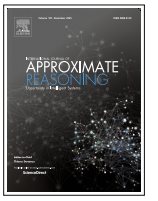
International Journal of Approximate Reasoning 189 (2026) 109590

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

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



Multi-granularity Knowledge Fusion for Feature Selection Using Granular-ball Entropy Uncertainty Measures

Kehua Yuan [a,1], Yuji Bai [a,1], Duoqian Miao [a,*], Weiping Ding [b,c], Yiyu Yao [d], Hongyun Zhang [a], Witold Pedrycz [e,f,g]

[a] School of Computer Science and Technology, Tongji University, Shanghai, 201804, PR China
[b] School of Information Science and Technology, Nantong University, Nantong, 226019, PR China
[c] Faculty of Data Science, City University of Macau, Macau, 999078, PR China
[d] Department of Computer Science, University of Regina, Regina, SK, S4S 0A2, Canada
[e] Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, T6G 2R3, Canada
[f] Systems Research Institute, Polish Academy of Sciences, Warsaw, 00-901, Poland
[g] Research Center of Performance and Productivity Analysis, Istinye University, Istanbul, Türkiye

ARTICLE INFO

Keywords:
Granular computing
Feature selection
Granular ball
Uncertainty measure
Knowledge fusion

ABSTRACT

Multi-granularity computing for knowledge discovery has emerged as a remarkable paradigm in data mining and machine learning. As a representative method, granular-ball computing has attracted considerable attention due to its efficiency and adaptability in handling complex data distributions. However, most existing granularity-based approaches focus on intra-granular mutual information while neglecting the heterogeneity and overlapping phenomena across granularities. This limitation often leads to imprecise knowledge space construction and inaccurate uncertainty estimation in feature evaluation. To overcome this problem, this study proposes a novel and high-efficiency multi-granularity knowledge fusion framework for feature selection, incorporating an enhanced granular-ball generation mechanism and a newly designed granular-ball entropy (GB-E) uncertainty measure. Specifically, we first develop an enhanced granular-ball generation mechanism to construct multi-granularity knowledge space by incorporating class distribution information, thus achieving more accurate and flexible data partitioning. Subsequently, by jointly analyzing the separation and aggregation among granular balls, a novel granular-ball entropy is proposed to quantify uncertainty in the multi-granularity knowledge space. Compared with existing uncertainty measure methods, it provides a dual-perspective uncertainty characterization and effectively improves the accuracy of granularity information fusion. Furthermore, two feature significance measures based on the proposed GB-E measure are introduced for feature evaluation, and then a corresponding feature selection method is developed. Extensive experiments on multiple public datasets demonstrate the proposed method's superior classification performance compared with several state-of-the-art approaches.

* Corresponding author.
E-mail addresses: yuan945527632@126.com, yuan945527632@163.com (K. Yuan), 2254306@tongji.edu.cn (Y. Bai), dqmiao@tongji.edu.cn (D. Miao), dwp9988@163.com (W. Ding), yiyu.yao@uregina.ca (Y. Yao), zhanghongyun@tongji.edu.cn (H. Zhang), wpedrycz@ualberta.ca (W. Pedrycz).
[1] Equal Contribution.

https://doi.org/10.1016/j.ijar.2025.109590
Received 5 August 2025; Received in revised form 14 October 2025; Accepted 28 October 2025
Available online 11 November 2025
0888-613X/© 2025 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

1. Introduction

High-dimensional data are increasingly prevalent in various real-world scenarios, where the curse of dimensionality presents fundamental challenges to effective analysis [1–3]. The existence of redundant, irrelevant, or noisy features not only escalates computational complexity but also impairs model interpretability and increases uncertainty. To address these issues, feature selection has emerged as an essential step to reduce data dimensionality and computational cost while improving model learning performance [4–6]. By identifying and retaining only the most informative features, it helps reduce data dimensionality, enhance model generalization, and improve interpretability. Various feature selection strategies have been proposed, including those based on information gain [7], graph structure [8], and loss optimization [9], all aiming to preserve essential information while discarding noise and redundancy in complex data.

Multi-granularity computing, as an emerging paradigm inspired by human cognitive strategies, has been widely studied and investigated in classification tasks [10–12], rule extraction [13–16], and feature selection [17–19]. It advocates a methodology of observation and learning from multiple granularities, levels, and perspectives, where coarser granularity captures global patterns or structures, and finer granularity focuses on detailed information. Within this paradigm, various granularity models have been developed, including fuzzy sets [20,21], three-way decisions [22,23], rough sets [24,25], granular ball computing [26], and concept-cognitive learning [27,28]. In particular, granular ball computing (GBC) provides an adaptive granulation strategy to approximate data distribution, which has been widely employed for uncertainty measures [29], feature selection [30], and classification tasks [31]. For example, Xia et al. proposed a noise-aware sampling strategy for imbalanced data during granular ball generation [26]. Substantially, an acceleration mechanism is also proposed to further reduce computational overhead by replacing repeated K-means iterations with a division-based strategy [32]. In the context of open feature selection, Cao et al. [29] proposed a continual selection framework by integrating continual learning. For multi-label tasks, Qian et al. [33] introduced a method that combines rough set theory with label distribution modeling. Sun et al. [30] extended GBC by incorporating mutual information into a fuzzy clustering method for partially labeled multi-label data. Moreover, multi-granularity fusion within GBC has also shown promise in improving model performance. Sajid et al. [34] fused granular ball representations with randomized neural networks to address sample imbalance issues. Dunkin et al. [35] applied GBC-based granularity fusion techniques to fault diagnosis tasks. Despite these advances, current GBC methods mostly rely on repeated binary partitioning within each ball, which limits their adaptability to underlying class distributions and often results in imprecise granulation and high computational complexity during the splitting process. Therefore, developing an adaptive and efficient partitioning strategy for data granulation constitutes one of the motivations of this paper.

Feature selection is an important issue in muti-granularity computing, where various granularity-based uncertainty measures have been extensively studied from both algebraic and information theory perspectives [36–40]. From the algebraic viewpoint, numerous uncertainty measures, such as positive region, rough degree, dependency degree, and discernibility function, have been proposed and widely applied for feature evaluation. For instance, Hu et al. [36,37] introduced the neighborhood rough set model and developed corresponding attribute reduction methods based on the positive region. Extending this idea, Wang et al. [41] incorporated neighborhood concepts into fuzzy rough sets to address uncertainty from a fuzzy perspective. In the context of hybrid data, Kumar et al. [42] proposed a fuzzy min-max neural network by integrating fuzzy-rough approximations to achieve scalable feature selection. In image processing tasks, Li et al. [43] designed a granularity-based information amount measure using separation and aggregation degrees to enhance reduction efficiency. Additionally, Guo et al. [44] introduced a gene distinctiveness measure for core gene selection in disease diagnosis under concept-cognitive learning. From the information theory viewpoint, entropy-based approaches have demonstrated strong effectiveness in quantifying uncertainty. For example, Sun et al. [45] proposed fuzzy entropy measures in a multi-granularity fuzzy approximation space to guide heterogeneous feature selection. Wan et al. [46] designed a series of mutual information-based uncertainty measures to capture feature interactions. In dynamic environments, Sang et al. developed a conditional entropy-based incremental mechanism for efficient feature updating. More recently, Yuan et al. [47,48] introduced a novel zentropy-based uncertainty measure that comprehensively models the multi-granularity hierarchy and inter-feature interactions. In particular, inspired by granular ball computing, several uncertainty-based feature selection approaches have been successively developed. To enhance the adaptability of neighborhood boundaries, Xia et al. [49] first defined a granular ball neighborhood rough set model for attribute reduction and classification based on the positive region. Chen et al. [50] further proposed a variable-precision FGBNRS model to improve tolerance in the approximation process and adopted a variable-precision dependency function for feature evaluation. From the information-theoretic viewpoint, Sun et al. [51] developed a granular ball-based fuzzy neighborhood entropy to evaluate feature uncertainty. To further consider feature interactions, Qian et al. [52] proposed a fuzzy correlation-redundancy evaluation method for partial-label feature selection based on granular ball generation. Despite their effectiveness in data granulation and uncertainty quantification, existing granular ball-based methods are primarily constructed upon generated granular balls, lacking a deeper investigation into the intrinsic structure of the multi-granularity granular ball knowledge space. From the above analysis, current studies, including the granular ball-based methods, have mainly focused on intra-granular information measures, with limited attention to heterogeneity and overlap across granular levels. This limitation hinders a more comprehensive understanding of uncertainty in multi-granularity knowledge representation. Therefore, a systematic and precise characterization of granularity uncertainty remains essential for effective knowledge representation and feature selection in multi-granularity spaces.

Inspired by the above issues, this paper proposes a multi-granularity knowledge fusion method for feature selection based on the granular ball entropy uncertainty measure. It applies the naive idea of granular ball splitting and the generation mechanism to construct multi-granularity knowledge spaces. By jointly fusing the uncertainty information from these multi-granularity granular ball structures, the proposed framework achieves a more adaptive and effective uncertainty characterization and measurement of information systems, which is multi-granularity knowledge fusion. Compared with existing methods, it can adaptively construct

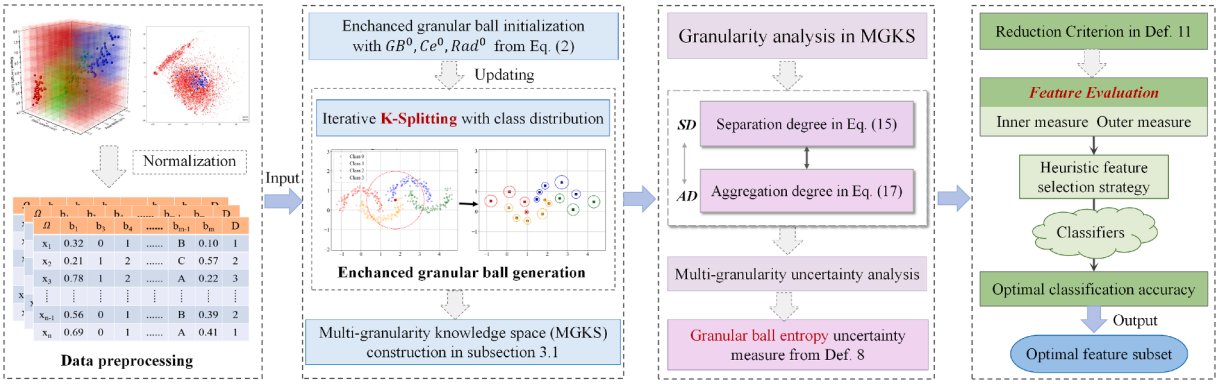


Fig. 1. The framework of the proposed FGB-E method comprises four main steps: data preprocessing, enhanced granular ball generation, granular ball entropy-based uncertainty measure, and feature selection. First, the input data is normalized to the unit interval in Step 1. Then, enhanced granular balls are generated by incorporating class distribution information to construct a multi-granularity knowledge space in step 2. Moreover, the granular ball entropy is introduced in step 3, which jointly considers separation and aggregation degrees among granules to quantify uncertainty within the multi-granularity space. Finally, Step 4 applies the proposed uncertainty measure to feature evaluation, enabling compelling and informative feature selection.

multi-granularity knowledge spaces in a data-driven manner, where the enhanced granular ball generation is automatically guided by the intrinsic data distribution. Moreover, the designed granular ball entropy jointly captures the separation and aggregation degrees among granules, thereby improving the precision of uncertainty quantification and reducing redundant information computation across granules. The overall framework is illustrated in Fig. 1. Specifically, the main contributions of this paper are summarized as follows.

- It provides an efficient multi-granularity knowledge fusion approach for feature selection based on the multi-granularity knowledge space and granular ball entropy uncertainty measure. This approach enables a more precise understanding of uncertainty from dual-perspectives while effectively avoiding redundant granularity information during knowledge modeling.
- It develops an enhanced granular ball generation mechanism for constructing the multi-granularity knowledge space. This mechanism adaptively partitions data according to class distributions within each granule, offering a more flexible and data-driven multi-granularity knowledge representation method than existing methods.
- It proposes a novel granular ball entropy to quantify uncertainty in multi-granularity knowledge spaces. By jointly considering granule separation and aggregation degrees, this method achieves dual-perspective uncertainty characterization and significantly improves the accuracy of granularity-based information fusion.
- It designs feature significance measures based on the proposed GB-E uncertainty measure and develops a corresponding feature selection method under the principle of information gain. Experimental results on multiple benchmark datasets demonstrate the effectiveness and superiority of the proposed method over several state-of-the-art approaches.

The remainder of this paper is organized as follows. Section 2 provides a brief review of related work. Section 3 presents the construction of a multi-granularity knowledge space and introduces the proposed granular ball entropy for multi-granularity uncertainty quantification. Section 4 develops the feature evaluation measures and the corresponding feature selection algorithm. Section 5 analyzes the experimental results on publicly available datasets. Finally, Section 6 concludes the paper and outlines directions for future work.

