Consistency-Enhanced Multi-Granularity Knowledge Modeling for Semi-Supervised Feature Selection

Kehua Yuan, Duoqian Miao, Weiping Ding, *Senior Member, IEEE*, Witold Pedrycz, *Life Fellow, IEEE*, Hongyun Zhang, and Yiyu Yao

***Abstract*—Optimizing the feature selection process by combining labeled data with unlabeled data is a core topic of semi-supervised feature selection (SFS). However, a handful of labeled data and an abundance of unlabeled data are bound to affect consistency among data, knowledge, and models. The central challenge of SFS still lies in robustness and uncertainty in data modeling and knowledge discovery. Accordingly, SFS needs to be concerned with three problems: 1) how to represent the knowledge structure of semi-supervised data, 2) how to enhance the robustness of knowledge modeling, and 3) how to effectively reason and measure uncertainty in modeling processing. To accomplish these, this study proposes a consistency-enhanced multi-granularity knowledge modeling (Ce-MGK) framework for semi-supervised feature selection. Specifically, an affinity-separation-aware soft learning mechanism with consistent object selection is established, which can effectively enhance the performance of subsequent multi-granularity knowledge acquisition based on fuzzy rough approximation modeling. Furthermore, multi-granularity structures and inter-level interactions are discussed for multi-level uncertainty analysis, i.e., zentropy uncertainty measure. Unlike existing measures, zentropy fully utilizes inter-level relationships among multi-granularity knowledge structures, which can effectively exploit hidden information in semi-supervised data. Finally, two inner and outer significance measures are designed for feature evaluation, and the optimal feature subset is determined based on consistent information gain. Extensive experiments on multiple benchmark datasets demonstrate the superior robustness and effectiveness of the Ce-MGK framework compared with state-of-the-art methods.**

Received 7 October 2025; revised 23 June 2026; accepted 21 July 2026. Date of publication 31 July 2026; date of current version 15 September 2026. This work was supported in part by the National Natural Science Foundation of China under Grant 624B2104, Grant 62576251, Grant 62376198, and Grant 62576247 and in part by the National Key Research and Development Program of China "Key Special Project on Cyberspace Security Governance" under Grant 2022YFB3104700. Recommended for acceptance by W. Zhao. *(Corresponding author: Duoqian Miao.)*

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

Weiping Ding is with the School of Artificial Intelligence and Computer Science, Nantong University, Nantong 226019, China (e-mail: dwp9988@163.com).

Witold Pedrycz is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2R3, Canada, also with the Institute of Systems Engineering, Macau University of Science and Technology, Taipa 999078, China, and also with the Systems Research Institute, Research Center of Performance and Productivity Analysis, Istinye University, Istanbul 34010, 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).

This article has supplementary downloadable material available at https://doi.org/10.1109/TKDE.2026.3719432, provided by the authors.

Digital Object Identifier 10.1109/TKDE.2026.3719432

***Index Terms*—Data mining, semi-supervised feature selection, granular computing, fuzzy rough set, uncertainty measure.**

I. Introduction

RECENTLY, volumes of data can be widely collected from diverse sources such as web crawlers, sensor networks, and database systems [1], [2]. However, annotating such data is usually time-consuming and costly, which results in a scarcity of labeled samples in practical applications [3], [4], [5]. When only a limited number of labeled samples are available, many classical supervised learning algorithms fail to effectively exploit the abundant unlabeled data, thereby resulting in unsatisfactory performance [6], [7], [8]. Consequently, various semi-supervised learning (SSL) techniques have been proposed to exploit the latent information and knowledge embedded in unlabeled data [9], [10], [11]. In particular, semi-supervised feature selection has been introduced to enhance the effectiveness of SSL by reducing redundant and noisy features in high-dimensional data scenarios [12], [13]. Over the past few years, SFS has witnessed rapid development. Nevertheless, two thorny issues remain: 1) **limited robustness in knowledge modeling**: reliance on abundant unlabeled data may compromise quality and stability, especially under noise or uncertain objects scenarios; and 2) **incomplete characterization of uncertainty**, incorporating unlabeled data increases modeling complexity and makes it challenging to capture the uncertainty inherent in data and models accurately.

As for the robust modeling issue, extensive efforts have been devoted to enhancing the robustness of semi-supervised feature selection by leveraging unlabeled data more effectively [14], [15], [16], [17]. For instance, Sheikhpour et al. [14] proposed a Hessian-based method that maintains local topology under a generalized uncorrelated constraint. To balance the relevance and redundancy of partially labeled data, Xu et al. [17] introduced a criterion balancing maximum relevance and minimum redundancy. From a representation learning perspective, Sheikhpour et al. [18] developed a semi-supervised feature selection method by integrating concept factorization with a linear regression loss. Tan et al. [19] further proposed an enhanced semi-supervised feature selection with condensed graph representation. In addition, the deep learning approaches, such as transductive semi-supervised deep hashing [20], self-supervised graph auto-encoder-based learning [21], and a self-adaptive weight redistribution strategy for sample and feature learning [22], are designed to leverage unlabeled data. Although the above methods improve stability, they generally rely on a single

1041-4347 © 2026 IEEE. All rights reserved, including rights for text and data mining, and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: TONGJI UNIVERSITY. Downloaded on November 11,2026 at 13:57:24 UTC from IEEE Xplore. Restrictions apply.

granularity of structural description or have limited interpretability due to the complex network structures, while ignoring the potential multi-granularity structure among features [23], [24]. In fact, real-world data often exhibit inherent multi-granularity structures, including but not limited to local object relationships and global class distributions. This single granularity limitation restricts their ability to comprehensively characterize data with inherently multi-level structures. In contrast, multi-granularity knowledge modeling, including rough approximation [10], [25], [26], three-way decisions [27], [28], concept-cognitive learning [29], [30], hierarchical clustering [31], [32], [33], offers a more flexible and interpretable paradigm for capturing knowledge at different levels. Particularly, fuzzy rough approximation has proven effective in modeling both fuzzy and rough aspects of data from multiple granular perspectives [34], [35], [36]. Nevertheless, existing methods typically represent knowledge at a single granularity and rely heavily on boundary sample modeling. Such modeling is sensitive to uncertainty and ambiguity, especially in the presence of noisy boundary samples. Therefore, fully leveraging unlabeled data to achieve robust multi-granularity knowledge modeling and acquisition is the first motivation of this study.

As for the uncertainty phenomenon, various uncertainty-based methods have been developed to strengthen the performance of semi-supervised models [37], [38], [39], [40], [41], [42], [43], [44]. For instance, Zhou et al. [38] proposed an information fusion strategy based on optimal scales, while An et al. [41] employed lower approximation and classification loss to mitigate the noisy labels' influence. To further improve computational efficiency, Yang et al. [39] designed a local selection algorithm based on approximation sets. Moreover, from the information theory perspective, Liu et al. [42] introduced fuzzy relevance and redundancy for semi-supervised feature selection, and Madhu et al. integrated cross-entropy with fuzzy graphs to formulate label propagation [43]. These methods typically represent knowledge at a single granularity formulation from algebraic or information theory, failing to capture the hierarchical structures and interactions in the modeling process. Although the zentropy-based measures have been introduced to model system uncertainty by integrating multi-granularity knowledge, they primarily depend on granularity structures induced by fully supervised scenarios and are limited to high-quality data [34], [37]. Therefore, in semi-supervised scenarios, how to comprehensively characterize modeling uncertainty across multiple granularities to further enhance model performance is another motivation of this study.

According to the above analysis, two issues of multi-granularity modeling in SFS can be obtained: 1) **effective multi-granularity knowledge characterization**, and 2) **accurate multi-level uncertainty analysis**. Inspired by these, this article develops a consistency-enhanced multi-granularity knowledge modeling framework for semi-supervised feature selection. Rather than relying on learned latent representations, Ce-MGK focuses on explicit multi-granularity knowledge modeling and multi-level uncertainty analysis, which provides stronger interpretability and is particularly suitable for scenarios with complex interactions and uncertainty. Specifically, the main contributions of this paper are as follows.

- It introduces a consistency-enhanced knowledge strategy that explicitly accounts for distributional differences, effectively mitigating inconsistency and improving the robustness of semi-supervised knowledge modeling.
- It develops a multi-granularity knowledge modeling framework based on fuzzy rough approximation, which enables more effective acquisition and representation of uncertainty in semi-supervised data.
- It proposes a multi-level zentropy measure that integrates granularity structures and interactions. Compared with existing approaches, it could better exploit unlabeled data and reduce the adverse effects of uncertainty on learning.
- It designs a semi-supervised feature selection method via consistency-enhanced multi-granularity knowledge. Extensive experiments validate its robustness and effectiveness compared with other state-of-the-art methods.

The remainder of this article is organized as follows. Section II provides the basic concepts. Section III introduces the proposed consistency-driven multi-granularity knowledge modeling. Section IV presents the proposed SFS method with multi-level zentropy measure. Section V reports the experimental results, and Section VI concludes the paper.

