Fuzzy Sets and Systems 486 (2024) 108971

Contents lists available at ScienceDirect

Fuzzy Sets and Systems

journal homepage: www.elsevier.com/locate/fss

Uncertainty measures and feature selection based on composite entropy for generalized multigranulation fuzzy neighborhood rough set

Xiaoyan Zhang *, Weicheng Zhao

College of Artificial Intelligence, Southwest University, Chongqing 400715, People's Republic of China

ARTICLE INFO

Keywords:
Feature selection
Fuzzy neighborhood rough set
Generalized multigranulation rough set
Entropy

ABSTRACT

With the continuous advancement of information technology, the information and data covered by various information systems become increasingly complex and diverse, it is essential to perform knowledge mining from multiple perspectives to extract valuable insights. Fuzzy neighborhood multigranulation rough set, as an excellent feature selection model, is capable of handling heterogeneous datasets more effectively, significantly improving learning efficiency. In this study, we investigate a feature selection method based on a generalized multigranulation fuzzy rough set (GMFNRS) in fuzzy decision systems. First, the concepts of fuzzy neighborhood rough sets and generalized multigranulation rough sets are introduced. Subsequently, the GMFNRS model is established to enable data mining and rule extraction from various perspectives. Secondly, from an informational perspective, the study investigates uncertainty measurement methods through fuzzy neighborhood joint entropy. Furthermore, a novel fuzzy neighborhood generalized composite entropy is proposed by integrating the GMFNRS model with uncertainty measures. Finally, a forward greedy feature selection algorithm is considered to extract essential information from complex datasets. Experimental results on 15 public datasets demonstrate that the proposed model effectively selects important features in fuzzy systems and exhibits excellent classification performance.

1. Introduction

In recent years, feature selection has emerged as a crucial step in data preprocessing and has found widespread applications in intelligent computing, machine learning, and artificial intelligence domains [42,7,2]. By removing irrelevant or redundant features while preserving the classification capacity of the knowledge base, feature selection allows for better extraction of essential features from high-dimensional data, thereby reducing time and space complexity [24,9,22]. Rough set theory, as a powerful tool for dealing with uncertainty, has garnered extensive research and attention in the field of feature selection [28]. However, the classical rough set models [15] are limited to handling symbolic data, and when dealing with continuous data, it is necessary to discretize it, which can result in the loss of valuable information [5]. Therefore, researchers have extended the rough set models further to handle diverse types of data. Among them, there has been a significant amount of research on feature selection using neighborhood rough set (NRS) and fuzzy rough set (FRS) approaches [37,25,30]. Hu et al. proposed a model based on NRS to handle heterogeneous data by

* Corresponding author.
E-mail addresses: zxy19790915@163.com (X. Zhang), 2504975527@qq.com (W. Zhao).

https://doi.org/10.1016/j.fss.2024.108971
Received 28 November 2023; Received in revised form 20 February 2024; Accepted 11 April 2024

Available online 17 April 2024

assigning different thresholds for each attribute [11]. Barman et al. introduced a novel approach to determine optimal thresholds for NRS in the context of hyperspectral band selection [3]. Zhang et al. introduced a novel feature selection algorithm called weighted neighborhood rough sets, incorporating different weights into neighborhood relations [41]. Xing et al. proposed a feature selection algorithm based on NRS to assess the ability to distinguish different groups by evaluating features [29]. While NRS effectively tackles the drawback of discretizing data in classical rough set models, it faces challenges when dealing with samples in a fuzzy context. Jensen et al. proposed a greedy feature selection algorithm in FRS by incorporating dependency functions [13]. Das et al. introduced a fuzzy graph technique based on fuzzy sets to assess the relevance and redundancy of features [6]. Chen et al. extended the FRS model by introducing a fuzzy discernibility matrix for attribute reduction [4]. Sang et al. proposed an active antinoise fuzzy dominance rough set model by identifying noise samples based on the density of noisy samples [19]. Dong et al. studied the incremental feature selection mechanism of fuzzy rough sets when both samples and features change simultaneously [8]. However, the FRS model has a major drawback in that it always relies on the nearest samples to calculate fuzzy upper and lower approximations, thereby increasing the vulnerability to noise in the dataset. For this issue, Wang et al. introduced the concept of using the min-max operation to calculate the membership degree of samples in fuzzy rough sets. They incorporated this concept into neighborhood rough sets, leading to the construction of the fuzzy neighborhood rough set model [27]. Yue et al. introduced the concept of fuzzy neighborhood coverage for conducting three-way classification [36]. However, the application of fuzzy neighborhood rough sets in describing features is limited, and there is a scarcity of research in this area.

In real life, it is often necessary to describe things from multiple different perspectives. However, most current models are based on a single binary relation [22], which means they describe objects from a single perspective. To address this issue, Qian et al. extended the single-granularity rough set model to a multi-granularity rough set model, enabling attribute reduction to be performed at multiple granularities [17]. Yao et al. investigated rough set models in multi-granular spaces [35]. However, the current research on two models, optimistic multi-granularity rough set and pessimistic multi-granularity rough set, has approximation operators that are either too loose or too strict, resulting in a lack of flexibility in the models. Xu et al. proposed the generalized multi-granularity rough set, which effectively addresses this issue and allows for more flexible descriptions of objects [32]. To better handle mixed data, additional relations are introduced in the multi-granularity model. Lin et al. conducted research on neighborhood-based multi-granularity rough sets [14]. Yang et al. developed a multi-granularity decision rough set model based on tolerance relations for incomplete information systems [34]. Sun et al. proposed a novel attribute reduction method based on neighborhood-based multi-granularity rough sets to handle incomplete neighborhood systems [22]. Zhang et al. introduced hesitant fuzzy tolerance relation into the multi-granularity framework and proposed two multi-granularity models [39]. However, there is limited research on further extending the generalized multi-granularity rough set model.

It is widely recognized that uncertainty measures have made rapid progress in the field of feature selection. By employing a distance-based approach, Wang et al. constructed a fuzzy rough set model to measure fuzzy dependency and attribute importance in decision systems [26]. Hu et al. proposed a matrix incremental-based approach to update the approximation space of neighborhood multi-granularity rough sets [10]. Peng et al. discussed uncertainty measures for feature selection based on the fuzzy symmetric relation of set-valued information systems [16]. The aforementioned studies primarily focus on uncertainty from an algebraic perspective, with a focus on describing the influence of features contained in the feature subset through uncertainty measures. Information entropy provides an intuitive measure, making it widely applied in the field of feature selection, along with its various derived forms. Zhang et al. investigated active incremental feature selection based on fuzzy rough set information entropy [42]. Aremu et al. proposed a feature engineering framework for asset data, which utilizes measures of correlation and relative entropy [2]. An et al. proposed a granularity entropy theory and designed a innovative feature selection algorithm [1]. Xu et al. developed a feature gene selection algorithm by utilizing the fuzzy neighborhood conditional entropy model [30]. Zeng et al. introduced two models of multi-granularity entropy to evaluate the significance of features [38]. Xu et al. introduced a novel feature selection approach for imbalanced fuzzy data by proposing a concept known as local composite entropy [33]. Huang et al. proposed a novel incremental feature selection algorithm that utilizes conditional entropy based on multi-source data [12]. Nonetheless, the assessment of feature importance based on information view is limited to describing the influence of features in uncertain classification. Integrating the aforementioned perspectives in feature selection to enhance the measurement quality in decision systems is an innovative and challenging task. Sun et al. introduced an innovative approach to attribute reduction in incomplete neighborhood decision systems based on NMRS, utilizing Lebesgue measure and entropy measure [22]. Zhang et al. introduced the concept of neighborhood composite entropy based on NRS to reflect the probability of granule pairs in the neighborhood [40]. Sun et al. proposed a novel attribute reduction algorithm that combines algebraic and information perspectives, based on neighborhood entropy [23]. Song et al. investigated uncertainty measures based on the divergence of cross-entropy [20]. Sun et al. proposed a feature subset selection method using fuzzy neighborhood pessimistic multigranulation entropy, specifically designed for heterogeneous datasets [21]. However, there has been a scarcity of research investigating the measurement of entropy in fuzzy contexts.

Building upon the above insights, to address the limitations of FNRS in singularly describing objects, we further investigate fuzzy neighborhood multigranulation rough sets from an algebraic perspective. Moreover, to the best of our knowledge, there is limited research on combining FNRS and GMRS for feature selection in fuzzy decision systems. Simultaneously, from an informational perspective, we enhanced the descriptive capability of information entropy in capturing feature uncertainty. Therefore, it is crucial to investigate uncertainty measures based on GMFNRS from these two aspects, and then develop a forward feature selection algorithm specifically designed for fuzzy decision systems dealing with complex datasets. In contrast to existing methods, our novel approach offers a comprehensive assessment of features, leading to a significant enhancement in the classification performance of the target. The main contributions of this paper include:

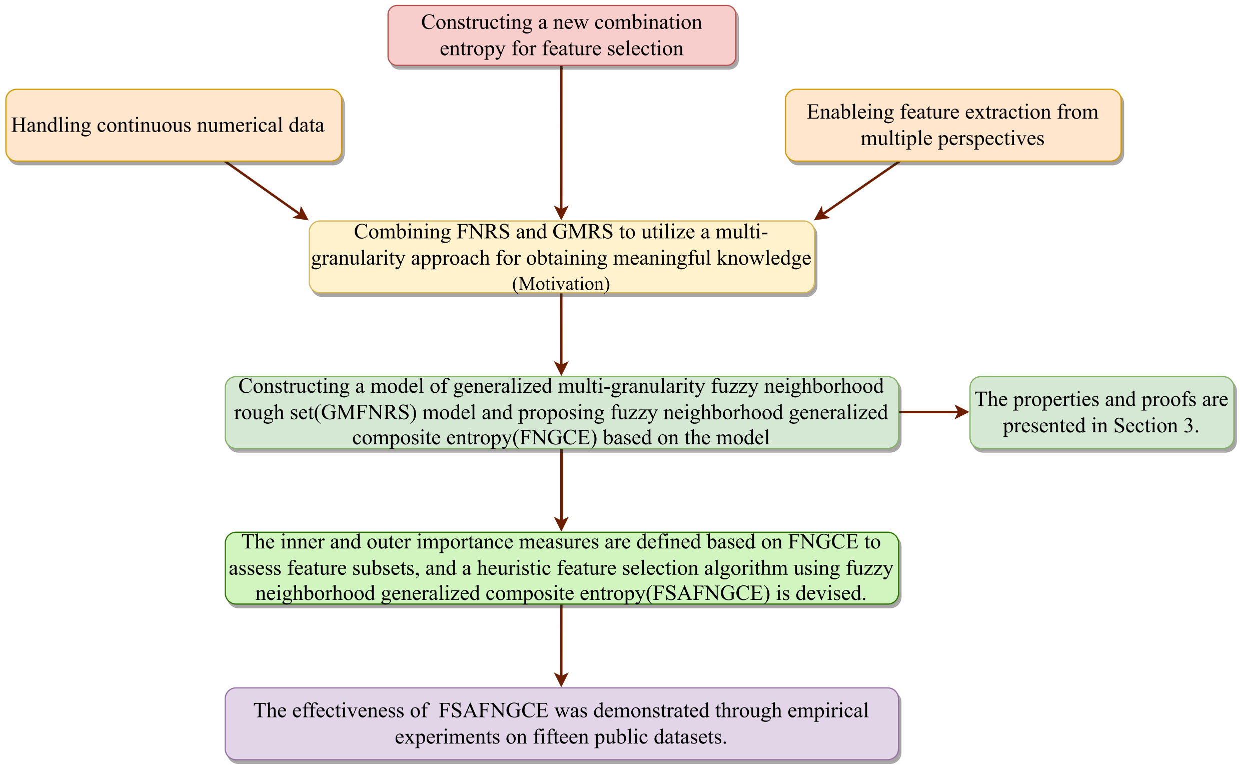


Fig. 1. The overall framework diagram of the paper.

(1) Building a fuzzy decision system that integrates FNRS with GMRS, a generalized multigranulation fuzzy neighborhood rough set (GMFNRS) model is proposed to better handle heterogeneous datasets and provide a comprehensive characterization of objects from multiple perspectives.
(2) By investigating the uncertainty measures of fuzzy neighborhood entropy, a fuzzy domain generalized composite entropy is proposed and a corresponding feature subset selection algorithm is designed.

