption is reasonable, and the algorithm achieves a favorable balance between effectiveness and efficiency. Collectively, these findings indicate that K-FSAMGE exhibits outstanding comprehensive performance in classification accuracy, feature reduction and computational efficiency.

4.5. Statistical test

To strengthen the comparison of different algorithms, we adopt the Friedman test to assess overall inter-algorithm performance differences and the Wilcoxon test to verify pairwise significance gaps. Some experiments on certain datasets took more than 24 h to complete or failed to generate valid results. Accordingly, all subsequent statistical tests were only conducted on datasets where all eight algorithms could produce valid outputs. The significance level is uniformly set at 0.05. A p-value greater than 0.05 implies that there is no statistically significant difference, whereas a p-value smaller than 0.05 indicates a significant performance gap.

Table 17 lists the average Friedman ranking values of all eight feature selection algorithms under KNN, DT, RF, and MLP classifiers. A lower average rank denotes superior classification performance, and K-FSAMGE achieves the minimum rank value across every classifier, which directly reflects its dominant performance among all competitors. Table 18 further presents the Friedman chi-square, F-statistic and corresponding p-values for each classifier. All p-values are far below the 0.05 significance threshold, so we reject the null hypothesis that all algorithms deliver equivalent performance. This confirms the existence of prominent overall performance discrepancies across the tested methods.

The pairwise p-values obtained from the Wilcoxon test are listed in Table 19. While most pairwise comparisons yield p-values less than 0.05, a small fraction of pairs return p-values slightly above 0.05. Collectively, K-FSAMGE exhibits statistically significant performance advantages over the vast majority of competing feature selection algorithms.

4.6. Parameter analysis

We systematically investigate the stability and tunability of the proposed algorithm through a sensitivity analysis of the key parameters: k_1 , k_2 , k_3 , δ , and α . Random search with ten-fold cross-validation is adopted to search for optimal parameter settings. Using the KNN classifier as an example, we examine how classification accuracy changes with variations in each parameter. The results are presented in Figs. 12–15.

Fig. 12 illustrates the effect of k_1 and k_2 . If k_1 is too small, the clustering granularity is coarse, causing some important features to be aggregated into the same cluster. This leads to a slight drop in accuracy due to loss of detailed information. If k_1 is too large, the number of clusters increases sharply, raising computational complexity and amplifying local noise. This results in minor performance fluctuations. A similar pattern is observed for k_2 . Therefore, it is essential to select an appropriate combination of k_1 and k_2 based on dataset characteristics for optimal classification performance.

Fig. 13 shows the performance curve of k_3 . When k_3 is too small, only a few neighboring samples are available for similarity calculation, resulting in lower classification accuracy. When k_3 is too large, distant heterogeneous samples are included, introducing irrelevant interference into feature evaluation. Even so, the overall fluctuation range of accuracy is far slighter than that of k_1 and k_2 . The curve gradually plateaus at $k_3 = 5$, and we adopt $k_3 = 5$ as a unified baseline for fair experimental comparison.

Fig. 14 shows the influence of δ . A too-small δ leads to insufficient neighboring samples, resulting in weakened feature discrimination. A too-large δ introduces heterogeneous samples into the neighborhood, contaminating the similarity calculation. The accuracy fluctuates only slightly and quickly stabilizes within a wide interval. We take $\delta = 0.5$ as the baseline parameter in our experiments.

Fig. 15 presents the trend for α . A too-low α sets an extremely loose similarity filter, retaining plenty of low-correlation features that increase redundant noise. A too-high α raises the similarity bar excessively, discarding many moderately relevant yet informative features. Throughout the entire value interval, accuracy only fluctuates gently. For consistent comparison across datasets, we use $\alpha = 0.5$ as the baseline setting.

Overall, k_1 and k_2 are highly sensitive parameters that dominate model performance, while k_3 , δ and α trigger only subtle accuracy shifts, demonstrating strong parameter robustness of our algorithm.