II. Preliminaries

This section provides a brief overview of some basic notions of this study, with more details available in [2], [34], [45].

A. Problem Definition

A dataset typically comprises three components: objects, features, and decisions. It can be formally represented as a triplet $\mathbb{DS} = (O, F, D)$, where $O = \{o_1, o_2, \ldots, o_w\}$ denotes a finite set of objects, $F = \{f_1, f_2, \ldots, f_m\}$ represents the set of conditional features, and $D = \{d\}$ corresponds to the decision feature. The decision partition induced by D is denoted as $O/D = \{D_1, D_2, \ldots, D_k\}$. In information theory, such a structure is referred to as a decision system. In practical scenarios, it is common for some objects to lack decision labels. When this occurs, the system is termed a partially labeled decision system, denoted as $\mathbb{PLDS} = (O, F, D)$, which satisfies the following conditions:

- $O = O^L \cup O^U$, where $O^L = \{o_1, o_2, \ldots, o_{w_1}\}$ is the labeled objects, $O^U = \{o_1, o_2, \ldots, o_{w_2}\}$ is the unlabeled objects, with $w = w_1 + w_2$;
- $F = \{f_1, f_2, \ldots, f_m\}$ denotes the conditional feature set;
- $D = D^L \cup D^U$, where D^L corresponds to the decision values of the labeled object set O^L, and D^U denotes the unlabeled decision values, remarked by $*$, corresponding to the objects in O^U.

Definition 1: [2] Given a decision system $\mathbb{DS} = (O, F, D)$, for $R^* \subseteq F$, R^* is defined as the optimal feature subset of F if it satisfies the following condition:

$$R^* = \arg\min_{R \subseteq F} \mathcal{J}(R), \tag{1}$$

Authorized licensed use limited to: TONGJI UNIVERSITY. Downloaded on November 11,2026 at 13:57:24 UTC from IEEE Xplore. Restrictions apply.

where $\mathcal{J}$ denotes the entropy-based uncertainty measure.

Designing a reasonable and effective uncertainty measure $\mathcal{J}$ that aligns with practical application scenarios remains a key challenge and focus in feature selection research.

B. Multi-Granularity Knowledge Modeling and Analysis

This subsection introduces the framework of multi-granularity knowledge modeling and its corresponding uncertainty analysis, where the representation based on fuzzy rough approximation is formally defined as follows.

Definition 2: [34] Let $\mathbb{DS} = (O, F, D)$ be a decision system, where $O/D = \{D_1, D_2, \ldots, D_s\}$. For $\forall S \subseteq F$ and $D_t \in O/D$, the fuzzy lower and uper approximations of D_t on S are described as

$$\underline{R_S}(D_t)(o) = inf_{p\in O} max\{1 - R_S(o,p), D_t(p)\},$$
$$\overline{R_S}(D_t)(o) = sup_{p\in O} min\{R_S(o,p), D_t(p)\}. \quad (2)$$

where R_S is the fuzzy relation induced by S. When R_S is a fuzzy similarity relation constructed by introducing a neighborhood radius, the corresponding model is referred to as a fuzzy neighborhood rough set.

To further quantify the certainty of the fuzzy rough set (FRS) model induced by the fuzzy relation R_S, the fuzzy dependency degree of D on S is defined as follows.

$$\gamma_S(D) = \frac{|POS_S(D)|}{|O|}, \quad (3)$$

where $POS_S(D) = \cup_{D_t/D} \underline{R_S}(D_t)$ denotes the fuzzy positive region of D with respect to S, and $|\cdot|$ represents the cardinality.

Moreover, from an information perspective, entropy-based measures, such as the representative fuzzy conditional entropy (FCE), have been proposed to evaluate certainty.

Definition 3: [45] Let $\mathbb{DS} = (O, F, D)$ be a decision system, where $O/D = \{D_1, D_2, \ldots, D_s\}$. For $\forall S \subseteq F$, the fuzzy conditional entropy of D related to S is described as follows.

$$\text{FCE}(D|S) = -\frac{1}{|O|}\sum_{j=1}^{w} \log\left(\frac{|[o_j]_{R_S} \cap D(o_j)|}{|[o_j]_{R_S}|}\right), \quad (4)$$

where $[o_j]_{R_S}$ denotes the fuzzy information granule of object o_j induced by the fuzzy relation R_S, and $D(o_j)$ represents the decision class that contains o_j.

The above entropy measure captures the amount of common information between conditional features and decisions within the fuzzy approximation space. However, similar to the fuzzy dependency degree in Eq. (3), FCE only provides a single level characterization based on object similarity, while ignoring the multi-granularity level structure and interaction relations in the fuzzy approximation process. These single methods inevitably lead to incomplete or imprecise uncertainty measures, especially in partially labeled scenarios.

III. The Proposed Multi-Granularity Knowledge Modeling

Considering the limited robustness of knowledge modeling approaches in SFS, this section proposes a consistency-driven multi-granularity knowledge acquisition method for partially labeled decision systems. Compared with existing methods, the proposed approach significantly enhances the model's effectiveness and robustness by introducing a strategic and consistent mechanism for object selection and learning.

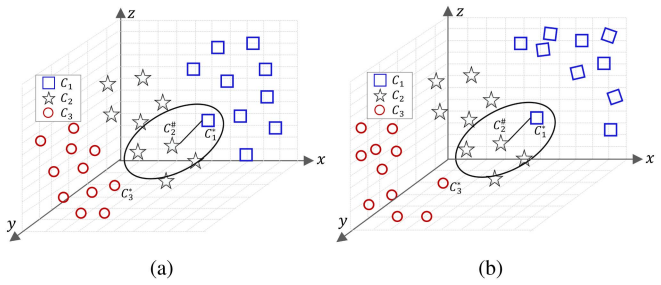


Fig. 1. The fuzzy rough approximation with uniform distribution in (a) and inconsistent object scenario in (b).

A. Motivations

According to the fuzzy rough approximation process described in Def. 2, for a given object $c_2^{\#}$, its fuzzy lower approximation to the target decision D_t is determined by the nearest objects outside D_t. Consequently, the lower approximation is highly sensitive to boundary objects with different labels, making it particularly vulnerable to noisy samples (see Fig. 1). Existing extensions, such as fuzzy neighborhood rough set models with neighborhood radii, enhance robustness to some extent but still struggle to mitigate the intrinsic effects of noise and class inconsistencies, especially in partially labeled scenarios. To address this challenge, this study proposes a consistency-driven multi-granularity knowledge acquisition method that improves both the accuracy and robustness of semi-supervised learning on uncertain objects.

From a systematic perspective, uncertainty in decision systems arises at multiple levels, including target decisions, fuzzy approximations, and specific object granules. This naturally reveals a multi-granularity and interactive structure within the fuzzy rough approximation process. Hence, the single granularity measures, such as the fuzzy positive region and fuzzy conditional entropy, are limited in capturing this complexity. Although fuzzy granule zentropy has been introduced to integrate the approximation and object granules, it fails to adequately capture inter-level interactions and remains constrained to fully supervised settings, which limits its applicability to real-world semi-supervised tasks. Therefore, it is essential to develop systematic uncertainty measures within multi-granularity knowledge structures to more effectively quantify and exploit uncertainty in semi-supervised learning.

B. Affinity-Separation-Aware Soft Learning

To improve classification accuracy and stability in the presence of uncertain samples caused by incomplete supervision,

Authorized licensed use limited to: TONGJI UNIVERSITY. Downloaded on November 11,2026 at 13:57:24 UTC from IEEE Xplore. Restrictions apply.

this subsection proposes an affinity-separation-aware soft learning(ASL) method that includes two important steps: 1) a consistent object selection strategy for determining those unlabeled objects compatible with the labeled data, and 2) a soft label learning method for processing unlabeled objects.

The consistent object selection is designed from a statistical perspective, where consistent objects are defined as those lying within the statistical region of at least one decision class.

Definition 4: Given a $\mathbb{PLDS} = (O, F, D)$, where $O = O^L \cup O^U$, $D = D^L \cup D^U$ and $O^L/D^L = \{D_1, D_2, \ldots, D_s\}$. Then, the consistent objects O^C are obtained as follows.

$$O^C = \{x \in O^U | Dis_F(x, \mathbf{C}_{Fi}) \leq \boldsymbol{\sigma}_{Fi}, \exists D_i \in O^L/D^L\}, \tag{5}$$

where $Dis_F(\cdot,\cdot)$ denotes the Euclidean distance in the feature space F, and $\mathbf{C}_{Fi}$ and σ_{Fi} represent the center and standard deviation of decision class D_i under F, respectively.

By restricting the learning process to such statistically consistent samples, the proposed mechanism reduces the influence of distant or ambiguous boundary samples, which tend to introduce high uncertainty. This facilitates the effective utilization of limited label information and provides a more reliable basis for subsequent soft label learning. The final set of consistent objects is defined as $O' = O^L \cup O^C$.