2. Preliminaries

This section introduces the fundamental concepts related to granular ball computing and entropy-based measures of this work. Further details can be found in [26,46,53].

2.1. Granular ball computing

Granular ball computing is an effective approach for characterizing knowledge across multiple granularities by adaptively generating a set of balls based on the underlying data distribution. Each ball is defined by its center and radius, and its purity reflects the degree of class homogeneity within the ball. Before the formal definition, we first define decision information systems.

Definition 1 ([26]). Let $DS = \langle Q, P, D\rangle$ be a decision information system, where $Q = \{u_1, u_2, \ldots, u_n\}$ denotes a finite universe of objects, $P = \{p_1, p_2, \ldots, p_m\}$ represents the set of conditional attributes, and $D = \{d\}$ is the decision attribute. The partition of Q induced by D is denoted as $Q/D = \{C_1, C_2, \ldots, C_s\}$. For any $u_i \in Q$ and $p_j \in P$, $p_j(u_i)$ denotes the value of attribute p_j for object u_i.

Definition 2 ([26]). Let DS = $\langle Q, P, D\rangle$ be a decision information system. A granular ball set $G = \{GB_1, GB_2, \dots, GB_W\}$ is constructed based on DS. For any granular ball $GB_i \in G$, its center and radius are computed as follows.

$$\begin{aligned} Ce(GB) &= \frac{1}{|GB|}\sum_{u_i \in GB} u_i, \\ Rad(GB) &= \frac{1}{|GB|}\sum_{u_i \in GB} Dis_P(u_i, Ce(GB)), \end{aligned} \tag{1}$$

where $|GB|$ denotes the number of objects in the ball, and $Dis_P(u_i, Center(GB))$ represents the distance between object u_i and the center of the ball. In this paper, the Euclidean distance is used as the distance metric.

According to the above definitions, the center and radius of a granular ball characterize the spatial distribution of objects within the ball. A granular ball is considered pure if most constituent objects belong to the same decision class. In such cases, the ball's label is assigned based on the majority class, i.e., the class with the highest occurrence probability among its objects. Subsequently, the corresponding purity evaluation measures will be obtained.

Definition 3 ([26]). Let DS = $\langle Q, P, D\rangle$ be a decision information system. A granular ball set $G = \{GB_1, GB_2, \dots, GB_W\}$ is constructed based on DS. For any granular ball $GB_i \in G$, the ball label and purity is obtained as follows.

$$\begin{aligned} Lab(GB) &= \arg\max_{i\in\{1,2,\dots,s\}} |C_i \cap GB|, \\ Pur(GB) &= \frac{|GB \cap D_{Lab(GB)}|}{|GB|}, \end{aligned} \tag{2}$$

where $1 \le Pur(GB) \le 1$, and the larger value denotes the more pure of GB.

The generation of granular balls aims to represent the data distribution by grouping nearby samples into compact spherical regions. Given a dataset $Q = \{u_1, u_2, \dots, u_n\}$, all the system is considered as a granular ball, then the initial granular ball center is first computed as $Ce_0 = \frac{1}{|Q|}\sum_{u_i \in Q} u_i$ and the corresponding radius is obtained by $Rad_0 = \frac{1}{|Q|}\sum_{u_i \in GB} Dis_P(u_i, Ce_0)$ from Definition 2. In the second iteration, two points are randomly selected as the centers of granular balls. All the objects are divided into different granular balls using the minimum distance values. If the purity computed by Eq. (2) satisfies the given threshold, the granular ball is stopped splitting; otherwise, it will be further binary splitting. A more detailed granular ball generation process is illustrated in Fig. 2. This process adopts a two-splitting strategy, where each granular ball is recursively divided into smaller sub-balls. The splitting continues until the resulting balls achieve a satisfactory homogeneity, which is determined by a predefined purity threshold. This approach ensures that each granular ball predominantly contains samples from a single decision class, allowing for a coarse-grained approximation of the underlying data distribution corresponding to each class.

However, in multi-class scenarios, directly applying binary splitting to impure balls without considering the specific distribution of multiple classes often leads to suboptimal partitioning. This can reduce the precision of the granularity and negatively impact subsequent analytical tasks. To overcome this limitation, this paper proposes an enhanced granular ball splitting method that adaptively accounts for class overlap, thereby improving the purity and representational quality of the resulting granular structure.

2.2. Entropy-based uncertainty measure

Information entropy, originally proposed by Shannon, is a fundamental measure used to quantify the uncertainty of a system based on its probability distribution. Given a discrete random variable $W = \{w_1, w_2, \dots, w_l\}$ and corresponding probability distribution $P = \{p_1, p_2, \dots, p_l\}$, the information entropy $E_n(X)$ is defined as follows.

$$E_n(X) = -\sum_{i=1}^{l} P_i log P_i, \tag{3}$$

To comprehensively evaluate the inherent uncertainty of data, various entropy-based measures have been proposed by incorporating granularity information. Representative examples include neighborhood mutual information and conditional entropy, which enable a finer characterization of uncertainty from a local perspective.

Definition 4 ([46]). Let DS = $\langle Q, P, D\rangle$ be a decision information system. For $R, T \subseteq A$, $g_R(x)$ and $g_T(x)$ are the neighborhood granule of x induced by R and T, respectively. Then, the neighborhood mutual information $Mul(R, T)$ between R and T is represented as follows.

$$Mul(R, T) = -\frac{1}{|Q|}\sum_{i=1}^{|Q|} log\frac{|g_R(x) \cap g_T(x)|}{|Q|}, \tag{4}$$

where $g_R(x) = \{y \in Q | Dis(x, y) \le g\}$ is the neighborhood granule of x under R, and g is a predefined neighborhood radius threshold.

Definition 5 ([53]). Let DS = $\langle Q, P, D\rangle$ be a decision information system. For $R \subseteq A$, $g_R(x)$ is the neighborhood granule of x induced by R. Then, the neighborhood conditional entropy $NCE(D|R)$ of D related to R is represented as follows.

$$Nce(D|R) = -\frac{1}{|Q|}\sum_{i=1}^{|Q|} log\frac{|g_R(x) \cap D(x)|}{|g_R(x)|}, \tag{5}$$

where $D(x)$ denotes the decision granule of x, i.e., the set of all objects in Q that share the same decision class as x.

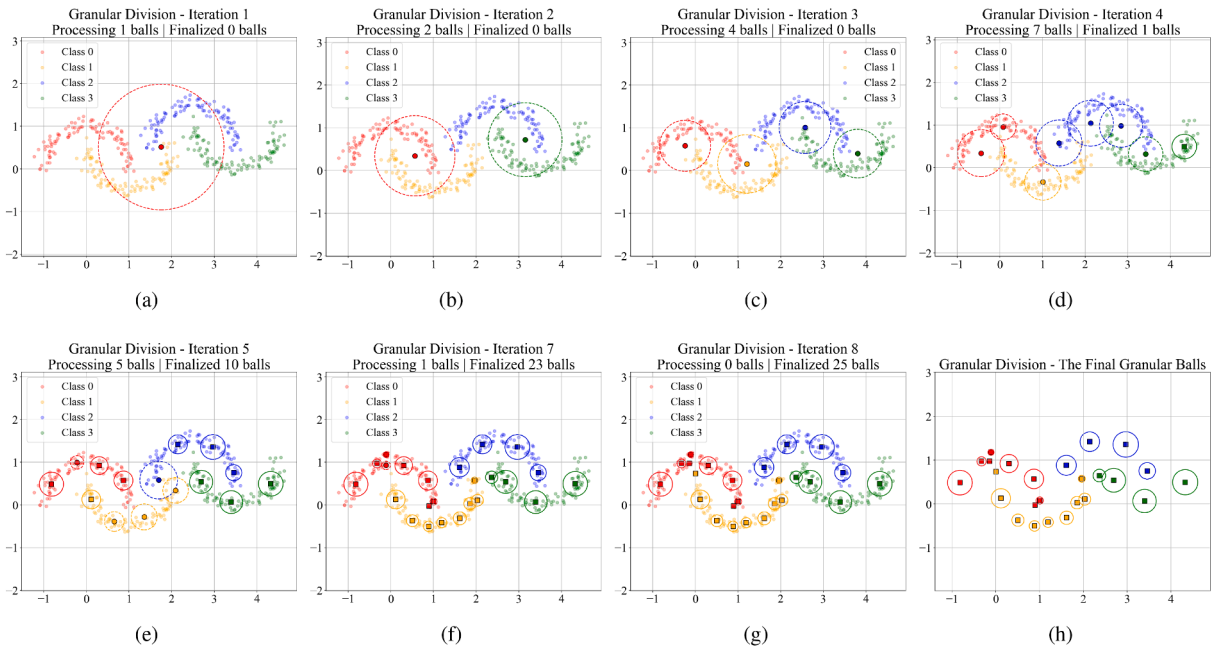


Fig. 2. Granular balls splitting generation process of the existing method, where four colors respectively correspond to four types of classes. (a) Random initial a granular ball in the first iteration; (b) The second iteration; (c) The third iteration; (d) The fourth iteration; (e) The fifth iteration; (f) The seventh iteration; (g) The final generated granular balls; (h) The extracted ball results.

Based on the above definitions of neighborhood mutual information and neighborhood conditional entropy, it can be observed that the existing entropy-based measures in classification tasks primarily emphasize quantifying the uncertainty within classes, focusing on the shared information within individual classes. For instance, neighborhood conditional entropy quantifies the homogeneity of decision labels within local neighborhoods, thereby capturing the consistency of samples belonging to the same class. However, these approaches often fail to adequately reflect inter-class heterogeneity, i.e., distinguishing information between different classes, which is crucial for effective classification. Therefore, developing an effective uncertainty measure is necessary to comprehensively characterize system uncertainty by integrating granularity properties.

3. Multi-granularity knowledge construction and measure with separation and aggregation

To address the limitations of existing granular ball-based data granulation methods, this section introduces an enhanced granular ball model aimed at improving both the accuracy and efficiency of the granulation process. In addition, a novel entropy measure, termed granular ball entropy, is proposed to effectively characterize and manage uncertainty by jointly considering both the separation between granules and the aggregation within granules.

3.1. Multi-granularity knowledge construction

The traditional granular ball method employs a binary splitting strategy in each iteration and terminates once the resulting granules meet a predefined purity threshold. However, this approach often leads to imprecise partitioning and increased complexity in the granulation process. To overcome these limitations, this subsection proposes a multi-granularity knowledge construction framework based on a K-splitting strategy, where K corresponds to the number of distinct classes present within the current granular ball.

Let $\mathrm{DS} = \langle Q, P, D\rangle$ be a decision information system, where $Q = \{u_1, u_2, \ldots, u_n\}$ denotes the universe consisting of n objects, $P = \{p_1, p_2, \ldots, p_m\}$ represents the set of m conditional attributes, and the decision partition is given by $Q/D = \{C_1, C_2, \ldots, C_s\}$. The proposed enhanced multi-granularity ball method is structured in three main steps: (1) initialization of the granular ball structure; (2) iterative refinement based on $K-$splitting strategy; and (3) enhanced granular ball generation. Each step is designed to improve the granulation accuracy while maintaining computational efficiency progressively.

- **Initialization:** Obtain the initial granular ball $\mathrm{GB}^0 = \{GB_1^0\}$ with center $\mathrm{Ce}^0 = \{Ce(GB_1^0)\}$ and radius $\mathrm{Rad}^0 = \{Rad(GB_1^0)\}$, which are defined as follows:

$$GB_1^0 = \{u \in Q \mid Dis_P(x, Ce(GB_1^0)) \le Rad(GB_1^0)\}, \tag{6}$$

where $Ce(GB_1^0) = \frac{1}{|Q|}\sum_{u_i \in Q} u_i$ and $Rad(GB_1^0) = \frac{1}{|Q|}\sum_{u_i \in Q} Dis_P(u_i, Ce(GB_1^0))$ are the center and radius of GB_1^0. This initialization ensures that all objects in the universe are included within the initial granular ball.

Fig. 3. The enhanced granular balls splitting generation process with iterative K-splitting, where four colors respectively correspond to four types of classes. (a) Initial a granular ball in the first iteration; (b) The second iteration; (c) The third iteration; (d) The fourth iteration; (e) The final generated granular balls; (f) The extracted ball results.