5. Conclusion and future work

This paper targets three core challenges summarized in the motivation framework: overly subjective granular division, insufficient quantification of granular interaction importance, and low computational efficiency under dynamic feature changes. To address this, we propose K-FSAMGE. This model uses the k-means algorithm to divide granules according to their weight values, resulting in a more reasonable granularity structure. Based on this model, we introduce MGE to accurately quantify inter-granular correlations and the significance of each granular layer. The entropy is then combined with WKN-RS to perform granularity reduction for feature selection. In addition, we analyze a dynamic update mechanism for feature changes and design two incremental feature selection algorithms to accelerate computation when features are added or deleted. Experimental results on 15 datasets demonstrate that K-FSAMGE achieves strong feature selection performance. Meanwhile, the two incremental variants drastically cut running time while maintaining nearly identical prediction accuracy.

Practically, our method boasts lightweight computation and strong interpretability, with no demand for massive training data or high-end computing devices. It fits perfectly for small and micro industries, such as agricultural product quality inspection, grassroots medical auxiliary diagnosis, small enterprise credit evaluation and low-cost equipment fault detection. Compared with data-hungry deep learning and complex metaheuristic algorithms, our rough-set-based scheme provides an affordable, transparent decision-making tool to resolve uncertain real-world industrial problems for small-scale businesses.

Despite the promising results achieved by the proposed method, several limitations remain and deserve further investigation. First, the current incremental feature selection schemes are designed only for feature variation and cannot effectively handle scenarios involving the addition or deletion of samples. Future research will focus on developing a unified bidirectional dynamic updating framework capable of simultaneously accommodating both feature-level and sample-level changes, thereby enhancing the adaptability and stability of the model in dynamic environments. Second, although parameter sensitivity analyses were conducted within reasonable ranges to select suitable hyperparameters, the model still relies on these user-specified ranges rather than being fully data-driven. Future work will explore adaptive parameter optimization strategies to enable autonomous tuning and further reduce human intervention. Third, while the proposed MGE framework demonstrates strong performance in uncertainty-aware feature evaluation, its capability for modeling highly nonlinear feature interactions remains limited. The distance-driven neighborhood mechanism of fuzzy rough sets relies on local similarity metrics, which struggles to fit sharply irregular decision boundaries built from intricate nonlinear feature couplings. From a comparative perspective, deep learning excels at hierarchical nonlinear fitting yet suffers from poor interpretability [43,44]. Meanwhile, metaheuristic optimizers deliver powerful global search but easily fall into redundant iterative computation. Our MGE can act as a transparent preprocessing module to screen informative features in advance. This synergy reduces input dimension for deep networks and narrows the search space for metaheuristic solvers, lowering computational burden and mitigating overfitting risk. Future work will explore the integration of MGE with deep learning and metaheuristic optimization techniques to construct hybrid feature selection frameworks [45,46]. Such hybrid architectures combine the nonlinear fitting strengths of advanced paradigms and the interpretable granular evaluation advantages of MGE, delivering stable performance on datasets with strong feature nonlinearity and complex decision boundaries.

CRediT authorship contribution statement

Xiaoyan Zhang: Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization; **Meixia Zhou:** Writing – review & editing, Writing – original draft, Validation, Software, Investigation, Formal analysis, 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.

Acknowledgement

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 of China (Grant No. CSTB2023NSCQ-MSX1063).

Data availability

All datasets used in this research are sourced from the UCI Machine Learning Repository (<https://archive.ics.uci.edu/>).