Proposition 1. Given a $\mathbb{PLDS} = (O, F, D)$, where $O = O^L \cup O^U$ and $D = D^L \cup D^U$. For $D_i \in O^L/D^L$, the $\mathbf{C}_{Fi}$ and σ^2_{Fi} are the class center and standard deviation. The selection rule $Dis_F(x, \mathbf{C}_{Fi}) \leq \boldsymbol{\sigma}_{Fi}$ is class-wise adaptive in the sense that the selection radius is determined by the local dispersion of class D_i, rather than by a shared global threshold.

Proof: According to Definition 4, σ_{Fi} is derived locally from the labeled samples of each class D_i. In imbalanced situations, the selection volume $V_i \propto \sigma^d_{Fi}$ varies the specific dispersion of each category rather than a global threshold, mitigating bias caused by class imbalance or intra-class variance.

Definition 5: Given a $\mathbb{PLDS} = (O', F, D)$, where $O' = O^L \cup O^C$ and $O^L/D^L = \{D_1, D_2, \ldots, D_s\}$. For $o_i \in O^C$ and $D_p \in O^L/D^L$, the classify affinity of object o_i to D_p is cacluated as follows.

$$\mathbb{M}(o_i, D_p) = \frac{\sum_{o_j \in D_p} e^{-\frac{Dis_F(o_i, o_j)}{1+\boldsymbol{\sigma}_{Fj}}}}{|D_j|}, \tag{6}$$

where $Dis_F(\cdot,\cdot)$ is consistent with the distance metric used in Eq. (5). The complete classification affinity vector of o_i with respect to all decision classes is then

$$\mathbb{M}(o_i) = \{\mathbb{M}(o_i, D_1), \mathbb{M}(o_i, D_2), \ldots, \mathbb{M}(o_i, D_s)\}. \tag{7}$$

This definition characterizes the classification affinity of each object based on its similarity to labeled instances in each decision class, taking into account both local density and dispersion. Moreover, to capture inter-class differences, inter-class separation is incorporated into the membership degree computation, which is defined as follows.

$$\mathbf{M}(o_i, D_p) = \frac{\mathbb{M}(o_i, D_p)}{\sum_{D_q \neq D_p} \mathbb{M}(o_i, D_q)}. \tag{8}$$

Finally, the normalized soft label of object o_i with respect to decision class D_p is computed as

$$\tilde{\mathbf{M}}(o_i, D_p) = \frac{\mathbf{M}(o_i, D_p)}{\sum_{D_q \in O^L/D^L} \mathbf{M}(o_i, D_q)}. \tag{9}$$

Then, the soft label of object $o_i \in O^C$ is obtained as $\tilde{\mathbf{M}}(o_i) = \{\tilde{\mathbf{M}}(o_i, D_1), \tilde{\mathbf{M}}(o_i, D_2), \ldots, \tilde{\mathbf{M}}(o_i, D_s)\}$. This designed soft label learning method effectively improves the classification results of unlabeled samples by capturing both the intra-class affinity and inter-class separation. For notational consistency, the soft labels for labeled objects $o_i \in O^L$ are also defined as $\tilde{\mathbf{M}}(o_i, D_p) = 1$ if $D^L(o_i) = D_p$, and $\tilde{\mathbf{M}}(o_i, D_p) = 0$ for all $D_p \neq D^L(o_i)$.

C. Consistency-Driven Knowledge Acquisition

Based on the ASL method, this subsection proposes a consistency-driven multi-granularity knowledge acquisition method under fuzzy rough approximation. Given a partially labeled decision system $\mathbb{PLDS} = (O', F, D)$, where $O' = O^L \cup O^C$ and $O^L/D^L = \{D_1, D_2, \ldots, D_s\}$. For any $o_i, o_j \in O'$, their fuzzy binary relation on the feature set $S \subseteq F$ is defined as follows.

$$\mathrm{R}^{\lambda}_S(o_i, o_j) = \begin{cases} e^{-\frac{Dis_S(o_i,o_j)}{\delta^2}}, & \text{if } e^{-\frac{Dis_S(o_i,o_j)}{\delta^2}} \geq \lambda, \\ 0, & \text{otherwise}, \end{cases} \tag{10}$$

where δ^2 is a tuning parameter set to the 20-th percentile of distances between o_i and all others according to [46]. Then, the fuzzy object granules of o_i is represented as follows.

$$[o_i]^{\lambda}_{\mathrm{R}_S} = \frac{\mathrm{R}^{\lambda}_S(o_i, o_1)}{o_1} + \frac{\mathrm{R}^{\lambda}_S(o_i, o_2)}{o_2} + \cdots + \frac{\mathrm{R}^{\lambda}_S(o_i, o_{|O'|})}{o_{|O'|}}, \tag{11}$$

where $\lambda \in [0, 1]$ is a neighborhood parameter that directly influences the size of fuzzy granules and approximation process.

Definition 6: Given a $\mathbb{PLDS} = (O', F, D)$, where $O' = O^L \cup O^C$ and $O^L/D^L = \{D_1, D_2, \ldots, D_s\}$. For $D_p \in O^L/D^L$, $S \subseteq F$ and $0 \leq \lambda \leq 1$, let R^{λ}_S be the corresponding fuzzy relation. Then, the consistency-driven fuzzy lower and upper approximations of D_p on S are calculated as follows.

$$\begin{aligned} \underline{\mathrm{R}^{\lambda}_S}(D_p)(o_i) &= \inf_{o_j \in O'} \max\left\{1 - \mathrm{R}^{\lambda}_S(o_i, o_j), \tilde{\mathbf{M}}(o_j, D_p)\right\}, \\ \overline{\mathrm{R}^{\lambda}_S}(D_p)(o_i) &= \sup_{o_j \in O'} \min\left\{\mathrm{R}^{\lambda}_S(o_i, o_j), \tilde{\mathbf{M}}(o_j, D_p)\right\}. \end{aligned} \tag{12}$$

where $\tilde{\mathbf{M}}(o_j, D_p)$ denotes the soft membership degree of o_j belonging to D_p, as defined in Section III-B.

The consistency-driven approximation process is shown in Fig. 2. Compared with existing FRS methods, it effectively reduces the ambiguous and uncertain samples during knowledge construction by incorporating an affinity-separation-aware object selection mechanism. Accordingly, the following probabilities are obtained.

Property 1. Given a $\mathbb{PLDS} = (O', F, D)$. For $D_p \in O^L/D^L$, $S \subseteq F$ and λ, $\underline{\mathrm{R}^{\lambda}_S}(D_p)$ and $\overline{\mathrm{R}^{\lambda}_S}(D_p)$ are the

Authorized licensed use limited to: TONGJI UNIVERSITY. Downloaded on November 11,2026 at 13:57:24 UTC from IEEE Xplore. Restrictions apply.

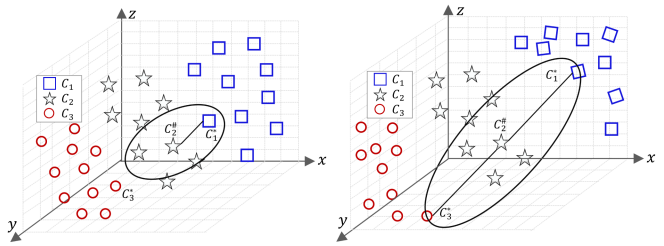


Fig. 2. The fuzzy rough approximation with uniform distribution in (a), while (b) could flexibly approximate targets with consistent information.

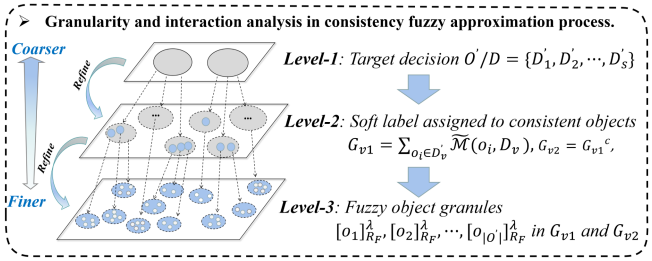


Fig. 3. Granularity and interaction analysis in fuzzy approximation process.

consistence-driven fuzzy approxiamtions of D_p on S. The following properties are obtained.

1) *Threshold monotonicity:* If $0 \le \lambda_1 \le \lambda_2 \le 1$, we have $\underline{\mathrm{R}}_S^{\lambda_1}(D_p) \subseteq \underline{\mathrm{R}}_S^{\lambda_2}(D_p)$ and $\overline{\mathrm{R}}_F^{\lambda_1}(D_p) \supseteq \overline{\mathrm{R}}_F^{\lambda_2}(D_p)$;
2) *Attribute monotonicity:* If $S_1 \subseteq S_2 \subseteq S$, we have $\underline{\mathrm{R}}_{S_1}^{\lambda}(D_p) \supseteq \underline{\mathrm{R}}_{S_2}^{\lambda}(D_p)$ and $\overline{\mathrm{R}}_{S_1}^{\lambda}(D_p) \subseteq \overline{\mathrm{R}}_{S_2}^{\lambda}(D_p)$;
3) *Granule inclusion:* For $\forall o \in O'$, if $[o]_{\mathrm{W}}^{\lambda} \subseteq [o]_{\mathrm{S}}^{\lambda}$, we have $\underline{\mathrm{R}}_W^{\lambda}(D_p) \supseteq \underline{\mathrm{R}}_S^{\lambda}(D_p)$ and $\overline{\mathrm{R}}_W^{\lambda}(D_p) \subseteq \overline{\mathrm{R}}_S^{\lambda}(D_p)$.

