Zentropy-Enhanced Multigranularity Knowledge Modeling for Robust Feature Selection

Kehua Yuan, Duoqian Miao, Witold Pedrycz, *Life Fellow, IEEE*, and Yiyu Yao

***Abstract*—Multigranularity knowledge modeling is an influential study for information processing and knowledge discovery in artificial intelligence (AI). A central research focus is the multigranularity representation and learning of knowledge structures. Among them, fuzzy rough sets (FRSs) have emerged as a representative method for characterizing uncertain knowledge. However, the existing FRS studies still exhibit two limitations: low robustness in knowledge acquisition and incomplete characterization of uncertainty. Hence, this article proposes a zentropy-enhanced multigranularity knowledge modeling framework for robust feature selection (ZeMG-FS). Specifically, we design a fast and adaptive multigranularity information granulation mechanism based on generalized granular-ball generation to effectively capture data distributions embedded in complex data. Then, the fuzzy rough approximation method is incorporated into the representation of multigranularity knowledge. Furthermore, we analyze the fundamental relationships and structures of the multigranularity knowledge model to introduce a novel multilevel zentropy. Unlike existing entropy measures, the primary consideration of the proposed zentropy is to match and enhance the performance of the proposed model. Finally, we design two feature evaluation criteria grounded in the model and apply them to feature selection. Extensive experiments demonstrate that our proposed methods achieve superior robustness and effectiveness compared with state-of-the-art approaches.**

***Index Terms*—Feature selection, fuzzy rough sets (FRSs), granular computing (GrC), granular-ball computing, uncertainty measure.**

I. Introduction

RECENTLY, the emergence of new technologies such as Sora and ChatGPT has considerably accelerated the development of artificial intelligence (AI). Simultaneously, developing a credible, reliable, and interpretable AI model has attracted renewed attention [1], [2], [3]. Nevertheless, in many real scenarios, uncertainties arising from data collection and information processing severely affect the credibility and reliability of AI models. Thus, uncertainty has become an unavoidable topic, particularly in complex data modeling with high-dimensional features [4], [5], [6]. Consequently, feature selection incorporating uncertainty analysis has emerged as a crucial research focus in data mining and knowledge discovery [7].

Feature selection using granular computing (GrC) theory is a common and effective approach for enhancing data quality and conducting uncertainty analysis in high-dimensional complex data [8]. The core idea of GrC is to represent and process data through information granules with different granularities by simulating the human cognitive process [9], [10], [11]. Among various GrC models, fuzzy rough sets (FRSs) [12], [13], [14] have been widely adopted for feature evaluation by combining qualitative and quantitative information. Nevertheless, the effectiveness of FRS-based methods for feature selection is still subject to two notable limitations: 1) *low robustness in knowledge acquisition* since the fuzzy rough approximation process is highly sensitive to boundary samples and noise and 2) *incomplete characterization of uncertainty* due to the intrinsic multigranularity nature of uncertainty in both data and modeling. To mitigate the first limitation, many FRS models are extended to improve the model's antinoise ability [15], [16], [17]. Particularly, the fuzzy neighborhood rough set (FNRS) is proposed to improve model robustness by introducing a neighborhood radius [18]. Similarly, Yuan et al. constructed a noise-aware neighborhood rough approximation model for uncertainty measure [19]. Ding et al. designed a fuzzy neighborhood rule for classification problems [20]. Moreover, Kumar and Prasad [21] integrated FRS with fuzzy neural networks to design a hybrid feature selection approach. Nonetheless, these approaches typically adopt a unitary and fixed granulation strategy that overlooks intrinsic data distributions, limiting their adaptability and granularity precision. The concept of granular all proposed by Xia et al. [22] offers a fast and adaptive mechanism for knowledge representation. However, its binary splitting strategy often fails to account for the actual class distribution and further influences suboptimal granulation. Therefore, enhancing the adaptability and robustness of FRSs under noisy conditions remains necessary.

Furthermore, as the second limitation, various FRS-based uncertainty measure methods have been investigated and successfully applied to feature selection [23], [24], [25]. For instance, Zhang et al. [23] introduced an instance selection

Received 16 December 2025; revised 28 January 2026; accepted 12 February 2026. This work was supported in part by the National Natural Science Foundation of China under Grant 624B2104, Grant 62576251, and Grant 62376198. This article was recommended by Associate Editor S. Tong. *(Corresponding author: Duoqian Miao.)*

Kehua Yuan and Duoqian Miao are with the School of Computer Science and Technology, Tongji University, Shanghai 201804, China (e-mail: yuan945527632@163.com; dqmiao@tongji.edu.cn).

Witold Pedrycz is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6R 2V4, Canada, also with the Systems Research Institute, Polish Academy of Sciences, 00-901 Warsaw, Poland, and also with the Research Center of Performance and Productivity Analysis, Istinye University, 34010 Istanbul, Türkiye (e-mail: wpedrycz@ualberta.ca).

Yiyu Yao is with the Department of Computer Science, University of Regina, Regina, SK S4S 0A2, Canada (e-mail: Yiyu.Yao@uregina.ca).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCYB.2026.3665802.

Digital Object Identifier 10.1109/TCYB.2026.3665802

method and proposed an importance degree measure considering fuzzy object granules. To handle multilabel tasks, Yin et al. [24] developed a graph-based fuzzy dependency degree to improve feature evaluation. In addition to algebraic approaches, several entropy-based measures have also been introduced based on information theory. For example, Zou and Dai [25] designed monotonic information measures for fuzzy−*beta* covering reduction. Addressing heterogeneous granulation, Yuan et al. [26] proposed a heterogeneous granulation method for multigranularity uncertainty measures. In multiscale contexts, Sang et al. [27] developed an optimal scale selection and feature evaluation based on fuzzy information. Despite these advancements, most existing approaches focus primarily on a single level of fuzzy object granules or approximations while neglecting the multigranularity structure and interactions among decision targets, fuzzy approximations, and fuzzy object granules. As a result, uncertain knowledge is described in an imprecise and limited manner, leading to suboptimal feature selection performance.

Zentropy-based uncertainty measure is an effective approach for characterizing uncertainty across multiple scales, where z is derived from the German term "zusstandssumme," which represents the sum of different scales. It provides a systematic perspective in which total system entropy arises from contributions at different levels, where entropy at a coarse scale can be further refined into entropies at finer scales. This hierarchical and decomposable nature of zentropy aligns well with the granularity structure and interdependencies inherent in multigranularity knowledge modeling, which has been fully demonstrated in uncertainty quantification [28], feature extraction [11], and material property prediction [29]. Nevertheless, existing zentropy methods are still heavily constrained to neighborhood-based approximation, overlooking the adequate characterization of fuzziness and exhibiting limited adaptability in granularity modeling. Inspired by the above insights, this study investigates a novel zentropy-enhanced multigranularity knowledge modeling approach for robust feature selection. The main contributions of this article are summarized as follows.

1) It designs an adaptive multigranularity information granulation mechanism based on generalized granular-ball generation, which effectively captures data distributions embedded in complex datasets.
2) It proposes a robust multigranularity knowledge modeling method via fuzzy rough approximation, which comprehensively depicts uncertain knowledge using adaptive boundary samples to enhance noise resistance.
3) It defines a new multilevel zentropy structure based on the proposed multigranularity knowledge model, which is better aligned with the granularity architecture than existing entropy measures.
4) It leverages zentropy to enhance the performance of multigranularity knowledge modeling in feature evaluation and classification tasks. Extensive experiments demonstrate the robustness and effectiveness of our proposed methods.

The remainder of this article is organized as follows. Section II reviews the basic concepts. Section III presents the adaptive multigranularity knowledge modeling. Section IV develops a robust feature selection approach. Section V demonstrates the effectiveness of the proposed methods. Section VI concludes this article and discusses future work.

II. Preliminaries

This section briefly overviews multigranularity computing, fuzzy rough approximation, and FRSs' sensitivity analysis.

A. Multigranularity Computing via Granular Ball

In the context of data analysis, a dataset consisting of an object set, a conditional feature set, and a decision attribute is typically referred to as a decision information system. It can be formally denoted as $\mathbb{DIS} = (O, C, D)$, where $O = \{x_1, x_2, \ldots, x_n\}$ is a finite universe of n objects, $C = \{c_1, c_2, \ldots, c_m\}$ is the set of m conditional features, and D is the decision attribute that partitions O into decision classes, i.e., $O/D = \{D_1, D_2, \ldots, D_w\}$.

Multigranularity computing is an effective paradigm for uncertainty modeling and knowledge acquisition. In particular, granular-ball computing provides an adaptive data granulation mechanism based on binary splitting. Given a decision system, the granular balls are constructed as follows.

Definition 1 [22]: Given a $\mathbb{DIS} = (O, C, D)$, where $O/D = \{D_1, D_2, \ldots, D_w\}$, let $B \subseteq C$ denote the resulting set of granular balls as $G_B = \{g_1, g_2, \ldots, g_s\}$. For each granular ball $g_i \in G_B$, its center and radius are defined as

$$\begin{aligned} \mathbf{Ce}(g_i) &= \frac{1}{|g_i|}\sum_{o_j \in g_i} o_j \\ \mathbf{Ra}(g_i) &= \frac{1}{|g_i|}\sum_{o_j \in g_i} \|o_j - \mathbf{Ce}(g_i)\|_B \end{aligned} \tag{1}$$

where $|\cdot|$ denotes the cardinality of a set and $\|\cdot\|_B$ denotes the Euclidean distance induced by B.

During the construction process, a granular ball continues to split until it becomes pure, i.e., the majority of its objects belong to the same decision class. The label and purity of a granular ball g_i are computed as

$$\begin{aligned} \mathbf{Lab}(g_i) &= \arg\max_{j \in \{1,2,\ldots,w\}} \left| g_i \cap D_j \right| \\ \mathbf{P}(g_i) &= \frac{|g_i \cap D(g_i)|}{|g_i|} \end{aligned} \tag{2}$$

where $0 \le \mathbf{P}(g_i) \le 1$. A higher value of $\mathbf{P}(g_i)$ indicates a higher consistency of decision labels within g_i.

Note that the granulation process often disregards the coexistence of multiple decision classes within a single granular ball, which may compromise the accuracy and efficiency of representation, particularly in multiclass scenarios. Hence, more precise generation mechanisms are crucial for knowledge representation in decision information systems.

B. Fuzzy Rough Approximation

In decision information systems, fuzzy rough approximation provides an effective framework for handling uncertainty by leveraging fuzzy logical operators. These operators quantify the certainty and possibility that an object belongs to a decision class by comparing the fuzzy similarity among objects and their membership degrees to decision classes.

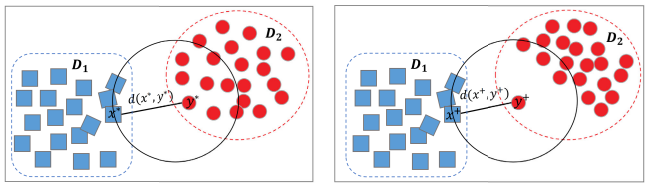

Fig. 1. Visualization of equal fuzzy lower approximation of x^* and x^+ with different distributions because of the noise distribution of boundary samples. Here, (a) has a uniform distribution and (b) is with a discordant object y^+.

Definition 2 [30]: Given a $\mathbb{DIS} = (O, C, D)$, where $O/D = \{D_1, D_2, \ldots, D_w\}$, for $D_p \in O/D$ and $B \subseteq C$, the fuzzy lower and upper approximations of D_p on B are defined as

$$\begin{aligned} \underline{R_B}\left(D_p\right)(o_i) &= \inf_{o_j \in O} \max\left\{1 - R_B\left(o_i, o_j\right), D_p\left(o_j\right)\right\} \\ \overline{R_B}\left(D_p\right)(o_i) &= \sup_{o_j \in O} \min\left\{R_B\left(o_i, o_j\right), D_p\left(o_j\right)\right\} \end{aligned} \tag{3}$$

where $R_B(o_i, o_j) = min_{b \in B}\{1 - abs(b(o_i) - b(o_j))\}$ is the fuzzy similarity relation induced by B and $D_p(o_j)$ denotes the membership degree of object o_j to D_p.

To evaluate the quality of approximation in an FRS model, the fuzzy dependency degree is employed. It measures the proportion of objects that can be accurately classified based on the conditional attributes B and is computed as follows:

$$\gamma_B = \frac{\left|\cup_{D_p \in O/D} \underline{R_B}\left(D_p\right)\right|}{|O|}. \tag{4}$$

A larger value of γ_B indicates a higher level of certainty in the classification, thereby reflecting stronger dependency between the conditional attributes and the decision attribute.