- **Iterative K-Splitting:** An adaptive partition strategy is applied to iteratively refine the generated granular balls to obtain a finer representation. Suppose there are w granular balls in the $t-1$th iteration, denoted by $\mathrm{GB}^{t-1} = \{GB_1^{t-1}, GB_2^{t-1}, \dots, GB_w^{t-1}\}$. During the tth iteration, each granular ball $GB_i^{t-1} \in \mathrm{GB}^{t-1}$ is evaluated based on its internal purity. If the purity of GB_i^{t-1} exceeds a predefined threshold, the ball is retained without further splitting. Otherwise, it is adaptively divided into K sub-granular balls, where K corresponds to the number of distinct classes contained within the current granular ball.

$$GB_i^{t-1}/D = \{gb_{i1}^t, gb_{i2}^t, \dots, gb_{iK}^t\}, \tag{7}$$

where $gb_{ij}^t = \{u \in GB_i^{t-1} \mid u \in C_j\}$, $j = 1, 2, \dots, K$ denotes the class types of objects in GB_i^{t-1}.

Based on this division, K new granular balls are formed in the tth iteration. To reduce computational complexity, the center of each sub-granule gb_{ij}^t is reused to define the granular ball $GB_{\mathbf{i}j}^t$ as:

$$GB_{\mathbf{i}}^t = \{GB_{\mathbf{i}1}^t, GB_{\mathbf{i}2}^t, \dots, GB_{\mathbf{i}K}^t\}, \quad \mathbf{i} = 1, 2, \dots, w, \tag{8}$$

where $GB_{\mathbf{i}j}^t$ is defined as:

$$GB_{\mathbf{i}j}^t = \{u \in GB_{\mathbf{i}}^{t-1} \mid Dis_P(u, Ce(GB_{\mathbf{i}j}^t)) \le Rad(GB_{\mathbf{i}j}^t)\}, \quad j = 1, 2, \dots, K, \tag{9}$$

with

$$Ce(GB_{\mathbf{i}j}^t) = \frac{1}{|gb_{ij}^t|} \sum_{x \in gb_{ij}^t} u, \quad Rad(GB_{\mathbf{i}j}^t) = \frac{1}{|gb_{ij}^t|} \sum_{u \in gb_{ij}^t} Dis_P(u, Ce(GB_{\mathbf{i}j}^t)). \tag{10}$$

The updated granule set after the tth iteration is then given by:

$$GB^t = GB_{\mathbf{1}}^t \cup GB_{\mathbf{2}}^t \cup \cdots \cup GB_{\mathbf{w}}^t. \tag{11}$$

- **Enhanced granular ball generation:** The iterative process continues until the purity of all granular balls meets the given threshold. Suppose r granular balls are finally obtained in dataset DS, denoted by $\mathrm{GB} = \{GB_1, GB_2, \dots, GB_r\}$. The corresponding centers and radii are:

$$\mathrm{Ce} = \{Ce(GB_1), Ce(GB_2), \dots, Ce(GB_r)\}, \quad \mathrm{Rad} = \{Rad(GB_1), Rad(GB_2), \dots, Rad(GB_r)\}. \tag{12}$$

The labels of the balls according to Eq. (2) are given by:

$$Lab(\mathrm{GB}) = (Lab(GB_1), Lab(GB_2), \dots, Lab(GB_r)). \tag{13}$$

Finally, the enhanced granular balls generated through the K-splitting strategy are represented as a set of tuples:

$$(GB_i, Ce(GB_i), Rad(GB_i), Lab(GB_i)), \quad i = 1, 2, \dots, r. \tag{14}$$

Based on the enhanced granular ball generation process described above, a Multi-Granularity Knowledge Space (MGKS) is constructed. Compared with existing binary splitting in [26] and [45], the class distribution-driven splitting and generation method effectively improves granulation imprecision and efficiency. In this MGKS, each enhanced granular ball is treated as a representative "point" that characterizes the corresponding region of the data space. This representation enables the collection of granular balls to effectively fit the underlying data distribution. The detailed procedure for generating enhanced granular balls is illustrated in Fig. 3 to provide a clearer understanding of this generation process.

3.2. GB-Entropy uncertainty measure

To comprehensively characterize uncertainty within the multi-granularity knowledge space, this subsection proposes an emerging GB-entropy measure that integrates both the separability and aggregability of information granules. By jointly considering these two perspectives, the proposed method provides a more balanced and informative representation of uncertainty in granular structures.

Definition 6. Let $\mathrm{DS} = \langle Q, P, D\rangle$ be a decision information system, where $Q/D = \{C_1, C_2, \dots, C_s\}$. Suppose the enhanced granular ball set generated by the K-splitting process is denoted as $\mathrm{GB} = \{GB_1, GB_2, \dots, GB_r\}$. For any $GB_i \in \mathrm{GB}$, its separation degree $DD(GB_i)$ with respect to other granular balls is defined as follows.

$$SD(GB_i) = \frac{\sum_{GB_j \in GB, Lab(GB_j) \neq Lab(GB_i)} |GB_j| Dis_P(GB_j, GB_i)}{\sum_{GB_j \in G, Lab(GB_j) \neq Label(GB_i)} |GB_j|}, \tag{15}$$

where $Dis_P(GB_j, GB_i)$ denotes the distance between granular balls GB_i and GB_j, calculated as:

$$Dis_P(GB_i, GB_j) = max(0, Dis_P(Cen(GB_i), Cen(GB_j)) - Rad(GB_i) - Rad(GB_j)), \tag{16}$$

where $Dis_P(Cen(GB_i), Cen(GB_j))$ denotes the distance between granular ball centers $Cen(GB_i)$ and $Cen(GB_j)$ under feature subset P. In this paper, the Euclidean distance is used as the distance metric.

A larger value of $SD(GB_i)$ indicates that GB_i is more distinguishable from other granules, thereby possessing stronger class-separability. Moreover, the aggregation degree is also given to characterize the membership degree of objects in each enhanced granular ball.

Definition 7. Let DS $= \langle Q, P, D\rangle$ be a decision information system, where $Q/D = \{C_1, C_2, \dots, C_s\}$, and GB $= \{GB_1, GB_2, \dots, GB_r\}$ is the set of enhanced granular balls obtained via the K-splitting strategy. For any $GB_i \in$ GB, the aggregation degree $AD(GB_i)$ of GB_i is defined as follows.

$$AD(GB_i) = \frac{1}{|GB_i|}\sum_{u\in GB_i}\mu(u, Lab(GB_i)), \tag{17}$$

where $\mu(u, Lab(GB_i))$ is the membership degree of u to decision class $Lab(GB_i)$, calculated as:

$$\mu(u, Lab(GB_i)) = \frac{1}{\sum_{j=1}^{s}\left(\frac{Dis_P(u,Ce(C_i))}{Dis_P(u,Ce(C_j))}\right)^{\frac{2}{m-1}}}, \tag{18}$$

where $m = 2$ is the fuzzification coefficient, and $\mathrm{Ce}(C_j)$ denotes the center of decision class C_j.

A higher value of $AD(GB_i)$ indicates stronger internal consistency between the granular ball and its assigned label, reflecting a higher degree of class-specific aggregation. Based on the above definitions, the separation and aggregation degrees jointly describe the information content of a granular ball from two complementary perspectives: inter-class separation and intra-class cohesion. To comprehensively quantify the information carried by each granular ball, we further define a granular ball entropy that integrates both separation and aggregation characteristics.

Definition 8. Let DS $= \langle Q, P, D\rangle$ be a decision information system, where $Q/D = \{C_1, C_2, \dots, C_s\}$ denotes the decision partition, and let GB $= \{GB_1, GB_2, \dots, GB_r\}$ be the set of enhanced granular balls obtained via the $K-$splitting process. The GB-entropy with respect to attribute set P is defined as follows.

$$GB\text{-}E_P(\mathrm{GB}) = -\sum_{GB_i\in \mathrm{GB}} p(GB_i)\log(p(GB_i)), \tag{19}$$

where $p(GB_i)$ denotes the probability of information amount induced by GB_i, computed as

$$p(GB_i) = \frac{\mathrm{I}(GB_i)}{\sum_{GB_j\in \mathrm{GB}}\mathrm{I}(GB_j)}, \tag{20}$$

and the information content of GB_i is given by

$$\mathrm{I}(GB_i) = |GB_i|\cdot AD(GB_i)\cdot SD(GB_i), \tag{21}$$

where $|GB_i|$ denotes the number of objects in GB_i, $AD(GB_i)$ is the aggregation degree, and $SD(GB_i)$ describes the distinguishing degree of GB_i.

According to the enhanced granular ball generation process and the construction of granular ball entropy, several interesting properties of the GB-E uncertainty measure can be derived as follows.

Proposition 1. *Let* DS $= \langle Q, P, D\rangle$ *be a decision information system, where* $Q/D = \{C_1, C_2, \dots, C_s\}$. *For a subset* $B \subseteq P$, *let* $GB\text{-}E_B(\mathrm{GB})$ *denote the granular ball entropy induced by* B. *Then, the following properties hold:*

(1) Non-positivity: $GB\text{-}E_P(\mathrm{GB}) \le 0$;
(2) Incomparability across subsets: For $Q \subseteq B \subseteq P$, $GB\text{-}E_B(\mathrm{GB})$ *and* $GB\text{-}E_Q(\mathrm{GB})$ *are incomparable;*
(3) Incomparability across granular ball groups: Given two groups of enhanced granular balls $GB^1 \subseteq GB^2$, $GB\text{-}E_B(\mathrm{GB}^1)$ *and* $GB\text{-}E_B(\mathrm{GB}^2)$ *are incomparable.*

Proof. These properties follow directly from the Definition 8. □

1) Based on the definition of the probability $p(GB_i)$ in Eq. (20), and since $p(GB_i) \in (0, 1]$, the entropy term $p(GB_i)\log p(GB_i)$ is non-positive. Therefore, the overall entropy $GB\text{-}E_P(\mathrm{GB}) \le 0$.
2) From Section 3.1, the process of enhanced granular ball generation is adaptive. Specifically, the centers and radius of the generated granular balls are determined in such a way that they do not satisfy an inclusion relationship across different feature subsets. As a result, for $Q \subseteq B \subseteq P$, the corresponding granular balls may differ in both structure and granularity, making $GB\text{-}E_B(\mathrm{GB})$ and $GB\text{-}E_Q(\mathrm{GB})$ generally incomparable.
3) Suppose $GB^1 \subseteq GB^2$, i.e., each granular ball in GB^1 is entirely contained in some ball of GB^2. Then, there exist $GB_i^1 \in GB^1$ and $GB_j^2 \in GB^2$ such that $GB_i^1 \subseteq GB_j^2$. This implies that

$$\frac{|GB_i^1|}{\sum_{GB_q^1\in GB^1,\mathrm{Lab}(GB_q^1)\neq \mathrm{Lab}(GB_i^1)}|GB_q^1|} \le \frac{|GB_j^2|}{\sum_{GB_q^2\in GB^2,\mathrm{Lab}(GB_q^2)\neq \mathrm{Lab}(GB_j^2)}|GB_q^2|},$$

and $Rad(GB_i^1) \le Rad(GB_j^2)$. However, the distances between balls satisfy $Dis_B(GB_i^1, GB_q^1) \ge Dis_B(GB_j^2, GB_q^2)$, depending on their spatial distribution. Hence, the separation degrees induced by GB^1 and GB^2 cannot be directly compared. Consequently, the entropy values $GB\text{-}E_B(\mathrm{GB}^1)$ and $GB\text{-}E_B(\mathrm{GB}^2)$ are also incomparable.

□

In summary, this subsection introduces a novel GB-E uncertainty measure that effectively captures the uncertainty inherent in the multi-granularity knowledge space. By jointly considering the separation degree and aggregation degree of each granular ball, the proposed measure provides a unified view of both inter-class discriminability and intra-class cohesion. This dual-perspective evaluation not only enhances the interpretability of granular structures but also lays a solid foundation for downstream tasks such as uncertainty modeling, feature selection, and decision-making within granular computing frameworks.

4. Feature selection based on granular ball uncertainty measure

To leverage the effectiveness of the proposed granular ball entropy in multi-granularity uncertainty modeling, a novel feature selection method is developed based on the GB-E measure. This method aims to identify an optimal feature subset by accurately quantifying feature significance within the constructed MGKS.

4.1. Feature evaluation and feature reduction

This subsection introduces two feature significance measures designed to evaluate and select informative features based on the principle of maximum information gain.