Proof: These properties directly follow from Definition 6 and the monotonicity of fuzzy relations and object granules.

1) From Eq. (10), when $\lambda_1 \le \lambda_2$, we have $\mathrm{R}_S^{\lambda_1} \supseteq \mathrm{R}_S^{\lambda_2}$. Then, it directly obtained $\underline{\mathrm{R}}_S^{\lambda_1}(D_p) \subseteq \underline{\mathrm{R}}_S^{\lambda_2}(D_p)$ and $\overline{\mathrm{R}}_S^{\lambda_1}(D_p) \supseteq \overline{\mathrm{R}}_S^{\lambda_2}(D_p)$ according to the fuzzy approximation in Definition 6.
2) When $S_1 \subseteq S_2 \subseteq S$, the fuzzy relation satisfies $\mathrm{R}_{S_1}^{\lambda} \subseteq \mathrm{R}_{s_2}^{\lambda}$, and the fuzzy approximations satisfy $\underline{\mathrm{R}}_{S_1}^{\lambda}(D_p) \supseteq \underline{\mathrm{R}}_{S_2}^{\lambda}(D_p)$ and $\overline{\mathrm{R}}_{S_1}^{\lambda}(D_p) \subseteq \overline{\mathrm{R}}_{S_2}^{\lambda}(D_p)$.
3) Similarly to 2), if the object granules satisfy $[o]_{\mathrm{W}}^{\lambda} \subseteq [o]_{\mathrm{S}}^{\lambda}$, the inclusion relation between fuzzy approximations hold.

D. Uncertainty Analysis and Measure

Based on the above consistency-driven fuzzy approximation process, the corresponding fuzzy dependency degree and fuzzy conditional entropy measures are substantially defined to characterize uncertainty in a partial decision system.

Definition 7: Given a $\mathbb{PLDS} = (O', F, D)$, where $O' = O^L \cup O^C$ and $O^L/D^L = \{D_1, D_2, \ldots, D_s\}$. For $S \subseteq F$ and $0 \le \lambda \le 1$, the consistency-driven fuzzy dependency degree of D on S is defined as follows.

$$\gamma_S^{\mathbf{c}}(D) = \frac{\bigcup_{D_p \in O^L/D^L} \underline{\mathrm{R}}_S^{\lambda}(D_p)}{|O'|}. \tag{13}$$

This dependency degree quantifies the proportion of objects that can be certainly classified based on their fuzzy lower approximations. It captures uncertainty based on approximation granules from an algebraic perspective.

Definition 8: Given a $\mathbb{PLDS} = (O', F, D)$, where $O' = O^L \cup O^C$ and $O^L/D^L = \{D_1, D_2, \ldots, D_s\}$. For $S \subseteq F$ and $0 \le \lambda \le 1$, the consistency-driven fuzzy conditional entropy of D related to S is defined as follows.

$$\mathbf{c}FCE(D|S) = -\frac{1}{|O'|}\sum_{i=1}^{|O'|} \log \frac{|[o_i]_{\mathrm{R}_S}^{\lambda} \cap \tilde{\mathbf{M}}(\cdot, D^*(o_i))|}{|[o_i]_{\mathrm{R}_S}^{\lambda}|}, \tag{14}$$

where $D^*(o_i) = \arg\max_{D_p \in O^L/D^L} \tilde{\mathbf{M}}(o_i, D^*)$ is the decision label induced by the maximum membership degree, and $\tilde{\mathbf{M}}(\cdot, D^*(o_i))$ is a vector with $|O'| \times 1$.

This entropy measure captures the information uncertainty in feature subset S with respect to the soft-labeled decision from fuzzy object granules. Compared with the consistency-driven fuzzy dependency degree, it offers a more detailed perspective grounded in entropy theory. However, for the multi-granularity structure of the fuzzy approximation process, developing accurate and interpretable uncertainty measures grounded in it remains essential for robust modeling in partially labeled environments.

IV. Semi-Supervised Feature Selection Method

According to the consistency-driven multi-granularity knowledge acquisition method developed in the previous section, this section introduces a new uncertainty measure, Multi-level Zentropy, to effectively quantify the structural uncertainty in partially labeled decision systems for semi-supervised feature selection.

A. Multi-Level Zentropy Uncertainty Measure

As illustrated in Fig. 3, the consistency-driven fuzzy approximation process exhibits a multi-granularity structure from coarse to fine granularity. It involves three hierarchical levels: target decision classes, soft labels assigned to consistent objects, and fuzzy granules. Each level captures uncertainty from a distinct perspective and jointly contributes to the overall approximation process in a complementary manner. This systematic modeling aligns with the zentropy concept.

Motivated by the above analysis, a new multi-level zentropy uncertainty measure integrating these multiple granular levels is proposed in partially labeled decision systems, which is based on the following theoretical assumptions:

1) *Multi-granularity knowledge structure:* Knowledge exhibits a multi-granularity structure, and the overall system uncertainty arises from the integration of uncertainties across these levels.

Authorized licensed use limited to: TONGJI UNIVERSITY. Downloaded on November 11,2026 at 13:57:24 UTC from IEEE Xplore. Restrictions apply.

2) *Granularity interactions:* Coarse-grained representations can be hierarchically decomposed into finer-grained components, and the uncertainties at different granular levels are interdependent and interact with each other.

Definition 9: Given a $\mathbb{PLDS} = (O', F, D)$, where $O' = O^L \cup O^C$ and $O^L/D^L = \{D_1, D_2, \ldots, D_s\}$. For $S \subseteq F$ and $0 \leq \lambda \leq 1$, the multi-level zentropy uncertainty measure of feature set S is defined as follows.

$$\mathbf{Z}^{\mathbf{M}}_{\mathbf{S},\lambda}(D) = -\sum_{v=1}^{s} P_v \log P_v + \sum_{v=1}^{s} P_v \mathbf{Z}_v, \tag{15}$$

where $P_v = \frac{|D'_v|}{|O'|}$ represents the probability of the v-th decision class over the set O', and $D'_v = \{o \in O' \mid \arg\max_{p=1,\ldots,s} \tilde{\mathbf{M}}(o, D_p) = v\}$ is the set of objects assigned to D_v according to maximum soft membership. The internal entropy $\mathbf{Z}_v$ captures uncertainty within D_v and is defined in the following equations.

At the second granularity level (i.e., the ASL in Fig. 3), the intra-class zentropy $\mathbf{Z}_v$ is given by:

$$\mathbf{Z}_v = -\sum_{t=1}^{2} P_{vt} \log P_{vt} + \sum_{t=1}^{2} P_{vt} \mathbf{Z}_{vt}, \tag{16}$$

where $P_{v1} = \frac{\sum_{o_i \in D'_v} \tilde{\mathbf{M}}(o_i, D_v)}{|D'_v|}$ and $P_{v2} = 1 - P_{v1}$ respectively denote the aggregated certainty and uncertainty of objects in D'_v with respect to D_v. Each $\mathbf{Z}_{vt}$ further quantifies the specific object's information and is defined as:

$$\mathbf{Z}_{vt} = -\sum_{o_g \in C_{vt}} P_{vtg} \log P_{vtg}, \tag{17}$$

where $P_{vtg} \log P_{vtg}$ denotes the aggregated entropy of objects in C_{vt}, $P_{vtg} = \frac{|[o_g]^{\lambda}_{\mathrm{R}_S} \cap \tilde{\mathbf{M}}(o_g, D_v)|}{|[o_g]^{\lambda}_{\mathrm{R}_S}|}$ reflects the certainty of o_g belonging to D_v. Specifically, the $C_{v1} = D'_v \cap D^{\#}_v$ consists of confidently assigned objects (via both ASL and consistency-driven approximation), where $D^{\#}_v = \{o \in O' \mid \arg\max_{p=1,\ldots,s} \underline{\mathrm{R}^{\lambda}_F}(D_p)(o) = v\}$. In contrast, $C_{v2} = D'_v \setminus C_{v1}$ captures those with higher classification uncertainty.

Compared with existing single granularity methods, such as fuzzy entropy or approximations, the proposed multi-level zentropy measure effectively integrates three granular levels and models their interactions to characterize system uncertainty. Specifically, coarse-level information could regularize local fluctuations at finer levels and avoid measurement bias induced by a single level, thereby achieving a more stable and reliable uncertainty measure for feature selection. Notably, when the model degenerates to a single granularity, the zentropy will be decomposed into the traditional fuzzy entropy.