C. Sensitivity and Structural Analysis of FRSs

In a decision information system, the fuzzy approximation process in FRSs is inherently sensitive to boundary objects between decision classes. As derived from Definition 2, the lower and upper approximations of D_p can be simplified as

$$\begin{aligned} \underline{R_B}\left(D_p\right)(o_i) &= \inf_{o_j \notin D_p} \left\{1 - R_B\left(o_i, o_j\right)\right\} \\ \overline{R_B}\left(D_p\right)(o_i) &= \sup_{o_j \in D_p} R_B\left(o_i, o_j\right) \end{aligned} \tag{5}$$

where the certainty of o_i belonging to D_p is determined by its nearest objects outside D_p and the possibility is characterized by its maximum similarity to objects within D_p.

Fig. 1 illustrates a boundary-based approximation mechanism, where different shapes denote various decision classes. In fact, noisy or misclassified samples can severely disturb this mechanism, especially under imbalanced conditions. To address this issue, several FRS-based methods introduce neighborhoods to control object granularity for improving model robustness. Despite their effectiveness, they mostly adopt a uniform neighborhood radius across all objects and classes, neglecting variations in local data distributions. This limitation underscores the need for an adaptive and noise-resilient FRS model that dynamically adjusts to the data structure.

In addition to the approximation mechanism, various FRS-based uncertainty measures have been proposed, such as fuzzy dependency degree and fuzzy conditional entropy. Specifically, the fuzzy dependency degree is an algebraic measure that evaluates classification certainty based on fuzzy lower approximations. In contrast, fuzzy conditional entropy is an information-theoretic method that quantifies the inconsistency between fuzzy object granules and decision labels. These are single-granularity methods based on approximations or object granules. Notably, the whole approximation process is multigranular from target decisions to fuzzy approximation and object granules. As such, the single-granularity measures fail to capture the hierarchical and interdependent nature of uncertainty propagation across different levels. Therefore, it is imperative to design more accurate and systematic uncertainty measures for enhancing the expressiveness and reliability of uncertainty in fuzzy approximation.

III. Adaptive Multigranularity Knowledge Modeling

To address the sensitivity of FRSs to boundary samples discussed in Section II-C, this section proposes an adaptive multigranularity knowledge modeling method based on generalized granular-ball representation. By explicitly incorporating the class distribution into the granulation process, the proposed method enables adaptive object granulation and significantly improves model robustness in complex decision environments.

A. Generalized Granular-Ball Generation

Regarding the multiclass splitting limitation of conventional binary splitting in multigranularity knowledge modeling, this section introduces a generalized granular-ball representation that incorporates class-aware splitting into the granulation process, thereby achieving more general and precise data granulation.

Definition 3: Given a $\mathbb{DIS} = (O, C, D)$, where $O/D = \{D_1, D_2, \ldots, D_w\}$, for $B \subseteq C$ and purity $\mathbf{P} \in (0, 1]$, the enhanced granular-ball representation $\mathbf{EG}_B^P = \{g_1, g_2, \ldots, g_r\}$ is constructed through the following iterative procedure.

1) *Initialization:* Begin with an initial granular ball $\mathbf{EG}_B^0 = \{g_1^0\}$, computed as follows:

$$g_1^0 = \left\{o_j \in O, \left\|o_j - \mathbf{Ce}\left(g_1^0\right)\right\|_B \leq \mathbf{Ra}\left(g_1^0\right)\right\} \tag{6}$$

where $\|\cdot\|_B$ denotes the Euclidean distance induced by B, and the center $\mathbf{Ce}(g_1^0)$ and radius $\mathbf{Ra}(g_1^0)$ of the initial granular ball are given by

$$\begin{aligned} \mathbf{Ce}\left(g_1^0\right) &= \frac{1}{|O|}\sum_{o_j \in O} o_j \\ \mathbf{Ra}\left(g_1^0\right) &= \frac{1}{|O|}\sum_{o_j \in O}\left\|o_j - \mathbf{Ce}\left(g_1^0\right)\right\|_B. \end{aligned} \tag{7}$$

This initialization generates a single granular ball based on the data center and its average distance from the data center, serving as the foundation for subsequent adaptive splitting.

2) *Class-Aware Splitting:* This iterative splitting process continues until each granular ball meets a predefined purity threshold. Assume that the current granular ball set at iteration $t - 1$ is given as follows:

$$\mathbf{EG}_B^{t-1} = \left\{g_1^{t-1}, g_2^{t-1}, \ldots, g_l^{t-1}\right\}.$$

For each $g_i^{t-1} \in \mathbf{EG}_B^{t-1}$, if the ball satisfies the purity requirement, it is retained. Otherwise, it is split into subballs based on its internal class distribution

$$g_i^t = \left\{g_{i1}^t, g_{i2}^t, \ldots, g_{ik_i}^t\right\} \tag{8}$$

where k_i is the number of decision classes present in g_i^{t-1}. Each subball is constructed as follows:

$$g_{ij}^t = \left\{o_j \in g_i^{t-1} \;\middle|\; \left\|o_j - \mathbf{Ce}\left(g_{ij}^t\right)\right\|_B \le \mathbf{Ra}\left(g_{ij}^t\right)\right\} \tag{9}$$

where the center and radius are given by

$$\begin{aligned}\mathbf{Ce}\left(g_{ij}^t\right) &= \frac{1}{|g_{ij}^t|}\sum_{o_j \in g_{ij}^t} o_j \\ \mathbf{Ra}\left(g_{ij}^t\right) &= \frac{1}{|g_{ij}^t|}\sum_{o_j \in g_{ij}^t} \left\|o_j - \mathbf{Ce}\left(g_{ij}^t\right)\right\|.\end{aligned} \tag{10}$$

The updated granular ball set at tth iteration is then

$$\mathbf{EG}_B^t = \bigcup_{i=1}^{l}\left\{g_{i1}^t, g_{i2}^t, \ldots, g_{ik_i}^t\right\}. \tag{11}$$

3) *Termination and Final Representation:* The iterative process terminates when all granular balls satisfy the desired purity condition $\mathbf{P}$. The generated enhanced granular-ball representation is presented as follows:

$$\mathbf{EG}_B^P = \{g_1, g_2, \ldots, g_r\} \tag{12}$$

where r is the number of final granular balls, and the corresponding centers and radii are

$$\begin{aligned}\mathbf{Ce} &= \{\mathbf{Ce}(g_1), \ldots, \mathbf{Ce}(g_r)\} \\ \mathbf{Ra} &= \{\mathbf{Ra}(g_1), \ldots, \mathbf{Ra}(g_r)\}.\end{aligned} \tag{13}$$

The label set of each ball is defined as

$$\mathbf{Lab}\left(\mathbf{EG}_B^P\right) = \{\mathbf{Lab}(g_1), \mathbf{Lab}(g_2), \ldots, \mathbf{Lab}(g_r)\} \tag{14}$$

where $\mathbf{Lab}(g_i)$ is determined according to (2).

Compared with existing granular-ball computing methods, the above splitting procedure is fast and adaptive, generating a set of granular balls that reflect class consistency. A more detailed visualization of the generation process is provided in Fig. 2. In practical modeling, each granular ball can be viewed as a representative region that captures local data characteristics, and the collective set of balls approximates the overall distribution in decision information systems.

B. Fuzzy Rough Knowledge Acquisition

Building on the generalized granular-ball representation introduced in Section III-A, this part presents a robust fuzzy rough knowledge acquisition model that leverages the class-aware granulation to improve the robustness of fuzzy rough approximation.

Definition 4: Given a $\mathbb{DIS} = (O, C, D)$, for $B \subseteq C$ and $P, \delta \in (0, 1]$, $\mathbf{EG}_B^P = \{g_1, g_2, \ldots, g_r\}$ is the generalized granular balls and $\mathbf{EG}_B^P/\mathbf{Lab}(\mathbf{EG}_B^P) = \{\mathbf{DL}_1, \mathbf{DL}_2, \ldots, \mathbf{DL}_w\}$. For any $\mathbf{DL}_i \in \mathbf{EG}_B^P/\mathbf{Lab}(\mathbf{EG}_B^P)$, the fuzzy lower and upper approximations are defined as

$$\begin{aligned}\underline{R_B^{P,\delta}}(\mathbf{DL}_i)(g_i) &= \inf_{g_j \in \mathbf{EG}_B^P} \max\left\{1 - R_B^\delta\left(g_i, g_j\right), \mathbf{DL}_i\left(g_j\right)\right\} \\ \overline{R_B^{P,\delta}}(\mathbf{DL}_i)(g_i) &= \sup_{g_j \in \mathbf{EG}_B^P} \min\left\{R_B^\delta\left(g_i, g_j\right), \mathbf{DL}_i\left(g_j\right)\right\}\end{aligned} \tag{15}$$

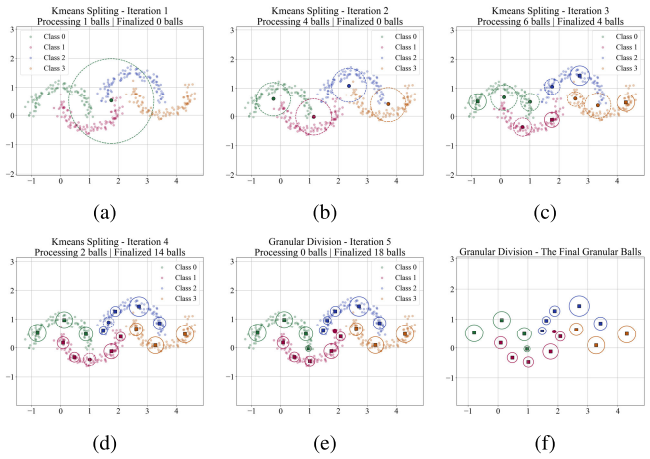

Fig. 2. Visualization example of generalized granular-ball generation, where (a)–(f) depicts the detailed split process. Data are first initialized into a single granular ball based on the initial center and radius in (a) afterward, an adaptive splitting process is performed by considering the class distribution within each subgranular ball. This iterative refinement continues until all granular balls meet the predefined purity. The final results are presented in (f).

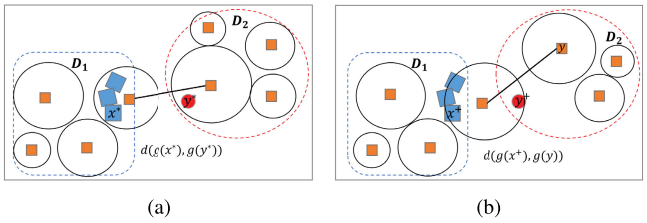

Fig. 3. Visualization of fuzzy rough approximation for $g(x^*)$ and $g(x^+)$ to D_1. Compared with their difference in terms of equality in Fig. 1, $g(x^+)$ to D_1 in (b) reasonably differs from $g(x^*)$ to D_1 under uniform distribution (a). This is because the influence of the discordant object y^+ is removed.

where $R_B^\delta(g_i, g_j)$ denotes the fuzzy relation between g_i and g_j on B and $[g_i]_{R_B^\delta} = (R_B^\delta(g_i, g_1))/g_1 + (R_B^\delta(g_i, g_2))/g_2 + \cdots + (R_B^\delta(g_i, g_r))/g_r$ is referred to the fuzzy object granule of g_i under B.

In this study, the fuzzy relation is defined as follows to evaluate the degree of similarity among different objects:

$$R_B^\delta\left(g_i, g_j\right) = \begin{cases} e^{-\frac{\|o_i - o_j\|_B}{\lambda^2}}, & \text{otherwise} \\ 0, & \text{if } e^{-\frac{\|o_i - o_j\|_B}{\lambda^2}} < \delta \end{cases} \tag{16}$$

where λ^2 is a tuning parameter set to the 20th percentile of distances between g_i and all others, inspired by [31].

In the above definition, $\underline{R_B^{P,\delta}}(\mathbf{DL}_i)(g_i)$ and $\overline{R_B^{P,\delta}}(\mathbf{DL}_i)(g_i)$ represent the certainty and possibility, respectively, of a granular ball g_i belonging to $\mathbf{DL}_i$. Compared with traditional fuzzy rough models, the proposed method exhibits enhanced robustness to noise by utilizing the enhanced granular-ball representation. A comparative visualization in Fig. 3 illustrates how this approach improves noise tolerance near decision boundaries, especially compared to the existing boundary object approximations in Fig. 1.

To further analyze the behavior of the above fuzzy rough approximation process under different conditions, several important properties are stated as follows.

Property 1: Given a $\mathbb{DIS} = (O, C, D)$. Given $P, \delta \in (0, 1]$, $\mathbf{EG}_B^P = \{g_1, g_2, \ldots, g_r\}$ is the generalized granular balls. Then,

for $B \subseteq C$ and $\mathbf{DL}_i \in \mathbf{EG}_B^P/\mathbf{Lab}(\mathbf{EG}_B^P)$, the following properties hold.