Definition 9. Let $\mathrm{DS} = \langle Q, P, D\rangle$ be a decision information system, where $Q/D = \{C_1, C_2, \dots, C_s\}$, and $GB\text{-}E_P(\mathrm{GB})$ denotes the GB-E uncertainty measure induced by attribute set P. For any attribute $p \in P$, the inner feature significance measure is defined as follows:

$$F_{\mathrm{inner}}(p, P, D) = GB\text{-}E_{P-\{p\}}(\mathrm{GB}) - GB\text{-}E_P(\mathrm{GB}). \tag{22}$$

The inner significance measure evaluates the contribution of a specific attribute within a given subset to uncertainty reduction. A higher value of $F_{\mathrm{inner}}(p, P, D)$ indicates that attribute p plays a more crucial role in preserving internal information. If $F_{\mathrm{inner}}(p, P, D) \leq 0$, the attribute is considered redundant. Core features are thus identified by selecting those attributes satisfying $F_{\mathrm{inner}}(p, P, D) > 0$. Similarly, the outer feature significance measure is defined as follows:

Definition 10. Let $\mathrm{DS} = \langle Q, P, D\rangle$ be a decision information system, where $Q/D = \{C_1, C_2, \dots, C_s\}$, and $GB\text{-}E_B(\mathrm{GB})$ denotes the GB-E uncertainty measure induced by subset $B \subseteq P$. For any attribute $p \in P - B$, the outer feature significance measure is defined as:

$$F_{\mathrm{outer}}(p, B, D) = GB\text{-}E_B(\mathrm{GB}) - GB\text{-}E_{B\cup\{p\}}(\mathrm{GB}). \tag{23}$$

The outer significance measure characterizes the importance of an attribute based on the information gain achieved by adding it to the current subset. A higher value of $F_{\mathrm{outer}}(p, B, D)$ indicates that the attribute p contributes more significantly to reducing the overall uncertainty and thus conveys greater informational value.

The ultimate goal of feature selection is to identify an informative subset of features that preserves the decision-making capability of the original information system. Given the effectiveness of granular ball entropy in characterizing the uncertainty and informativeness of features, we further establish a reduction criterion as follows.

Definition 11. Let $\mathrm{DS} = \langle Q, P, D\rangle$ be a decision information system, where $Q/D = \{C_1, C_2, \dots, C_s\}$, and $GB\text{-}E_P(\mathrm{GB})$ denotes the GB-E uncertainty measure induced by the full attribute set P. A subset $R \subseteq P$ is called a feature reduct if it satisfies the following two conditions:

1) $GB\text{-}E_R(\mathrm{GB}) \leq GB\text{-}E_P(\mathrm{GB})$;
2) For each $r \in R$, it holds that $GB\text{-}E_{R-\{r\}}(\mathrm{GB}) > GB\text{-}E_R(\mathrm{GB})$.

According to the above definition, condition 1) guarantees that the selected subset R retains at least as much discriminative information as the original complete set P, ensuring that the overall representational capability of the system is preserved. Condition 2) provides that each feature in the reduct is indispensable and contributes meaningfully to the descriptive power of the reduced system.

4.2. Algorithm design and its analysis

Building upon the feature evaluation and reduction principles established in Section 4.1, this subsection presents a heuristic forward feature selection algorithm, as detailed in Algorithm 1. The proposed algorithm consists of the following three key stages:

- **Core Feature Identification**: Core features are first extracted by evaluating the inner significance of each attribute with respect to the full feature set P. Specifically, attributes satisfying $F_{\mathrm{inner}}(p, P, D) > 0$ are retained, as they contribute positively to reducing uncertainty within the system.
- **Outer Feature Selection**: Informative features are then incrementally selected based on the outer significance measure. In each iteration, the feature p_o is selected such that

$$p_o = \arg\max_{p\in P-B} F_{\mathrm{outer}}(p, B, D),$$

where B is the currently selected feature subset. This strategy ensures the inclusion of features that offer the most significant additional information gain.

Table 1
Time complexity analysis of proposed FGB-E method.

Steps	Description	Time complexity
Step I	Core feature identification	$O(nm^2(k+r))$
Step II	Outer feature selection	$O\left(\sum_{i=1}^{l_2} n(l_1+i)(m-l_1-i-1)(k+r)\right)$
Step III	Redundancy elimination	$O(n(l_1+l_2)(l_1+l_2-1)(k+r))$

- **Redundancy Elimination**: Given the non-monotonic property of the GB-E uncertainty measure, a refinement step is performed to eliminate any redundant features. This post-processing step guarantees that each feature in the final subset is indispensable, satisfying the minimality condition specified in Definition 11 while maintaining the original discriminative power.

The time complexity of Algorithm 1 is analyzed as follows. In Step I, the enhanced granular ball generation requires a computational cost of $O(nmk)$, where k denotes the maximum number of granular balls generated during each iteration. The computation of GB-E measure incurs a complexity of $O(nmr)$, where r represents the number of enhanced granular balls generated at the final stage. Additionally, the complexity of the inner important feature selection is $O(n(m-1)km + nr(m-1)m)$. To avoid ambiguity, it is noted that r also refers to the number of randomly generated granular balls in each run. Thus, the total time complexity of Step I is $O(nm^2(k+r))$. Suppose there are l_1 and l_2 features selected in Steps I and II, the iterative time complexity of Step II is given by $O\left(\sum_{i=1}^{l_2} n(l_1+i)(m-l_1-i-1)(k+r)\right)$. Similarly, the redundant feature elimination in Step III has a computational cost of $O(n(l_1+l_2)(l_1+l_2-1)(k+r))$. The complexity of each step is shown in Table 1. Considering that the maximum values of k and r are bounded by the number of classes s and the number of instances n, respectively, the overall time complexity of the proposed algorithm is $O\left((k+r)(nm^2 + \sum_{i=1}^{l_2} n(l_1+i)(m-l_1-i-1) + n(l_1+l_2)(l_1+l_2-1))\right)$.

Algorithm 1: Feature selection algorithm based on GB-E uncertainty measure(FGB-E).

Input: Decision information system DS $= \langle Q, P, D\rangle$, purity $Pur \in (0,1]$;
Output: Feature reduction R.
`// Stage I: Core feature identification`
Initialization the granular ball GB^0 with $Ce(GB_1^0)$ and $Rad(GB_1^0)$ in Eq. (6);
Obtain the enchanced granular ball GB $= \{GB_1, GB_2, \ldots, GB_r\}$ according to Section 3.1;
Calculate the GB-E uncertainty measure $GB\text{-}E_P(\mathrm{GB})$ according to Definition 8;
for $p \in P$ **do**
  Update the enchanced granular ball GB according to Section 3.1;
  Calculate the GB-E uncertainty measure $GB\text{-}E_{P-\{p\}}(\mathrm{GB})$ and inner significant measure $F_{\mathrm{inner}}(p, P, D)$ from Eq. (22);
  if $F_{inner}(p, P, D) > 0$ **then**
    $R \leftarrow R \cup \{p\}$;
Calculate the GB-Entropy $GB\text{-}E_R(\mathrm{GB})$;
`// Stage II: Outer feature selection`
while $GB\text{-}E_R(\mathrm{GB}) > GB\text{-}E_P(\mathrm{GB})$ **do**
  for $p \in P - B$ **do**
    Update the enchanced granular ball GB induced by $R \cup \{p\}$;
    Calculate the GB-E uncertainty measure $GB\text{-}E_{R\cup\{p\}}(\mathrm{GB})$ and outer significant measure $F_{\mathrm{outer}}(p, R, D)$ from Eq. (23);
  $p_o = \arg\max_{p\in P-R} F_{\mathrm{outer}}(p, R, D)$;
  $R \leftarrow R \cup \{p_o\}$;
`// Stage III: Redundancy elimination`
for $r \in R$ **do**
  Update the enchanced granular ball GB induced by $R - \{r\}$;
  Calculate the GB-E uncertainty measure $GB\text{-}E_{R-\{r\}}(\mathrm{GB})$;
  if $GB\text{-}E_{R-\{r\}}(\mathrm{GB}) \le GB\text{-}E_R(\mathrm{GB})$ **then**
    $R \leftarrow R - \{r\}$;
return R.

Example 1. An example of a decision information system is illustrated in Table 2 to demonstrate the feature-selection process in Algorithm 1. In this system the universe consists of nine objects $Q = \{u_1, u_2, \ldots, u_9\}$, the feature set is $P = \{p_1, p_2, p_3, p_4\}$, and the decision partition is $Q/D = \{\{u_1, u_2, u_3\}, \{u_4, u_5, u_6\}, \{u_7, u_8, u_9\}\}$.

As described in Section 3.1, the enhanced granular ball initialization Eq. (6) yields the initial centers

$$Ce_1 = (0.2600, 0.2933, 0.2433, 0.2800), Ce_2 = (0.5667, 0.5533, 0.5633, 0.5600), Ce_3 = (0.9667, 0.9767, 0.9767, 0.9867),$$

Table 2
An example of a decision information system.

Q	p_1	p_2	p_3	p_4	D
u_1	0.29	0.25	0.22	0.31	1
u_2	0.26	0.28	0.27	0.28	1
u_3	0.23	0.35	0.24	0.25	1
u_4	0.58	0.56	0.54	0.54	2
u_5	0.53	0.57	0.59	0.58	2
u_6	0.59	0.53	0.56	0.56	2
u_7	0.97	0.97	0.95	0.97	3
u_8	0.93	1.00	1.00	0.99	3
u_9	1.00	0.96	0.98	1.00	3

Table 3
Granular-ball results for the example in Table 2.

Q	Center	Radius	$d(u, Ce_1)$	$d(u, Ce_2)$	$d(u, Ce_3)$
u_1			**0.0623**	0.5888	1.4212
u_2	(0.2600,0.2933,0.2433,0.2800)	0.0528	**0.0281**	0.5750	1.4092
u_3			**0.0680**	0.5884	1.4173
u_4			0.5647	**0.0354**	0.8530
u_5	(0.5667,0.5533,0.5633,0.5600)	0.0388	0.5969	**0.0481**	0.8205
u_6			0.5863	**0.0329**	0.8341
u_7			1.3954	0.8159	**0.0289**
u_8	(0.9667,0.9767,0.9767,0.9867)	0.0394	1.4212	0.8428	**0.0497**
u_9			1.4293	0.8490	**0.0395**

where all values are rounded to four decimal places. Using these centers, the corresponding radii and the distances between objects and centers are obtained and summarized in Table 3.

With purity $Pur = 0.8$, the final granular balls are

$$GB_1 = \{u_1, u_2, u_3\}, \quad GB_2 = \{u_4, u_5, u_6\}, \quad GB_3 = \{u_7, u_8, u_9\},$$

as given by Eq. (9). Based on these granulation results, the separation degrees (SD) and aggregation degrees (AD) of each granular ball (computed via Eqs. (15) and (17)) are

$$SD(GB_1) = 0.904914, \quad SD(GB_2) = 0.620832, \quad SD(GB_3) = 1.035646;$$

$$AD(GB_1) = 0.999572, \quad AD(GB_2) = 0.999742, \quad AD(GB_3) = 0.999944.$$

Using Definition 8, the granular-ball entropy under feature set P is

$$GB\text{-}E_P(GB) = -\sum_{GB_i \in \mathrm{GB}} p(GB_i) \log_2 p(GB_i) = 1.5541,$$

(where the probabilities and the entropy are rounded to four decimal places).

We also compute the entropy after deleting each single attribute

$$GB\text{-}E_{P\setminus\{p_1\}}(GB) = 1.5541, \quad GB\text{-}E_{P\setminus\{p_2\}}(GB) = 1.5547,$$

$$GB\text{-}E_{P\setminus\{p_3\}}(GB) = 1.5535, \quad GB\text{-}E_{P\setminus\{p_4\}}(GB) = 1.5541.$$

According to Eq. (22), the Step I core inner feature identification yields $R = \{p_2\}$, and the corresponding entropy is $GB\text{-}E_R(GB) = 1.5519$. Since the $GB\text{-}E_R(GB) < GB\text{-}E_P(GB)$, the final reduct for this example is $R = \{p_2\}$. If multiple features are selected, eliminating redundant features will be necessary to ensure the importance of the selected features.

5. Experimental analysis

To verify the effectiveness and robustness of the proposed FGB-E method, we conduct extensive experiments on twelve publicly available public datasets. All experiments are performed in a consistent computing environment with the following specifications: operating system-Microsoft Windows 10; processor-Intel(R) Core(TM) i7-6800K CPU@3.40GHz ×12; memory-62.7 GB; programming platform-MATLAB.

5.1. Experimental design

This subsection outlines the experimental design to provide a comprehensive and fair evaluation of the proposed FGB-E method. Specifically, we describe the datasets employed in the experiments, the benchmark methods selected for comparison, and the general experimental settings. These elements collectively ensure that the proposed method's performance is rigorously assessed across various tasks and conditions.

Table 4
Basic description of twelve datasets.