Property 2. Given a $\mathbb{PLDS} = (O', F, D)$. For $S \subseteq F$ and $0 \leq \lambda \leq 1$, $\mathbf{Z}^{\mathbf{M}}_{\mathbf{S},\lambda}(D)$ is the multi-level zentropy measure on S. The following properties hold:

1) *Non-negativity:* $\mathbf{Z}^{\mathbf{M}}_{\mathbf{S},\lambda}(D) \geq 0$;
2) *Non-monotonicity:* For parameters $0 \leq \lambda_1 \leq \lambda_2 \leq 1$, the $\mathbf{Z}^{\mathbf{M}}_{\mathbf{S},\lambda_1}(D)$ and $\mathbf{Z}^{\mathbf{M}}_{\mathbf{S},\lambda_2}(D)$ are incomparable;
3) *Non-monotonicity:* For subsets $S_1 \subseteq S_2 \subseteq S$, the $\mathbf{Z}^{\mathbf{M}}_{\mathbf{S_1},\lambda}(D)$ and $\mathbf{Z}^{\mathbf{M}}_{\mathbf{S_2},\lambda}(D)$ are incomparable;
4) *Unit Vector:* For any $o_i \in O'$, if both $[o_i]^{\lambda}_{\mathrm{R}_S}$ and $\tilde{\mathbf{M}}(o_i)$ are unit vectors, $\mathbf{Z}^{\lambda}_S(D) = -\sum_{v=1}^{s} P_v \log P_v$.

Proof: These properties can be derived from Defs. 6 and 9.

1) It can be directly obtained from Definition 9.
2) As λ increases, the fuzzy neighborhoods $[o_i]^{\lambda}_{\mathrm{R}_S}$ become less strict. However, the certainty degrees P_{vtg} do not increase or decrease monotonically with λ. Hence, $\mathbf{Z}^{\mathbf{M}}_{\mathbf{S},\lambda}(D)$ does not exhibit monotonic behavior with λ.
3) Similarly to 2), the probability of P_{vtg} has not monotonic variztion with feature subsets variations. Therefore, $\mathbf{Z}^{\mathbf{M}}_{\mathbf{S_1},\lambda}(D)$ and $\mathbf{Z}^{\mathbf{M}}_{\mathbf{S_2},\lambda}(D)$ are incomparable.
4) If $[o_i]^{\lambda}_{\mathrm{R}_S}$ is a unit vector and $\tilde{\mathbf{M}}(o_i)$ also assigns full membership to a single decision class, then $P_{vtg} = 1$ for the corresponding object and 0 for others. This leads to $\mathbf{Z}_{vt} = 0$, $P_{v1} = 1$, $P_{v2} = 0$, and thus $\mathbf{Z}_v = 0$. Therefore, the multi-level zentropy measure simplifies $\mathbf{Z}^{\mathbf{M}}_{\mathbf{S},\lambda}(D) = -\sum_{v=1}^{s} P_v \log P_v$.

B. Feature Evaluation and Selection

In this subsection, a feature selection method based on the ASL mechanism and the multi-level zentropy measure is proposed, where the goal is to identify a subset that preserves the original uncertainty information as much as possible.

Definition 10: Given a $\mathbb{PLDS} = (O', F, D)$. For $0 \leq \lambda \leq 1$, a $R \subseteq F$ is called a reduct of F if it satisfies the following.

1) $\mathbf{Z}^{\mathbf{M}}_{\mathbf{R},\lambda}(D) \leq \mathbf{Z}^{\mathbf{M}}_{\mathbf{F},\lambda}(D)$;
2) For $r \in R$, $\mathbf{Z}^{\mathbf{M}}_{\mathbf{R}-\{\mathbf{r}\},\lambda}(D) \leq \mathbf{Z}^{\mathbf{M}}_{\mathbf{R},\lambda}(D)$.

To evaluate the contribution of features during the selection process, two significance measures are defined as follows.

Definition 11: Given a $\mathbb{PLDS} = (O', F, D)$. For $f \in F$ and $0 \leq \lambda \leq 1$, the inner significance of f to F is cacluated as

$$IS(f, F, D) = \mathbf{Z}^{\mathbf{M}}_{\mathbf{F}-\{\mathbf{f}\},\lambda}(D) - \mathbf{Z}^{\mathbf{M}}_{\mathbf{F},\lambda}(D), \tag{18}$$

where the important ones are obtained by $IS(f, F, D) > 0$. Similarly, for $S \in F$ and $0 \leq \lambda \leq 1$, the outer significance of $f \in F - S$ to S is cacluated as follows.

$$OS(f, S, D) = \mathbf{Z}^{\mathbf{M}}_{\mathbf{S},\lambda}(D) - \mathbf{Z}^{\mathbf{M}}_{\mathbf{S}\cup\{\mathbf{f}\},\lambda}(D). \tag{19}$$

In the reduction procedure, features with positive inner significance are initially selected. Then, other relevant features are identified using outer significance. Due to the non-monotonic behavior of the proposed zentropy measure, a final reduction step is required to remove redundant features and ensure the minimality of the resulting subset. The detailed procedure is summarized in Algorithm 1 and Fig. 4.

In Algorithm 1, the first ASL mechanism step identifies consistent objects across w samples, m features, and s decision classes. This computation about object consistency results in a complexity of $O(wms)$. Subsequently, soft membership vectors of w objects across s classes have an $O(w'^2\, ms)$ complexity. In the second step, the reduction criterion $\mathbf{Z}^{\mathbf{M}}_{\mathbf{F},\lambda}(D)$ involves constructing a fuzzy relation matrix among the w' consistent objects under the feature F, which incurs a complexity of $O(w'^2\, m)$. For the inner important feature selection, it needs to compare m feature groups with $m-1$ features, leading to a complexity of $O(w'^2 m(m-1))$. Suppose there are l_1

Authorized licensed use limited to: TONGJI UNIVERSITY. Downloaded on November 11,2026 at 13:57:24 UTC from IEEE Xplore. Restrictions apply.

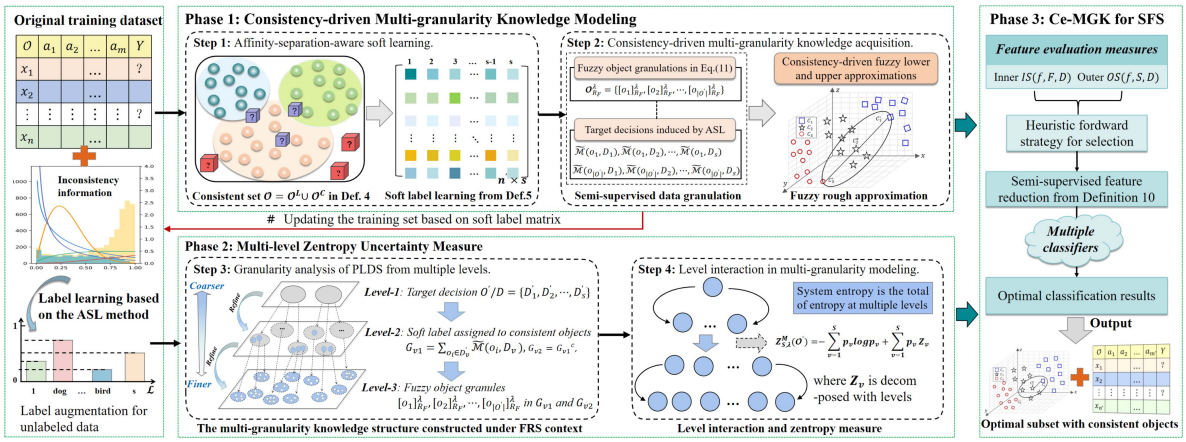


Fig. 4. The overall framework of proposed Ce-MGK method in semi-supervised scenario. Firstly, an ASL mechanism is introduced to estimate soft labels for the consistent objects and adopted for consistency-driven multi-granularity knowledge modeling. Then, by analyzing the granular level relationships among approximation processing, the multi-level zentropy measure is proposed for a comprehensive uncertainty characterization. Finally, two significance measures are defined to evaluate feature importance and employed for semi-supervised feature selection.

Algorithm 1: The Proposed Ce-MGK Algorithm.

Input : $\mathbb{PLDS} = (O^{'}, F, D)$ and parameter $0 \leq \lambda \leq 1$.
Output : The feature reduct R.