The remaining structure of this paper is as follows. In section 2, we provide a comprehensive review of the concepts related to FNRS, NMRS, and GMRS. In section 3, we present the construction of the GMFNRS model and introduce the concept of fuzzy neighborhood generalized composite entropy. Section 4 introduce a heuristic greedy selection algorithm and analyze its complexity accordingly. In section 5, we performed experimental analysis on a set of 15 datasets to evaluate the effectiveness and robustness of the proposed model. Finally, we summarize the findings of this study and presents future research directions in Section 6. Moreover, the overall framework diagram of the paper is shown in Fig. 1.

2. Related work and foundations

In this section, we review some fundamental concepts, including fuzzy neighborhood rough set, neighborhood multi-granulation rough set, and generalized multigranulation rough set.

2.1. Fuzzy neighborhood rough set

A quadruple $\mathcal{FDS} = (U, AT \cup D, g)$ is designated as a fuzzy decision system with a decision, where $U = \{x_1, x_2, \dots, x_n\}$ is a non-empty finite sample set. $AT = \{a_1, a_2, \dots, a_t\}$ represents a non-empty finite set of condition attributes, $D = \{d\}$ represents a non-empty finite set of decision attributes and contains only one decision attribute, and additionally, $AT \cap D = \emptyset$. $g : U \times (AT \cup D) \to V$ is a mapping function, and $g_a(x) \in [0, 1]$ represents the attribute value of object x for attribute a.

Given a fuzzy decision system $\mathcal{FDS} = (U, AT \cup \{d\}, g)$, $N \subseteq AT$ induces a fuzzy binary relation R_N on the domain U. For any $x, y \in U$, when R_N is a fuzzy similarity relation, the following properties hold:

1) Reflexivity: $R_N(x, x) = 1$
2) Symmetry: $R_N(x, y) = R_N(y, x)$

The classical fuzzy rough set determines the classification of samples through the maximum-minimum operations. However, when data contains noise, it can introduce errors in calculating the membership degrees of samples to different decision classes, thereby reducing the accuracy of sample classification. To address this issue, a parameterized fuzzy relation is introduced to determine the inclusion relationship between samples and decisions.

For a fuzzy decision system $\mathcal{FDS} = (U, AT \cup \{d\}, g)$, the fuzzy neighborhood radius of a sample is denoted as α, which is used to describe the similarity between two objects; for any $a \in AT$, and $\forall x, y \in U$, the fuzzy neighborhood similarity relation between two objects x and y on a can be represented as

$$R_a(x,y) = \begin{cases} 0, & |g_a(x) - g_a(y)| > \alpha \\ 1 - |g_a(x) - g_a(y)|, & |g_a(x) - g_a(y)| \le \alpha \end{cases} \tag{1}$$

The fuzzy neighborhood similarity matrix $[x]_a(y) = R_a(x,y)$ can be obtained from Equation (1). Similarly, for $N \subseteq AT$, the fuzzy neighborhood similarity matrix based on N is denoted as $[x]_N(y) = R_N(x,y) = \min_{a \in N}\left([x]_a(y)\right)$.

Given $\mathcal{FDS} = (U, AT \cup \{d\}, g)$, $N \subseteq AT, \forall x, y \in U$, the fuzzy neighborhood granule of object x with respect to N is represented as

$$\widetilde{\alpha}_N(x) = [x]_N^\alpha(y) = \begin{cases} 0, & R_N(x,y) < 1 - \alpha \\ R_N(x,y), & R_N(x,y) \ge 1 - \alpha \end{cases} \tag{2}$$

Definition 1. For any $X \subseteq U$, $\widetilde{\alpha}_N(x)$ is the fuzzy neighborhood granule of $x \in U$. Then the fuzzy neighborhood lower and upper approximations under N are respectively represented as

$$\begin{aligned} &\underline{R_{FN}^\alpha}(X) = \{x | \widetilde{\alpha}_N(x) \subseteq X, x \in U\}, \\ &\overline{R_{FN}^\alpha}(X) = \{x | \widetilde{\alpha}_N(x) \cap X \neq \emptyset, x \in U\}. \end{aligned} \tag{3}$$

The pair $\left(\underline{R_{FN}^\alpha}(X), \overline{R_{FN}^\alpha}(X)\right)$ is called fuzzy neighborhood rough set (**FNRS**).

For any $d_j \in U/d = \{d_1, d_2, \ldots, d_s\}$, the fuzzy neighborhood positive region of $\{d\}$ under N, as well as its dependency degree, are represented as

$$\begin{aligned} POS_N^\alpha(d) &= \bigcup_{j=1}^{s} \underline{R_{FN}^\alpha}(d_j), \\ \Upsilon_N^\alpha(d) &= \frac{\left|POS_N^\alpha(d)\right|}{|U|}. \end{aligned} \tag{4}$$

2.2. Neighborhood multi-granulation rough set

To overcome the limitation of classical rough set theory in handling numerical data, we utilize the Euclidean distance to describe the neighborhood relation defined on $N \subseteq AT$ for any two objects in the domain U. When the distance between two objects is less than or equal to the given neighborhood radius δ, they are regarded as identical. $[x]_N^\delta$ represents the neighborhood class of x on N and is denoted as

$$[x]_N^\delta = \left\{y | \Delta_N(x,y) \le \delta, y \in U\right\} \tag{5}$$

Definition 2. Let $\mathcal{FDS} = (U, AT \cup \{d\}, g)$ be a fuzzy decision system. Suppose $X \subseteq U$, $N \subseteq AT$ and $N = \{N_1, N_2, \ldots, N_r\}$. Then the optimistic neighborhood multigranulation lower and upper approximations of X with respect to N_i be respectively represented as

$$\begin{aligned} &\underline{OM_{\sum_{i=1}^r R_{N_i}}}(X) = \left\{x \in U | \bigvee_{i=1}^{r} \left([x]_{N_i}^\delta \subseteq X\right)\right\}, \\ &\overline{OM_{\sum_{i=1}^r R_{N_i}}}(X) = \left\{x \in U | \bigwedge_{i=1}^{r} \left([x]_{N_i}^\delta \cap X \neq \emptyset\right)\right\}, \end{aligned} \tag{6}$$

where "$\bigvee$" denotes *or*, "$\bigwedge$" denotes *and*, $\delta_{R_{N_i}}$ represents the neighborhood class of x on $N_i \subseteq N$, then the pair $\left(\underline{OM_{\sum_{i=1}^r R_{N_i}}}(X), \overline{OM_{\sum_{i=1}^r R_{N_i}}}(X)\right)$ is called optimistic multigranulation neighborhood rough set (**OMNRS**).

Assuming $d_j \in U/d = \{d_1, d_2, \ldots, d_s\}$ in OMNRS, the optimistic positive region of $\{d\}$ relative to N and its dependency degree can be respectively represented as

$$\begin{aligned} POS_N^O(d) &= \bigcup_{j=1}^{s} \underline{OM_{\sum_{i=1}^r R_{N_i}}}(d_j), \\ \Upsilon_N^O(d) &= \frac{\left|POS_D^O(d)\right|}{|U|}, \end{aligned} \tag{7}$$

where $N_i \subseteq A$, $i = 1, 2, \ldots, r$.

Definition 3. Let $\mathcal{FDS} = (U, AT \cup \{d\}, g)$ be a fuzzy decision system. Suppose $X \subseteq U$, $N \subseteq AT$ and $N = \{N_1, N_2, \ldots, N_r\}$. Then the pessimistic neighborhood multigranulation lower and upper approximations of X with respect to N_i be respectively represented as

$$\begin{aligned} \underline{PM_{\sum_{i=1}^r R_{N_i}}}(X) &= \left\{ x \in U \mid \bigwedge_{i=1}^{r} \left([x]_{N_i}^{\delta} \subseteq X \right) \right\}, \\ \overline{PM_{\sum_{i=1}^r R_{N_i}}}(X) &= \left\{ x \in U \mid \bigvee_{i=1}^{r} \left([x]_{N_i}^{\delta} \cap X \neq \emptyset \right) \right\}. \end{aligned} \tag{8}$$

The pair $\left(\underline{PM_{\sum_{i=1}^r R_{N_i}}}(X), \overline{PM_{\sum_{i=1}^r R_{N_i}}}(X)\right)$ is called pessimistic multigranulation neighborhood rough set (**PMNRS**).

Assuming $d_j \in U/d = \{d_1, d_2, \ldots, d_s\}$ in PMNRS, the pessimistic positive region of $\{d\}$ relative to N and its dependency degree can be respectively represented as

$$\begin{aligned} POS_N^P(d) &= \bigcup_{j=1}^{s} \underline{PM_{\sum_{i=1}^r R_{N_i}}}(d_j), \\ \Upsilon_N^P(d) &= \frac{\left|POS_N^P(d)\right|}{|U|}, \end{aligned} \tag{9}$$

where $N_i \subseteq N$, $i = 1, 2, \ldots, r$.

2.3. Generalized multigranulation rough set

An object $x \in U$ is considered a pessimistic lower approximation element only if its equivalence class is a subset of the target concept's equivalence class in all granularities; and if the equivalence class of x is contained in only one granularity, it can be considered a member of the optimistic lower approximation. From the above perspective, pessimism is too strict, and optimistic ones are too lenient, making them both unsuitable for practical applications. Therefore, scholars have proposed the generalized multi-granulation rough set model, which effectively addresses this problem.

Let $\mathcal{FDS} = (U, AT \cup \{d\}, g)$ be a fuzzy decision system, $A \subseteq AT$, $A_i = \{A_1, A_2, \ldots, A_r\}$, let $P_X^{R_{A_i}}(x)$ be the support characteristic function of x describing the equivalence class $[x]_{A_i}$ and the inclusion relationship of X, which is represented as

$$P_X^{R_{A_i}}(x) = \begin{cases} 1, & [x]_{A_i} \subseteq X \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

Definition 4. For a support characteristic function $P_X^{R_{A_i}}(x)$ $(i = 1, 2, \ldots, r)$, and the threshold $\beta \in (0.5, 1]$, then the generalize multi-granulation lower and upper approximations are represented as

$$\begin{aligned} \underline{GM^{\beta}_{\sum_{i=1}^r R_{A_i}}}(X) &= \left\{ x \in U \mid \frac{\sum_{i=1}^r P_X^{R_{A_i}}(x)}{r} \geq \beta \right\}, \\ \overline{GM^{\beta}_{\sum_{i=1}^r R_{A_i}}}(X) &= \left\{ x \in U \mid \frac{\sum_{i=1}^r \left(1 - P_{\sim X}^{R_{A_i}}(x)\right)}{r} \geq 1 - \beta \right\}, \end{aligned} \tag{11}$$

where β is the ratio between the value of the support function and r, as β increases, our requirements become stricter. Then the pair $\left(\underline{GM^{\beta}_{\sum_{i=1}^r R_{A_i}}}(X), \overline{GM^{\beta}_{\sum_{i=1}^r R_{A_i}}}(X)\right)$ is called generalize multigranulation rough set (**GMRS**).

The generalized multigranulation positive region of $\{d\}$ relative to A and its dependency degree can be respectively represented as

$$POS_A^{G,\beta}(d) = \bigcup_{j=1}^{s} \underline{GM^{\beta}_{\sum_{i=1}^{r} R_{A_i}}}(d_j),$$
$$\Upsilon_A^{G,\beta}(d) = \frac{\left|POS_A^{G,\beta}(d)\right|}{|U|}, \tag{12}$$

where $A_i \subseteq A$, $i = 1,2,\dots,r$, $d_j \in U/d$ and $j = 1,2,\dots,s$.

3. Uncertainty measures of the GMFNRS model based on fuzzy neighborhood entropy

Optimistic multi-Granulation rough set and pessimistic multi-Granulation rough set were proposed by Xu et al. [32]. However, the optimistic and pessimistic scenarios are not applicable in real-life situations. Thus, we propose the generalized multi-granulation fuzzy neighborhood rough set, which offers a higher degree of flexibility by incorporating a threshold. This allows us to strike a balance between optimism and pessimism, making the model more adaptable and applicable in real-world scenarios.