No.s	Datasets	Abbreviations	Objects	Features	Classes
1	Breast cancer coimbra	Bcco	116	9	2
2	Cardiotocography	Card	2126	20	3
3	Climate model simulation crashes	Cmsc	540	17	2
4	Colon	Colo	62	2000	2
5	German	Germ	1000	20	2
6	Heart	Hear	270	13	2
7	ILPD	Ilpd	579	10	2
8	Mice protein expression	Mpex	1077	68	8
9	Segmentation	Sege	2310	19	7
10	Spectf heart	Spec	267	44	2
11	Thyroid	Thyr	7200	22	3
12	WDBC	Wdbc	569	30	2

5.1.1. Datasets

In the experiments, twelve datasets from the UCI Machine Learning Repository[1] are employed for performance comparison. Detailed information about these datasets is summarized in Table 4. Prior to processing, all datasets are normalized to the unit interval $[0, 1]$ using the following equation:

$$\widehat{f}(u_i, p_j) = \frac{f(u_i, p_j) - \min(p_j)}{\max(p_j) - \min(a_j)}, \tag{24}$$

where $\widehat{f}(u_i, p_j)$ and $f(u_i, p_j)$ denote the normalized and original values of object u_i with respect to feature p_j, respectively. Here, $\max(p_j)$ and $\min(p_j)$ represent all objects' maximum and minimum values of feature p_j.

5.1.2. Benchmark comparison methods

To evaluate the performance of the proposed method, seven representative algorithms are selected as benchmark methods. Detailed descriptions of these methods are provided below.

- FHFS [43]: This method is based on refined justifiable granularity and further designs granularity information measures to feature selection from an uncertainty perspective.
- ARSC [54]: This approach constructs a fuzzy dominance approximation model to mitigate the influence of noisy samples and further selects the optimal feature subset by considering both class separability and attribute correlation.
- GBNRS [49]: This method introduces the concept of granular balls into Pawlak's rough set to define the granular-ball rough set model and perform feature reduction based on the positive region.
- FNRS [41]: This method incorporates neighborhood concepts to reconstruct fuzzy lower and upper approximations and applies them to feature selection based on fuzzy dependency.
- FRMR [55]: This is a fuzzy rough set-based feature selection method that integrates fuzzy dependency and classification error for feature selection.
- FScNCE [56]: This approach proposes neighborhood combination entropy to describe feature uncertainty from an information-theoretic perspective.
- MEFS [57]: This method develops multi-scale fuzzy information measures, including fuzzy entropy and mutual information, which are employed for feature selection until the overall information amount remains stable.

5.1.3. General setting

It is noteworthy that the purity degree Pur of enhanced granular balls is a critical parameter that directly affects both the construction of the multi-granularity knowledge space and the GB-E uncertainty measure, thereby influencing the feature selection process. To investigate the impact of this parameter, Pur varies from 0.60 to 0.99 with an increment of 0.01. Meanwhile, three classifiers, including K-Nearest Neighbors (KNN with $K = 3$), Naive Bayes (NB), and Support Vector Machine (SVM), are employed to evaluate the performance of different methods. To ensure a fair comparison, all experiments adopt ten-fold cross-validation. Each dataset is partitioned into ten subsets, where the nine subsets are used for training and the remaining one for testing.

5.2. Performance comparison and analysis

In this subection, we illustrate the effectiveness of proposed FGB-E by comparing it with other seven representative algorithms. The main comparison and analysis mainly consists three aspects including computational time, feature number, and classification performance.

[1] https://www.uci.edu/

Table 5
Eight compared methods on twelve datasets (s): computing overhead.

Datasets	FHFS	ARSC	GBNRS	FNRS	FRMR	FScNCE	MEFS	FGB-E
Bcco	32.11	3177.93	10860.90	**16.17**	1099.19	16.43	165.08	537.13
Card	3.58	79.43	580.10	**1.09**	35.15	1.53	2.65	36.90
Cmsc	0.32	1.30	26.57	**0.22**	0.91	0.47	0.73	4.55
Colo	6101.46	441521.88	104079.24	**3.33**	332.03	34.59	340.15	1100.74
Germ	10.59	225.10	2039.41	**2.06**	228.78	4.04	21.94	89.29
Hear	1.16	7.83	149.45	**0.31**	7.24	0.80	2.48	10.39
Ilpd	1.51	14.91	261.77	**0.37**	8.43	1.07	2.38	22.23
Mpex	424.14	15201.95	100461.63	98.91	3397.80	**21.91**	1872.79	1516.31
Sege	68.76	6670.50	16030.43	44.23	1206.24	**14.46**	439.34	337.53
Spec	12.27	159.66	1607.69	**0.92**	58.30	1.76	40.44	21.91
Thyr	**71.14**	64611.77	460.26	190.12	9056.34	118.09	2971.42	388.43
Wdbc	11.46	230.11	1578.59	**2.61**	161.26	3.68	46.64	7.65
Ave	561.54	44325.20	19844.67	30.03	1299.30	**18.24**	492.17	339.42

Table 6
Number of selected features among compared methods under KNN classifier.

Datasets	RAW	FHFS	ARSC	GBNRS	FNRS	FRMR	FScNCE	MEFS	FGB-E
Bcco	9	5	7	**1**	2	9	8	9	7
Card	20	12	20	**2**	19	20	20	9	5
Cmsc	17	10	16	**1**	14	9	17	4	9
Colo	2000	1200	31	4	6	29	**1**	16	595
Germ	20	12	20	**3**	**3**	20	20	20	19
Hear	13	8	12	**4**	5	13	12	13	9
Ilpd	10	6	7	3	**1**	**1**	10	9	5
Mpex	68	41	29	**1**	44	65	66	65	21
Sege	19	11	13	**1**	15	17	16	18	10
Spec	44	26	14	**2**	14	41	22	18	12
Thyr	22	13	**3**	5	15	21	20	17	4
Wdbc	30	18	29	**1**	21	29	28	29	14
Ave	189.33	113.50	16.75	**2.33**	13.25	22.83	20.00	18.92	59.17

5.2.1. Computational time

This subsection records and analyzes the computational time of eight compared methods. Table 5 presents the computational time (in seconds) required by eight competing methods across twelve benchmark datasets. It is evident that the proposed method, FGB-E, achieves a favorable balance between efficiency and performance. Although not always the fastest on every dataset, FGB-E consistently demonstrates competitive running time, especially when compared to computationally intensive baselines like ARSC and GBNRS.

Specifically, FGB-E significantly outperforms ARSC and GBNRS in terms of efficiency on large-scale datasets such as Colo and Mpex, where their execution times reach several orders of magnitude higher than FGB-E. For instance, on the Colo dataset, FGB-E requires only 1100.74 s, while ARSC and GBNRS take 441521.88 and 104079.24 s, respectively. Similar trends are observed on Mpex and Sege datasets.

Although methods like FNRS and FScNCE are faster on certain small-scale datasets, FGB-E provides a more consistent trade-off across different dataset scales. Meanwhile, the average running time of FGB-E (339.42s) is substantially lower than the other five traditional methods (e.g., GBNRS: 19844.67s, ARSC: 44325.20s), highlighting its scalability and practical applicability in real-world scenarios.

5.2.2. Number of selected features

This subsection compares the number of features selected by different methods across various classifiers, with the corresponding results presented in Tables 6–8. In these tables, “RAW” denotes the original datasets without feature selection. Overall, all feature selection methods reduce the dimensionality compared to RAW, confirming the presence of redundant features in the original data and demonstrating the effectiveness of these methods in feature reduction.

Across different classifiers, GBNRS consistently selects the fewest features in most cases and achieves the lowest average number of selected features. Specifically, it records the smallest feature count in 8, 6, and 9 datasets under the three classifiers, respectively. However, selecting only one or two features may lead to insufficient discriminative information, resulting in suboptimal classification performance. Similarly, a trend was also observed for FNRS under the NB classifier.

In contrast, the proposed FGB-E method generally selects a more appropriate number of features, yielding a better balance between dimensionality reduction and classification accuracy, except on the Colon dataset. Note that the Colo dataset is a small-sample and high-dimensional dataset, where many features are weakly informative but jointly contribute to discriminative representation. In this situation, there are marginal features that slightly reduce uncertainty and are retained. Excluding Colon, FGB-E achieves average

Table 7
Number of selected features among compared methods under NB classifier.

Datasets	RAW	FHFS	ARSC	GBNRS	FNRS	FRMR	FScNCE	MEFS	FGB-E
Bcco	9	5	7	**1**	5	9	8	9	6
Card	20	12	20	2	**1**	20	8	9	3
Cmsc	17	10	16	**1**	9	9	17	4	9
Colo	2000	1200	31	4	6	29	**1**	16	738
Germ	20	12	20	**3**	7	20	20	20	16
Hear	13	8	12	4	**2**	13	12	13	12
Ilpd	10	6	7	3	**1**	10	8	9	**1**
Mpex	68	41	29	**1**	42	65	66	65	36
Sege	19	11	13	**1**	11	17	15	18	9
Spec	44	26	14	2	**1**	**1**	**1**	40	2
Thyr	22	13	3	5	13	21	20	21	**1**
Wdbc	30	18	29	**1**	13	29	28	13	14
Ave	189.33	113.50	16.75	**2.33**	9.25	20.25	17.00	19.75	70.58

Table 8
Number of selected features among compared methods under SVM classifier.

Datasets	RAW	FHFS	ARSC	GBNRS	FNRS	FRMR	FScNCE	MEFS	FGB-E
Bcco	9	5	7	**1**	4	9	9	9	7
Card	20	12	20	**2**	18	20	20	20	8
Cmsc	17	10	16	**1**	11	9	17	17	16
Colo	2000	1200	31	4	6	29	**1**	16	816
Germ	20	12	20	**3**	12	20	20	20	19
Hear	13	8	12	**4**	5	13	12	13	8
Ilpd	10	6	7	3	7	**1**	10	10	**1**
Mpex	68	41	29	**1**	52	65	66	65	52
Sege	19	11	13	**1**	15	17	15	18	15
Spec	44	26	14	**2**	14	41	42	40	14
Thyr	22	13	**3**	5	13	21	20	21	4
Wdbc	30	18	29	**1**	22	29	28	27	14
Ave	189.33	113.50	16.75	**2.33**	14.92	22.83	21.67	23.00	81.17

Fig. 4. The number of selected features among compared methods under three classifiers.

Table 9
Classification accuracy of compared methods on twelve datasets under KNN classifier(%).

Datasets	RAW	FHFS	ARSC	GBNRS	FNRS	FRMR	FScNCE	MEFS	FGB-E
Bcco	70.68±8.23	70.08±9.21	75.83±7.94	64.92±15.48	67.12±14.93	66.44±16.25	67.27±10.40	66.44±7.19	**76.74±9.13**
Card	79.12±2.90	79.87±2.80	78.50±2.13	58.94±3.02	77.90±2.77	78.27±2.65	78.60±2.49	85.42±2.21	**87.25±2.84**
Cmsc	92.96±2.28	93.70±1.29	93.15±1.52	90.74±1.24	93.15±2.32	92.78±1.84	92.78±1.84	94.07±2.73	**94.81±2.73**
Colo	70.71±10.83	78.09±17.05	85.71±16.03	79.05±17.45	77.38±11.46	72.62±20.30	54.76±20.45	**86.67±13.15**	75.48±12.20
Germ	75.10±3.11	75.00±1.49	74.70±1.89	71.20±3.05	69.40±4.17	75.00±3.06	74.40±4.97	75.00±4.47	**75.70±1.89**
Hear	82.22±7.96	82.59±6.77	77.04±10.88	79.63±7.25	81.11±3.68	80.37±5.25	81.11±3.68	80.74±5.47	**84.07±7.00**
Ilpd	69.61±4.42	68.06±5.05	72.02±5.73	71.16±4.29	71.68±2.52	71.50±0.87	68.74±3.79	69.09±5.73	**72.37±3.29**
Mpex	96.84±1.47	95.73±2.71	96.75±1.92	27.30±4.31	97.22±1.51	97.21±1.57	96.94±1.58	96.57±1.85	**97.49±0.76**
Sege	94.89±0.79	94.76±1.42	94.81±1.41	62.68±3.15	95.02±1.14	94.94±0.79	95.06±1.36	95.02±1.12	**95.97±1.23**
Spec	78.62±5.77	76.35±6.38	77.89±7.01	78.26±7.31	80.14±5.85	75.34±7.92	74.99±10.55	78.65±6.34	**81.00±8.70**
Thyr	**91.85±0.84**	79.52±0.84	91.77±1.12	50.17±0.83	91.74±0.76	91.75±0.83	91.40±0.80	91.65±1.13	50.17±0.93
Wdbc	96.66±2.27	97.01±2.49	**97.89±2.00**	88.40±3.90	96.49±2.34	97.01±2.99	97.19±1.23	96.66±2.92	97.54±1.89
Ave±Std	83.45±4.21	83.84±4.78	85.09±4.92	70.21±6.40	83.67±4.79	82.96±5.33	81.32±4.83	85.04±4.49	**86.26±4.31**
Rank	5.17	5.08	4.50	7.00	4.67	5.75	5.92	4.33	**1.75**

feature counts of 10.45, 9.91, and 14.36 under the three classifiers, respectively, which outperforms the other 6, 5, and 6 methods. These results demonstrate effective and stable reduction ability in feature selection across different classifiers, where a more intuitive manifestation is presented in Fig. 4.