1) For $B_1 \subseteq B_2$, $\underline{R_{B_1}^{P,\delta}}(\mathbf{DL}_i)$ and $\underline{R_{B_2}^{P,\delta}}(\mathbf{DL}_i)$ (and similarly the upper approximations) are incomparable.
2) For $0 \le P_1 \le P_2 \le 1$, $\underline{R_B^{P_1,\delta}}(\mathbf{DL}_i)$ and $\underline{R_B^{P_2,\delta}}(\mathbf{DL}_i)$ (and similarly upper approximations) are incomparable.
3) For $0 \le \delta_1 \le \delta_2 \le 1$, $\underline{R_B^{P,\delta_1}}(\mathbf{DL}_i) \subseteq \underline{R_B^{P,\delta_2}}(\mathbf{DL}_i)$ and $\overline{R_B^{P,\delta_1}}(\mathbf{DL}_i) \supseteq \overline{R_B^{P,\delta_2}}(\mathbf{DL}_i)$.
4) For each $g_i \in \mathbf{EG}_B^P$, if $[g_i]_{R_1} \subseteq [g_i]_{R_2}$, then $\underline{R_1}(\mathbf{DL}_i) \supseteq \underline{R_2}(\mathbf{DL}_i)$ and $\overline{R_1}(\mathbf{DL}_i) \subseteq \overline{R_2}(\mathbf{DL}_i)$.

Proof: These properties directly follow the definitions of the generalized granular-ball structure and fuzzy approximation model, along with the monotonicity of fuzzy relations.

Moreover, to evaluate the approximation quality of the proposed model, the fuzzy dependency degree is also defined.

Definition 5: Given a $\mathbb{DIS} = (O, C, D)$ and $P, \delta \in (0, 1]$, for $B \subseteq C$, $\mathbf{EG}_B^P = \{g_1, g_2, \ldots, g_r\}$ is the generalized granular balls, and $\mathbf{EG}_B^P/\mathbf{Lab}(\mathbf{EG}_B^P) = \{\mathbf{DL}_1, \mathbf{DL}_2, \ldots, \mathbf{DL}_w\}$. The fuzzy dependency degree of $\mathbf{Lab}(\mathbf{EG}_B^P)$ on B is defined as

$$\gamma_B^{P,\delta}\left(\mathbf{Lab}\left(\mathbf{EG}_B^P\right)\right) = \frac{\left|\bigcup_{\mathbf{DL}_i \in \mathbf{EG}_B^P/\mathbf{Lab}\left(\mathbf{EG}_B^P\right)} \underline{R_B^{P,\delta}}(\mathbf{DL}_i)\right|}{\left|\mathbf{EG}_B^P\right|}. \quad (17)$$

This measure is based on the fuzzy rough lower approximation, which evaluates model quality through approximate accuracy and maintains a positive trend in its values.

IV. Robust Feature Selection Method

Considering the incomplete representation of uncertainty inherent in traditional fuzzy rough approximations, this section proposes a new multilevel zentropy uncertainty analysis method that could capture hierarchical and interactional structures embedded within the multigranularity structure.

A. Multilevel Zentropy Uncertainty Analysis

The fuzzy rough approximation is inherently progressed from coarser target decisions to finer fuzzy approximations and finally to object granules. Each level reflects its scale uncertainty, where coarse representations summarize finer structures, and finer ones provide detailed insight. This hierarchical and nested nature aligns with the conceptual structure of zentropy [28], [29]. Inspired by this, a novel multilevel zentropy measure is proposed to advance uncertainty analysis.

Definition 6: Given a $\mathbb{DIS} = (O, C, D)$, for $B \subseteq C$ and $P, \delta \in (0, 1]$, $\mathbf{EG}_B^P = \{g_1, g_2, \ldots, g_r\}$ is the generalized granular balls and $\mathbf{EG}_B^P/\mathbf{Lab}(\mathbf{EG}_B^P) = \{\mathbf{DL}_1, \mathbf{DL}_2, \ldots, \mathbf{DL}_w\}$. The multilevel zentropy uncertainty measure is defined as

$$\mathbb{Z}_B^{P,\delta}(D) = -\sum_{i=1}^{w} P_i \log P_i + \sum_{i=1}^{w} P_i \mathbb{Z}_i \quad (18)$$

where $P_i = |\mathbf{DL}_i|/|\mathbf{EG}_B^P|$ denotes the proportion of granular balls labeled as class i and $\mathbb{Z}_i$ represents the uncertainty at the finer levels, decomposing with the same (18).

$\mathbb{Z}_i$ reflecting uncertainty within each decision group $\mathbf{DL}_i$ is considered from certain and uncertain granular regions

$$\mathbb{Z}_i = -\sum_{j=1}^{2} P_{ij} \log P_{ij} + \sum_{j=1}^{2} P_{ij} \mathbb{Z}_{ij} \quad (19)$$

where $P_{ij} = |G_{ij}|/|\mathbf{DL}_i|$ and the objects of certainly assignmented to class i is $G_{i1} = \{g \in \mathbf{DL}_i \mid \arg\max_{k=\{1,2,\ldots,w\}} \underline{R_B^{P,\delta}}(\mathbf{DL}_k)(g) = i\}$, while $G_{i2} = \mathbf{DL}_i - G_{i1}$ represents the uncertain ones.

Furthermore, $\mathbb{Z}_{ij}$ captures the fine-grained uncertainty at the fuzzy object granule level

$$\mathbb{Z}_{ij} = -\sum_{g_t \in G_{ij}} P_{ijt} \log P_{ijt} \quad (20)$$

where $P_{ijt} = |[g_t]_{R_B^\delta} \cap \mathbf{DL}_i|/|[g_t]_{R_B^\delta}|$ quantifies the proportion of objects in g_t's fuzzy neighborhood that belongs to decision class i. Thus, $\mathbb{Z}_{ijt} = -P_{ijt} \log P_{ijt}$ characterizes the uncertainty within the fuzzy object granule $[g_t]_{R_B^\delta}$.

Overall, this measure effectively integrates hierarchical granularity and interlevel interactions among target decisions, approximation granules, and object granules. Compared with existing uncertainty measures in FRSs, it offers a comprehensive quantification of uncertainty across these three fuzzy layers. Its variation under different data distributions and parameters is further discussed in the following.

Property 2: Given a $\mathbb{DIS} = (O, C, D)$, for $B \subseteq C$ and $P \in (0, 1]$, the multilevel zentropy uncertainty measure $\mathbb{Z}_B^{P,\delta}(D)$ satisfies the following properties.

1) **Nonnegativity**: $\mathbb{Z}_B^{P,\delta}(D) \ge 0$.
2) **Nonmonotonicty**: For $B_1 \subseteq B_2 \subseteq C$, the $\mathbb{Z}_{B_1}^{P,\delta}(D)$ and $\mathbb{Z}_{B_2}^{P,\delta}(D)$ are nonmonotonic.
3) *Nonmonotonicty:* For $0 < P_1 \le P_2 \le 1$, $\mathbb{Z}_B^{P_1,\delta}(D)$ and $\mathbb{Z}_B^{P_2,\delta}(D)$ are nonmonotonic.
4) *Nonmonotonicty:* For $0 < \delta_1 \le \delta_2 \le 1$, $\mathbb{Z}_B^{P,\delta_1}(D)$ and $\mathbb{Z}_B^{P,\delta_2}(D)$ are nonmonotonic.
5) *Degeneracy:* For $\forall g_t \in \mathbf{EG}_B^P$, if $[g_t]_{R_B^\delta}$ is an unit vector, then we have $\mathbb{Z}_B^{P,\delta}(D) = -\sum_{i=1}^{w} P_i \log P_i$.

Proof: These properties can be proved from Definitions 4 and 6 and Property 1.

1) It is directly obtained from Definition 6.
2) According to Property 1, feature subset inclusion does not ensure the monotonicity of approximations or fuzzy granules. Hence, these measures are incomparable.
3) Similar to 2), changes in the purity P lead to structural differences in $\mathbf{EG}_B^P$ and the resulting uncertainty, making the $\mathbb{Z}_B^{P,\delta}(D)$ values nonmonotonic.
4) It can be proved similar to 2) and 3).
5) When $[g_t]_{R_B^\delta}$ is a unit vector (i.e., $R_B^\delta(g_t, g_t) = 1$ and $R_B^\delta(g_t, g_j) = 0$ for $g_j \ne g_t$), the uncertainty in $\mathbb{Z}_{ij}$ is zero. Meanwhile, the generalized granular balls in $\mathbf{DL}_i$ are divided into this target decision, leading to $P_{i1} = 1$ and $P_{i2} = 0$. Therefore, the full measure degenerates to top-level entropy.

B. Feature Evaluation and Selection

Based on the proposed multilevel zentropy uncertainty measure, this section designs two significance measures for feature evaluation and further develops a robust feature selection method by integrating information gain analysis.

Definition 7: Given a $\mathbb{DIS} = (O, C, D)$, for $P, \delta \in (0, 1]$ and $c \in C$, the inner significance of c related to C is defined as

$$Inn(c, C, D) = \mathbb{Z}_{C-\{c\}}^{P,\delta}(D) - \mathbb{Z}_C^{P,\delta}(D) \quad (21)$$

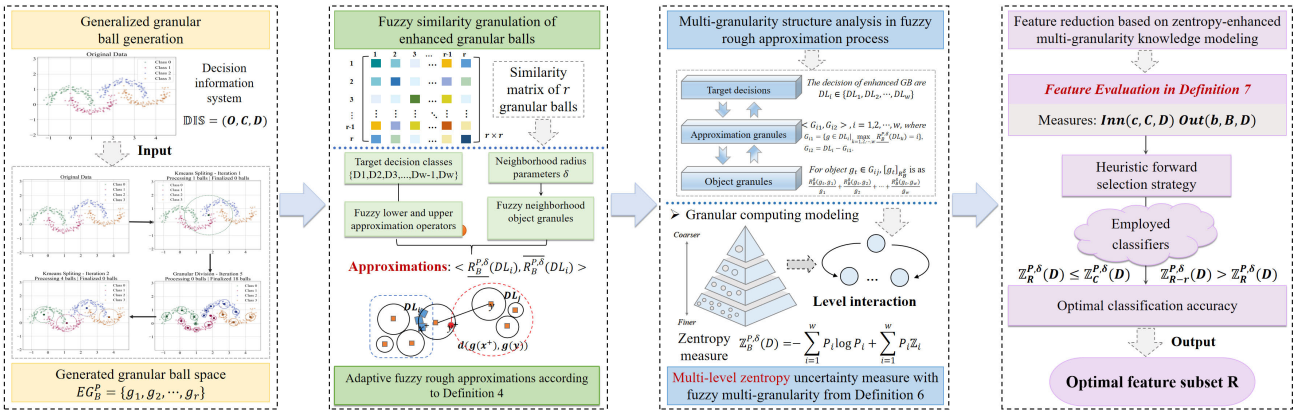

Fig. 4. Framework of the proposed method.

where $Inn(c, C, D) > 0$ implies that feature c is important for preserving uncertainty information in the system.

Similarly, for a feature subset $B \subset C$ and $\forall b \in C - B$, the outer significance of b with respect to B is defined as

$$Out(b, B, D) = \mathbb{Z}_B^{P,\delta}(D) - \mathbb{Z}_{B\cup\{b\}}^{P,\delta}(D) \tag{22}$$

where a larger $Out(b, B, D)$ value indicates a greater information gain brought by feature b.

Definition 8: Given a $\mathbb{DIS} = (O, C, D)$, for $P, \delta \in (0, 1]$, a feature subset $R \subseteq C$ is called a **reduct** if it satisfies the following conditions.

1) $\mathbb{Z}_R^{P,\delta}(D) \le \mathbb{Z}_C^{P,\delta}(D)$.
2) For $\forall r \in R$, $\mathbb{Z}_{R-\{r\}}^{P,\delta}(D) > \mathbb{Z}_R^{P,\delta}(D)$.

Combining the above feature reduction definitions, the developed zentropy-enhanced multigranularity knowledge modeling framework for robust feature selection (ZeMG-FS) is presented in Algorithm 1. In this algorithm, three steps are involved: core feature identification, outer feature selection, and redundant feature deletion. In the first step, generalized granular balls are generated with a time complexity of $O(nmw)$, where n, m, and w denote the numbers of objects, features, and decision classes, respectively. The construction of fuzzy relations among n objects under m features requires $O(n^2m)$ operations, which dominate the computational cost of the zentropy measure. Therefore, the time complexity of core feature evaluation over m features is $O(n^2(m - 1)m)$. Suppose that there are l_1 and l_2 selected features in Steps I and II, respectively. The time complexity of lines 14–20 is $O\left(\sum_{i=1}^{l_2} n^2(l_1 + i)(m - l_1 - i + 1)\right)$. In addition, to ensure the indispensability of each selected feature in Step III, all $l_1 + l_2$ features must be evaluated, resulting in a time complexity of $O((l_1 + l_2)(l_1 + l_2 - 1)n^2)$. Therefore, the overall time complexity of the algorithm is bounded by $O(n^2m^2)$.