3.1. GMFNRS model

Given $\mathcal{FDS} = (U, AT \cup \{d\}, g)$, $N \subseteq AT$, $N_i = \{N_1, N_2, \dots, N_r\}$, let $P_X^{R_{N_i}}(x)$ be the support characteristic function of x describing the fuzzy neighborhood granules $\widetilde{\alpha}_{N_i}(x)$ and the inclusion relationship of X, which is represented as

$$P_X^{R_{N_i}}(x) = \begin{cases} 1, & \widetilde{\alpha}_{N_i}(x) \subseteq X \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

Definition 5. For a threshold $\beta \in (0.5, 1]$, the generalize multigranulation fuzzy neighborhood lower and upper approximations are represented as

$$\underline{GM^{\alpha,\beta}_{\sum_{i=1}^{r} R_{N_i}}}(X) = \left\{ x \in U \mid \frac{\sum_{i=1}^{r} P_X^{R_{N_i}}(x)}{r} \geq \beta \right\}, \tag{14}$$

$$\overline{GM^{\alpha,\beta}_{\sum_{i=1}^{r} R_{N_i}}}(X) = \left\{ x \in U \mid \frac{\sum_{i=1}^{r} \left(1 - P_{\sim X}^{R_{N_i}}(x)\right)}{r} \geq 1 - \beta \right\}. \tag{15}$$

The pair $\left(\underline{GM^{\alpha,\beta}_{\sum_{i=1}^{r} R_{N_i}}}(X), \overline{GM^{\alpha,\beta}_{\sum_{i=1}^{r} R_{N_i}}}(X) \right)$ is called generalize multigranulation fuzzy neighborhood rough set (**GMFNRS**).

And the generalized multigranulation fuzzy neighborhood positive region of $\{d\}$ relative to N and its dependency degree can be respectively represented as

$$POS_N^{G,\beta}(d) = \bigcup_{j=1}^{s} \underline{GM^{\alpha,\beta}_{\sum_{i=1}^{r} R_{N_i}}}(d_j), \tag{16}$$

$$\Upsilon_N^{G,\beta}(d) = \frac{\left|POS_N^{G,\beta}(d)\right|}{|U|}, \tag{17}$$

where $N_i \subseteq N$, $i = 1,2,\dots,r$, $d_j \in U/d$ and $j = 1,2,\dots,s$.

Proposition 1. *Given $\mathcal{FDS} = (U, AT \cup \{d\}, g)$ with any $X \subseteq U$, $P \subseteq Q \subseteq AT$, the following properties hold:*

1) $POS_P^{G,\beta}(d) \subseteq POS_Q^{G,\beta}(d)$.
2) $\Upsilon_P^{G,\beta}(d) \leq \Upsilon_Q^{G,\beta}(d)$.

Proof. 1) For any $P \subseteq Q$ and $x, y \in U$, suppose that $P_i \subseteq P \subseteq AT$, $Q_i \subseteq Q \subseteq AT$, $i = 1,2,\dots,|P|$ and $j = 1,2,\dots,|Q|$, according to the concept of fuzzy similarity relation, it follows that $R_P \subseteq R_Q$, and $\widetilde{\alpha}_P(x) \subseteq \widetilde{\alpha}_Q(x)$ holds. For any $X \subseteq U$, from Eq. (14), $\underline{GM^{\alpha,\beta}_{\sum_{i=1}^{r} R_{P_i}}}(X) \subseteq \underline{GM^{\alpha,\beta}_{\sum_{i=1}^{r} R_{Q_i}}}(X)$ can be obtained. Therefore, it follows that $POS_P^{G,\beta}(d) \subseteq POS_Q^{G,\beta}(d)$.

2) Clearly, according to 1), it satisfies $\Upsilon_P^{G,\beta}(d) \leq \Upsilon_Q^{G,\beta}(d)$.

Table 1
A fuzzy decision system.

U	a_1	a_2	a_3	a_4	a_5	a_6	d
x_1	0.58	0.13	1	0.78	0.16	0.51	1
x_2	0.45	0.4	0.73	0.57	0	0.82	1
x_3	0.4	0.33	0.8	0.63	0.1	0.77	1
x_4	0.36	0.48	0.45	0.59	0.24	0.2	2
x_5	0.85	0.39	0.51	0.29	0.02	0.15	2
x_6	0.55	0.63	0.42	0.75	0.88	0.07	2

Example 1. A fuzzy information system with decisions is illustrated in Table 1, where $\alpha = 0.3$. The domain of we studied is $U = \{x_1, x_2, x_3, x_4, x_5, x_6\}$, and N consists of six conditional attributes $\{a_1, a_2, a_3, a_4, a_5, a_6\}$, with d being the decision attribute. We have $U/d = \{d_1, d_2\}$, where $d_1 = \{x_1, x_2, x_3\}$ and $d_2 = \{x_4, x_5, x_6\}$. To facilitate subsequent calculations, we divide the conditional attributes into granules, with every two attributes forming one granule, that is, $N_1 = \{a_1, a_2\}$, $N_2 = \{a_3, a_4\}$ and $N_3 = \{a_5, a_6\}$.

To obtain the approximation in the generalized multi-granularity fuzzy neighborhood, we first calculate the fuzzy neighborhood similarity matrix between any two objects for each attribute, and then derive the fuzzy neighborhood granule matrix for each granularity. If the calculated fuzzy neighborhood similarity is 0, it indicates that there is no fuzzy similarity relationship between these two objects. In other words, one object is not in the fuzzy neighborhood class of the other.

According to $N_1 = \{a_1, a_2\}$, and the fuzzy neighborhood similarity matrices $[x]_{a_1}(y)$ and $[x]_{a_2}(y)$ for attributes a_1 and a_2, the fuzzy neighborhood granule matrix M_1 is:

$$[x]_{a_1}(y) = \begin{bmatrix} 1 & 0.87 & 0.82 & 0.78 & 0.73 & 0.97 \\ 0.87 & 1 & 0.95 & 0.91 & 0 & 0.9 \\ 0.82 & 0.95 & 1 & 0.96 & 0 & 0.85 \\ 0.78 & 0.91 & 0.96 & 1 & 0 & 0.81 \\ 0.73 & 0 & 0 & 0 & 1 & 0.7 \\ 0.97 & 0.9 & 0.85 & 0.81 & 0.7 & 1 \end{bmatrix}, [x]_{a_2}(y) = \begin{bmatrix} 1 & 0.73 & 0.8 & 0 & 0.74 & 0 \\ 0.73 & 1 & 0.93 & 0.92 & 0.99 & 0.77 \\ 0.8 & 0.93 & 1 & 0.85 & 0.94 & 0.7 \\ 0 & 0.92 & 0.85 & 1 & 0.91 & 0.85 \\ 0.74 & 0.99 & 0.94 & 0.91 & 1 & 0.76 \\ 0 & 0.77 & 0.7 & 0.85 & 0.76 & 1 \end{bmatrix}.$$

$$M_1 = \min\left([x]_{a_1}(y), [x]_{a_2}(y)\right) = \begin{bmatrix} 1 & 0.73 & 0.8 & 0 & 0.73 & 0 \\ 0.73 & 1 & 0.93 & 0.91 & 0 & 0.77 \\ 0.8 & 0.93 & 1 & 0.85 & 0 & 0.7 \\ 0 & 0.91 & 0.85 & 1 & 0 & 0.81 \\ 0.73 & 0 & 0 & 0 & 1 & 0.7 \\ 0 & 0.77 & 0.7 & 0.81 & 0.7 & 1 \end{bmatrix}.$$

Then, the fuzzy neighborhood classes on N_1 are:

$$\widetilde{\alpha}_{N_1}(x_1) = \{x_1, x_2, x_3, x_5\}, \quad \widetilde{\alpha}_{N_1}(x_2) = \{x_1, x_2, x_3, x_4, x_6\}, \quad \widetilde{\alpha}_{N_1}(x_3) = \{x_1, x_2, x_3, x_4, x_6\},$$
$$\widetilde{\alpha}_{N_1}(x_4) = \{x_2, x_3, x_4, x_6\}, \quad \widetilde{\alpha}_{N_1}(x_5) = \{x_1, x_5, x_6\}, \quad \widetilde{\alpha}_{N_1}(x_6) = \{x_2, x_3, x_4, x_5, x_6\}.$$

According to $N_2 = \{a_3, a_4\}$, the fuzzy neighborhood granule matrix M_2 is:

$$M_2 = \begin{bmatrix} 1 & 0.73 & 0.8 & 0 & 0 & 0 \\ 0.73 & 1 & 0.93 & 0.72 & 0.72 & 0 \\ 0.8 & 0.93 & 1 & 0 & 0 & 0 \\ 0 & 0.72 & 0 & 1 & 0.7 & 0.84 \\ 0 & 0.72 & 0 & 0.7 & 1 & 0 \\ 0 & 0 & 0 & 0.84 & 0 & 1 \end{bmatrix}.$$

Similarly, the fuzzy neighborhood classes on N_2 are:

$$\widetilde{\alpha}_{N_2}(x_1) = \{x_1, x_2, x_3\}, \quad \widetilde{\alpha}_{N_2}(x_2) = \{x_1, x_2, x_3, x_4, x_5\}, \quad \widetilde{\alpha}_{N_2}(x_3) = \{x_1, x_2, x_3\},$$
$$\widetilde{\alpha}_{N_2}(x_4) = \{x_2, x_4, x_5, x_6\}, \quad \widetilde{\alpha}_{N_2}(x_5) = \{x_2, x_4, x_5\}, \quad \widetilde{\alpha}_{N_2}(x_6) = \{x_4, x_6\}.$$

According to $N_3 = \{a_5, a_6\}$, the fuzzy neighborhood granule matrix M_3 is:

$$M_3 = \begin{bmatrix} 1 & 0 & 0.74 & 0 & 0 & 0 \\ 0 & 1 & 0.9 & 0 & 0 & 0 \\ 0.74 & 0.9 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.78 & 0 \\ 0 & 0 & 0 & 0.78 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

We obtain the fuzzy neighborhood classes on N_3 are:

$$\widetilde{\alpha}_{N_3}(x_1)=\{x_1,x_3\},\quad \widetilde{\alpha}_{N_3}(x_2)=\{x_2,x_3\},\quad \widetilde{\alpha}_{N_3}(x_3)=\{x_1,x_2,x_3\},$$
$$\widetilde{\alpha}_{N_3}(x_4)=\{x_4,x_5\},\quad \widetilde{\alpha}_{N_3}(x_5)=\{x_4,x_5\},\quad \widetilde{\alpha}_{N_3}(x_6)=\{x_6\}.$$

Assuming $\beta=0.6$, we can obtain that $\underline{GM^{\alpha,\beta}_{\sum_{i=1}^{r}R_{N_i}}}(d_1)=\{x_1,x_3\}$, $\underline{GM^{\alpha,\beta}_{\sum_{i=1}^{r}R_{N_i}}}(d_2)=\{x_6\}$.

3.2. Fuzzy neighborhood composite entropy

The rough set model is employed to quantify the uncertainty arising from the disparity between lower approximation and upper approximation. This measure of uncertainty solely captures the uncertainty associated with the features present in the attribute subset. Therefore, in order to enhance decision analysis, we propose a joint entropy based on the generalized multigranulation fuzzy neighborhood rough set.

Given $\mathcal{FDS}=(U,AT\cup\{d\},g)$, $U=\{x_1,x_2,\ldots,x_n\}$, $N\subseteq AT$, where $U/d=\{d_1,d_2,\ldots,d_s\}$. The fuzzy decision induced by $\{d\}$ for the samples can be represented as

$$\widetilde{FD}=\{\widetilde{FD}_1,\widetilde{FD}_2,\ldots,\widetilde{FD}_s\} \tag{18}$$

where $\widetilde{FD}_j=\{\widetilde{FD}_j(x_1),\widetilde{FD}_j(x_2),\ldots,\widetilde{FD}_j(x_n)\}$ represents the fuzzy neighborhood decision equivalence class of objects partitioned by $\{d\}$. When $k=1,2,\ldots,n$, $i=1,2,\ldots,r$, for any $x,y\in U$, $\widetilde{FD}_j(x_k)$ represents the degree of membership of $x_k\in U$ to $\widetilde{FD}_j$ on N and represented as

$$\widetilde{FD}_j(x_k)=\frac{\left|\left[x_k\right]_N(y)\cap d_j\right|}{\left|\left[x_k\right]_N(y)\right|} \tag{19}$$

where $\left[x_k\right]_N(y)$ is the fuzzy neighborhood similarity degree of objects on N, $d_j\in U/d$, and $j=1,2,\ldots,s$.