Initialize $R \leftarrow \emptyset$;
Step I: Affinity-separation-aware soft learning
Obtain the consistent objects $O^{'}$ according to Def. 4;
forall $o_i \in O^{'}$ **do**
  forall $D_p \in O^L/D^L$ **do**
    Compute the classify affinity $\mathbb{M}(o_i, D_p)$;
  end
  forall $D_p \in O^L/D^L$ **do**
    Compute $\mathbf{M}(o_i, D_p)$ with inter-classes differences;
  end
end
Obtain the affinity soft label $\tilde{\mathbf{M}}$;
Step II: Feature selection with multi-level zentropy measure
Calculate the reduction ceteria $\mathbf{Z}^{\mathbf{M}}_{\mathbf{F},\lambda}(D)$ by Def. 9;
forall $f \in F$ **do**
  Calculate inner measure $IS(f, F, D)$ by Eq. (18);
  if $IS(f, F, D) > 0$ **then**
    $R \leftarrow R \cup \{f\}$;
  end
end
Calculate the zentropy measure of $\mathbb{PLDS}$ on R, i.e., $\mathbf{Z}^{\mathbf{M}}_{\mathbf{R},\lambda}(D)$;
while $\mathbf{Z}^{\mathbf{M}}_{\mathbf{R},\lambda}(D) > \mathbf{Z}^{\mathbf{M}}_{\mathbf{F},\lambda}(D)$ **do**
  forall $f \in F - R$ **do**
    Calculate outer measure $OS(f, S, D)$ by Eq. (19);
  end
  Select $f_o = argmax_{f \in F-R} OS(f, S, D)$;
  $R \leftarrow R \cup \{f_o\}$;
end
for $r \in R$ **do**
  Calculate $\mathbf{Z}^{\mathbf{M}}_{\mathbf{R}-\{\mathbf{r}\},\lambda}(D)$;
  if $\mathbf{Z}^{\mathbf{M}}_{\mathbf{R}-\{\mathbf{r}\},\lambda}(D) \leq \mathbf{Z}^{\mathbf{M}}_{\mathbf{R},\lambda}(D)$ **then**
    $R \leftarrow R - \{r\}$;
  end
end

and l_2 inner and outer essential features, the time complexity of codes 22-28 is $O(\sum_{i=1}^{l_2} |w^{'2}(l_1 + i)(m - l_1 - i + 1)|)$. Finally, the elimination of redundant features with complexity $O(|w^{'2}(l_2 - 1)l_2|)$. Therefore, the overall time complexity of this algorithm is $O(w^{'2}m(m-1))$.

V. Experimental Analysis

This section presents numerical experiments to evaluate the effectiveness of the proposed methods in $\mathbb{PLDS}$. All experiments are conducted on a computer running Microsoft Windows 10, equipped with an Intel(R) Core(TM) i7-6800 K CPU@3.40 GHz (12 cores), 62.7 GB RAM, and implemented using MATLAB R2020a.

TABLE I
Detailed Information About Datasets

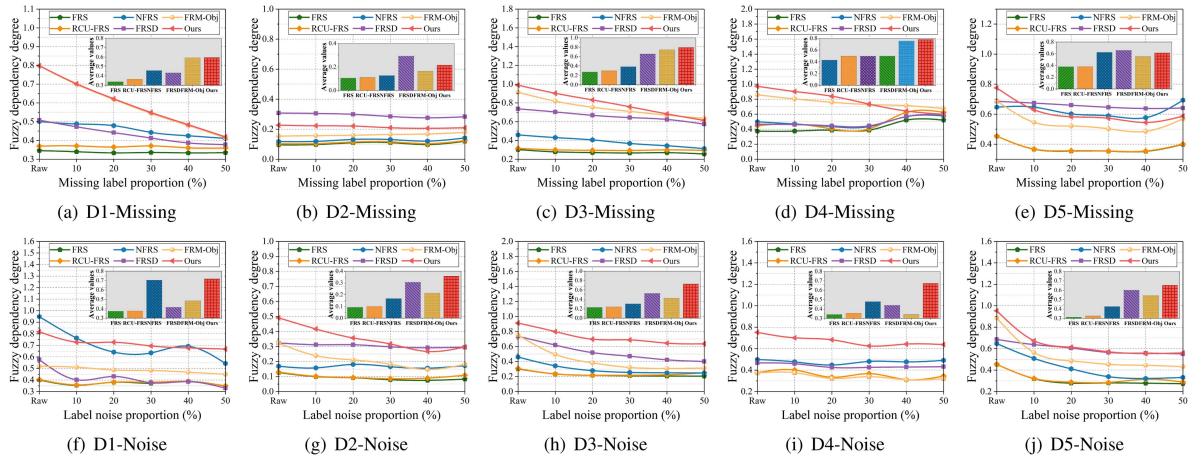
No.s	Dataset	Abbreviation	Object	Feature	Classe
D1	Arrhythmia	Arrh	450	277	13
D2	Cardiotocography	Card	2126	21	3
D3	Climate Model Simulation Crashes	CMSC	504	18	2
D4	Colonok	Colo	62	2001	2
D5	Ionosphere	Iono	351	34	2
D6	Leukemia	Leuk	72	7071	2
D7	Lymphography	Lymp	96	4027	9
D8	Nursery	Nurs	12960	9	5
D9	Parkinsons	Park	197	23	2
D10	South German Credit	SoGC	1000	21	2
D11	Spectfheart	Spec	267	45	2
D12	WarpAR10P	WarP	130	2400	10
D13	Wdbc	Wdbc	569	31	2
D14	Avila	Avil	20867	11	12
D15	Autos	Auto	159	25	6
D16	Landsat	Land	2000	36	6
D17	Page	Page	548	11	5
D18	Yeast	Yeas	1483	9	10

A. General Setting

1) Datasets: Eighteen benchmark datasets from the UCI Machine Learning Repository and scikit-feature library are used, with their details summarized in Table I. All conditional features are normalized to the [0,1] interval using the max-min method to ensure scale consistency.

2) Baselines: To demonstrate the effectiveness of the proposed methods, four FRS models, including the classical FRS (FRS) [47], RCU-measure-based FRS (RCU-FRS) [36], noise-aware FRS (NFRS) [48], and fuzzy-decision-based FRS (FRSD) [25], are first employed to make robust evaluations of the proposed consistency-driven FRS model. Meanwhile, the ablation model without consistent object selection (FRM-Obj) is evaluated. Then, eight representative feature selection methods are involved for comparative evaluation, namely

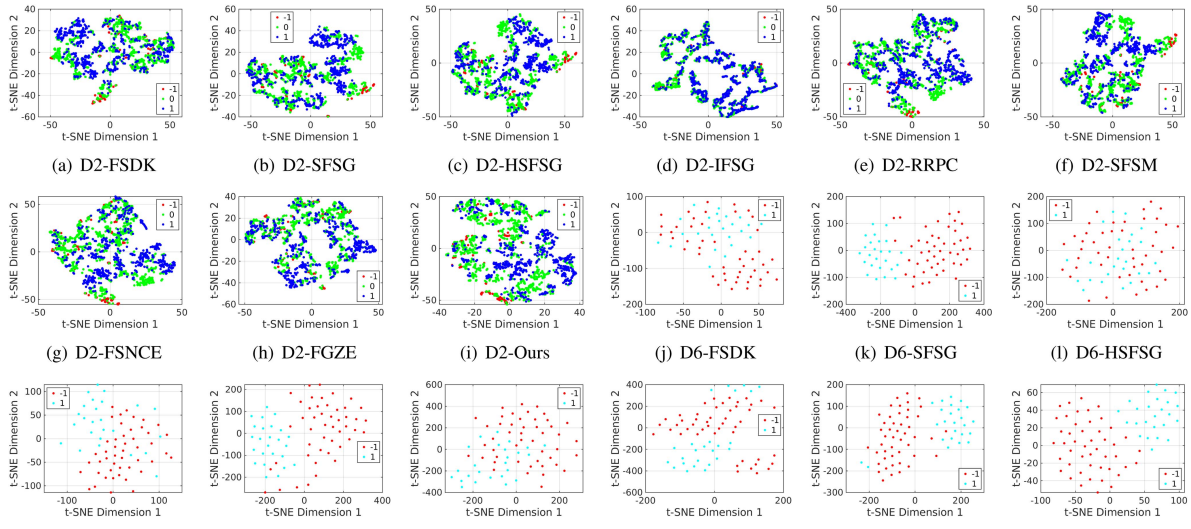
Authorized licensed use limited to: TONGJI UNIVERSITY. Downloaded on November 11,2026 at 13:57:24 UTC from IEEE Xplore. Restrictions apply.



(a) D1-Missing (b) D2-Missing (c) D3-Missing (d) D4-Missing (e) D5-Missing

(f) D1-Noise (g) D2-Noise (h) D3-Noise (i) D4-Noise (j) D5-Noise

Fig. 5. The fuzzy dependency degree of different fuzzy rough approximation models in different missing label and label noise proportions.