References

- [1] Y. Li, T. Li, H. Liu, Recent advances in feature selection and its applications, *Knowl. Inf. Syst.* 53 (2017) 551–577.
- [2] Z. Jensen, E. Kim, et al., A machine learning approach to zeolite synthesis enabled by automatic literature data extraction, *ACS Cent. Sci.* 5 (5) (2019) 892.
- [3] N.D. Galvão, H.D.F. Marin, Data mining: a literature review, *Acta Paul. Enferm.* 22 (2009) 686–690.
- [4] G. Zhao, Z. Peng, et al., Angular histograms: frequency-based visualizations for large, high dimensional data, *IEEE Trans. Vis. Comput. Graph.* 17 (12) (2011) 2572–2580.
- [5] K.C. Sharmili, A. Chilambuchelvan, A hybridization approach for optimal feature subset selection in high dimensional data, *Int. J. Uncertain. Fuzziness Knowl. Based Syst.* 26 (6) (2018) 949–970.
- [6] B. Jin, X. Xu, Forecasts of wholesale soybean oil price indices via gaussian process regressions, *Int. J. Big Data Min. Glob. Warm.* 7 (1) (2025) 2550001.
- [7] X. Xu, Corn cash price forecasting, *Am. J. Agric. Econ.* 102 (4) (2020) 1297–1320.
- [8] B. Jin, X. Xu, Chinese energy security index price forecasting through the neural network, *Innov. Emerg. Technol.* 12 (2025) 2550036.
- [9] P. Dhal, B. Pradhan, U. Fiore, S.A.J. Francis, D.S. Roy, A clinical diabetes prediction based support system based on the multi-objective metaheuristic inspired fine tuning deep network, *Inf. Fusion* 122 (2025) 103188.
- [10] X. Xu, Y. Zhang, Price forecasts of ten steel products using Gaussian process regressions, *Eng. Appl. Artif. Intell.* 126 (2023) 106870.
- [11] B. Jin, X. Xu, Predictions of residential property price indices for China via machine learning models, *Qual. Quant.* 59 (2) (2025) 1481–1513.
- [12] M.I. Jordan, T.M. Mitchell Machine learning: Trends, perspectives, and prospects, *Science* 349 (2015) 255–260.
- [13] Z. Guo, T.J. Wang, et al. Prompt-based weakly-supervised vision-language pre-training, *Pattern Recognit. Lett.* 197 (2025) 8–15.
- [14] X. Xu, Y. Zhang, Corn cash price forecasting with neural networks, *Comput. Electron. Agric.* 184 (2021) 106120.
- [15] P. Dhal, C. Azad, A multi-objective evolutionary feature selection approach for the classification of multi-label data in: 2022 2nd International conference on advance computing and innovative technologies in engineering (ICACITE), 2022, pp. 1986–1989.
- [16] P. Dhal, C. Azad, A multi-stage multi-objective GWO based feature selection approach for multi-label text classification in: 2022 2nd International conference on intelligent technologies (CONIT), 2022, pp. 1–5.
- [17] P. Dhal, C. Azad, Zone oriented binary multi-objective charged system search based feature selection approach for multi-label classification, *Expert Syst.* 42 (2) (2025) e13803.
- [18] P. Dhal, D. Mishra, B. Pradhan, A deep ensemble-based framework for the prediction of oral cancer through histopathological images, *Appl. Soft Comput.* 179 (2025) 113258.
- [19] Z. Pawlak, Rough sets, *Int. J. Comput. Inf. Sci.* 11 (1982) 341–356.
- [20] Y. Cheng, Dynamic maintenance of approximations under fuzzy rough sets, *Int. J. Mach. Learn. Cybern.* 9 (12) (2018) 2011–2026.
- [21] J.L. Tu, R.X. Duan, D. De Cun, An approach for construction of fault trees using rough set and fuzzy clustering algorithm, *Appl. Mech. Mater.* 1021 (34) (2010) 186–191.
- [22] Q. Hu, D. Yu, et al., Neighborhood rough set based heterogeneous feature subset selection, *Inf. Sci.* 178 (18) (2008) 3577–3594.
- [23] L. D'eer, M. Restrepo, et al., Neighborhood operators for covering-based rough sets, *Inf. Sci.* 336 (2016) 21–44.