5.2.3. Classification performance

This subsection compares the classification performance of compared methods across twelve benchmark datasets under three classifiers. The detailed results are presented in Tables 9–11, where average accuracy and standard deviation are denoted as Ave and Std, respectively. For clarity, the best results in each setting are highlighted in bold.

Table 10
Classification accuracy of compared methods on twelve datasets under NB classifier(%).

Datasets	RAW	FHFS	ARSC	GBNRS	FNRS	FRMR	FScNCE	MEFS	FGB-E
Bcco	67.20±6.95	68.41±17.26	65.61±18.13	61.89±11.31	63.41±18.65	66.21±10.85	70.61±11.73	63.64±9.60	**77.20±15.89**
Card	53.86±6.86	66.60±2.63	49.67±3.28	58.37±2.59	56.91±3.77	50.05±2.81	67.54±4.06	57.95±4.12	**70.79±3.73**
Cmsc	92.04±1.25	93.52±1.57	92.04±1.76	91.48±0.96	94.07±2.28	93.15±1.96	92.04±1.52	94.07±1.91	**94.26±2.04**
Colo	68.09±11.89	71.43±18.68	72.62±19.06	80.95±11.50	64.76±4.02	68.33±12.75	64.76±4.02	**88.81±10.29**	74.52±12.65
Germ	70.70±1.64	71.20±2.49	70.50±1.58	70.70±1.83	71.50±2.22	70.40±2.07	70.20±2.15	70.40±2.41	**71.80±2.15**
Hear	80.74±7.77	78.52±6.49	79.26±3.58	78.52±6.72	78.15±6.86	81.48±7.61	81.11±7.89	80.74±9.37	**81.48±6.76**
Ilpd	66.83±4.49	71.52±6.26	67.68±7.08	71.51±4.64	71.16±2.42	67.02±4.59	70.81±4.75	67.52±5.40	**71.85±1.57**
Mpex	84.67±3.54	83.00±3.37	85.53±4.03	32.49±3.53	85.52±2.58	86.07±4.64	85.61±2.48	84.41±2.56	**88.40±2.96**
Sege	90.17±1.71	89.83±1.84	90.09±1.26	61.47±2.58	**92.86±1.65**	90.26±2.47	91.21±2.08	89.83±2.32	92.68±1.59
Spec	72.99±7.45	74.54±8.71	71.94±9.05	75.27±6.45	**79.42±1.75**	**79.42±1.75**	**79.42±1.75**	73.35±9.85	**79.42±1.75**
Thyr	88.75±0.77	87.44±1.07	88.70±0.99	70.97±1.02	88.86±1.05	88.78±0.55	88.68±0.72	**88.92±0.80**	70.97±1.58
Wdbc	94.03±1.88	**95.26±1.85**	93.84±2.53	88.58±5.18	94.55±2.10	94.02±2.79	94.36±3.54	94.91±2.25	95.26±2.87
Ave±Std	77.95±4.64	79.86±5.96	77.93±5.99	70.11±5.21	78.87±4.05	78.37±4.55	80.14±3.86	79.97±5.04	**82.79±4.50**
Rank	5.92	4.33	5.92	6.58	4.50	5.17	4.33	5.33	**1.33**

Table 11
Classification accuracy of compared methods on twelve datasets under SVM classifier(%).

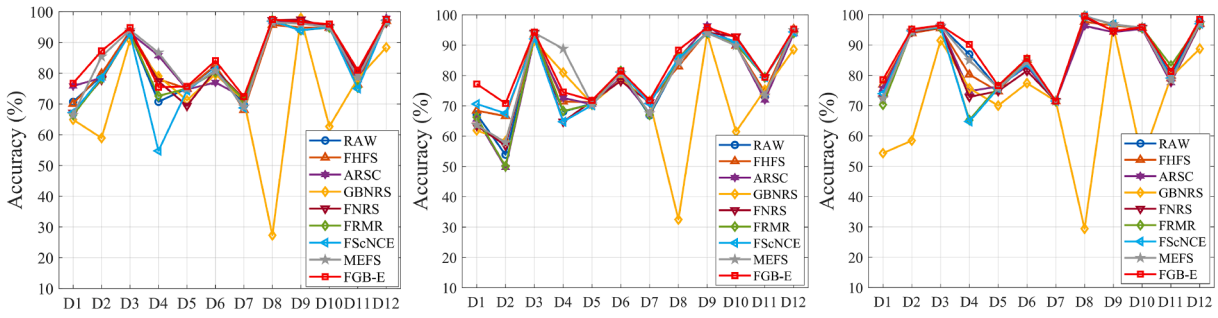
Datasets	RAW	FHFS	ARSC	GBNRS	FNRS	FRMR	FScNCE	MEFS	FGB-E
Bcco	73.94±16.10	76.97±10.93	75.53±17.11	54.32±7.80	72.35±7.89	70.30±15.93	73.94±11.94	72.35±12.18	**78.56±10.76**
Card	94.59±1.51	93.89±0.86	94.26±1.01	58.47±3.33	94.40±1.27	94.78±2.05	94.69±1.61	94.45±1.15	**95.20±1.42**
Cmsc	96.30±2.14	95.56±2.79	96.30±2.76	91.48±0.96	95.74±2.32	95.74±2.15	95.93±2.10	96.30±2.31	**96.48±3.32**
Colo	86.91±17.18	80.24±13.23	74.76±13.34	75.71±11.71	72.86±26.66	65.24±17.57	64.76±4.02	85.00±14.59	**90.24±11.53**
Germ	75.60±3.03	74.70±1.34	76.30±3.92	70.00±0.00	74.80±4.49	76.30±4.69	75.70±4.17	75.70±3.33	**76.60±4.30**
Hear	84.07±7.82	85.19±5.52	82.96±5.58	77.41±6.40	81.48±5.24	84.44±10.15	83.70±6.58	84.44±8.87	**85.56±5.08**
Ilpd	**71.50±0.87**	**71.50±0.87**	**71.50±0.87**	**71.50±0.87**	**71.50±0.87**	**71.50±0.87**	**71.50±0.87**	**71.50±0.87**	**71.50±0.87**
Mpex	99.44±0.65	97.68±1.26	96.29±1.80	29.43±2.77	99.07±0.98	**99.53±0.49**	99.53±0.66	99.44±0.48	99.53±0.66
Sege	95.50±1.73	95.67±1.00	95.33±1.15	52.34±1.31	95.63±1.36	95.67±1.57	95.63±1.68	95.71±1.33	**95.89±1.52**
Spec	79.80±8.04	79.07±8.00	77.55±5.47	79.42±1.75	78.31±3.65	**83.16±3.08**	79.07±7.72	78.32±8.05	81.28±2.96
Thyr	83.42±0.95	**84.41±0.87**	83.28±0.63	70.97±1.14	83.26±0.96	83.20±1.15	83.37±1.08	83.34±0.86	70.97±1.11
Wdbc	97.72±2.20	97.01±1.44	96.66±2.10	88.75±3.14	96.84±1.81	97.02±2.62	97.02±2.03	97.19±3.53	**98.42±1.75**
Ave±Std	87.68±5.14	86.99±3.96	85.97±4.62	68.08±3.64	85.64±4.74	85.86±5.14	85.69±3.64	87.27±4.75	**88.65±3.68**
Rank	3.58	4.92	5.83	7.33	6.17	3.58	4.08	3.42	**1.67**

The classification accuracy of the compared methods under the KNN classifier is recorded in Table 9. It can be observed that the proposed FGB-E consistently achieves superior performance on most datasets. Specifically, FGB-E attains the highest accuracy on 9 out of 12 datasets, except on the Colo, Thyr, and Wdbc datasets. Notably, it outperforms the second-best method by a considerable margin on datasets such as Card, Sege, and Spec, demonstrating its effectiveness in capturing discriminative information. Moreover, FGB-E achieves the highest average accuracy of 86.26%, with a standard deviation of 4.31, outperforming all compared methods. In contrast, methods like GBNRS show relatively low performance (Ave = 70.21%), indicating limited generalizability under the limited selected features.

Table 10 reports the classification accuracy of all competing methods on twelve benchmark datasets using the NB classifier. Similarly, the proposed FGB-E performs better under this classifier, obtaining the best results on 8 out of 12 datasets, including Bcco, Card, Cmsc, Germ, Hear, Ilpd, Mpex, and Wdbc. Notably, on the Bcco and Card datasets, FGB-E significantly outperforms the second-best methods by margins of 6.59% and 4.19%, respectively. Although several methods such as FScNCE, FNRS, and MEFS occasionally show competitive results (e.g., FScNCE on Spec and Hear), they generally exhibit greater performance fluctuations, as reflected by higher standard deviations on certain datasets. In contrast, FGB-E achieves high accuracy and maintains a relatively low standard deviation on most datasets, indicating its robustness and stability.

Moreover, the classification accuracy of the SVM classifier is summarized in Table 11. Overall, the proposed FGB-E method achieves the highest average accuracy of 88.65%, outperforming all compared methods. Specifically, FGB-E obtains the best performance on 9 out of 12 datasets, except for the Mpex and Spec on FRMR, and Thyr on FHFS. In contrast, methods such as GBNRS show relatively poor performance across most datasets, especially on Card, Sege, and Mpex, where accuracy drops significantly. Although the number of selected features of Colo dataset is high, these features jointly form a compact representation that preserves the class-discriminative structure, as verified by the consistent improvement in classification accuracy across multiple classifiers compared with most methods, especially it achieves the first and second highest accuracy on the SVM and NB classifiers. In addition, from the whole experimental dataset, the proposed FGB-E also achieves the best overall performance, with the highest average accuracy of 88.65%. These results collectively demonstrate the effectiveness and robustness of FGB-E across diverse datasets and classification tasks.

Particularly, the average accuracy and ranking of the compared methods under the three classifiers are presented in Fig. 5. As shown in these figures, the average accuracy of FGB-E across different classifiers is consistently higher than that of the other methods. Meanwhile, the corresponding rankings of FGB-E are consistently lower, indicating its superior classification performance. Based on



(a) KNN (b) NB (c) SVM

Fig. 5. Classification performance of compared methods on twelve benchmark datasets under three classifiers.

Table 12
Computational time of the compared methods on four text and image datasets.

Datasets	Abbreviations	Train	Test	Dimension	Class	Type
Wiki	IWiki	2173	693	128	10	Text
Pascal sentences	IPS	700	300	1024	20	Text
Wiki	TWiki	2173	693	10	10	Image
Pascal sentences	TPS	700	300	100	20	Image

Table 13
Computational time (s) of the compared methods on four text and image datasets.

Datasets	FHFS	ARSC	GBNRS	FNRS	FRMR	FScNCE	MEFS	Our
TWiki	**26.87**	2328.16	15908.26	26.96	314.10	115.55	108.70	496.69
TPS	998.32	20380.10	211671.92	**137.67**	1671.54	427.25	1505.31	1518.77
TWiki	1954.19	1736.67	2307.76	**489.47**	4154.00	1709.01	1806.37	1018.41
IPS	139252.04	–	4983.61	**499.51**	4198.38	2990.76	7526.56	11546.00
Ave	35557.86	8148.31	58717.89	**288.40**	2584.51	1310.64	2736.74	3644.97

the above results and analysis, the proposed FGB-E demonstrates better classification capabilities with appropriately selected features, further illustrating its effectiveness in feature selection.

5.3. Comparison on text and image datasets

This subsection evaluates the proposed method on both image and text datasets. Four benchmark datasets adopted from [58] and [59] are employed for the experiments, and their detailed descriptions are summarized in Table 12. All methods are trained on the training set and validated on the testing set.

The computational time and the number of selected features are reported in Tables 13 and 14, where the best results are highlighted in bold. Note that ARSC could not complete the computation on the IPS dataset within a reasonable time, so the result is denoted by "-". As shown in Table 13, the FHFS and FNRS achieve the lowest computational time on one and three datasets, respectively, mainly because they select a smaller number of features. In contrast, the proposed FGB-E method demonstrates a balanced performance, achieving competitive efficiency while significantly outperforming FHFS, ARSC, and GBNRS regarding overall stability and accuracy.

Moreover, the numbers of selected features under different classifiers are presented in Table 14. It can be observed that GBNRS and FRMR tend to select a minimal number of features, sometimes even only one, which fails to capture sufficient discriminative information and thus results in poor performance. By contrast, the proposed FGB-E method adaptively selects a more flexible number of features, leading to improved performance across classifiers. Specifically, it outperforms FHFS under the KNN classifier, surpasses FHFS and ARSC under NB, and achieves superior results to FHFS under SVM.