Definition 6. Let $\widetilde{\alpha}_N(x)$ be the fuzzy neighborhood granule of object x, the fuzzy neighborhood entropy of N and the fuzzy neighborhood joint entropy regarding N and $\{d\}$ are respectively represented as

$$\begin{aligned} FNE(N)&=-\sum_{k=1}^{n}\frac{\left|\widetilde{\alpha}_N(x_k)\right|}{|U|}\log_2\frac{\left|\widetilde{\alpha}_N(x_k)\right|}{|U|},\\ FNE(N,d)&=-\sum_{j=1}^{s}\sum_{k=1}^{n}\frac{\left|\widetilde{\alpha}_N(x_k)\cap FD_j\right|}{|U|}\log_2\frac{\left|\widetilde{\alpha}_N(x_k)\cap FD_j\right|}{|U|}\end{aligned} \tag{20}$$

where $x_k\in U$, $k=\{1,2,\ldots,n\}$, FD_j represents the fuzzy neighborhood decision equivalence class of objects, $j=\{1,2,\ldots,s\}$, and $\left|\widetilde{\alpha}_N(x_k)\cap FD_j\right|$ denotes the count of objects in $\widetilde{\alpha}_N(x_k)$ with a membership degree is not greater than FD_j.

Definition 7. Given $\mathcal{FDS}=(U,AT\cup\{d\},g)$, $U=\{x_1,x_2,\ldots,x_n\}$, $N\subseteq AT$, $N_i=\{N_1,N_2,\ldots,N_r\}$ and $U/d=\{d_1,d_2,\ldots,d_s\}$, the fuzzy neighborhood generalized composite entropy (**FNGCE**) regarding $\{d\}$ in relation to N_i is represented as

$$FNGCE(N,d)=\left(1-\Upsilon^{G,\beta}_N(d)\right)\left[-\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{n}\frac{\left|\widetilde{\alpha}_{N_i}(x_k)\cap FD_j\right|}{|U|}\log_2\frac{\left|\widetilde{\alpha}_{N_i}(x_k)\cap FD_j\right|}{|U|}\right] \tag{21}$$

where $\Upsilon^{G,\beta}_N(d)$ denotes the generalized multigranulation fuzzy neighborhood dependency degree of $\{d\}$ on N_i, $i=\{1,2,\ldots,r\}$, $\widetilde{\alpha}_{N_i}(x_k)$ represents the fuzzy neighborhood granules of $x_k\in U$ under N_i, and $\left|\widetilde{\alpha}_{N_i}(x_k)\cap FD_j\right|$ denotes the count of objects in $\widetilde{\alpha}_{N_i}(x_k)$ with a membership degree is not greater than FD_j, $j=\{1,2,\ldots,s\}$.

Proposition 2. *Given $\mathcal{FDS}=(U,AT\cup\{d\},g)$ with any $N_i\subseteq N\subseteq AT$, $N_i=\{N_1,N_2,\ldots,N_r\}$, then one has $FNGCE(N,d)=\sum_{i=1}^{r}\left(1-\Upsilon^{G,\beta}_N(d)\right)\times FNE(N_i,d)\geq 0$.*

Proof. According to the definition of fuzzy neighborhood decision classes, it can be inferred $\frac{\left|\widetilde{\alpha}_{N_i}(x_k)\cap FD_j\right|}{|U|}\in[0,1]$ for all target decision, $\log_2\frac{\left|\widetilde{\alpha}_{N_i}(x_k)\cap FD_j\right|}{|U|}\leq 0$ for $i=\{1,2,\ldots,r\}$. Furthermore, based on the definition of generalized multi-granularity fuzzy neighborhood dependency, $\Upsilon^{G,\beta}_N(d)\geq 0$ holds, then $FNGCE(N,d)\geq 0$. □

Proposition 3. *Given $\mathcal{FDS}=(U,AT\cup\{d\},g)$ with any $X\subseteq U$, $P\subseteq Q\subseteq AT$, then $FNGCE(P,d)\leq FNGCE(Q,d)$.*

Algorithm 1 Feature selection algorithm based on fuzzy neighborhood generalized composite entropy (FSAFNGCE).

Input:
1. A fuzzy decision system $\mathcal{FDS} = (U, AT \cup \{d\}, g)$, where $U/d = \{d_1, d_2, \dots, d_s\}$;
2. The radius of fuzzy neighborhood α, the threshold $\beta \in (0.5, 1]$.

Output: An optimal attribute subset C.

1: Initialize $C \leftarrow \emptyset$, $B \leftarrow \emptyset$.
2: Compute $FNGCE(AT, d)$ of feature set AT.
3: Compute $FNGCE(AT - N_i, d)$ for $N_i \subseteq AT$.
4: **for** $i = 1$ to $|AT|$ **do**
5:     Compute $SIM^{in}(N_i, AT, d)$.
6:     **if** $SIM^{in}(N_i, AT, d) > 0$ **then**
7:         $C \leftarrow C \cup N_i$.
8:     **end if**
9: **end for**
10: Let $B \leftarrow AT - C$.
11: **while** $FNGCE(C, d) \neq FNGCE(AT, d)$ **do**
12:     **for** $j = 1$ to $|B|$ **do**
13:         Compute $SIM^{out}(N_j, C, d)$.
14:     **end for**
15:     Choose $N_j = argmax_{N_j \subseteq B} SIM^{out}(N_j, C, d)$.
16:     $C \leftarrow C \cup N_j$ and $B \leftarrow B - N_j$.
17:     Compute $FNGCE(C, d)$.
18: **end while**
19: **for** $k = 1$ to $|C|$ **do**
20:     **if** $FNGCE(C - N_k, d) = FNGCE(C, d)$ **then**
21:         $C \leftarrow C - N_k$
22:     **end if**
23: **end for**
24: **return** The optimal attribute subset C.

Proof. For any $P \subseteq Q \subseteq AT$, $X \subseteq U$, $x \in U$, according to Proposition 1, $\Upsilon_P^{G,\beta}(d) \leq \Upsilon_Q^{G,\beta}(d)$. From Proposition 2 in [21], it follows that $FNE(P, d) \leq FNE(Q, d)$. When $\widetilde{\alpha}_P(x) = \widetilde{\alpha}_Q(x)$, one has $\Upsilon_P^{G,\beta}(d) = \Upsilon_Q^{G,\beta}(d)$ and $FNE(P, d) = FNE(Q, d)$. Thus, $FNGCE(P, d) \leq FNGCE(Q, d)$ holds. □

4. Feature selection algorithm based on fuzzy neighborhood generalized composite entropy

In the fuzzy neighborhood generalized composite entropy, the degree of dependence can be adjusted by controlling the threshold value β, thereby determining the level of strictness required and allowing the rough set model to be more flexible and adaptable to real-world situations.

Definition 8. Let $\mathcal{FDS} = (U, AT \cup \{d\}, g)$, $N' \subseteq N \subseteq AT$, for $\forall N_i \subseteq N', i = \{1, 2, \dots, r\}$, when $FNGCE(N', d) = FNGCE(N, d)$ and $FNGCE(N', d) \neq FNGCE(N - N_i, d)$, then N_i represents the reduction of N with respect to $\{d\}$.

According to Definition 8, we ensure the effectiveness of the reduction. Firstly, we guarantee that the discernibility of the remaining feature subsets after reduction is the same as that of the original feature set. Secondly, the remaining feature subsets after reduction are all essential and cannot be omitted. Next, we define two attribute importance measures to assist us in selecting the necessary attribute subsets.

Let $\mathcal{FDS} = (U, AT \cup \{d\}, g)$ be a fuzzy decision system, $N' \subseteq N \subseteq AT$, $N_i = \{N_1, N_2, \dots, N_r\}$, $i = \{1, 2, \dots, r\}$, for $\forall N_i \subseteq N'$, the FNGCE-based inner significance measure of attribute subset N_i in relation to N is represented as

$$SIM^{in}(N_i, N, d) = FNGCE(N - N_i, d) - FNGCE(N, d) \tag{22}$$

If the value of $SIM^{in}(N_i, N, d)$ is higher, it indicates that the attribute subset N_i is more important compared to N. Hence, we choose the attribute subset of $SIM^{in}(N_i, N, d) > 0$ as the essential feature subset from the attribute set.

And for $\forall N_i \subseteq N - N'$, the FNGCE-based outer significance measure of attribute subset N_i in relation to N is represented as

$$SIM^{out}(N_i, N, d) = FNGCE(N, d) - FNGCE(N \cup N_i, d) \tag{23}$$

The outer importance plays a crucial role in the selection of feature subsets by identifying significant features that impact the decision outcome while eliminating redundant features with negligible effects. We iteratively choose the most optimal feature subsets that meet the criteria of $argmax_{N_i \subseteq N} SIM^{out}(N_i, N, d)$, ultimately obtaining the reduction.

Algorithm 1 presents the Feature selection algorithm based on fuzzy neighborhood generalized composite entropy (FSAFNGCE), designed on the basis of SIM^{in} and SIM^{out}. The time complexity for calculating the FNGCE of the original conditional attribute set in step 3 is approximately $\mathcal{O}(|AT|^2)$. The first loop, in steps 5–10, selects the attribute subset belonging to the core from the original conditional attribute set based on inner significance measures, and the second loop, in steps 12–19, sequentially adds the attributes with the highest external importance measures to the necessary feature subset, the time complexity of these two loops is $\mathcal{O}(|AT|^3)$

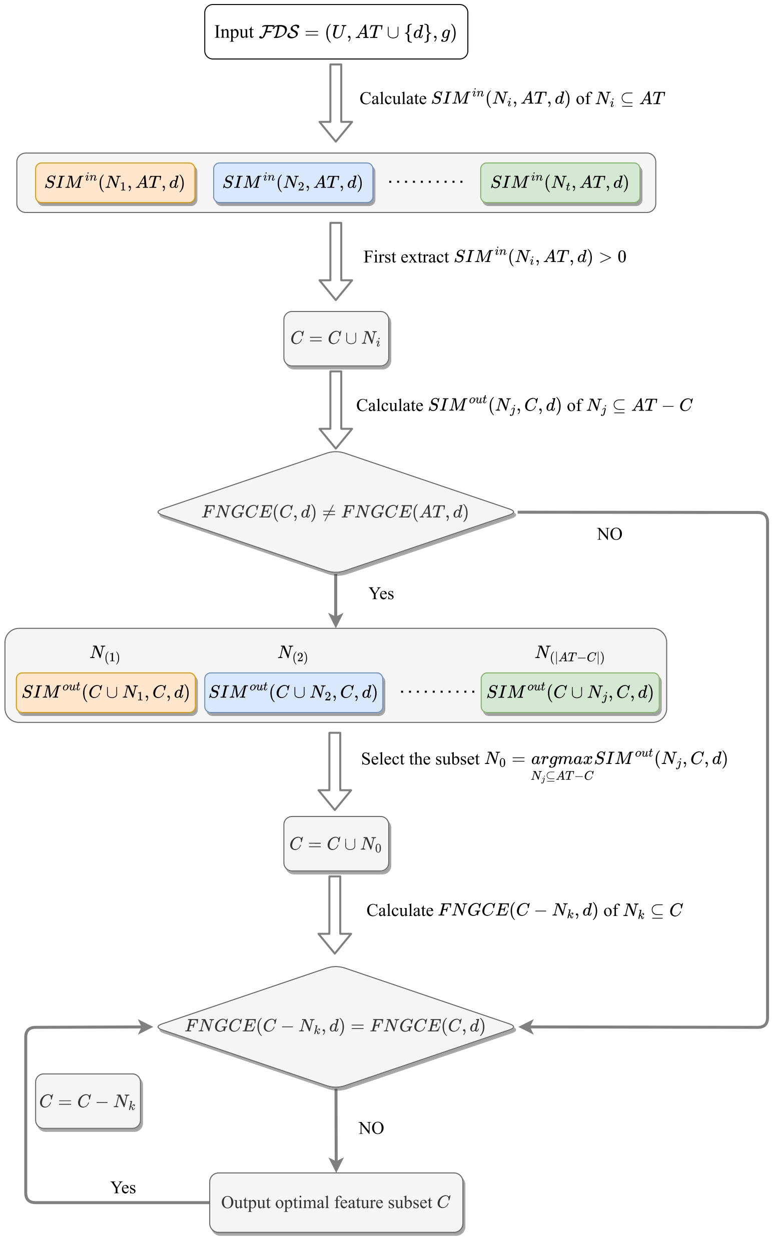


Fig. 2. The process of selecting optimal subsets based on $FNGCE$.

and $\mathcal{O}(|B|^2)$, respectively. The time complexity involved in obtaining the optimal reduction subset by eliminating redundant feature subsets in steps 20–24 is $\mathcal{O}(|C|^2)$ (Fig. 2).