Algorithm 1 Proposed Feature Selection Algorithm

Input : $\mathbb{DIS} = (O, C, D)$, purity threshold P, radius δ.
Output : Selected feature subset R.
1 Initialize $R \leftarrow \emptyset$;
2 ***Step I: Core feature identification***
3 Generate enhanced granular balls $\mathbf{EG}_C^{P,\delta}$ using Definition 3;
4 Compute the multi-level zentropy uncertainty measure of the full set: $\mathbb{Z}_C^{P,\delta}(D)$;
5 **forall** $c \in C$ **do**
6 Generate $\mathbf{EG}_{C-\{c\}}^{P,\delta}$;
7 Compute $Inn(c, C, D)$ using Eq. (21);
8 **if** $Inn(c, C, D) > 0$ **then**
9 $R \leftarrow R \cup \{c\}$;
10 **end**
11 **end**
12 Calculate the multi-level zentropy uncertainty measure on R, i.e., $\mathbb{Z}_R^{P,\delta}(D)$;
13 ***Step II: Outer important feature selection***
14 **while** $\mathbb{Z}_R^{P,\delta}(D) > \mathbb{Z}_C^{P,\delta}(D)$ **do**
15 **forall** $b \in C - R$ **do**
16 Compute $Out(b, R, D)$ using Eq. (22);
17 **end**
18 Select $b^* = \arg\max_{b\in C-R} Out(b, R, D)$;
19 $R \leftarrow R \cup \{b^*\}$;
20 **end**
21 ***Step III: Redundant feature deletion***
22 **forall** $r \in R$ **do**
23 **if** $\mathbb{Z}_{R-\{r\}}^{P,\delta}(D) \le \mathbb{Z}_R^{P,\delta}(D)$ **then**
24 $R \leftarrow R - \{r\}$;
25 **end**
26 **end**
Return : Reduct R

C. Overall Procedure

The overall procedure of the proposed method is illustrated in Fig. 4, which comprises four interconnected phases: generalized granular-ball generation, robust fuzzy rough approximation modeling, multigranularity uncertainty analysis, and feature selection. In the first phase, a generalized granular-ball generation method is developed. Unlike existing strategies, it integrates class label information to better reflect the local data distribution, enabling adaptive splitting and a robust foundation for multigranularity knowledge modeling.

In the second phase, a robust fuzzy rough approximation model is constructed. Fuzzy object granules are generated by introducing a neighborhood δ to the fuzzy similarity matrix, and then, the approximation space is obtained within a pair of fuzzy operators. Compared with the classical FRS model, this model significantly improves the tolerance of generalized boundary granular balls.

In the third phase, a multigranularity structure is established in the fuzzy approximation process. By analyzing interactions among target decisions, approximation granules, and object granules, a new multilevel zentropy uncertainty measure is

TABLE I
DETAILS ABOUT EMPLOYED 12 DATASETS

No.s	Dataset	Abbreviation	Object	Feature	Class
D1	Cardiotocography DatasetS3	Card	2126	21	3
D2	Climate Model Simulation Crashes	Cmsc	504	18	2
D3	Colon	Colo	62	2001	2
D4	Ionosphere	Iono	351	34	2
D5	Mice Protein Expression Dataset	MPED	1077	68	8
D6	Musk	Musk	476	168	2
D7	Segmentation	Sege	2310	20	7
D8	South German Credit	SoGC	1000	21	2
D9	Spambase	Spam	4601	58	2
D10	Thyroid	Thyr	7200	22	3
D11	Urban Land Cover	ULC	168	148	9
D12	Visegrad Group Companies Dataset	VGCD	85	450	6

designed to quantify the multigranularity uncertainty in the GFRS model accurately.

In the final phase, a robust feature selection approach is proposed by leveraging the zentropy uncertainty measure. Specifically, two significance criteria are defined to assess the individual contributions of features. A forward selection mechanism guided by the maximum information gain is then employed to construct the optimal feature subset iteratively.

V. EXPERIMENTAL RESULTS AND ANALYSIS

This section presents a series of numerical experiments to verify the robustness and effectiveness of the proposed method. All the experiments are performed on a computer with Intel[1] Core[2] i7-6800K ×12, memory: 62.7 GB, and MATLAB R2020a.

A. General Setting

1) Datasets: Twelve public datasets from the UCI Machine Learning Repository and the scikit-feature library are employed to make comparisons, where the details are summarized in Table I. Meanwhile, all the conditional attributes are normalized to the unit interval using the following equation:

$$\widehat{v}\left(o_i, c_j\right) = \frac{v\left(o_i, c_j\right) - min\left(c_j\right)}{max\left(c_j\right) - min\left(c_j\right)} \tag{23}$$

where $v(o_i, c_j)$ denotes the value of object o_i under feature c_j, and $\min(c_j)$ and $\max(c_j)$ represent the minimum and maximum values of c_j over all objects.

2) Compared Methods: To evaluate the antinoise ability of the proposed fuzzy rough knowledge acquisition model, five FRS methods, including classific FRS (CFRS) [23], noise-aware FRS (NFRS) [32], FNRS [18], FRS with fuzzy decision (FRSD) [14], and RCU-based FRS (RFRS) [13], are adopted to make comparisons. In addition, seven other representative feature selection methods are employed to illustrate the effectiveness of ZeMG-FS in classification performance. These details are given as follows.

1) *FNCE [33]:* It introduces neighborhood granulation and proposes combination entropy for feature selection.
2) *FSNC [34]:* It applies stripped neighborhood covers to feature selection, where fuzzy dependency degree is employed as an important index.
3) *FSZUM [28]:* It analyzes the structure relationship among different granularities and develops a zentropy-based uncertainty measure for feature selection.
4) *MFEFS [35]:* It proposes multiscale fuzzy entropy for uncertain information processing and uses it for feature selection from a multiscale viewpoint.
5) *VCOS [27]:* It defines a variable-consistency selection mechanism for optimal scales and proposes fuzzy combination entropy for feature evaluation.
6) *AGM-FS [1]:* It defines the margin-based weight optimization function for feature selection and obtains the final subset based on the average granule margin measure.
7) *SFSS [9]:* It constructs a separability measure by integrating the relationship between objects and decisions for feature selection from a reduction perspective.

3) Comparison Setups: Four classifiers—KNN, NB, SVM, and DT—are employed to evaluate the classification performance of the selected features. For each dataset, tenfold cross-validation is used to ensure a fair and reliable comparison. To analyze the effects of the parameters on the proposed method, the purity P is set from 0.8 to 1 with a step of 0.05, and the radius δ is set from 0 to 0.5 with a step of 0.05.

B. Robustness Analysis of Fuzzy Approximation

As discussed in Section III-B, the proposed fuzzy approximation model enhances noise tolerance by flexibly adjusting boundary regions. To verify the robustness of the proposed fuzzy approximation, this section compares it with several representative FRS-based models under varying noise conditions. Specifically, the fuzzy dependency degree, as defined in Definition 5, is employed as the evaluation criterion to measure the approximation accuracy of each model. For each dataset, the noised data are obtained as follows:

$$v^*\left(o_i, c_j\right) = \begin{cases} 0, & \widehat{v}\left(o_i, c_j\right) + z_{ij} < 0 \\ \widehat{v}\left(o_i, c_j\right) + z_{ij}, & 0 \le \widehat{v}\left(o_i, c_j\right) + z_{ij} \le 1 \\ 1, & \widehat{v}\left(o_i, c_j\right) + z_{ij} > 1 \end{cases} \tag{24}$$

where $z_{ij} \in [0, 1]$ is the Gaussian random noise.

Table II presents the approximate accuracy of all compared methods across ten noise levels, where the noise level refers to the property of added noise objects, and the best results are presented in bold. As observed, the proposed approximation model achieves the highest average values on nine out of the 12 datasets, except for D3, D6, and D8, demonstrating its strong approximation ability and robustness in noisy environments. This is because the proposed generalized granular-ball granulation reduces discordant objects neglected in the approximation process, effectively improving the noise tolerance and enhancing the fuzzy approximation accuracy. Furthermore, it exhibits lower variance across noise levels than CFRS, NFRS, and RFRS do in most situations, indicating more stable performance under perturbation.

A visual comparison is also provided in Fig. 5, which clearly illustrates the superior resilience of our method to noise interference. These results collectively verify that the proposed method significantly enhances noise robustness.

[1]Registered trademark.
[2]Trademarked.

TABLE II
APPROXIMATION ACCURACY OF DIFFERENT FRS MODELS IN NOISY ENVIRONMENT

No.s	FRSs	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	Ave
D1	CFRS	0.19	0.23	0.26	0.30	0.33	0.36	0.40	0.43	0.46	0.49	0.34±0.10
	NFRS	0.20	0.24	0.29	0.33	0.37	0.41	0.45	0.49	0.53	0.58	0.39±0.12
	FNRS	0.10	0.15	0.20	0.26	0.31	0.36	0.41	0.46	0.51	0.56	0.33±0.15
	FRSD	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44±0.00
	RFRS	0.12	0.16	0.20	0.24	0.28	0.32	0.36	0.40	0.44	0.48	0.30±0.11
	Our	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.97	0.87	0.85	**0.97**±0.05
D2	CFRS	0.57	0.58	0.59	0.60	0.61	0.61	0.62	0.64	0.65	0.67	0.61±0.03
	RFRS	0.68	0.70	0.71	0.72	0.73	0.74	0.74	0.76	0.77	0.78	0.73±0.03
	FNRS	0.92	0.92	0.92	0.93	0.93	0.94	0.94	0.95	0.95	0.96	0.94±0.01
	NFRS	0.89	0.88	0.89	0.89	0.90	0.90	0.90	0.91	0.92	0.96	0.90±0.02
	FRSD	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84±0.00
	Our	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.92	0.90	0.97	**0.98**±0.04
D3	CFRS	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**1.00**±0.00
	NFRS	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**1.00**±0.00
	FNRS	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**1.00**±0.00
	FRSD	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54±0.00
	RFRS	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**1.00**±0.00
	Our	0.81	0.81	0.76	1.00	0.63	0.69	1.00	1.00	0.67	0.78	0.82±0.13
D4	CFRS	0.42	0.43	0.44	0.45	0.47	0.47	0.49	0.52	0.53	0.55	0.48±0.04
	RFRS	0.57	0.59	0.62	0.64	0.66	0.68	0.70	0.72	0.75	0.77	0.67±0.06
	FNRS	0.48	0.52	0.56	0.58	0.61	0.63	0.66	0.68	0.71	0.74	0.62±0.08
	NFRS	0.84	0.85	0.85	0.86	0.87	0.88	0.89	0.93	0.93	0.95	0.88±0.04
	FRSD	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54±0.00
	Our	1.00	1.00	1.00	0.90	0.90	0.91	0.87	0.93	0.95	0.93	**0.94**±0.04
D5	CFRS	0.29	0.32	0.34	0.37	0.40	0.43	0.46	0.48	0.51	0.54	0.41±0.08
	NFRS	0.51	0.54	0.57	0.59	0.62	0.64	0.67	0.70	0.73	0.75	0.63±0.08
	FNRS	0.04	0.09	0.14	0.19	0.24	0.29	0.34	0.39	0.44	0.49	0.27±0.14
	FRSD	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13±0.00
	RFRS	0.43	0.46	0.49	0.52	0.55	0.58	0.61	0.64	0.67	0.70	0.56±0.09
	Our	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**1.00**±0.00
D6	CFRS	0.59	0.61	0.63	0.65	0.67	0.70	0.72	0.74	0.76	0.78	0.68±0.06
	RFRS	0.90	0.91	0.91	0.92	0.92	0.93	0.94	0.95	0.95	0.96	0.93±0.02
	FNRS	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**1.00**±0.00
	NFRS	0.96	0.97	0.97	0.97	0.97	0.98	0.98	0.98	0.98	0.99	0.97±0.01
	FRSD	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51±0.00
	Our	1.00	0.95	0.98	0.95	0.88	0.97	0.94	0.94	0.97	1.00	0.96±0.03
D7	CFRS	0.26	0.29	0.32	0.35	0.38	0.40	0.43	0.46	0.49	0.51	0.39±0.08
	NFRS	0.25	0.29	0.33	0.37	0.41	0.45	0.48	0.52	0.56	0.59	0.42±0.11
	FNRS	0.06	0.11	0.16	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.28±0.14
	FRSD	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15±0.00
	RFRS	0.19	0.23	0.26	0.30	0.33	0.36	0.39	0.43	0.46	0.49	0.34±0.09
	Our	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**1.00**±0.00
D8	CFRS	0.55	0.56	0.58	0.60	0.62	0.64	0.65	0.67	0.68	0.70	0.62±0.05
	RFRS	0.52	0.55	0.58	0.60	0.62	0.64	0.66	0.69	0.70	0.73	0.63±0.06
	FNRS	0.92	0.93	0.95	0.95	0.96	0.96	0.98	0.98	0.99	0.99	**0.96**±0.02
	NFRS	0.75	0.76	0.78	0.79	0.81	0.82	0.84	0.86	0.87	0.88	0.82±0.04
	FRSD	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58±0.00
	Our	0.99	0.99	0.97	0.99	0.97	0.96	0.94	0.90	0.86	0.85	0.94±0.05
D9	CFRS	0.10	0.13	0.16	0.18	0.21	0.23	0.25	0.28	0.30	0.32	0.22±0.07
	NFRS	0.10	0.14	0.18	0.22	0.26	0.30	0.34	0.38	0.42	0.46	0.28±0.12
	FNRS	0.03	0.07	0.10	0.15	0.19	0.23	0.28	0.33	0.39	0.45	0.22±0.13
	FRSD	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18±0.00
	RFRS	0.06	0.09	0.12	0.14	0.17	0.19	0.22	0.25	0.27	0.30	0.18±0.08
	Our	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**1.00**±0.00
D10	CFRS	0.18	0.23	0.26	0.31	0.34	0.38	0.42	0.46	0.49	0.52	0.36±0.11
	RFRS	0.10	0.14	0.19	0.23	0.28	0.32	0.36	0.40	0.44	0.48	0.29±0.12
	FNRS	0.95	0.96	0.96	0.96	0.96	0.97	0.97	0.97	0.97	0.98	**0.97**±0.01
	NFRS	0.26	0.30	0.34	0.38	0.42	0.46	0.51	0.55	0.58	0.63	0.44±0.12
	FRSD	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86±0.00
	Our	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.88	0.89	0.90	**0.97**±0.05
D11	CFRS	0.47	0.49	0.52	0.54	0.57	0.59	0.62	0.64	0.66	0.69	0.58±0.07
	NFRS	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.99	0.99	0.98±0.00
	FNRS	0.09	0.14	0.18	0.23	0.28	0.32	0.37	0.42	0.48	0.52	0.30±0.14
	FRSD	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12±0.00
	RFRS	0.91	0.92	0.93	0.93	0.93	0.93	0.94	0.95	0.95	0.95	0.94±0.01
	Our	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**1.00**±0.00
D12	CFRS	0.16	0.20	0.24	0.29	0.33	0.37	0.42	0.45	0.49	0.54	0.35±0.12
	RFRS	0.11	0.16	0.20	0.25	0.30	0.34	0.39	0.43	0.47	0.52	0.32±0.13
	FNRS	0.10	0.14	0.19	0.25	0.30	0.34	0.38	0.42	0.46	0.51	0.31±0.13
	NFRS	0.21	0.25	0.29	0.34	0.44	0.45	0.49	0.52	0.56	0.60	0.41±0.13
	FRSD	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33±0.00
	Our	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.97	0.87	0.85	**0.97**±0.05