(a) D2-FSDK (b) D2-SFSG (c) D2-HSFSG (d) D2-IFSG (e) D2-RRPC (f) D2-SFSM

(g) D2-FSNCE (h) D2-FGZE (i) D2-Ours (j) D6-FSDK (k) D6-SFSG (l) D6-HSFSG

(m) D6-IFSG (n) D6-RRPC (o) D6-SFSM (p) D6-FSNCE (q) D6-FGZE (r) D6-Ours

Fig. 6. The visualization t-SNE results of selected features induced by different methods, where (a)–(i) shows the compared results on the D2 dataset, (j)–(r) shows the compared results on the D6 dataset.

unsupervised feature selection via fast sparse discriminative K-means (FSDK) [49], infinite feature selection with graph modeling (IFSG) [50], hessian-driven semi-supervised selection with generalized uncorrelated constraint (HSFSG) [14], semi-supervised feature selection with a generalized uncorrelated constraint (SFSG) [24], semi-supervised feature selection with relevance-redundancy analysis using Pearson's correlation (RRPC) [17], semi-supervised feature selection with heuristic strategies (SFSM) [15], neighborhood combination entropy criterion (FSNCE) [45], and fuzzy-granule-based zentropy selection (FGZE) [34]. Moreover, the comparison with FGZE highlights the advantage of the proposed multi-level zentropy measure in capturing and quantifying uncertainty.

3) Comparison Setups: Three widely used classifiers, including K-nearest neighbor (KNN, $K = 3$), Naive Bayes (NB), and decision tree (DT, Gini index), are adopted as benchmark classifiers to evaluate different feature selection methods in semi-supervised scenarios. Meanwhile, considering the important influences of neighborhood parameter λ in the whole Ce-MGK framework, it varies from 0 to 1 with a step size of 0.05. Furthermore, 10-fold cross-validation is employed to ensure a fair evaluation on each dataset.

TABLE II
THE AVERAGE COMPUTATIONAL TIME (SECONDS) OF DIFFERENT METHODS UNDER TWO MISSING LABEL RATIOS

No.s	FSDK	SFSG	HSFSG	IFSG	RRPC	SFSM	FSNCE	FGZE	Ours
D1	55.00	66.50	74.36	12.50	233.00	62.50	87.07	1759.65	100.89
D2	14.69	14.95	12.35	3.59	13.06	12.14	29.42	123.10	342.07
D3	3.57	3.69	3.25	0.47	3.61	3.96	6.80	9.34	15.92
D4	6.50	271.62	55.70	19.50	68788	301.00	75.13	130.75	62.24
D5	16.36	4100	574.49	100.00	567100	1650.00	154.88	749.81	34.34
D6	7.60	7.15	7.44	2.33	13.98	9.46	43.09	20.61	988.59
D7	8.00	8.50	5.44	2.25	8.00	8.00	2.89	4.82	67.12
D8	31.05	31.25	27.45	5.56	32.83	33.78	57.82	351.82	33.07
D9	168.51	199.47	182.67	32.36	162.96	168.35	232.06	453.89	1111
D10	2.16	1.99	2.25	0.43	1.98	2.05	3.36	5.07	8.50
D11	3.47	3.47	3.31	0.56	3.82	3.43	7.63	17.22	13.13
D12	29.00	112.00	155.03	85.50	116245	647.50	251.89	1484	38.98
D13	4.54	4.79	4.61	2.42	4.50	4.70	11.46	7.99	107.56
D14	389.93	1320	607.84	1008	758.51	315.75	1988	3479.80	2605
D15	34.75	21.16	22.73	37.13	29.57	21.58	6.88	15.04	37.61
D16	110.60	96.13	96.62	171.53	158.44	12.56	74.32	51.37	91.96
D17	21.77	21.61	22.77	31.25	31.94	4.62	6.08	4.83	17.58
D18	73.01	60.82	60.66	71.40	68.24	10.69	20.72	8.77	15.29
Ave	54.47	352.52	106.61	88.14	41869.86	181.78	169.97	482.08	316.15

Authorized licensed use limited to: TONGJI UNIVERSITY. Downloaded on November 11,2026 at 13:57:24 UTC from IEEE Xplore. Restrictions apply.

TABLE III
AVERAGE SELECTED FEATURES NUMBER OF COMPARED METHODS ACROSS THREE CLASSIFIERS UNDER 30% AND 60% MISSING PROPORTIONS

No.s	Number of selected features under 30% missing labels										No.s	Number of selected features under 60% missing labels									
	Raw	FSDK	SFSG	HSFSG	IFSG	RRPC	SFSM	FSNCE	FGZE	Ours		Raw	FSDK	SFSG	HSFSG	IFSG	RRPC	SFSM	FSNCE	FGZE	Ours
D1	276	42.33	42.33	42.33	27.00	42.33	42.33	30.00	118.33	42.33	D1	276	33.67	33.67	33.67	27.00	33.67	33.67	48.00	94.00	33.67
D2	20	7.00	7.00	7.00	3.00	7.00	7.00	15.33	9.00	7.00	D2	20	10.67	10.67	10.67	3.00	10.67	10.67	17.67	7.00	10.67
D3	17	8.33	8.33	8.33	3.00	8.33	8.33	9.67	12.33	8.33	D3	17	7.33	7.33	7.33	3.00	7.33	7.33	16.00	12.33	7.33
D4	2000	1.33	1.33	1.33	335	1.33	1.33	1.00	13.00	1.33	D4	2000	2.33	2.33	2.33	335	2.33	2.33	1.00	8.00	2.33
D5	33	7.00	7.00	7.00	4.00	7.00	7.00	14.67	10.00	7.00	D5	33	5.33	5.33	5.33	4.00	5.33	5.33	26.67	15.67	5.33
D6	7070	3.00	3.00	3.00	1115	3.00	3.00	1.00	6.00	3.00	D6	7070	4.33	4.33	4.33	1115	4.33	4.33	1.00	4.00	4.33
D7	18	9.33	9.33	9.33	3.00	9.33	9.33	1.00	12.00	9.33	D7	18	10.33	10.33	10.33	3.00	10.33	10.33	10.00	10.33	10.33
D8	8	7.00	7.00	7.00	1.00	7.00	7.00	7.00	1.00	7.00	D8	8	7.00	7.00	7.00	1.00	7.00	7.00	7.00	3.00	7.00
D9	22	9.33	9.33	9.33	3.00	9.33	9.33	8.67	6.33	9.33	D9	22	5.33	5.33	5.33	3.00	5.33	5.33	10.00	8.00	5.33
D10	20	9.33	9.33	9.33	3.00	9.33	9.33	14.00	16.00	9.33	D10	20	7.33	7.33	7.33	3.00	7.33	7.33	13.67	11.00	7.33
D11	44	7.33	7.33	7.33	8.00	7.33	7.33	13.67	24.67	7.33	D11	44	3.67	3.67	3.67	8.00	3.67	3.67	27.00	3.00	3.67
D12	2400	11.00	11.00	11.00	386	11.00	11.00	1.00	65.33	11.00	D12	2400	10.33	10.33	10.33	386	10.33	10.33	1.00	57.00	10.33
D13	30	15.00	15.00	15.00	6.00	15.00	15.00	28.00	2.00	15.00	D13	30	10.00	10.00	10.00	6.00	10.00	10.00	26.67	2.00	10.00
D14	11	2.33	2.33	2.33	1.00	2.33	2.33	4.33	1.00	2.33	D14	11	3.00	3.00	3.00	1.00	3.00	3.00	8.67	1.00	3.00
D15	25	5.33	5.33	5.33	3.00	5.33	5.33	6.33	2.00	5.33	D15	25	4.00	4.00	4.00	3.00	4.00	4.00	10.67	1.00	4.00
D16	9	17.00	17.00	17.00	5.00	17.00	17.00	36.00	1.00	17.00	D16	9	13.00	13.00	13.00	5.00	13.00	13.00	35.67	1.00	13.00
D17	11	6.00	6.00	6.00	1.00	6.00	6.00	5.33	1.00	6.00	D17	11	6.00	6.00	6.00	1.00	6.00	6.00	4.67	1.00	6.00
D18	36	3.67	3.67	3.67	1.00	3.67	3.67	8.00	1.00	3.67	D18	36	4.33	4.33	4.33	1.00	4.33	4.33	8.00	1.00	4.33
Ave	669	9.54	9.54	9.54	106.00	9.54	9.54	11.39	16.78	9.54	Ave	669	8.22	8.22	8.22	106.00	8.22	8.22	15.19	13.35	8.22

B. Robustness Analysis

This section primarily illustrates the robustness of the proposed consistency-driven fuzzy rough approximation in comparison to several representative FRS models under scenarios involving missing labels and label noise, with ratios all ranging from 10% to 50% in steps of 10%. The fuzzy dependency degree defined in Definition 7 is adopted as the evaluation metric, where a larger value indicates a stronger capability of the model to capture reliable approximations in the presence of incomplete or corrupted supervision.

The fuzzy dependency degrees of the compared FRS models are reported in Tables A.I and A.II and visualized in Figs. A.1 and A.2 of the support materials. As observed, the proposed model achieves the best performance in 14 and 16 cases under missing label and label noise settings, respectively, with the highest overall mean values of 0.561 and 0.507, compared with 0.489 for FRM-Obj and 0.374 for FRSD. Although FRSD enhances approximation through fuzzy decision strategies, its performance remains unstable owing to the lack of consistent knowledge modeling. In particular, the comparison with the ablation FRM-Obj model illustrates the effectiveness of consistent selection in fuzzy rough approximation. Due to the space limitation, the results of some datasets are shown in Fig. 5, where curves depict trends and bars denote averaged values. It can be seen that the proposed model consistently outperforms the competitors in most cases, demonstrating its robustness and superior approximation capability in semi-supervised and noise scenarios.