5. Experimental decision and analysis

In this section, we have designed a series of experiments to validate the efficiency and effectiveness of the proposed algorithm (FSAFNGCE) through the following three aspects: (1) the classification accuracy performance of different feature selection algorithms under different classifiers; (2) time comparison for selection algorithm reduction; (3) the impact of different combinations of parameters α and β on the algorithm reduction results.

5.1. Experimental design

Four feature selection algorithms were chosen as comparative methods, and they are detailed below.

Table 2
Dataset description.

Nos.	Datasets	Abbreviation	Samples	Features	Class
1	Wine	Wine	178	13	3
2	Seeds	Seeds	210	7	3
3	Wisconsin diagnostic breast cancer	Wdbc	569	31	2
4	Hill with noise	Hill	606	100	2
5	Breast cancer	Breast	699	10	2
6	Wine quality-red	Wine-red	1599	11	6
7	Cardiotocography	Card	2126	20	3
8	Statlog (Image Segmentation)	Image	2310	19	7
9	Spambase	Spambase	4601	57	2
10	Page-blocks	Page	5473	10	5
11	Musk (Version 2)	Musk2	6598	165	2
12	Electrical Grid Stability Simulated Data	EGSS	10000	13	2
13	DLBCL-Harvard	DLBCL	77	7130	2
14	LungCancer-BrighamAndWomenHospital	LungCancer-1	181	12533	2
15	LungCancer-DanaFarberCancerInstitute	LungCancer-2	203	12600	5

(1) Infinite feature selection (Inf-UFS) [18]: The feature subset is treated as paths in the graph in the filter feature selection algorithm, effectively addressing issues of relevance and redundancy.
(2) The interval-valued feature selection (IGUFS) [31]: A graph-theory-based feature selection method was employed, and its optimization was achieved using the properties of matrix power series, resulting in a significant reduction in time complexity.
(3) Multigranularity entropy-based feature selection [38]: Extended the classical entropy model to a multigranularity entropy model and proposed two MGE models.

Moreover, all experiments were performed on PyCharm 2020.3.5 (Community Edition) with the following configuration. Windows 10 operating system, with the processor being Intel(R) Core(TM) i5-8250U @1.60 GHz 1.80 GHz, and 8.0 GB memory. We downloaded 15 datasets from the UCI and DBC databases, which are summarized in Table 2. To capture objects from multiple perspectives, simulating a multi-granularity context, in the following experiments, we formed feature subsets in the low-dimensional UCI datasets by combining every two features and in the three high-dimensional datasets by combining every four features. When the features cannot be evenly divided, the remaining features are grouped into a feature subset.

To evaluate our algorithm, as well as the four comparative algorithms and the classification performance of the original data, using three classification methods, including decision trees (DT), k-nearest neighbors (KNN), and support vector machines (SVM). During the classification experiments, we employed a ten-fold cross-validation approach for assessment. Additionally, provide the feature selection times for the five algorithms, and compare and analyze them. Furthermore, the impact of different parameter combinations on algorithm classification performance was explored. As two parameters, α controls the proportion of fuzzy neighborhood classes, and β determines the level of strictness in constructing composite entropy. Based on the definitions, varying parameter combinations can result in different classification outcomes. Thus, we set the range for parameter α from 0 to 0.5 with a step size of 0.05. Simultaneously, β was set to values between 0.6 and 0.9 with a step size of 0.1. Finally, a hypothesis test was conducted to delve deeper into the distinctions among our algorithm and the other algorithms.

5.2. Experimental analysis

The feature selection method, particularly under multi-granularity entropy, exhibited excessive time consumption on low-dimensional datasets, with the longest running for over 2 million seconds, approximately equivalent to one month. As a result, experiments on three high-dimensional feature datasets were terminated.

Table 3, Table 4, and Table 5 present the classification accuracy for the original data as well as data reduced using the four comparative methods and our proposed algorithm, under the DT, SVM, and KNN classifiers. The bolded numbers within each row of the three tables indicate the highest classification accuracy achieved for the corresponding dataset across all six algorithms. We consider the first 12 low-dimensional feature datasets as normal datasets, while the remaining three datasets (DLBCL, LungCancer-1 and lungCancer-2) have high-dimensional features. As seen in the tables, the classification accuracy of our proposed FSAFNGCE is notably higher than that of the other methods, except for two datasets with DT classifiers, four datasets with SVM classifiers, and four datasets with KNN classifiers. Especially with the DT classifier, our proposed FSAFNGCE method remains superior in classification accuracy compared to other algorithms, except for the Breast dataset where it is only 0.73% lower than the PMGE method, and the Image dataset where it is only 0.56% lower than the IGUFS algorithm. Furthermore, the FSAFNGCE algorithm demonstrates significantly improved and the highest average classification accuracy across the three classifiers on low-dimensional datasets. Moreover, for the three high-dimensional datasets, except for the DLBCL dataset under the SVM classifier, our algorithm outperforms the other two selection algorithms in terms of accuracy in SVM and in both DT and KNN classifiers. Table 6 provides a comparison of the reduction time for the five algorithms across seven datasets. The bold numbers in the table represent the minimum time spent by the algorithm proposed in this paper across seven datasets. While our algorithm is only faster than the other three on three of the low-dimensional datasets, it is still faster than OMGE and PMGE on the remaining nine datasets. Although the

Table 3
The accuracy of data reduction based on DT (%).

Datasets	ODP	Inf-UFS	IGUFS	OMGE	PMGE	FNGCE
Wine	89.38±8.37	83.17±8.57	90.42±6.69	91.47±8.21	89.90±5.43	**92.22±6.67**
Seeds	90.00±6.88	83.81±6.1	91.43±7.00	90.00±6.88	89.52±7.62	**93.33±5.71**
Wdbc	92.98±3.59	79.07±5.75	93.86±3.25	94.03±3.06	92.79±2.42	**94.38±3.31**
Hill	53.14±3.41	52.81±4.08	52.83±5.47	52.81±2.93	54.14±4.82	**55.30±5.03**
Breast	93.56±2.27	93.41±2.39	92.09±3.86	94.15±2.06	**95.17±1.85**	94.44±1.91
Wine-red	59.04±2.90	59.04±3.45	55.97±2.83	58.85±2.40	58.04±3.44	**59.73±2.34**
Card	83.16±3.14	62.70±3.27	77.71±2.87	66.74±3.33	81.42±3.22	**83.16±2.34**
Image	93.90±1.14	87.19±1.56	**94.72±1.30**	70.65±2.30	85.97±2.28	94.16±1.15
Spambase	91.81±0.59	85.24±1.70	88.02±0.80	87.94±1.59	65.46±2.30	**92.07±0.54**
Page	96.53±1.12	95.80±0.85	94.28±0.77	96.49±0.77	93.55±1.15	**96.73±0.72**
Musk2	96.51±0.60	46.74±2.15	38.13±3.91	18.52±2.30	36.95±3.53	**96.80±0.60**
EGSS	99.98±0.04	77.18±1.08	77.85±1.39	65.97±1.09	82.42±1.27	**99.99±0.01**
Average	86.67±2.84	75.51±3.41	78.94±3.35	73.97±3.08	77.11±3.28	**87.69±2.53**
DLBCL	76.61±16.28	79.46±15.07	78.39±12.33	–	–	**85.89±10.82**
LungCancer-1	90.64±6.51	91.32±10.02	88.98±6.50	–	–	**93.98±6.17**
LungCancer-2	87.14±6.36	85.71±4.08	63.02±5.71	–	–	**90.57±7.24**

Table 4
The accuracy of data reduction based on SVM (%).

Datasets	ODP	Inf-UFS	IGUFS	OMGE	PMGE	FNGCE
Wine	94.97±4.62	86.54±8.33	97.75±3.72	96.63±3.71	94.93±4.62	**98.33±2.55**
Seeds	93.33±4.36	90.48±5.22	93.33±5.30	93.33±4.36	93.33±4.36	**94.29±3.56**
Wdbc	94.02±3.94	86.47±3.43	**95.78±2.1**	95.43±3.44	95.25±3.05	95.08±2.92
Hill	56.27±4.45	**57.75±3.14**	55.78±3.76	56.27±4.45	54.78±4.50	56.27±4.54
Breast	96.19±1.88	94.44±3.17	95.03±2.08	96.19±1.88	95.91±1.92	**96.63±1.74**
Wine-red	62.98±3.05	59.16±3.00	56.28±2.30	62.98±3.05	62.85±3.74	**63.10±4.06**
Card	90.92±1.87	65.05±3.64	84.05±1.72	75.40±3.13	90.83±1.70	**92.19±2.20**
Image	97.06±1.04	86.62±2.62	**97.97±0.75**	69.13±2.87	88.92±1.61	97.49±0.54
Spambase	93.61±0.86	86.33±1.55	86.09±1.16	89.02±0.85	64.23±2.41	**93.94±0.87**
Page	90.72±0.96	94.62±0.47	93.81±0.69	92.91±0.97	93.82±0.60	**94.65±1.13**
Musk2	91.75±1.78	41.59±1.56	40.42±1.73	24.16±1.98	40.48±1.86	**91.77±1.80**
EGSS	99.17±0.30	79.58±1.44	81.06±1.43	66.44±1.25	92.84±0.56	**99.39±0.23**
Average	88.42±2.43	77.39±3.13	81.45±2.23	76.49±2.66	80.68±2.58	**89.43±2.18**
DLBCL	73.93±11.72	88.21±13.49	**96.07±6.02**	–	–	77.86±9.98
LungCancer-1	82.89±8.71	98.33±3.56	97.11±4.97	–	–	**98.89±3.70**
LungCancer-2	88.19±3.81	92.14±4.91	67.93±6.58	–	–	**92.71±6.05**

time is somewhat longer on these nine datasets compared to Inf-UFS and IGUFS, our algorithm outperforms these two in terms of classification accuracy across the three classifiers. Furthermore, on the three high-dimensional datasets, DLBCL and LungCancer-1 exhibit significantly lower time complexity compared to Inf-UFS and IGUFS. Even though the time on LungCancer-2 is not the shortest, it is still competitive with these two algorithms. Moreover, the classification accuracy on all three classifiers, except for one dataset under SVM, surpasses that of Inf-UFS and IGUFS, with differences ranging from 0.56% to 29.07%. Across different classifiers and datasets, it is evident that our proposed algorithm excels at selecting essential features, resulting in a substantial enhancement in classification accuracy, as indicated by the data presented in Tables 3, 4, and 5. It consistently demonstrates strong classification performance, displaying a competitive edge.

5.3. Statistical testing

To further examine whether there are significant differences in the classification performance among different algorithms, we employed the Wilcoxon test method to validate the effectiveness of algorithm comparison. At a significance level of 0.05, for each hypothesis test, the null hypothesis suggests no significant difference between our proposed algorithm and other algorithms. The P-values obtained from the Wilcoxon test are presented in Table 7.

When rejecting the null hypothesis, it implies a significant distinction between our algorithm and the others. From the table, it's evident that the feature selection algorithm proposed in this paper, whether tested in DT, SVM, or KNN classifiers, consistently rejects the null hypothesis. It suggests that FSAFNGCE displays a significant distinction from other feature selection algorithms.

5.4. Experimental parameter

Furthermore, to delve into the impact of the fuzzy domain parameter α and information level parameter β on the classification accuracy of the proposed FSAFNGCE algorithm under different combinations, we provided the optimal parameter combinations for

Table 5
The accuracy of data reduction based on KNN (%).