TABLE III
COMPUTATIONAL TIME OF COMPARED METHODS ON 12 DATASETS (S)

No.s	VCOS	FSNC	MFEFS	FSZUM	FNCE	AGM-FS	SFSS	Our
D1	3844	166.5	273.4	48.20	44.50	146.10	4.08	663.4
D2	61.90	14.68	1.90	3.13	7.30	6.30	0.98	144.1
D3	15.78	76.43	466.3	21.55	108.9	530.00	1.28	237.2
D4	32.70	25.83	13.50	6.68	8.75	17.30	2.15	46.70
D5	1973	874.4	1918	90.45	62.05	343.10	61.50	2055
D6	490.2	770.6	5594	33.18	53.85	471.90	19.93	582.5
D7	4554	140.9	699	84.05	50.60	132.70	8.48	624.0
D8	81.25	51.50	91.75	3.05	14.83	31.00	2.78	346.4
D9	8319	9560	28095	2915	1005	3691.60	58.53	6754
D10	13018	2214	6472	897.0	534.1	1329.80	8.18	875.9
D11	50.95	177.1	368.1	50.68	36.50	96.10	24.08	137.8
D12	37.38	27.63	48.23	16.70	11.85	3.80	1.98	110.8
Ave	2706.53	1174.97	3670.09	347.47	161.52	566.64	16.16	1048.10

C. Performance Comparison of ZeMG-FS Algorithm

This section mainly evaluates the effectiveness of the proposed method from computational time, selected feature number, classification performance, and statistical validation.

1) Computational Time: The computational times (in seconds) of all methods compared on the 12 datasets are listed in Table III. On average, our method ZeMG-FS exhibits a lower runtime than the other three methods—VCOS, FSNC, and MFEFS. Moreover, the SFSS achieves the shortest time because it neglects the multigranularity structure and approximation. Overall, the proposed ZeMG-FS achieves competitive efficiency in feature selection.

2) Selected Feature Number: Tables IV–VII present the feature numbers selected through different methods under four classifiers. On the KNN classifier, the proposed method selects the fewest features on six datasets, while FSNC, MFEFS, FSZUM, and FNCE obtain the lowest values 1×, 2×, 1×, and 2×, respectively. Similarly, ZeMG-FS performs better 8×, 9×, and 7*times* under the NB, SVM, and DT classifiers, respectively. Moreover, the average feature number is the lowest for these four classifiers, demonstrating that the proposed ZeMG-FS effectively selects appropriate features for learning tasks.

3) Classification Performance: The classification accuracy results for the four classifiers are reported in Tables IV–VII, where the best results are highlighted in bold.

Overall, ZeMG-FS consistently outperforms the other methods across most datasets. Specifically, it achieves the highest accuracy on nine datasets, showing lower performance for the KNN classifier only on the Colo, Musk, and Thyr datasets. In contrast, FSNC, MFEFS, FNCE, and AGM-FS are excellent on only one or two datasets. Notably, despite selecting 20 and 23 features, respectively, MFEFS and FNCE yield the lowest accuracy on the Iono dataset, further underscoring the importance of selecting appropriate features.

Similar trends are observed with the NB, SVM, and DT classifiers, where ZeMG-FS achieves the highest accuracy on ten, ten, and 11 datasets, respectively. Other methods, such as MFEFS, VCOS, FSZUM, and AGM-FS, perform better on only one or two datasets. Poor classification caused by selecting redundant features is evident in methods, such as

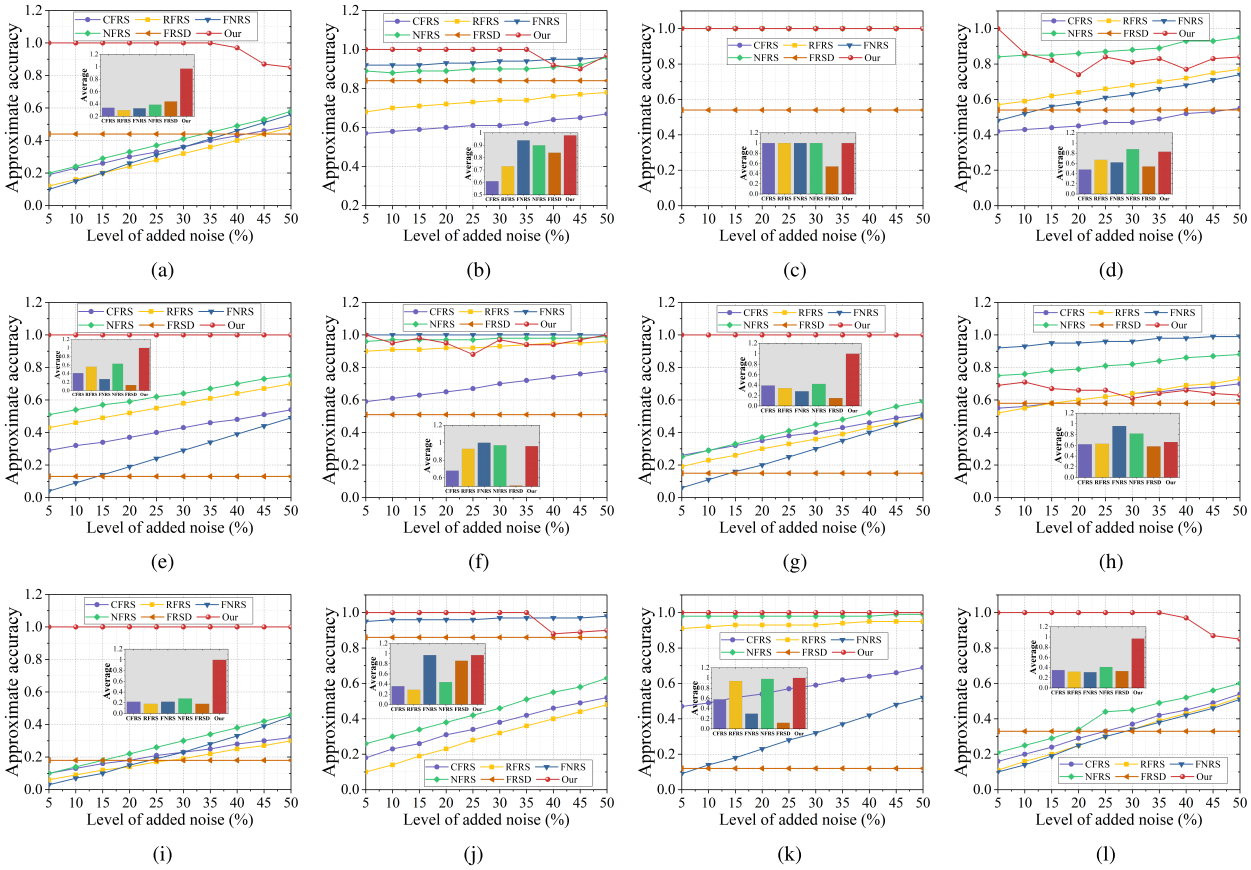

Fig. 5. Approximation accuracy of different FRS models on ten different noise levels. (a) D1. (b) D2. (c) D3. (d) D4. (e) D5. (f) D6. (g) D7. (h) D8. (i) D9. (j) D10. (k) D11. (l) D12.

TABLE IV

CLASSIFICATION ACCURACY AND SELECTED FEATURE NUMBER OF COMPARED METHODS WITH THE KNN CLASSIFIER (%)