- [24] X. Zhang, Z. Jiang, W. Xu, Feature selection using a weighted method in interval-valued decision information systems, *Appl. Intell.* 53 (9) (2023) 9858–9877.
- [25] D. Chen, L. Zhang, et al., A novel algorithm for finding reducts with fuzzy rough sets, *IEEE Trans. Fuzzy Syst.* 20 (2) (2011) 385–389.
- [26] W. Xu, W. Li, X. Zhang, Generalized multigranulation rough sets and optimal granularity selection, *Granul. Comput.* 2 (4) (2017) 271–288.
- [27] A. Stotland, A. A. Pomeransky, et al., The information entropy of quantum-mechanical states, *EPL* 67 (5) (2004) 700–706.
- [28] H.E. Montgomery, Information entropy and uncertainty relations, *Chem. Educ.* 7 (6) (2002) 334–338.
- [29] W. Xu, Q. Bu, Matrix-based incremental feature selection method using weight-partitioned multigranulation rough set, *Inf. Sci.* 681 (2024) 121219.
- [30] 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.
- [31] M. Nordin, et al., Dynamic data migration for structured AMR solvers, *Int. J. Parallel Program.* 35 (5) (2007) 477–491.
- [32] Y. Yanyan, Z. Degang, et al., Incremental feature selection by sample selection and feature-based accelerator, *Appl. Soft Comput.* 121 (2022) 108800.
- [33] W. Ning, Z. Enhui, A new method for feature selection based on weighted nearest neighborhood rough set, *Expert Syst. Appl.* 238 (2024) 122324.
- [34] B. Sang, H. Chen, L. Yang, D. Zhou, T. Li, W. Xu, Incremental attribute reduction approaches for ordered data with time-evolving objects, *Knowl. Based Syst.* 212 (2021) 106583.
- [35] K. Yuan, D. Miao, Y. Yao, H. Zhang, X. Zhao, Feature selection using zentropy-based uncertainty measure, *IEEE Trans. Fuzzy Syst.* 32 (4) (2023) 2246–2260.
- [36] Q. Hu, D. Yu, Neighborhood entropy, in: 2009 International Conference on Machine Learning and Cybernetics, vol.3, 2009, pp. 1776–1782.
- [37] Y. Zhou, X. Zhang, Z. Mo, Conditional neighborhood entropy with granulation monotonicity and its relevant attribute reduction, *J. Comput. Res. Dev.* 55 (11) (2018) 2395–2405.
- [38] X. Weihua, H. Man, et al., Graphbased unsupervised feature selection for interval-valued information system, *IEEE Trans. Neural Netw. Learn. Syst.* 35 (9) (2023) 12576–12589.
- [39] W. Xu, Y. Zhang, Y. Qian, A novel unsupervised feature selection for high-dimensional data based on FCM and k-nearest neighbor rough sets, *IEEE Trans. Neural Netw. Learn. Syst.* 36 (2024) 10889–10898.
- [40] A.M. Ghaith, N. Slawomir, et al., Fast genetic algorithm for feature selection – a qualitative approximation approach, *Expert Syst. Appl.* 211 (2023) 118528.
- [41] K. Robindro, S.S. Devi, et al., Hybrid distributed feature selection using particle swarm optimization-mutual information, *Data Sci. Manage.* 7 (1) (2024) 64–73.
- [42] K. Chen, K. Liu, P. Wang, J. Liang, Z. Yang, K. Yu, An efficient evolutionary feature selection algorithm with divide-and-conquer strategy for classification, *Expert Syst. Appl.* 302 (2026) 130416.
- [43] B. Jin, X. Xu, Machine learning-based forecasts of residential property prices in Hangzhou city, Zhejiang province, China, *Neural Comput. Appl.* 37 (2025) 4971–4988.

- [44] X. Xu, Y. Zhang, Individual time series and composite forecasting of the Chinese stock index, *Mach. Learn. Appl.* 5 (2021) 100035.
- [45] X. Xu, Contemporaneous and Granger causality among US corn cash and futures prices, *Eur. Rev. Agric. Econ.* 46 (4) (2019) 663–695.
- [46] B. Jin, X. Xu, Machine learning gold price forecasting, *Int. J. Manage. Sci. Eng. Manage.* 20 (2) (2025) 273–285.