The classification accuracies of all compared methods on the four datasets are listed in Table 15. As observed, the proposed method consistently achieves the highest accuracy across most tasks, except for the TPS dataset under the KNN classifier. These results demonstrate the robustness and effectiveness of the proposed method in handling high-dimensional text and image data.

5.4. Statistical tests

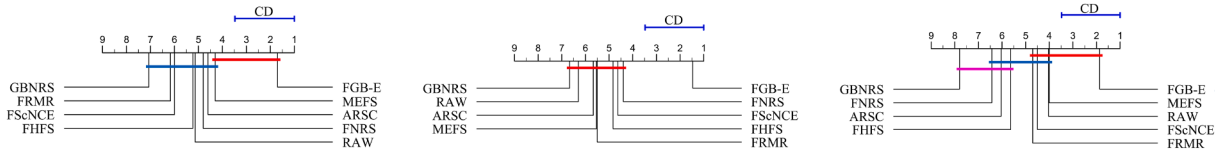
To further validate whether significant differences exist among the compared methods in terms of classification performance, statistical tests were conducted across all datasets and classifiers. First, the Friedman test [60] is employed to assess whether there exist differences in all methods. The null hypothesis assumes that all methods perform equally well, and it can be rejected when the p-value is less than the given significance level $\alpha = 0.1$. The p-values obtained from the Friedman test under the three classifiers are

Table 14
Number of selected features among the compared methods on four text and image datasets.

No.s	Datasets	RAW	FHFS	ARSC	GBNRS	FNRS	FRMR	FScNCE	MEFS	Our
KNN	TWiki	10	6	10	**1**	10	10	10	10	9
	TPS	100	60	87	**12**	87	54	79	88	71
	TWiki	128	77	85	4	9	**1**	97	107	65
	IPS	1024	614	–	6	12	**1**	7	82	195
	Ave	315.50	189.25	60.67	5.75	29.50	**16.50**	48.25	71.75	85.00
NB	TWiki	10	6	10	**1**	10	10	10	10	9
	TPS	100	60	87	**12**	40	54	67	86	16
	TWiki	128	77	85	4	9	**1**	97	107	65
	IPS	1024	614	–	6	12	**1**	7	82	151
	Ave	315.50	189.25	60.67	5.75	17.75	**16.50**	45.25	71.25	60.25
SVM	TWiki	10	6	10	**1**	10	10	10	10	9
	TPS	100	60	87	**12**	88	54	81	88	65
	TWiki	128	77	85	4	9	**1**	97	107	65
	IPS	1024	614	–	6	12	**1**	7	82	151
	Ave	315.50	189.25	60.67	5.75	29.75	16.50	48.75	71.75	72.50

Table 15
Classification accuracy of the compared methods on four text and image datasets under three classifiers(%).

No.s	Datasets	RAW	FHFS	ARSC	GBNRS	FNRS	FRMR	FScNCE	MEFS	Our
KNN	TWiki	66.40±0.06	66.24±0.07	65.67±0.04	14.43±0.03	66.40±0.06	66.40±0.06	66.40±0.06	66.40±0.06	**66.53±0.04**
	TPS	43.67±0.08	42.67±0.11	43.67±0.08	18.33±0.05	44.33±0.12	39.00±0.09	42.33±0.11	**44.67±0.08**	44.33±0.08
	TWiki	17.04±0.04	17.61±0.03	18.03±0.04	17.46±0.04	16.30±0.05	10.83±0.03	16.89±0.04	17.34±0.06	**18.46±0.04**
	IPS	14.33±0.07	14.33±0.06	–	5.67±0.06	8.33±0.05	5.67±0.05	6.00±0.04	10.33±0.06	**14.67±0.07**
NB	TWiki	66.81±0.04	65.37±0.06	67.23±0.06	21.33±0.05	66.81±0.04	66.81±0.04	66.81±0.04	66.81±0.04	**67.83±0.06**
	TPS	14.00±0.05	29.67±0.08	17.67±0.11	23.33±0.08	33.00±0.09	25.67±0.04	16.33±0.08	15.00±0.06	**33.00±0.10**
	TWiki	15.15±0.03	14.87±0.05	15.72±0.04	15.59±0.04	16.45±0.04	9.80±0.03	17.18±0.04	15.73±0.06	**18.04±0.03**
	IPS	7.00±0.07	7.33±0.04	–	6.00±0.04	7.00±0.07	4.33±0.03	8.00±0.05	6.33±0.04	**8.67±0.04**
SVM	TWiki	65.09±0.05	63.33±0.06	65.25±0.04	19.77±0.03	65.09±0.05	65.09±0.05	65.09±0.05	65.09±0.05	**65.51±0.04**
	TPS	53.67±0.08	47.33±0.09	47.67±0.10	6.67±0.04	53.00±0.07	48.67±0.05	53.00±0.13	53.00±0.09	**54.33±0.1**
	TWiki	23.11±0.04	23.11±0.06	22.69±0.03	22.53±0.04	17.17±0.05	13.41±0.04	23.64±0.04	23.42±0.05	**23.66±0.06**
	IPS	**16.33±0.06**	16.00±0.05	–	0.67±0.01	4.33±0.04	0.33±0.01	0.67±0.02	14.00±0.07	**16.33±0.09**



(a) KNN (b) NB (c) SVM

Fig. 6. The post-hoc Nemenyi test results of compared methods under three classifiers.

1.03×10^{-5}, 8.01×10^{-6}, and 2.00×10^{-8}, which are significantly smaller than 0.1, indicating that there exists a statistical difference among all the methods.

To further determine which specific methods differ significantly from each other, the post-hoc Nemenyi test is further performed. The null hypothesis for the Nemenyi test assumes that the two methods have equal performance. This hypothesis is rejected when the absolute difference in their average ranks exceeds a certain threshold known as the critical difference (CD). The CD is calculated using the following formula:

$$CD = q_\alpha \sqrt{\frac{K(K+1)}{6N}}, \tag{25}$$

where q_α is the critical value based on the Studentized range statistic, $K = 9$ is the number of methods, and $N = 15$ is the number of benchmark datasets.

The results of the Nemenyi test are illustrated in Fig. 6. As shown in this figure, the proposed FGB-E method ranks highest and shows statistically significant improvements over 7, 9, and 4 other methods. These findings confirm that FGB-E not only achieves superior average classification accuracy but also demonstrates statistically significant advantages across multiple datasets and classifiers.

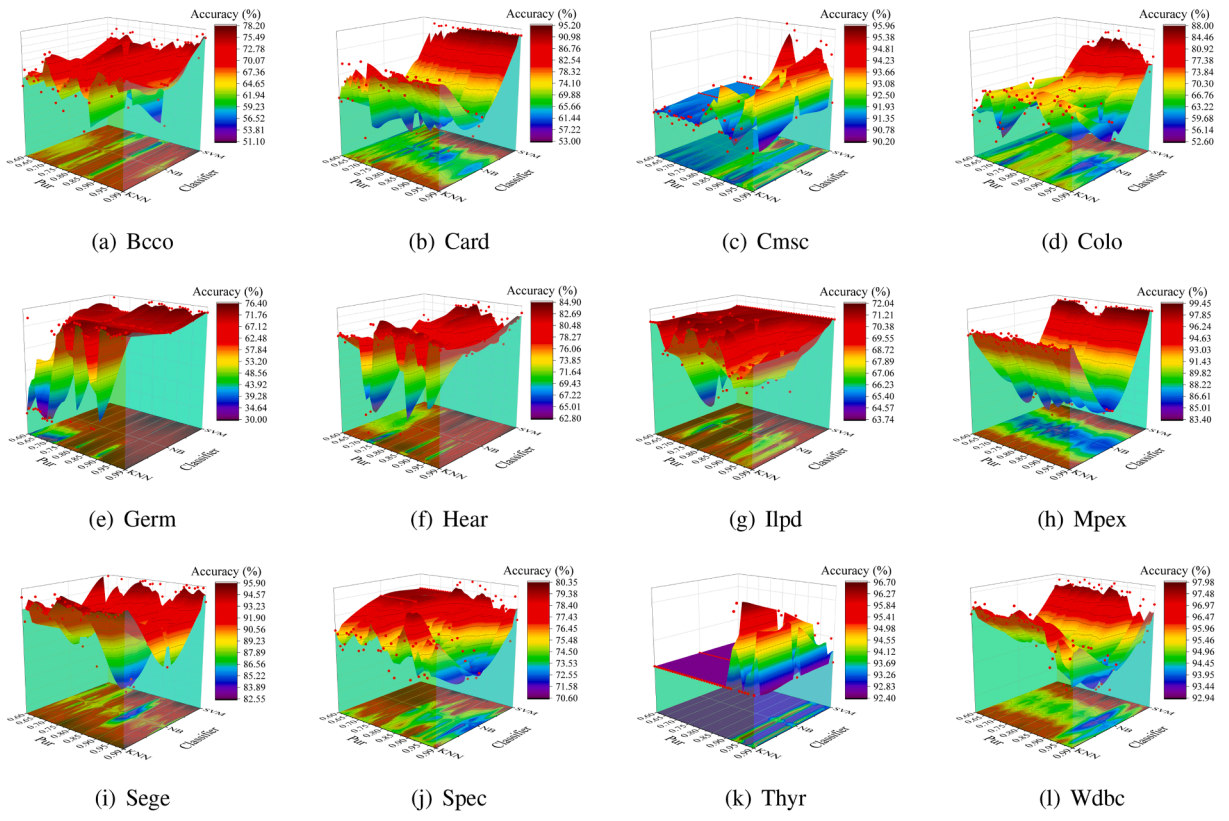


Fig. 7. Clasification performance of the proposed FGB-E method under different purity parameter and classifiers.

5.5. *Parameter sensitivity analysis*

As discussed in Section 3, the purity parameter *Pur* of the enhanced granular ball plays a critical role in the granulation process and the computation of the GB-E uncertainty measure, thereby affecting the overall feature selection procedure. To investigate the sensitivity of FGB-E to the parameter *Pur*, this subsection analyzes the variation in classification performance under different purity settings. In the experiments, *Pur* is varied from 0.60 to 0.99 with a step size of 0.01. The classification accuracy of FGB-E is evaluated under KNN, NB, and SVM classifiers.

As shown in Fig. 7, the classification accuracy of FGB-E exhibits noticeable fluctuations with changes in the purity parameter, indicating a clear sensitivity to its value, especially on datasets Bcco, Colo, Germ, and Thyr. In general, the accuracy increases with rising purity at first, reaches a peak at an intermediate value, and then declines when the parameter becomes either too high or too low. This trend suggests that extreme settings may lead to underfitting or overfitting in the granulation structure. Moreover, the optimal value *Pur* induced by the highest accuracy varies across different datasets and classifiers. This observation further emphasizes the necessity of parameter tuning to adapt the model to specific data characteristics and learning algorithms.

In conclusion, the purity parameter *Pur* significantly affects both the granular ball generation and the downstream feature selection. Therefore, appropriate parameter optimization is essential to ensure the best possible classification performance of the proposed FGB-E method.

6. Conclusion

Multi-granularity computing models, particularly those based on granular ball computing, have been widely applied in feature selection and uncertainty reasoning due to their efficiency and adaptability in complex data granulation. However, most existing granularity-based approaches are limited in addressing heterogeneity and overlapping phenomena among different granules. To tackle these challenges, this paper proposes a novel multi-granularity knowledge fusion framework for feature selection by integrating an enhanced granular ball generation mechanism and GB-E uncertainty measure. The proposed granular ball generation strategy enables the adaptive construction of multi-granularity knowledge spaces by incorporating class distribution information during each splitting process. In addition, the GB-E uncertainty measure jointly considers granule separation and aggregation, offering a more accurate and comprehensive characterization of uncertainty and knowledge representation. Based on these advances, a feature selection method is developed and successfully applied to extract informative features across multiple datasets.

This study enhances the understanding of multi-granularity knowledge fusion and uncertainty-driven feature selection by leveraging granular ball computing. The core idea is to integrate multiple perspectives of information to quantify uncertainty while mitigating granule overlap. Although the proposed enhanced granular ball generation mechanism demonstrates improved efficiency and precision in data granulation, it primarily focuses on numerical data, which limits its applicability to heterogeneous data scenarios. Therefore, developing a more flexible and generalizable granular ball generation mechanism is a valuable direction for future research. In addition, the current framework is designed for static environments, which may restrict its use in dynamic or evolving data contexts. Hence, an incremental granulation mechanism for dynamic feature selection should be further explored. Moreover, extending this framework to support multi-modal data fusion is another promising direction worth exploring.

CRediT authorship contribution statement

Kehua Yuan: Writing – original draft, Methodology, Investigation, Conceptualization; **Yuji Bai:** Methodology, Investigation, Conceptualization; **Duoqian Miao:** Writing – review & editing, Methodology, Investigation; **Weiping Ding:** Writing – review & editing, Methodology; **Yiyu Yao:** Writing – review & editing, Methodology; **Hongyun Zhang:** Writing – review & editing, Methodology; **Witold Pedrycz:** Writing – review & editing, Methodology.