Datasets	ODP	Inf-UFS	IGUFS	OMGE	PMGE	FNGCE
Wine	94.97±9.77	87.09±8.24	**97.78±5.09**	96.08±4.35	95.52±4.85	96.67±6.67
Seeds	92.86±3.84	87.14±8.53	91.90±5.65	92.86±3.84	92.86±3.84	**93.33±3.16**
Wdbc	96.84±1.53	84.87±6.61	96.49±1.92	95.25±2.24	95.26±2.48	**97.37±1.62**
Hill	51.01±6.47	50.18±7.19	**51.5±5.90**	51.01±6.47	49.36±7.31	51.34±6.50
Breast	96.64±2.26	94.73±3.33	95.47±2.55	**96.64±2.26**	95.47±2.63	96.35±2.69
Wine-red	57.60±2.92	57.41±2.91	54.22±2.87	57.60±2.92	57.66±3.42	**58.10±3.71**
Card	80.99±2.51	66.18±2.10	80.71±2.03	71.72±3.77	**81.93±2.85**	81.23±3.01
Image	96.32±1.03	81.60±2.59	96.28±1.24	70.22±3.36	87.88±2.27	**96.75±1.12**
Spambase	90.22±1.21	80.63±2.39	82.96±1.42	86.68±1.50	40.73±3.21	**90.41±1.14**
Page	95.52±0.50	94.04±0.62	93.75±0.64	94.61±0.73	93.75±0.64	**95.72±0.87**
Musk2	96.57±0.87	41.15±1.92	39.35±1.43	21.14±1.88	40.24±1.30	**96.71±0.73**
EGSS	90.06±0.94	78.72±0.94	79.72±1.57	63.15±1.73	88.00±1.04	**90.19±0.93**
Average	86.63±2.82	75.31±3.95	80.01±2.69	74.75±2.92	76.56±2.99	**87.01±2.68**
DLBCL	85.71±13.39	87.32±14.8	86.21±16.04	–	–	**88.57±14.22**
LungCancer-1	91.73±5.09	92.84±3.50	92.84±4.29	–	–	**98.36±2.50**
LungCancer-2	90.12±5.90	90.60±3.56	64.95±8.07	–	–	**94.02±3.75**

Table 6
The reduction time of different algorithms(s).

Datasets \ Algorithms	Inf-UFS	IGUFS	OMGE	PMGE	FNGME
Wine	1.180517912	1.803053379	199.2320955	32.94587159	**0.384840727**
Seeds	0.773453236	1.204103708	106.3903317	6.037920475	**0.333144665**
Breast	3.502872705	5.165860415	610434.4502	287.4281721	**3.432308197**
DLBCL	1.88×10^4	3.05×10^4	–	–	$\mathbf{0.61 \times 10^4}$
LungCancer-1	1.09×10^5	1.23×10^5	–	–	$\mathbf{0.69 \times 10^5}$
LungCancer-2	1.45×10^5	$\mathbf{1.13 \times 10^5}$	–	–	$\mathbf{1.18 \times 10^5}$

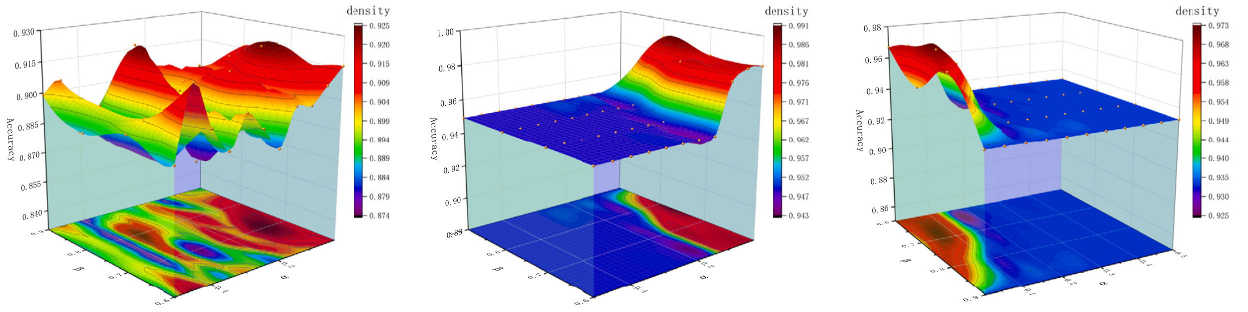
Table 7
P value of Wilcoxon test on three classifiers.

	ODP	Inf-UFS	IGUFS	OMGE	PMGE
KNN	8.36×10^{-3}	6.71×10^{-3}	0.02	0.03	0.04
SVM	9.79×10^{-4}	6.71×10^{-3}	0.01	5.85×10^{-3}	9.77×10^{-4}
DT	9.81×10^{-4}	6.10×10^{-5}	1.83×10^{-4}	4.88×10^{-4}	9.77×10^{-4}

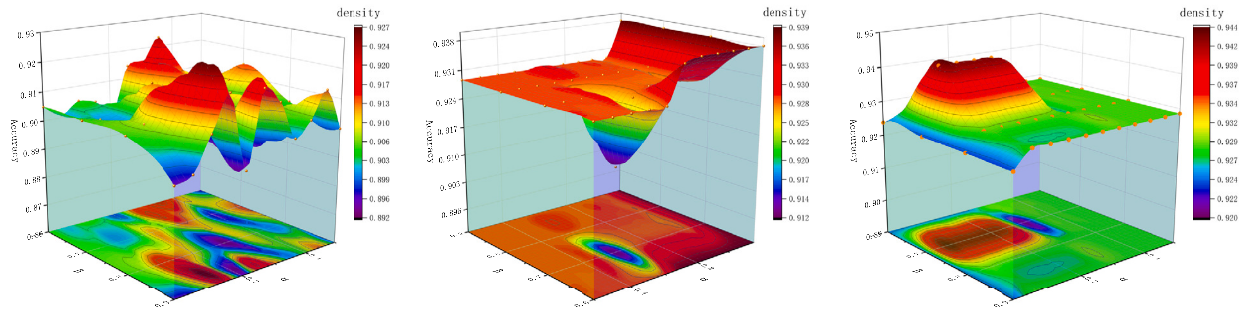
all datasets under different classifiers, as shown in Table 8. Additionally, we plotted the classification accuracy obtained by various parameter combinations for the 15 datasets under three classifiers, as illustrated in Figs. 3, 4, 5 and 6, where each row represents the classification accuracy results of the same dataset under different classifiers, and each column represents the same classifiers, from left to right, DT, SVM, and KNN classifiers, respectively. According to Table 8, varying datasets have distinct optimal parameter combinations that lead to the best classification results. For instance, observations indicate that datasets like Seeds, Wdbc, Wine-red, and Breast require a larger fuzzy neighborhood radius to achieve optimal classification performance. Conversely, datasets such as Image, Musk2, EGSS, and three high-dimensional feature datasets need a smaller α. By manipulating the confidence level parameter β, the stringency required for each dataset can be controlled, aiming to achieve the optimal classification performance for each dataset.

The illustrations clearly demonstrate the significance of exploring different combinations of parameters α and β. This exploration reveals that varying parameter combinations lead to noticeable differences in the classification accuracy results. And with the increase in the values of the fuzzy neighborhood parameter and the information level, in most datasets, the classification accuracy tends to stabilize after decreasing with the increase of α and β. This is because as the control parameter β increases, as defined in this paper, the strictness gradually enhances, demanding higher standards for the data. This results in minimal fluctuations in the entropy value. For every dataset, we have the flexibility to choose the most appropriate combination of α and β parameters to attain a relatively possible classification performance. For instance, under the DT classifier, the Seeds dataset exhibited relatively optimal performance at the combination $\alpha = 0.15$ and $\beta = 0.9$. Similarly, the SVM classifier demonstrated good classification performance for the Page dataset with the combination $\alpha = 0.20$ and $\beta = 0.6$. Additionally, the KNN classifier showcased relatively best classification accuracy for the high-dimensional feature dataset Lungcancer-2 at $\alpha = 0.05$ and $\beta = 0.6$. Therefore, opting for suitable combinations of α and β parameters is crucial to maximize the relative optimal classification performance of the FSAFNGCE algorithm.

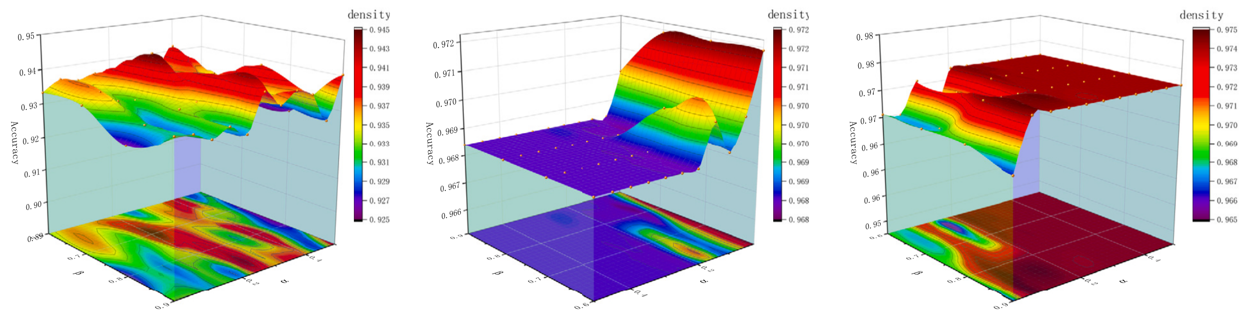
In conclusion, the FSAFNGCE algorithm demonstrates effective performance across the DT, SVM, and KNN classifiers, highlighting its efficiency as a feature selection approach.



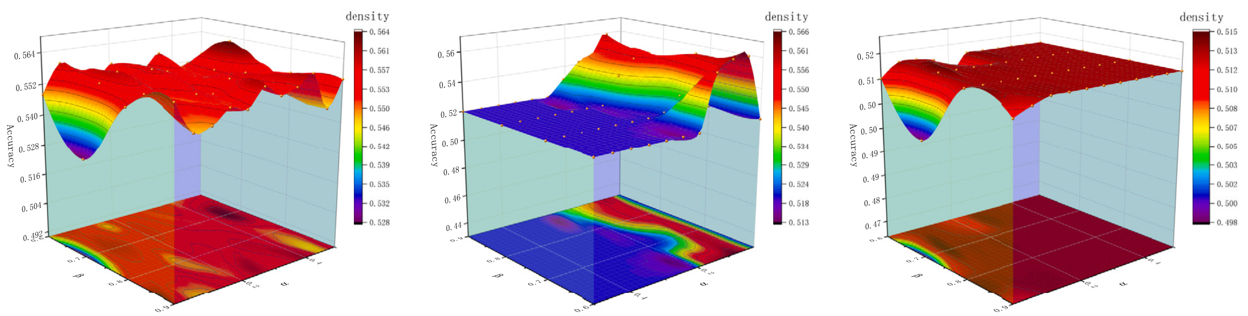
(a) Wine



(b) Seeds

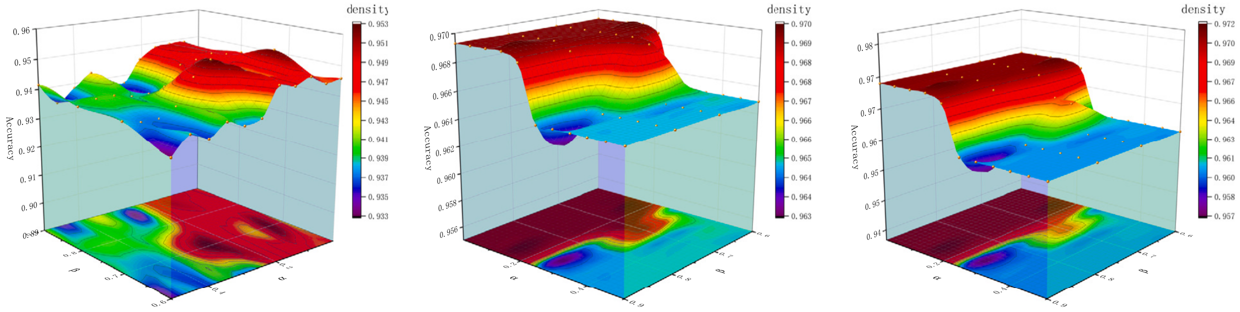


(c) Wdbc

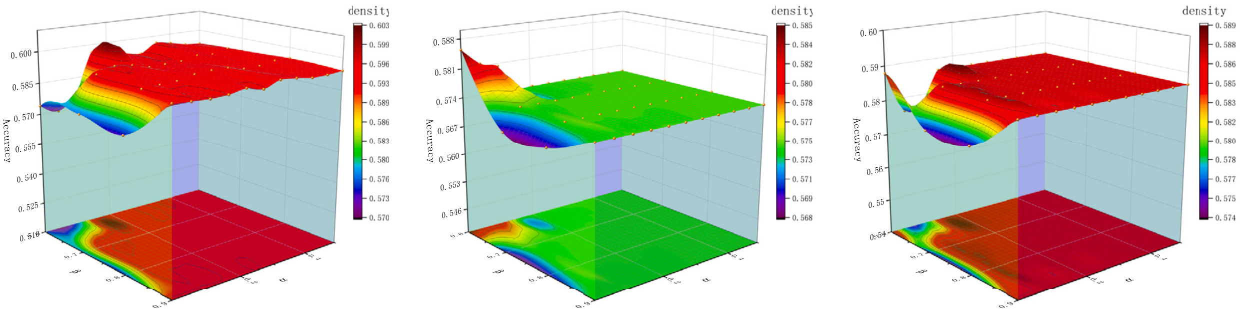


(d) Hill

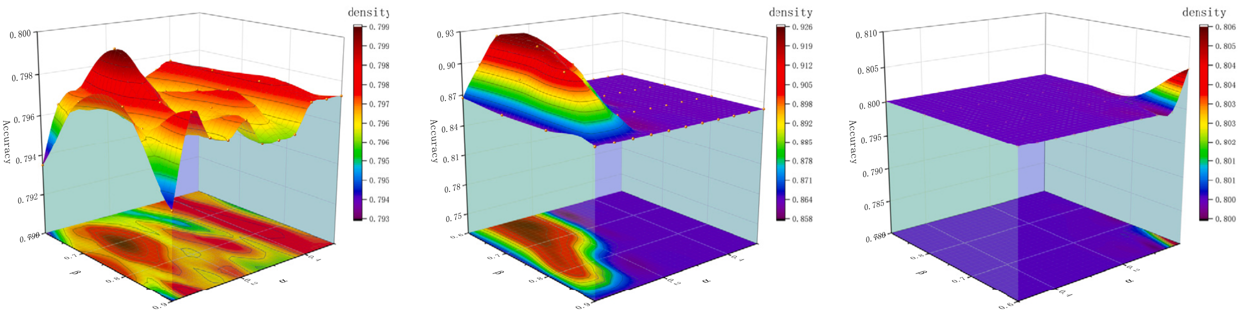
Fig. 3. Classification accuracy of Wine, Seeds, Wdbc and Hill datasets under DT (left column), SVM (middle column) and KNN (right column) classifier based on various parameters α and β.