No.s	Raw		VCOS		FSNC		MFEFS		FSZUM		FNCE		AGM-FS		SFSS		ZeMG-FS	
	Accuracy	Num	Accuracy	Num	Accuracy	Num	Accuracy	Num	Accuracy	Num	Accuracy	Num	Accuracy	Num	Accuracy	Num	Accuracy	Num
D1	81.28$_{\pm 0.02}$	20	81.33$_{\pm 0.03}$	20	81.23$_{\pm 0.02}$	17	87.06$_{\pm 0.02}$	9	83.02$_{\pm 0.01}$	16	81.04$_{\pm 0.03}$	20	81.65$_{\pm 0.02}$	19	71.50$_{\pm 0.03}$	**3**	**91.02**$_{\pm 0.02}$	**3**
D2	92.22$_{\pm 0.02}$	17	92.04$_{\pm 0.04}$	17	92.96$_{\pm 0.03}$	5	94.07$_{\pm 0.03}$	**4**	94.44$_{\pm 0.02}$	6	92.04$_{\pm 0.02}$	17	93.52$_{\pm 0.03}$	6	93.89$_{\pm 0.04}$	6	**95.74**$_{\pm 0.03}$	6
D3	77.38$_{\pm 0.15}$	2000	84.29$_{\pm 0.15}$	18	80.95$_{\pm 0.22}$	15	85.95$_{\pm 0.17}$	16	81.90$_{\pm 0.17}$	3	53.33$_{\pm 0.21}$	**1**	**91.67**$_{\pm 0.12}$	83	80.71$_{\pm 0.18}$	**1**	84.89$_{\pm 0.15}$	**1**
D4	86.04$_{\pm 0.05}$	33	85.75$_{\pm 0.06}$	28	89.46$_{\pm 0.05}$	**8**	90.02$_{\pm 0.05}$	20	86.04$_{\pm 0.03}$	33	86.30$_{\pm 0.06}$	23	85.77$_{\pm 0.06}$	33	89.44$_{\pm 0.08}$	10	**91.17**$_{\pm 0.04}$	10
D5	99.16$_{\pm 0.01}$	68	99.72$_{\pm 0.00}$	25	99.63$_{\pm 0.00}$	66	99.72$_{\pm 0.01}$	65	99.63$_{\pm 0.01}$	45	99.63$_{\pm 0.01}$	**21**	99.63$_{\pm 0.01}$	65	99.44$_{\pm 0.01}$	29	**99.81**$_{\pm 0.00}$	29
D6	82.97$_{\pm 0.04}$	166	80.05$_{\pm 0.06}$	59	83.56$_{\pm 0.08}$	76	83.62$_{\pm 0.05}$	107	78.37$_{\pm 0.07}$	**8**	82.55$_{\pm 0.06}$	77	**88.24**$_{\pm 0.04}$	122	73.75$_{\pm 0.04}$	44	84.67$_{\pm 0.05}$	44
D7	96.06$_{\pm 0.01}$	19	96.88$_{\pm 0.01}$	12	96.45$_{\pm 0.01}$	14	96.58$_{\pm 0.01}$	18	96.58$_{\pm 0.01}$	19	96.28$_{\pm 0.01}$	15	96.49$_{\pm 0.01}$	19	84.50$_{\pm 0.03}$	**3**	**97.14**$_{\pm 0.01}$	**3**
D8	71.20$_{\pm 0.04}$	20	72.20$_{\pm 0.05}$	20	**73.10**$_{\pm 0.05}$	20	71.70$_{\pm 0.04}$	20	72.50$_{\pm 0.03}$	20	**73.10**$_{\pm 0.03}$	20	72.40$_{\pm 0.05}$	15	72.90$_{\pm 0.04}$	**14**	**73.10**$_{\pm 0.04}$	**14**
D9	90.29$_{\pm 0.01}$	57	90.26$_{\pm 0.01}$	55	90.00$_{\pm 0.02}$	55	90.65$_{\pm 0.02}$	**42**	90.29$_{\pm 0.01}$	57	90.24$_{\pm 0.02}$	56	89.74$_{\pm 0.01}$	56	89.94$_{\pm 0.01}$	49	**90.87**$_{\pm 0.02}$	49
D10	93.83$_{\pm 0.00}$	21	93.93$_{\pm 0.01}$	20	93.93$_{\pm 0.01}$	20	96.47$_{\pm 0.00}$	17	93.83$_{\pm 0.00}$	21	93.93$_{\pm 0.01}$	20	93.83$_{\pm 0.02}$	19	95.72$_{\pm 0.01}$	**1**	95.82$_{\pm 0.01}$	**1**
D11	78.57$_{\pm 0.07}$	147	76.18$_{\pm 0.10}$	47	80.37$_{\pm 0.11}$	88	83.42$_{\pm 0.11}$	51	76.84$_{\pm 0.15}$	22	77.94$_{\pm 0.08}$	76	82.13$_{\pm 0.11}$	73	81.62$_{\pm 0.10}$	**16**	**83.46**$_{\pm 0.10}$	**16**
D12	47.42$_{\pm 0.07}$	49	48.71$_{\pm 0.06}$	49	48.72$_{\pm 0.07}$	23	50.02$_{\pm 0.07}$	32	50.31$_{\pm 0.07}$	49	46.41$_{\pm 0.09}$	19	40.58$_{\pm 0.10}$	1	46.44$_{\pm 0.11}$	**6**	**50.32**$_{\pm 0.08}$	**6**
Ave	83.04$_{\pm 0.04}$	218.08	83.45$_{\pm 0.05}$	30.83	84.20$_{\pm 0.06}$	33.92	85.77$_{\pm 0.05}$	33.42	83.65$_{\pm 0.05}$	24.92	81.07$_{\pm 0.05}$	30.42	84.64$_{\pm 0.05}$	42.58	81.65$_{\pm 0.06}$	**15.17**	**86.50**$_{\pm 0.04}$	**15.17**

FNCE, on the Musk and ULC datasets. Moreover, the average results are illustrated in Fig. 6, further demonstrating the superiority of generalized granular-ball generation and the multilevel zentropy measure in feature selection.

4) Statistical Validation: To assess whether significant differences exist among the compared methods, the Friedman test [36] is conducted at a significance level of $\alpha = 0.1$. The null hypothesis assumes that all methods perform equivalently; this hypothesis is rejected if the P value is less than the α level. The P values of the Friedman tests are 1.76×10^{-6}, 0.0016, 0.01, and 2.28×10^{-6} for the KNN, NB, SVM, and DT classifiers, respectively, significantly lower than 0.1. This indicates significant differences among these algorithms.

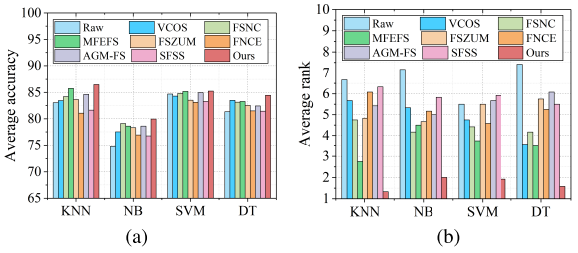

Fig. 6. Average classification performance of compared methods. (a) Accuracy. (b) Rank.

To further identify significant differences between any two methods, the Nemenyi post hoc test [36] is employed.

TABLE V
CLASSIFICATION ACCURACY AND SELECTED FEATURE NUMBER OF COMPARED METHODS WITH THE NB CLASSIFIER (%)

No.s	Raw		VCOS		FSNC		MFEFS		FSZUM		FNCE		AGM-FS		SFSS		ZeMG-FS	
	Accuracy	Num	Accuracy	Num	Accuracy	Num	Accuracy	Num	Accuracy	Num	Accuracy	Num	Accuracy	Num	Accuracy	Num	Accuracy	Num
D1	$50.28_{\pm 0.04}$	20	$53.81_{\pm 0.03}$	18	$69.05_{\pm 0.03}$	17	$58.28_{\pm 0.04}$	9	$68.30_{\pm 0.03}$	16	$67.31_{\pm 0.04}$	8	$64.16_{\pm 0.03}$	19	$67.22_{\pm 0.03}$	**3**	$\mathbf{71.54}_{\pm 0.03}$	**3**
D2	$92.04_{\pm 0.03}$	17	$92.22_{\pm 0.05}$	17	$93.70_{\pm 0.05}$	8	$94.44_{\pm 0.03}$	**4**	$94.44_{\pm 0.02}$	5	$92.41_{\pm 0.04}$	17	$94.07_{\pm 0.03}$	6	$93.89_{\pm 0.04}$	6	$\mathbf{95.00}_{\pm 0.03}$	6
D3	$61.43_{\pm 0.21}$	2000	$88.81_{\pm 0.13}$	18	$85.48_{\pm 0.09}$	4	$\mathbf{91.67}_{\pm 0.12}$	16	$83.81_{\pm 0.17}$	5	$65.48_{\pm 0.19}$	**1**	$82.14_{\pm 0.14}$	83	$79.29_{\pm 0.19}$	**1**	$80.71_{\pm 0.19}$	**1**
D4	$91.17_{\pm 0.05}$	33	$90.61_{\pm 0.08}$	28	$91.17_{\pm 0.05}$	33	$91.44_{\pm 0.05}$	25	$91.48_{\pm 0.06}$	33	$90.61_{\pm 0.06}$	33	$90.89_{\pm 0.06}$	33	$90.58_{\pm 0.03}$	**21**	$\mathbf{92.30}_{\pm 0.04}$	**21**
D5	$85.14_{\pm 0.05}$	68	$86.35_{\pm 0.02}$	**25**	$85.89_{\pm 0.04}$	67	$85.70_{\pm 0.04}$	65	$85.23_{\pm 0.05}$	68	$85.79_{\pm 0.03}$	66	$86.06_{\pm 0.05}$	65	$84.21_{\pm 0.02}$	58	$\mathbf{87.56}_{\pm 0.03}$	58
D6	$83.16_{\pm 0.05}$	166	$79.40_{\pm 0.06}$	59	$81.73_{\pm 0.06}$	106	$82.77_{\pm 0.05}$	107	$76.88_{\pm 0.06}$	**8**	$81.92_{\pm 0.04}$	100	$82.32_{\pm 0.05}$	122	$74.57_{\pm 0.06}$	66	$\mathbf{83.67}_{\pm 0.09}$	66
D7	$90.09_{\pm 0.02}$	19	$92.25_{\pm 0.02}$	11	$91.95_{\pm 0.01}$	12	$90.39_{\pm 0.03}$	18	$90.48_{\pm 0.02}$	19	$91.60_{\pm 0.02}$	15	$90.22_{\pm 0.03}$	19	$82.99_{\pm 0.02}$	**3**	$\mathbf{93.38}_{\pm 0.01}$	**3**
D8	$70.40_{\pm 0.05}$	20	$71.00_{\pm 0.05}$	20	$71.20_{\pm 0.03}$	15	$70.90_{\pm 0.05}$	20	$70.80_{\pm 0.04}$	14	$71.30_{\pm 0.03}$	**7**	$70.70_{\pm 0.07}$	15	$71.30_{\pm 0.05}$	12	$\mathbf{71.40}_{\pm 0.03}$	12
D9	$54.16_{\pm 0.03}$	57	$54.16_{\pm 0.03}$	53	$53.97_{\pm 0.03}$	55	$54.12_{\pm 0.03}$	57	$54.27_{\pm 0.03}$	55	$54.18_{\pm 0.03}$	56	$54.88_{\pm 0.02}$	56	$55.27_{\pm 0.03}$	**44**	$\mathbf{55.42}_{\pm 0.03}$	**44**
D10	$94.06_{\pm 0.01}$	21	$94.08_{\pm 0.01}$	20	$94.08_{\pm 0.01}$	20	$94.08_{\pm 0.01}$	20	$94.08_{\pm 0.01}$	20	$94.08_{\pm 0.01}$	20	$94.00_{\pm 0.01}$	19	$95.78_{\pm 0.01}$	**1**	$\mathbf{95.96}_{\pm 0.00}$	**1**
D11	$83.93_{\pm 0.10}$	147	$82.06_{\pm 0.14}$	47	$83.97_{\pm 0.09}$	88	$\mathbf{86.40}_{\pm 0.10}$	51	$83.31_{\pm 0.06}$	22	$83.90_{\pm 0.08}$	76	$83.97_{\pm 0.14}$	73	$74.96_{\pm 0.09}$	**16**	$81.65_{\pm 0.07}$	**16**
D12	$41.57_{\pm 0.09}$	49	$45.43_{\pm 0.07}$	29	$47.12_{\pm 0.08}$	24	$43.22_{\pm 0.05}$	34	$46.72_{\pm 0.07}$	38	$44.70_{\pm 0.12}$	24	$49.98_{\pm 0.07}$	1	$50.91_{\pm 0.09}$	**1**	$\mathbf{50.99}_{\pm 0.08}$	**1**
Ave	$74.79_{\pm 0.06}$	218.08	$77.52_{\pm 0.06}$	28.75	$79.11_{\pm 0.05}$	37.42	$78.62_{\pm 0.05}$	35.50	$78.32_{\pm 0.05}$	25.25	$76.94_{\pm 0.06}$	35.25	$78.62_{\pm 0.06}$	42.58	$76.75_{\pm 0.05}$	**19.33**	$\mathbf{79.97}_{\pm 0.05}$	**19.33**

TABLE VI
CLASSIFICATION ACCURACY AND SELECTED FEATURE NUMBER OF COMPARED METHODS WITH THE SVM CLASSIFIER (%)