Data availability

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

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.

Acknowledgment

This paper is supported in part by the National Natural Science Foundation of China under Grant 624B2104, Grant 62576251, and Grant 62376198, in part by the National Key Research and Development Program of China “Key Special Project on Cyberspace Security Governance” under Grant 2022YFB3104700.

References

[1] F. Bi, T. He, Y. Ong, et al., Graph linear convolution pooling for learning in incomplete high-dimensional data, IEEE Trans. Knowl. Data Eng. 37 (4) (2025) 1838–1852.
[2] B. Ahadzadeh, M. Abdar, F. Safara, et al., SFE: a simple, fast, and efficient feature selection algorithm for high-dimensional data, IEEE Trans. Evol. Comput. 27 (6) (2023) 1896–1911.
[3] J. Zhang, W. Wu, H.L. Ee, et al., On optimal scale combinations in generalized multi-scale set-valued ordered information systems, Int. J. Approximate Reasoning 183 (2025) 109429.
[4] K. Yuan, D. Miao, W. Pedrycz, et al., Multi-granularity data analysis with zentropy uncertainty measure for efficient and robust feature selection, IEEE Trans. Cybern. 55 (2) (2025) 740–752.
[5] Y. Chen, W. Ding, J. Huang, et al., Multigranularity fuzzy autoencoder for discriminative feature selection in high-dimensional data, 2025. https://doi.org/10.1109/TNNLS.2025.3569893
[6] K. Liu, T. Li, X. Yang, et al., SemiFREE: semi-supervised feature selection with fuzzy relevance and redundancy, IEEE Trans. Fuzzy Syst. 31 (10) (2023) 3384–3396.
[7] T. Yin, H. Chen, J. Wan, et al., Exploiting feature multi-correlations for multilabel feature selection in robust multi-neighborhood fuzzy β covering space, Inf. Fusion 104 (2024) 102150.
[8] W. Li, W. Xu, X. Zhang, et al., Updating approximations with dynamic objects based on local multigranulation rough sets in ordered information systems, Artif. Intell. Rev. 55 (2022) 1821–1855.
[9] Q. Zhang, D. Miao, Q. Zhang, et al., Dynamic frequency selection and spatial interaction fusion for robust person search, Inf. Fusion 124 (2025) 103314.
[10] X. Liang, Y. Qian, Q. Guo, et al., AF: an association-based fusion method for multi-modal classification, IEEE Trans. Pattern Anal. Mach. Intell. 44 (2021) 9236–9254.
[11] M. Cai, D. Qi, C. Huang, et al., Prototype-based fuzzy rough sets for outlier detection, Fuzzy Sets Syst. 517 (2025) 109460.
[12] X. Liang, Y. Qian, Q. Guo, et al., AF: an association-based fusion method for multi-modal classification, IEEE Trans. Pattern Anal. Mach. Intell. 44 (2022) 9236–9254.
[13] R. Li, C. Zhang, D. Li, et al., Improved evidential three-way decisions in incomplete multi-scale information systems, Int. J. Approximate Reasoning 181 (2025) 109417.
[14] X. Liang, Q. Guo, Y. Qian, et al., Evolutionary deep fusion method and its application in chemical structure recognition, IEEE Trans. Evol. Comput. 25 (5) (2021) 883–893.
[15] M. Cai, J. Wang, F. Xu, Anchor graph based connectivity peaks clustering, Knowl. Based Syst. 319 (2025) 113498.
[16] T. Zhan, S. Ma, W. Li, et al., Exponential stability of fractional order switched systems with mode-dependent impulses and its application, IEEE Trans. Cybern. 52 (11) (2022) 11516–11525.
[17] H. Wang, W. Li, T. Zhan, et al., Multi-granulation-based optimal scale selection in multi-scale information systems, Int. J. Mach. Learn. Cybern. 12 (7) (2021) 2135–2150.
[18] T. Yang, Y. Deng, B. Yu, et al., Local feature selection for large-scale data sets with limited labels, IEEE Trans. Knowl. Data Eng. 35 (7) (2023) 7152–7163.
[19] K. Yuan, D. Miao, W. Pedrycz, et al., Ze-HFS: zentropy-based uncertainty measure for heterogeneous feature selection and knowledge discovery, IEEE Trans. Knowl. Data Eng. 36 (11) (2024) 7326–7339.
[20] Y. Huang, T. Deng, G. Yang, et al., Heterogeneous space co-sparse representation: leveraging fuzzy dependency and feature reconstruction for feature selection, Appl. Soft Comput. 175 (2025) 113080.
[21] D. Guo, W. Xu, Y. Qian, et al., M-FCCL: memory-based concept- cognitive learning for dynamic fuzzy data classification and knowledge fusion, Inf. Fusion 100 (2023) 101962.

[22] W. Xu, D. Guo, J. Mi, et al., Two-way concept-cognitive learning via concept movement viewpoint, IEEE Trans. Neural Netw. Learn. Syst. 34 (10) (2023) 6798–6812.
[23] Y. Yao, Three-way granular computing, rough sets, and formal concept analysis, Int. J. Approximate Reasoning 116 (2020) 106–125.
[24] X. Wei, B. Pang, J. Mi, Axiomatic characterizations of l-valued rough sets using a single axiom, Inf. Sci. 283 (310) (2021) 283–310.
[25] Y. Li, Y. Liu, H. Zhang, et al., Occlusion-aware transformer with second-order attention for person re-identification, IEEE Trans. Image Process. 33 (2024) 3200–3211.
[26] S. Xia, S. Zheng, G. Wang, et al., Granular ball sampling for noisy label classification or imbalanced classification, IEEE Trans. Neural Netw. Learn. Syst. 34 (4) (2023) 2144–2155.
[27] D. Guo, W. Xu, W. Ding, et al., Concept-cognitive learning survey: mining and fusing knowledge from data, Inf. Fusion 109 (2024) 102426.
[28] D. Guo, W. Xu, Y. Qian, et al., Fuzzy-granular concept-cognitive learning via three-way decision: performance evaluation on dynamic knowledge discovery, IEEE Trans. Fuzzy Syst. 32 (2023) 1409–1423.
[29] X. Cao, X. Yang, S. Xia, et al., Open continual feature selection via granular-ball knowledge transfer, IEEE Trans. Knowl. Data Eng. 36 (12) (2024) 8967–8980.
[30] L. Sun, Q. Zhang, W. Ding, et al., FCPFS: fuzzy granular ball clustering-based partial multilabel feature selection with fuzzy mutual information, IEEE Trans. Emerging Top. Comput. Intell. 9 (1) (2025) 590–606.
[31] S. Xia, Y. Liu, X. Ding, et al., Granular ball computing classifiers for efficient, scalable and robust learning, Inf. Sci. 483 (2019) 136–152.
[32] S. Xia, X. Dai, G. Wang, et al., An efficient and adaptive granular-ball generation method in classification problem, IEEE Trans. Neural Netw. Learn. Syst. 35 (2024) 5319–5331.
[33] W. Qian, F. Xu, J. Qian, et al., Multi-label feature selection based on rough granular-ball and label distribution, Inf. Sci. 650 (2023) 119698.
[34] M. Sajid, A. Quadir, M. Tanveer, et al., GB-RVFL: fusion of randomized neural network and granular ball computing, Pattern Recognit. 159 (2025) 111142.
[35] F. Dunkin, X. Li, C. Hu, et al., Like draws to like: a multi-granularity ball-intra fusion approach for fault diagnosis models to resists misleading by noisy labels, Adv. Eng. Inf. 60 (2024) 102425.
[36] Q. Hu, D. Yu, Z. Me, Neighborhood classifiers, Expert Syst. Appl. 34 (2008) 866–876.
[37] D. Yu, Q. Hu, P. Zhu, Rule learning for classification based on neighborhood covering reduction, Inf. Sci. 181 (24) (2011) 5457–5467.
[38] H. Zhang, Q. Sun, K. Dong, Information-theoretic partially labeled heterogeneous feature selection based on neighborhood rough sets, Int. J. Approximate Reasoning 154 (2023) 200–217.
[39] Z. Yuan, H. Chen, T. Li, Exploring interactive attribute reduction via fuzzy complementary entropy for unlabeled mixed data, Pattern Recognit. 127 (2022) 108651.
[40] 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.
[41] C. Wang, Y. Qi, M. Shao, et al., A fitting model for feature selection with fuzzy rough sets, IEEE Trans. Fuzzy Syst. 25 (4) (2017) 741–753.
[42] A. Kumar, P.S. V. S.S. Prasad, Scalable feature subset selection with fuzzy rough sets and fuzzy min-max neural network in hybrid decision system, IEEE Trans. Fuzzy Syst. 33 (2) (2025) 669–679.
[43] W. Li, C. Zhai, W. Xu, Fuzzy measures and choquet integrals based on fuzzy covering rough sets, IEEE Trans. Fuzzy Syst. 30 (2022) 2360–2374.
[44] D. Guo, W. Xu, Fuzzy-based concept-cognitive learning: an investigation of novel approach to tumor diagnosis analysis, Inf. Sci. 639 (2023) 118998.
[45] L. Sun, L. Wang, W. Ding, et al., Feature selection using fuzzy neighborhood entropy-based uncertainty measures for fuzzy neighborhood multigranulation rough sets, IEEE Trans. Fuzzy Syst. 29 (1) (2021) 19–33.
[46] J. Wan, H. Chen, T. Li, et al., Interactive and complementary feature selection via fuzzy multigranularity uncertainty measures, IEEE Trans. Cybern. 53 (2) (2023) 1208–1221.
[47] K. Yuan, D. Miao, H. Zhang, et al., An efficient and robust feature selection approach based on zentropy measure and neighborhood-aware model, 2025. https://doi.org/10.1109/TNNLS.2025.3565320
[48] K. Yuan, D. Miao, Y. Yi, et al., Feature selection using zentropy-based uncertainty measure, IEEE Trans. Fuzzy Syst. 32 (2024) 2246–2260.
[49] S. Xia, C. Wang, G. Wang, et al., GBRS: a unified granular-ball learning model of pawlak rough set and neighborhood rough set, IEEE Trans. Neural Netw. Learn. Syst. 36 (1) (2025) 1719–1733.
[50] Y. Chen, Z. Huang, J. Li, Fuzzy neighborhood based variable-precision granular-ball rough sets with applications to feature selection, Fuzzy Sets Syst. 512 (2025) 109382.
[51] L. Sun, H. Liang, W. Ding, et al., Granular ball fuzzy neighborhood rough sets-based feature selection via multiobjective mayfly optimization, IEEE Trans. Fuzzy Syst. 32 (2024) 6112–6124.
[52] W. Qian, J. Li, X. Cai, et al., Granular ball-based partial label feature selection via fuzzy correlation and redundancy, Inf. Sci. 709 (2025) 122047.
[53] B. Sang, H. Chen, L. Yang, et al., Incremental feature selection using a conditional entropy based on fuzzy dominance neighborhood rough sets, IEEE Trans. Fuzzy Syst. 30 (6) (2022) 1683–1697.
[54] B. Sang, L. Yang, H. Chen, et al., Robust attribute reduction exploring class-separability and attribute-correlation for ordered decision systems, IEEE Trans. Syst. Man Cybern. Syst. 55 (6) (2025) 3941–3953.
[55] C. Wang, Y. Qian, W. Ding, et al., Feature selection with fuzzy-rough minimum classification error criterion, IEEE Trans. Fuzzy Syst. 30 (2022) 2930–2942.
[56] P. Zhang, T. Li, Z. Yuan, et al., Heterogeneous feature selection based on neighborhood combination entropy, IEEE Trans. Neural Netw. Learn. Syst. 35 (2024) 3514–3527.
[57] Z. Wang, H. Chen, Z. Yuan, et al., Multiscale fuzzy entropy-based feature selection, IEEE Trans. Fuzzy Syst. 31 (2023) 3248–3262.
[58] C. Rashtchian, P. Young, M. Hodosh, et al., Collecting image annotations using Amazon's mechanical Turk, in: Proc. Workshop Creating Speech Lang., Data with Amazon's Mechanical Turk, Workshop Creating Speech Lang., Data With Amazon's Mechanical Turk, 2010, pp. 139–147.
[59] N. Rasiwasia, et al., A new approach to cross-modal multimedia retrieval, in: Proc. 18th ACM Int. Conf. Multimedia, 18th ACM Int. Conf. Multimedia, 2010, pp. 251–260.
[60] J. Demšar, Statistical comparisons of classifiers over multiple data sets, J. Mach. Learn. Res. 7 (2006) 1–30.