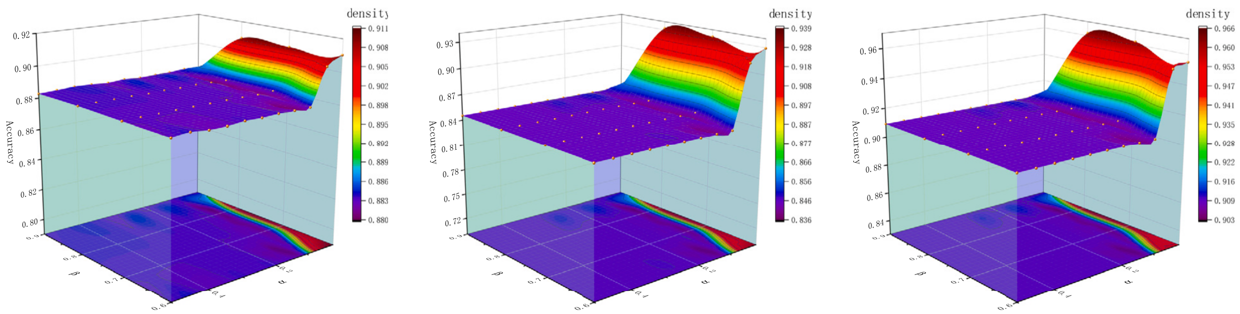
(a) Breast



(b) Wine-red

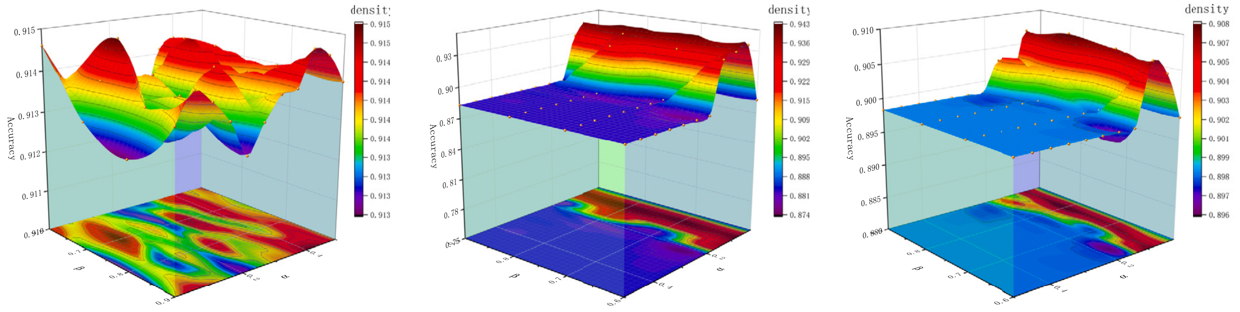


(c) Card

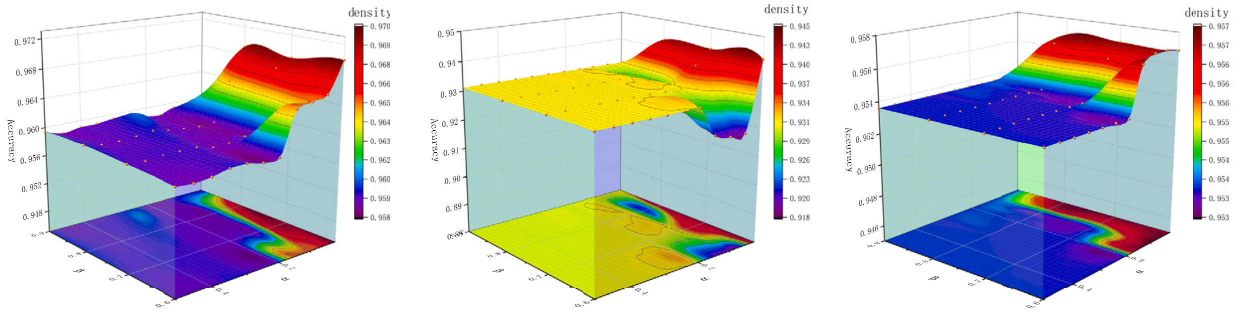


(d) Image

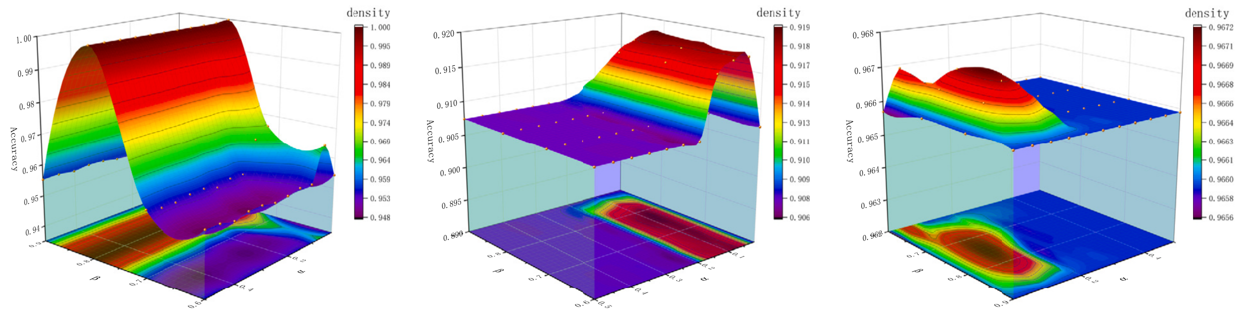
Fig. 4. Classification accuracy of Breast, Wine-red, Card and Image datasets under DT (left column), SVM (middle column) and KNN (right column) classifier based on various parameters α and β.



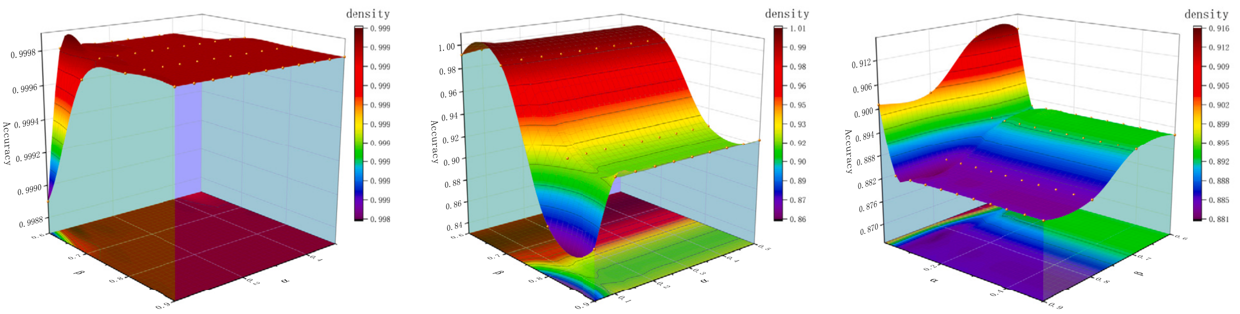
(a) Spambase



(b) Page

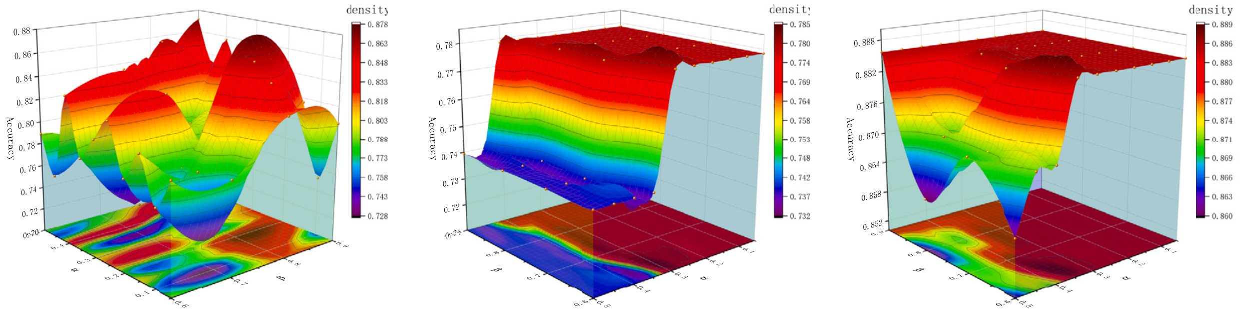


(c) Musk2

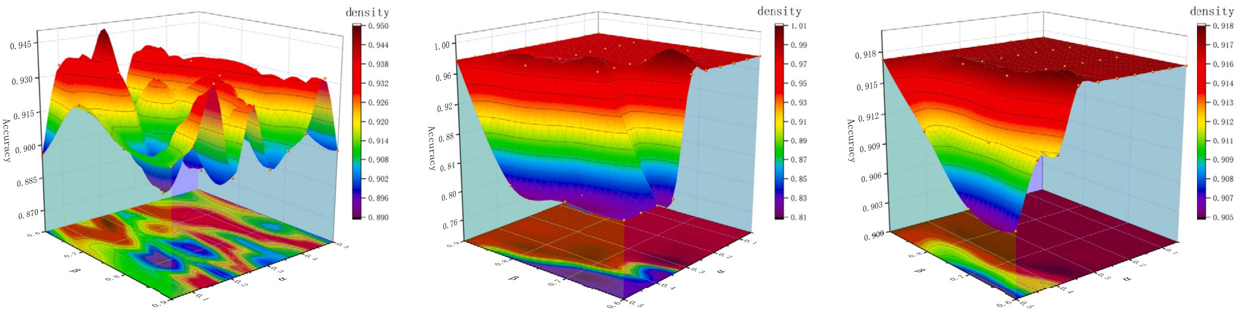


(d) EGSS

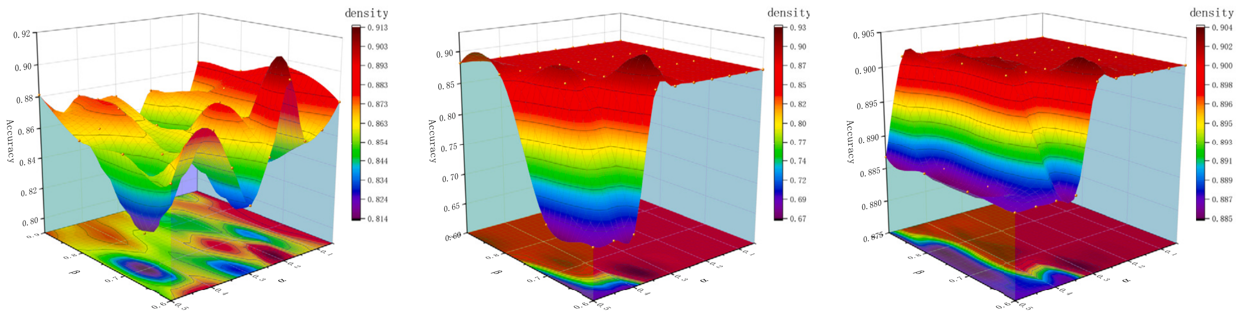
Fig. 5. Classification accuracy of Spambase, Page, Musk2 and EGSS datasets under DT (left column), SVM (middle column) and KNN (right column) classifier based on various parameters α and β.