No.s	Raw		VCOS		FSNC		MFEFS		FSZUM		FNCE		AGM-FS		SFSS		ZeMG-FS	
	Accuracy	Num	Accuracy	Num	Accuracy	Num	Accuracy	Num	Accuracy	Num	Accuracy	Num	Accuracy	Num	Accuracy	Num	Accuracy	Num
D1	$85.37_{\pm 0.02}$	20	$85.65_{\pm 0.02}$	20	$85.80_{\pm 0.02}$	20	$85.75_{\pm 0.02}$	20	$\mathbf{85.84}_{\pm 0.03}$	20	$85.70_{\pm 0.01}$	20	$85.56_{\pm 0.02}$	19	$85.70_{\pm 0.03}$	**17**	$\mathbf{85.84}_{\pm 0.02}$	**17**
D2	$95.00_{\pm 0.03}$	17	$95.00_{\pm 0.03}$	17	$95.00_{\pm 0.03}$	17	$\mathbf{95.37}_{\pm 0.03}$	11	$94.44_{\pm 0.03}$	**5**	$95.00_{\pm 0.03}$	17	$94.81_{\pm 0.03}$	6	$95.00_{\pm 0.03}$	6	$\mathbf{95.37}_{\pm 0.03}$	6
D3	$83.57_{\pm 0.16}$	2000	$84.29_{\pm 0.16}$	18	$83.81_{\pm 0.15}$	4	$\mathbf{88.57}_{\pm 0.14}$	16	$82.14_{\pm 0.16}$	3	$64.05_{\pm 0.21}$	**1**	$87.14_{\pm 0.12}$	83	$85.00_{\pm 0.15}$	**1**	$84.29_{\pm 0.12}$	**1**
D4	$87.74_{\pm 0.05}$	33	$88.61_{\pm 0.05}$	28	$88.59_{\pm 0.05}$	33	$88.59_{\pm 0.07}$	27	$88.60_{\pm 0.06}$	33	$88.32_{\pm 0.06}$	33	$87.18_{\pm 0.05}$	33	$87.22_{\pm 0.08}$	**23**	$\mathbf{89.45}_{\pm 0.04}$	**23**
D5	$97.77_{\pm 0.02}$	68	$97.77_{\pm 0.02}$	68	$97.96_{\pm 0.01}$	65	$97.96_{\pm 0.02}$	65	$97.77_{\pm 0.02}$	68	$97.87_{\pm 0.01}$	66	$97.77_{\pm 0.01}$	65	$97.68_{\pm 0.02}$	**59**	$\mathbf{98.05}_{\pm 0.01}$	**59**
D6	$82.54_{\pm 0.05}$	166	$79.84_{\pm 0.05}$	59	$81.09_{\pm 0.07}$	110	$81.54_{\pm 0.10}$	107	$74.16_{\pm 0.07}$	**7**	$82.14_{\pm 0.05}$	100	$82.32_{\pm 0.06}$	122	$78.17_{\pm 0.04}$	44	$\mathbf{82.57}_{\pm 0.05}$	44
D7	$92.90_{\pm 0.01}$	19	$92.90_{\pm 0.01}$	12	$92.94_{\pm 0.01}$	16	$92.86_{\pm 0.02}$	18	$92.81_{\pm 0.02}$	19	$\mathbf{93.03}_{\pm 0.02}$	16	$92.99_{\pm 0.02}$	19	$91.99_{\pm 0.02}$	**11**	$\mathbf{93.03}_{\pm 0.02}$	**11**
D8	$75.50_{\pm 0.06}$	20	$76.00_{\pm 0.05}$	20	$75.10_{\pm 0.03}$	19	$75.60_{\pm 0.03}$	20	$75.80_{\pm 0.05}$	14	$75.60_{\pm 0.03}$	20	$75.10_{\pm 0.05}$	15	$75.80_{\pm 0.05}$	**14**	$\mathbf{76.80}_{\pm 0.04}$	**14**
D9	$90.15_{\pm 0.01}$	57	$\mathbf{90.24}_{\pm 0.01}$	53	$90.02_{\pm 0.01}$	55	$90.15_{\pm 0.01}$	57	$90.15_{\pm 0.01}$	57	$90.18_{\pm 0.01}$	56	$90.07_{\pm 0.01}$	56	$90.15_{\pm 0.01}$	**52**	$89.85_{\pm 0.01}$	**52**
D10	$93.68_{\pm 0.01}$	21	$93.40_{\pm 0.01}$	20	$93.40_{\pm 0.01}$	20	$93.40_{\pm 0.01}$	20	$93.68_{\pm 0.01}$	21	$93.40_{\pm 0.01}$	20	$93.61_{\pm 0.02}$	19	$93.18_{\pm 0.01}$	**1**	$\mathbf{93.68}_{\pm 0.01}$	**1**
D11	$80.99_{\pm 0.08}$	147	$76.80_{\pm 0.11}$	47	$82.72_{\pm 0.09}$	88	$82.10_{\pm 0.09}$	91	$75.63_{\pm 0.10}$	**29**	$80.88_{\pm 0.08}$	76	$81.58_{\pm 0.09}$	73	$68.46_{\pm 0.13}$	33	$\mathbf{82.76}_{\pm 0.06}$	33
D12	$51.01_{\pm 0.08}$	49	$51.02_{\pm 0.08}$	49	$51.37_{\pm 0.12}$	22	$51.33_{\pm 0.07}$	30	$51.31_{\pm 0.09}$	9	$51.31_{\pm 0.07}$	25	$51.27_{\pm 0.08}$	1	$51.37_{\pm 0.12}$	**1**	$\mathbf{51.37}_{\pm 0.10}$	**1**
Ave	$84.69_{\pm 0.05}$	218.08	$84.29_{\pm 0.05}$	34.25	$84.82_{\pm 0.05}$	39.08	$85.27_{\pm 0.05}$	40.17	$83.53_{\pm 0.05}$	**23.75**	$83.12_{\pm 0.05}$	37.50	$84.95_{\pm 0.05}$	42.58	$83.31_{\pm 0.06}$	**21.83**	$\mathbf{85.25}_{\pm 0.04}$	**21.83**

TABLE VII
CLASSIFICATION ACCURACY AND SELECTED FEATURE NUMBER OF COMPARED METHODS WITH THE DT CLASSIFIER (%)

No.s	Raw		VCOS		FSNC		MFEFS		FSZUM		FNCE		AGM-FS		SFSS		ZeMG-FS	
	Accuracy	Num	Accuracy	Num	Accuracy	Num	Accuracy	Num	Accuracy	Num	Accuracy	Num	Accuracy	Num	Accuracy	Num	Accuracy	Num
D1	$84.05_{\pm 0.02}$	20	$85.09_{\pm 0.02}$	20	$85.41_{\pm 0.03}$	20	$86.60_{\pm 0.03}$	9	$85.09_{\pm 0.03}$	20	$85.37_{\pm 0.03}$	20	$84.29_{\pm 0.02}$	19	$70.70_{\pm 0.04}$	**3**	$\mathbf{87.82}_{\pm 0.02}$	**3**
D2	$90.56_{\pm 0.03}$	17	$91.67_{\pm 0.05}$	17	$92.41_{\pm 0.04}$	17	$93.70_{\pm 0.03}$	**4**	$93.15_{\pm 0.05}$	5	$92.04_{\pm 0.03}$	17	$91.85_{\pm 0.03}$	6	$91.85_{\pm 0.04}$	6	$\mathbf{94.07}_{\pm 0.03}$	6
D3	$73.57_{\pm 0.21}$	2000	$80.71_{\pm 0.16}$	18	$83.57_{\pm 0.16}$	4	$\mathbf{84.05}_{\pm 0.18}$	16	$82.38_{\pm 0.12}$	3	$65.48_{\pm 0.19}$	1	$82.14_{\pm 0.17}$	83	$83.81_{\pm 0.11}$	**1**	$\mathbf{84.05}_{\pm 0.10}$	**1**
D4	$89.46_{\pm 0.06}$	33	$90.60_{\pm 0.05}$	28	$91.46_{\pm 0.05}$	**8**	$91.71_{\pm 0.05}$	16	$90.88_{\pm 0.03}$	30	$90.31_{\pm 0.03}$	31	$89.75_{\pm 0.08}$	33	$90.03_{\pm 0.05}$	11	$\mathbf{92.33}_{\pm 0.04}$	11
D5	$85.24_{\pm 0.02}$	68	$86.26_{\pm 0.03}$	**25**	$86.17_{\pm 0.04}$	66	$86.35_{\pm 0.04}$	65	$85.52_{\pm 0.05}$	45	$85.61_{\pm 0.05}$	64	$84.41_{\pm 0.04}$	65	$85.42_{\pm 0.04}$	60	$\mathbf{86.90}_{\pm 0.04}$	60
D6	$78.51_{\pm 0.10}$	166	$80.47_{\pm 0.05}$	59	$79.21_{\pm 0.06}$	11	$76.71_{\pm 0.08}$	107	$74.57_{\pm 0.08}$	7	$79.82_{\pm 0.04}$	77	$77.29_{\pm 0.04}$	122	$78.56_{\pm 0.09}$	41	$\mathbf{81.94}_{\pm 0.06}$	41
D7	$95.28_{\pm 0.01}$	19	$96.36_{\pm 0.01}$	11	$95.93_{\pm 0.02}$	15	$96.02_{\pm 0.02}$	17	$95.67_{\pm 0.02}$	19	$96.02_{\pm 0.01}$	16	$96.10_{\pm 0.01}$	19	$96.06_{\pm 0.01}$	**10**	$\mathbf{97.06}_{\pm 0.01}$	**10**
D8	$70.50_{\pm 0.05}$	20	$71.20_{\pm 0.04}$	20	$70.50_{\pm 0.05}$	20	$70.50_{\pm 0.05}$	20	$70.50_{\pm 0.05}$	20	$70.50_{\pm 0.05}$	20	$68.00_{\pm 0.07}$	15	$70.70_{\pm 0.04}$	**14**	$\mathbf{72.10}_{\pm 0.06}$	**14**
D9	$91.59_{\pm 0.01}$	57	$91.81_{\pm 0.01}$	53	$91.81_{\pm 0.01}$	56	$91.59_{\pm 0.01}$	57	$91.59_{\pm 0.01}$	57	$91.52_{\pm 0.01}$	56	$92.24_{\pm 0.01}$	56	$91.68_{\pm 0.01}$	**50**	$\mathbf{92.61}_{\pm 0.01}$	**50**
D10	$99.51_{\pm 0.00}$	21	$99.56_{\pm 0.00}$	20	$99.56_{\pm 0.00}$	20	$99.56_{\pm 0.00}$	20	$99.51_{\pm 0.00}$	21	$99.56_{\pm 0.00}$	20	$99.53_{\pm 0.00}$	19	$96.01_{\pm 0.01}$	**1**	$96.13_{\pm 0.01}$	**1**
D11	$73.31_{\pm 0.12}$	147	$80.29_{\pm 0.11}$	**10**	$77.54_{\pm 0.12}$	13	$77.39_{\pm 0.07}$	91	$76.73_{\pm 0.09}$	19	$76.76_{\pm 0.09}$	76	$79.19_{\pm 0.07}$	73	$77.39_{\pm 0.08}$	46	$\mathbf{80.40}_{\pm 0.07}$	46
D12	$44.48_{\pm 0.08}$	49	$\mathbf{48.06}_{\pm 0.07}$	49	$44.17_{\pm 0.08}$	23	$45.14_{\pm 0.07}$	28	$45.12_{\pm 0.06}$	37	$45.12_{\pm 0.05}$	**17**	$44.47_{\pm 0.05}$	1	$45.17_{\pm 0.07}$	**1**	$\mathbf{48.06}_{\pm 0.09}$	**1**
Ave	$81.34_{\pm 0.06}$	218.08	$83.51_{\pm 0.05}$	27.50	$83.14_{\pm 0.06}$	22.75	$83.28_{\pm 0.05}$	37.50	$82.56_{\pm 0.05}$	23.58	$81.51_{\pm 0.05}$	34.58	$82.44_{\pm 0.05}$	42.58	$81.45_{\pm 0.05}$	**20.33**	$\mathbf{84.46}_{\pm 0.04}$	**20.33**

It examines the average rank between all method pairs, rejecting the null hypothesis of equivalence if the absolute difference between the pairs exceeds the critical distance (CD) computed as follows:

$$CD = q_\alpha \sqrt{\frac{N_m (N_m + 1)}{6 N_d}} \tag{25}$$

where $q_\alpha = 2.855$ when $N_d = 12$ and $N_m = 9$ denote the dataset and method numbers.

The Nemenyi test results are illustrated in Fig. 7. The proposed ZeMG-FS method consistently ranks first across all classifiers. Notably, it shows statistical significance over 7, 4, 4, and 5 compared methods under the KNN, NB, SVM, and DT classifiers. These results demonstrate the overall superiority of ZeMG-FS in feature selection and classification tasks.

D. Parameter Sensitivity Analysis

As discussed in Section III, the purity threshold P and neighborhood radius δ are two critical parameters that directly affect the fuzzy rough approximation and uncertainty measure, thereby influencing the performance of the ZeMG-FS algorithm. To evaluate the sensitivity of the model to these parameters, experiments are conducted on four benchmark datasets using the KNN classifier as an example.

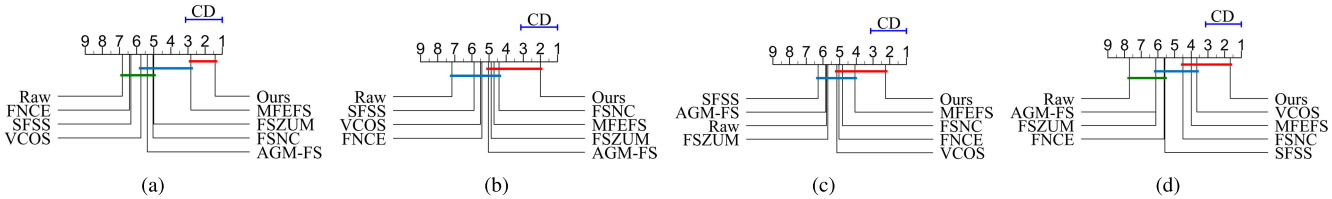

Fig. 7. Nemenyi post hoc test results of compared methods on different classifiers. (a) KNN. (b) NB. (c) SVM. (d) DT.