(a) DLBCL



(b) LungCancer-1



(c) LungCancer-2

Fig. 6. Classification accuracy of DLBCL, LungCancer-1 and LungCancer-2 datasets under DT (left column), SVM (middle column) and KNN (right column) classifier based on various parameters α and β.

6. Conclusion

The optimistic multi-granulation rough set model requires only one condition to be satisfied, while the pessimistic ones require all conditions to be satisfied. Both optimistic and pessimistic rough set models fail to meet practical requirements. Therefore, we introduce the generalized multi-granulation rough set model, which allows adjusting the strictness level based on people's preferences through an information level parameter. By studying fuzzy neighborhood joint entropy and integrating it into the GMFNRS model, a method for uncertainty measurement based on fuzzy neighborhood generalized composite entropy is proposed. This entropy measure is specifically designed to address the uncertainty and fuzziness present in fuzzy decision systems that handle mixed data. Furthermore, we have developed a corresponding heuristic feature selection algorithm. The classification results on 15 publicly available datasets demonstrate that our proposed model is capable of selecting feature subsets that exhibit superior classification performance. However, due to the subset-based nature of the proposed feature selection algorithm, it cannot perfectly extract essential features. Some redundant attributes may still be included in the selected features. In future work, it is imperative to delve deeper into uncertainty measurement methods for GMFNRS to enhance our classification efficiency. Based on the research findings of this paper, future development can be proceeded in two aspects. On the one hand, the importance of each feature varies, but in this study, each feature is assigned the same weight by default. Therefore, it is worth considering assigning different weights to attributes. This approach can also be applied to medical decision-making problems, where different physiological indicators may have varying

Table 8
Optimal parameter combinations under different classifiers.

	DT		*SVM*		*KNN*	
Data sets	α	β	α	β	α	β
Wine	0.35	0.8	0.10	0.7	0.05	0.7
Seeds	0.15	0.9	0.25	0.6	0.25	0.7
Wdbc	0.25	0.8	0.15	0.6	0.10	0.6
Hill	0.45	0.7	0.05	0.6	0.05	0.6
Breast	0.05	0.7	0.25	0.7	0.20	0.6
Wine-red	0.15	0.6	0.10	0.6	0.10	0.6
Card	0.10	0.7	0.10	0.6	0.05	0.7
Image	0.05	0.6	0.05	0.6	0.05	0.6
Spambase	0.30	0.7	0.05	0.7	0.05	0.8
Page	0.05	0.7	0.20	0.6	0.15	0.6
Musk2	0.05	0.6	0.05	0.7	0.10	0.7
EGSS	0.05	0.6	0.05	0.6	0.05	0.6
DLBCL	0.50	0.9	0.05	0.6	0.05	0.6
LungCancer-1	0.20	0.6	0.05	0.7	0.05	0.6
LungCancer-2	0.20	0.6	0.05	0.6	0.05	0.6

levels of importance, optimizing treatment strategies accordingly. Therefore, it is worth considering assigning different weights to attributes. This approach can also be applied to medical decision-making problems, where different physiological indicators may have varying levels of importance, optimizing treatment strategies accordingly.

CRediT authorship contribution statement

Xiaoyan Zhang: Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Weicheng Zhao:** Writing – original draft, Visualization, Validation, Software, 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 article.

Data availability

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

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, 62376229) and the Chongqing Natural Science Foundation (Grant NO. CSTB2023NSCO-MSX1063).

References

[1] S. An, Q. Xiao, C. Wang, S. Zhao, Granularity self-information based uncertainty measure for feature selection and robust classification, Fuzzy Sets Syst. 470 (2023) 108658.
[2] O.O. Aremu, R.A. Cody, D. Hyland-Wood, P.R. McAree, A relative entropy based feature selection framework for asset data in predictive maintenance, Comput. Ind. Eng. 145 (2020) 106536.
[3] B. Barman, S. Patra, A novel technique to detect a suboptimal threshold of neighborhood rough sets for hyperspectral band selection, Soft Comput. 23 (2019) 13709–13719.
[4] D. Chen, L. Zhang, S. Zhao, Q. Hu, P. Zhu, A novel algorithm for finding reducts with fuzzy rough sets, IEEE Trans. Fuzzy Syst. 20 (2011) 385–389.
[5] L. Chen, D. Chen, H. Wang, Fuzzy kernel alignment with application to attribute reduction of heterogeneous data, IEEE Trans. Fuzzy Syst. 27 (2018) 1469–1478.
[6] A.K. Das, B. Chakraborty, S. Goswami, A. Chakrabarti, A fuzzy set based approach for effective feature selection, Fuzzy Sets Syst. 449 (2022) 187–206.
[7] W. Ding, C.T. Lin, M. Prasad, Z. Cao, J. Wang, A layered-coevolution-based attribute-boosted reduction using adaptive quantum-behavior pso and its consistent segmentation for neonates brain tissue, IEEE Trans. Fuzzy Syst. 26 (2017) 1177–1191.
[8] L. Dong, R. Wang, D. Chen, Incremental feature selection with fuzzy rough sets for dynamic data sets, Fuzzy Sets Syst. 467 (2023) 108503.
[9] N.L. Giang, T.T. Ngan, T.M. Tuan, H.T. Phuong, M. Abdel-Basset, A.R.L. De Macedo, V.H.C. De Albuquerque, et al., Novel incremental algorithms for attribute reduction from dynamic decision tables using hybrid filter–wrapper with fuzzy partition distance, IEEE Trans. Fuzzy Syst. 28 (2019) 858–873.
[10] C. Hu, L. Zhang, B. Wang, Z. Zhang, F. Li, Incremental updating knowledge in neighborhood multigranulation rough sets under dynamic granular structures, Knowl.-Based Syst. 163 (2019) 811–829.
[11] Q. Hu, D. Yu, J. Liu, C. Wu, Neighborhood rough set based heterogeneous feature subset selection, Inf. Sci. 178 (2008) 3577–3594.
[12] Y. Huang, K. Guo, X. Yi, Z. Li, T. Li, Matrix representation of the conditional entropy for incremental feature selection on multi-source data, Inf. Sci. 591 (2022) 263–286.
[13] R. Jensen, Q. Shen, Fuzzy–rough attribute reduction with application to web categorization, Fuzzy Sets Syst. 141 (2004) 469–485.

[14] G. Lin, Y. Qian, J. Li, Nmgrs: neighborhood-based multigranulation rough sets, Int. J. Approx. Reason. 53 (2012) 1080–1093.
[15] Z. Pawlak, Rough sets, Int. J. Comput. Inf. Sci. 11 (1982) 341–356.
[16] Y. Peng, Q. Zhang, Uncertainty measurement for set-valued data and its application in feature selection, Int. J. Fuzzy Syst. 24 (2022) 1735–1756.
[17] Y. Qian, J. Liang, Y. Yao, C. Dang, Mgrs: a multi-granulation rough set, Inf. Sci. 180 (2010) 949–970.
[18] G. Roffo, S. Melzi, U. Castellani, A. Vinciarelli, M. Cristani, Infinite feature selection: a graph-based feature filtering approach, IEEE Trans. Pattern Anal. Mach. Intell. 43 (2020) 4396–4410.
[19] B. Sang, W. Xu, H. Chen, T. Li, Active anti-noise fuzzy dominance rough feature selection using adaptive k-nearest neighbors, IEEE Trans. Fuzzy Syst. (2023).
[20] Y. Song, Q. Fu, Y.F. Wang, X. Wang, Divergence-based cross entropy and uncertainty measures of Atanassov's intuitionistic fuzzy sets with their application in decision making, Appl. Soft Comput. 84 (2019) 105703.
[21] L. Sun, L. Wang, W. Ding, Y. Qian, J. Xu, Feature selection using fuzzy neighborhood entropy-based uncertainty measures for fuzzy neighborhood multigranulation rough sets, IEEE Trans. Fuzzy Syst. 29 (2020) 19–33.
[22] L. Sun, L. Wang, W. Ding, Y. Qian, J. Xu, Neighborhood multi-granulation rough sets-based attribute reduction using Lebesgue and entropy measures in incomplete neighborhood decision systems, Knowl.-Based Syst. 192 (2020) 105373.
[23] L. Sun, L. Wang, J. Xu, S. Zhang, A neighborhood rough sets-based attribute reduction method using Lebesgue and entropy measures, Entropy 21 (2019) 138.
[24] L. Sun, X. Zhang, Y. Qian, J. Xu, S. Zhang, Feature selection using neighborhood entropy-based uncertainty measures for gene expression data classification, Inf. Sci. 502 (2019) 18–41.
[25] A. Tan, W.Z. Wu, Y. Qian, J. Liang, J. Chen, J. Li, Intuitionistic fuzzy rough set-based granular structures and attribute subset selection, IEEE Trans. Fuzzy Syst. 27 (2018) 527–539.
[26] C. Wang, Y. Huang, M. Shao, X. Fan, Fuzzy rough set-based attribute reduction using distance measures, Knowl.-Based Syst. 164 (2019) 205–212.
[27] C. Wang, M. Shao, Q. He, Y. Qian, Y. Qi, Feature subset selection based on fuzzy neighborhood rough sets, Knowl.-Based Syst. 111 (2016) 173–179.
[28] C. Wang, Y. Wang, M. Shao, Y. Qian, D. Chen, Fuzzy rough attribute reduction for categorical data, IEEE Trans. Fuzzy Syst. 28 (2019) 818–830.
[29] Y. Xing, P. Kochunov, T.G. van Erp, T. Ma, V.D. Calhoun, Y. Du, A novel neighborhood rough set-based feature selection method and its application to biomarker identification of schizophrenia, IEEE J. Biomed. Health Inform. 27 (2022) 215–226.
[30] J. Xu, Y. Wang, H. Mu, F. Huang, Feature genes selection based on fuzzy neighborhood conditional entropy, J. Intell. Fuzzy Syst. 36 (2019) 117–126.
[31] W. Xu, M. Huang, Z. Jiang, Y. Qian, Graph-based unsupervised feature selection for interval-valued information system, IEEE Trans. Neural Netw. Learn. Syst. (2023).
[32] W. Xu, W. Li, X. Zhang, Generalized multigranulation rough sets and optimal granularity selection, Granul. Comput. 2 (2017) 271–288.
[33] W. Xu, K. Yuan, W. Li, W. Ding, An emerging fuzzy feature selection method using composite entropy-based uncertainty measure and data distribution, IEEE Trans. Emerg. Top. Comput. Intell. 7 (2022) 76–88.
[34] X. Yang, X. Wang, J. Kang, Multi-granularity decision rough set attribute reduction algorithm under quantum particle swarm optimization, IET Cyber-Syst. Robot. 4 (2022) 25–37.
[35] Y. Yao, Y. She, Rough set models in multigranulation spaces, Inf. Sci. 327 (2016) 40–56.
[36] X.D. Yue, Y.F. Chen, D.Q. Miao, H. Fujita, Fuzzy neighborhood covering for three-way classification, Inf. Sci. 507 (2020) 795–808.
[37] L.A. Zadeh, Probability measures of fuzzy events, J. Math. Anal. Appl. 23 (1968) 421–427.
[38] K. Zeng, K. She, X. Niu, Multi-granulation entropy and its applications, Entropy 15 (2013) 2288–2302.
[39] H. Zhang, J. Zhan, Y. He, Multi-granulation hesitant fuzzy rough sets and corresponding applications, Soft Comput. 23 (2019) 13085–13103.
[40] P. Zhang, T. Li, Z. Yuan, C. Luo, K. Liu, X. Yang, Heterogeneous feature selection based on neighborhood combination entropy, IEEE Trans. Neural Netw. Learn. Syst. (2022).
[41] X. Zhang, Z. Jiang, W. Xu, Feature selection using a weighted method in interval-valued decision information systems, Appl. Intell. 53 (2023) 9858–9877.
[42] X. Zhang, C. Mei, D. Chen, Y. Yang, J. Li, Active incremental feature selection using a fuzzy-rough-set-based information entropy, IEEE Trans. Fuzzy Syst. 28 (2019) 901–915.