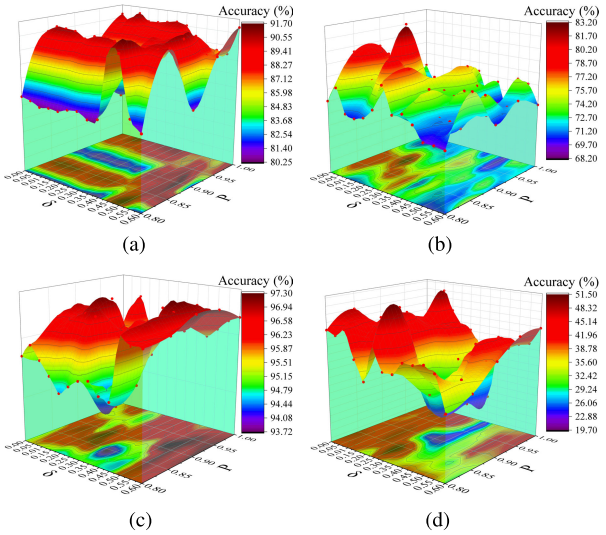

Fig. 8. Classification accuracy variation of ZeMG-FS on different parameters. (a) Card. (b) ULC. (c) Sege. (d) VGCD.

Fig. 8 illustrates the variation in classification accuracy at different values of purity P and radius δ. ZeMG-FS performance is notably affected by these parameters, particularly on the ULC and VGCD datasets, indicating its sensitivity to these parameters. Furthermore, the optimal parameter settings vary across datasets. Therefore, the appropriate selection of the P and δ values is essential for achieving optimal performance on different datasets.

VI. Conclusion

ZeMG-FS is a robust feature selection method for knowledge discovery from high-dimensional data. It adopts threefold ideas.

1) Generalized granular-ball generation for fast and adaptive information granulation.
2) Multigranularity fuzzy-rough knowledge modeling for accurate knowledge representation.
3) A multilevel zentropy structure to enhance model performance in robust feature selection. Theoretical and empirical studies show that ZeMG-FS achieves superior noise resistance and classification performance.

This study investigates a novel robust feature selection method by designing a zentropy-enhanced multigranularity knowledge modeling approach. Although the proposed ZeMG-FS method can significantly enhance the robustness and computational efficiency of feature selection, it remains limited to open and dynamic scenarios. Hence, exploring how to combine sustainable learning and evolutionary computing theory into robust knowledge modeling and feature selection deserves further investigation.

References

[1] C. Gao, J. Zhou, X. Wang, and W. Pedrycz, "Granule margin-based feature selection in weighted neighborhood systems," *IEEE Trans. Cybern.*, vol. 55, no. 5, pp. 2151–2164, May 2025.

[2] K. Yuan, D. Miao, W. Ding, W. Pedrycz, and Y. Yao, "Robust semi-supervised feature selection with multi-granularity zentropy modeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Dec. 24, 2025, doi: 10.1109/TPAMI.2025.3647921.

[3] W. Li, B. Yang, W. Pedrycz, C. Zhang, and T. Zhan, "Adaptive hyper-box granulation with justifiable granularity for feature selection," *IEEE Trans. Knowl. Data Eng.*, vol. 37, no. 12, pp. 6847–6862, Dec. 2025.

[4] D. Yang, Q. Zhang, G. Zong, H. Sun, and Y. Zhao, "Dynamic event-triggered model reference adaptive control for uncertain switched systems," *IEEE Trans. Cybern.*, vol. 56, no. 1, pp. 487–496, Jan. 2026.

[5] X. Liang, Y. Qian, Q. Guo, H. Cheng, and J. Liang, "AF: An association-based fusion method for multi-modal classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 12, pp. 9236–9254, Dec. 2022.

[6] S. Liu, J. Wang, G. Min, and J. Hu, "MG-WEP: Multi-granularity workload ensemble and variational inference for multivariate computing power prediction," *IEEE Trans. Netw. Sci. Eng.*, vol. 13, pp. 5471–5488, 2026.

[7] R. Sheikhpour, K. Berahmand, M. Mohammadi, and H. Khosravi, "Sparse feature selection using hypergraph Laplacian-based semi-supervised discriminant analysis," *Pattern Recognit.*, vol. 157, Jan. 2025, Art. no. 110882.

[8] D. Guo et al., "Concept-cognitive learning survey: Mining and fusing knowledge from data," *Inf. Fusion*, vol. 109, Sep. 2024, Art. no. 102426.

[9] M. Hu, E. C. C. Tsang, Y. Guo, and W. Xu, "Fast and robust attribute reduction based on the separability in fuzzy decision systems," *IEEE Trans. Cybern.*, vol. 52, no. 6, pp. 5559–5572, Jun. 2022.

[10] W. Li et al., "Granular-ball regeneration clustering with principle of justifiable granularity," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 10, pp. 18173–18187, Oct. 2025.

[11] K. Yuan, D. Miao, W. Pedrycz, H. Zhang, and L. Hu, "Multigranularity data analysis with zentropy uncertainty measure for efficient and robust feature selection," *IEEE Trans. Cybern.*, vol. 55, no. 2, pp. 740–752, Feb. 2025.

[12] D. Guo, W. Xu, Y. Qian, and W. Ding, "Fuzzy-granular concept-cognitive learning via three-way decision: Performance evaluation on dynamic knowledge discovery," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 3, pp. 1409–1423, Mar. 2024.

[13] P. Liang, D. Lei, K. Chin, and J. Hu, "Feature selection based on robust fuzzy rough sets using kernel-based similarity and relative classification uncertainty measures," *Knowledge-Based Syst.*, vol. 255, Nov. 2022, Art. no. 109795.

[14] C. Wang, Y. Qian, W. Ding, and X. Fan, "Feature selection with fuzzy-rough minimum classification error criterion," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 8, pp. 2930–2942, Aug. 2022.

[15] A. Theerens and C. Cornelis, "Fuzzy rough sets based on fuzzy quantification," *Fuzzy Sets Syst.*, vol. 473, Dec. 2023, Art. no. 108704.

[16] M. Cai and Z. Xiang, "Adaptive fuzzy finite-time control for a class of switched nonlinear systems with unknown control coefficients," *Neurocomputing*, vol. 162, pp. 105–115, Aug. 2015.

[17] D. Guo, W. Xu, Y. Qian, and W. Ding, "M-FCCL: Memory-based concept-cognitive learning for dynamic fuzzy data classification and knowledge fusion," *Inf. Fusion*, vol. 100, Dec. 2023, Art. no. 101962.

[18] C. Wang, M. Shao, Q. He, Y. Qian, and Y. Qi, "Feature subset selection based on fuzzy neighborhood rough sets," *Knowledge-Based Syst.*, vol. 111, pp. 173–179, Nov. 2016.

[19] K. Yuan, D. Miao, H. Zhang, and W. Pedrycz, "An efficient and robust feature selection approach based on zentropy measure and neighborhood-aware model," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 9, pp. 16351–16365, Sep. 2025.

[20] W. Ding et al., "A novel spark-based attribute reduction and neighborhood classification for rough evidence," *IEEE Trans. Cybern.*, vol. 54, no. 3, pp. 1470–1483, Mar. 2024.
[21] A. Kumar and 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.*, vol. 33, no. 2, pp. 669–679, Feb. 2025.
[22] S. Xia, X. Lian, G. Wang, X. Gao, J. Chen, and X. Peng, "GBSVM: An efficient and robust support vector machine framework via granular-ball computing," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 5, pp. 9253–9267, May 2025.
[23] X. Zhang, C. Mei, J. Li, Y. Yang, and T. Qian, "Instance and feature selection using fuzzy rough sets: A bi-selection approach for data reduction," *IEEE Trans. Fuzzy Syst.*, vol. 31, no. 6, pp. 1981–1994, Jun. 2023.
[24] T. Yin et al., "A robust multilabel feature selection approach based on graph structure considering fuzzy dependency and feature interaction," *IEEE Trans. Fuzzy Syst.*, vol. 31, no. 12, pp. 4516–4528, Dec. 2023.
[25] X. Zou and J. Dai, "Unified feature selection approach for complex data based on fuzzy β-covering reduction via information granulation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 9, pp. 17380–17394, Sep. 2025.
[26] K. Yuan, D. Miao, W. Pedrycz, W. Ding, and H. Zhang, "Ze-HFS: Zentropy-based uncertainty measure for heterogeneous feature selection and knowledge discovery," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 11, pp. 7326–7339, Nov. 2024.
[27] B. Sang, L. Yang, W. Xu, H. Chen, T. Li, and W. Li, "VCOS: Multi-scale information fusion to feature selection using fuzzy rough combination entropy," *Inf. Fusion*, vol. 117, May 2025, Art. no. 102901.
[28] K. Yuan, D. Miao, Y. Yao, H. Zhang, and X. Zhao, "Feature selection using zentropy-based uncertainty measure," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 4, pp. 2246–2260, Apr. 2024.
[29] Z.-K. Liu, S.-L. Shang, J. Du, and Y. Wang, "Parameter-free prediction of phase transition in $PbTiO_3$ through combination of quantum mechanics and statistical mechanics," *Scripta Mater.*, vol. 232, Jul. 2023, Art. no. 115480.
[30] Y. Yao, "Combination of rough and fuzzy sets based on α-level sets," in *Rough Sets and Data Mining: Analysis of Imprecise Data*. Boston, MA, USA: Springer, 1997, pp. 301–321.
[31] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. A. Raffel, "MixMatch: A holistic approach to semi-supervised learning," in *Proc. 33rd Int. Conf. Neural Inf. Process. Syst.*, 2019, pp. 5049–5059.
[32] X. Yang, H. Chen, T. Li, and C. Luo, "A noise-aware fuzzy rough set approach for feature selection," *Knowl.-Based Syst.*, vol. 250, Aug. 2022, Art. no. 109092.
[33] P. Zhang, T. Li, Z. Yuan, C. Luo, K. Liu, and X. Yang, "Heterogeneous feature selection based on neighborhood combination entropy," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 3, pp. 3514–3527, Mar. 2024.
[34] N. N. Thuy and S. Wongthanavasu, "A novel feature selection method for high-dimensional mixed decision tables," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 7, pp. 3024–3037, Jul. 2022.
[35] Z. Wang, H. Chen, Z. Yuan, J. Wan, and T. Li, "Multiscale fuzzy entropy-based feature selection," *IEEE Trans. Fuzzy Syst.*, vol. 31, no. 9, pp. 3248–3262, Sep. 2023.
[36] J. Demsar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, no. 1, pp. 1–30, 2006.

Kehua Yuan received the M.Sc. degree from the College of Artificial Intelligence, Southwest University, Chongqing, China, in 2022. She is currently pursuing the Ph.D. degree with the School of Computer Science and Technology, Tongji University, Shanghai, China.

She has widely published on these subjects at high-level international journals, such as IEEE TRANSACTIONS ON CYBERNETICS, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON FUZZY SYSTEMS, and IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS. Her current research interests include granular computing, uncertainty analysis, and machine learning.

Duoqian Miao is a Professor with the School of Computer Science and Technology at Tongji University, Shanghai, China. He has published more than 200 academic papers, which include prestigious journals and international conferences. His main research interests include rough sets, machine learning, and intelligent systems.

Mr. Miao is a fellow of the International Rough Set Society and the Director of Chinese Association for Artificial Intelligence (CAAI). He is currently an Associate Editor of the *Information Sciences*, the *CAAI Transactions on Intelligence Technology*, and the *International Journal of Approximate Reasoning*.

Witold Pedrycz (Life Fellow, IEEE) received the M.Sc., Ph.D., and D.Sc. degrees from the Silesian University of Technology, Gliwice, Poland, in 1977, 1980, and 1984, respectively.

He is a Professor with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada. He is also with the Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland, where he was elected as a Foreign Member in 2009. He also holds an appointment as a Special Professor with the School of Computer Science, University of Nottingham, Nottingham, U.K. His current research interests include computational intelligence, fuzzy modeling and granular computing, knowledge discovery and data mining, fuzzy control, pattern recognition, knowledge-based neural networks, relational computing, and software engineering.

Dr. Pedrycz was elected as a Foreign Member of the Polish Academy of Sciences in 2009. In 2012, he was elected as a fellow of the Royal Society of Canada. He was a recipient of the prestigious Norbert Wiener Award from the IEEE Systems, Man, and Cybernetics Society in 2007, the IEEE Canada Computer Engineering Medal, the Cajastur Prize for Soft Computing from the European Centre for Soft Computing, the Killam Prize, and the Fuzzy Pioneer Award from the IEEE Computational Intelligence Society. He is also the Editor-in-Chief of the *Information Sciences*, the *WIREs Data Mining and Knowledge Discovery*, and the *Granular Computing*.

Yiyu Yao received the B.E. degree in computer science from Xi'an Jiaotong University, Xi'an, China, in 1983, and the M.Sc. and Ph.D. degrees in computer science from the University of Regina, Regina, SK, Canada, in 1988 and 1991, respectively.

He is a Professor of computer science with the University of Regina. He proposed a theory of three-way decision, a decision-theoretic rough set model, and a triarchic theory of granular computing. He has published more than 400 articles. His research interests include three-way decision, granular computing, rough sets, formal concept analysis, information retrieval, data mining, and Web intelligence.

Dr. Yao was a Highly Cited Researcher from 2015 to 2019.