C. Performance Comparison of SFS Algorithms

This section presents a comparative study of the Ce-MGK with eight other representative feature selection methods under the 30% and 60% missing label proportions. The detailed results are in Tables III–V, and the best results are in bold.

1) Computational Time and Number of Selected Features: The average computational time of different methods under 30% and 60% missing label ratios is recorded in Table II. Overall, the Ce-MGK achieves competitive efficiency, outperforming SFSG, RRPC, and FGZE from a whole perspective. Compared with other single-granularity methods, Ce-MGK takes additional computational cost of multi-granularity knowledge structures. Notably, since the exhaustive pairwise feature comparisons and sorting operations at each iteration, RRPC exhibits substantially higher computational time than other methods, especially for high-dimensional datasets.

The average number of selected features obtained by different methods under the 30% and 60% missing label proportions across three classifiers is summarized in Table III. Among them, FSDK, SFSG, HSFSG, RRPC, and SFSM are ranking-based approaches, and are therefore directly consistent to the proposed method for a fair comparison. As observed, Ce-MGK achieves the smallest average number of selected features while maintaining remarkable stability across both missing-label settings. In contrast, IFSG and FGZE tend to select substantially larger subsets on several datasets, such as D4, D6, and D12, and present high variability across datasets. It is worth noting that selecting either too many or too few features may introduce redundant information or overlook informative features, which doubtlessly influences the performance of subsequent tasks. Overall, the proposed Ce-MGK could consistently identify compact and informative feature subsets with partial labels.

Moreover, the t-SNE method [51] is employed to visualize the feature spaces obtained by the nine compared methods. Due to space limitations, the Card and Leuk datasets are selected as illustrative examples. As shown in the Fig. 6, the features selected by the proposed Ce-MGK exhibit clearer category boundaries and demonstrate a stronger ability to distinguish between different classes.

2) Classification Performance: The classification accuracy of compared methods under 30% and 60% missing label proportions is summarized in Tables IV–V. When the missing label proportion is 30%, the proposed Ce-MGK achieves the highest accuracy on 17, 15, and 16 datasets with KNN, NB, and DT classifiers, respectively, while methods such as FSDK, SFSG, HSFSG, and IFSG only outperform on one or two datasets. In particular, on the high-dimensional datasets D4, D6, and

Authorized licensed use limited to: TONGJI UNIVERSITY. Downloaded on November 11,2026 at 13:57:24 UTC from IEEE Xplore. Restrictions apply.

TABLE IV
CLASSIFICATION ACCURACY OF DIFFERENT METHODS UNDER 30% MISSING PROPORTIONS (%)

No.s	Datasets	FSDK	SFSG	HSFSG	IFSG	RRPC	SFSM	FSNCE	FGZE	Ab-Ent	Ab-Appr	Ab-Obj	Ours
KNN	D1	57.11±0.09	64.67±0.07	**65.56**±0.06	53.56±0.05	60.22±0.09	63.33±0.07	59.33±0.07	49.99±0.06	60.22±0.07	63.11±0.08	63.33±0.10	63.78±0.07
	D2	80.01±0.02	83.07±0.02	82.88±0.03	60.25±0.02	76.39±0.05	82.73±0.04	81.47±0.02	81.93±0.03	83.63±0.01	83.30±0.02	82.55±0.02	**85.61**±0.03
	D3	90.37±0.04	93.52±0.03	94.07±0.04	89.63±0.04	87.41±0.03	90.74±0.04	90.19±0.03	90.35±0.04	91.67±0.05	88.70±0.05	92.41±0.05	**94.26**±0.03
	D4	48.81±0.16	86.90±0.13	85.24±0.17	78.57±0.16	88.57±0.21	69.05±0.30	42.86±0.26	84.05±0.13	77.62±0.15	88.57±0.08	76.90±0.18	**88.81**±0.10
	D5	82.06±0.07	89.75±0.04	89.46±0.06	83.48±0.03	89.46±0.05	89.16±0.07	85.18±0.08	86.93±0.05	89.75±0.05	88.33±0.05	91.74±0.03	**93.17**±0.03
	D6	78.52±0.11	78.52±0.06	64.44±0.12	63.33±0.08	78.89±0.08	78.52±0.06	65.71±0.17	95.71±0.07	88.57±0.11	90.18±0.12	87.50±0.08	**95.89**±0.07
	D7	45.95±0.12	52.86±0.12	57.24±0.17	55.33±0.08	66.71±0.15	64.86±0.09	79.10±0.13	79.67±0.09	79.19±0.13	76.52±0.15	78.52±0.12	**79.81**±0.07
	D8	91.08±0.01	88.71±0.01	89.04±0.01	19.61±0.13	89.58±0.01	91.03±0.01	90.87±0.01	61.81±0.02	91.06±0.01	91.06±0.01	91.06±0.01	**91.15**±0.01
	D9	93.39±0.05	90.79±0.04	90.74±0.06	85.13±0.10	86.63±0.09	93.74±0.05	93.79±0.06	90.56±0.07	94.89±0.05	93.45±0.07	94.87±0.05	**95.87**±0.04
	D10	72.00±0.03	73.00±0.04	72.10±0.03	70.00±0.04	72.80±0.04	72.70±0.06	72.90±0.04	71.70±0.04	73.00±0.05	73.10±0.03	72.00±0.03	**73.10**±0.04
	D11	77.22±0.08	76.81±0.07	75.68±0.10	78.26±0.07	71.18±0.07	78.21±0.08	71.88±0.08	77.55±0.09	72.64±0.07	68.19±0.11	75.67±0.07	**81.27**±0.07
	D12	26.15±0.14	67.69±0.10	43.85±0.12	45.38±0.11	40.00±0.15	37.69±0.12	16.92±0.10	60.00±0.16	65.38±0.15	63.85±0.10	50.00±0.15	**68.46**±0.16
	D13	94.20±0.02	95.78±0.02	95.61±0.02	91.92±0.01	85.78±0.06	95.08±0.02	96.84±0.02	93.59±0.02	96.84±0.02	96.31±0.02	96.31±0.03	**97.19**±0.02
	D14	20.13±0.05	49.34±0.01	49.34±0.01	48.89±0.01	42.76±0.01	48.89±0.01	72.33±0.14	17.95±0.01	51.63±0.01	51.90±0.01	51.90±0.01	**73.75**±0.16
	D15	57.88±0.10	61.04±0.14	59.67±0.16	47.88±0.15	65.46±0.09	71.00±0.06	71.17±0.13	14.27±0.08	68.54±0.07	71.63±0.14	71.71±0.13	**72.88**±0.10
	D16	87.20±0.02	87.80±0.02	88.15±0.01	82.15±0.03	86.95±0.03	87.30±0.02	88.55±0.04	37.76±0.04	86.70±0.03	86.95±0.02	87.45±0.02	**89.05**±0.02
	D17	93.43±0.03	93.07±0.03	92.70±0.04	89.04±0.03	94.53±0.04	93.06±0.02	93.97±0.03	89.16±0.05	94.89±0.03	91.25±0.05	94.89±0.03	**95.07**±0.02
	D18	29.06±0.03	24.27±0.02	23.94±0.03	25.49±0.04	26.24±0.04	28.12±0.04	29.74±0.03	21.02±0.04	30.41±0.04	26.71±0.03	25.63±0.03	**30.41**±3.75
	Ave	68.03±0.07	75.42±0.05	73.32±0.07	64.88±0.07	72.75±0.07	74.18±0.06	72.38±0.08	66.89±0.06	77.59±0.06	77.39±0.06	76.91±0.06	**85.26**±0.06
No.s	Datasets	FSDK	SFSG	HSFSG	IFSG	RRPC	SFSM	FSNCE	FGZE	Ab-Ent	Ab-Appr	Ab-Obj	Ours
NB	D1	51.11±0.07	61.11±0.07	58.44±0.05	59.11±0.07	60.67±0.10	62.22±0.07	62.44±0.05	52.95±0.05	62.44±0.07	66.22±0.07	65.56±0.07	**66.22**±0.06
	D2	**69.05**±0.03	67.12±0.02	67.45±0.03	61.24±0.03	65.14±0.02	66.99±0.04	57.86±0.05	65.38±0.03	47.13±0.10	70.18±0.04	37.48±0.15	68.53±0.04
	D3	91.30±0.03	93.33±0.04	93.33±0.04	91.48±0.04	91.48±0.03	92.22±0.04	92.04±0.03	91.67±0.04	92.04±0.03	92.22±0.03	93.15±0.03	**94.07**±0.03
	D4	64.76±0.24	77.38±0.16	56.90±0.20	80.24±0.15	67.38±0.19	75.48±0.15	65.24±0.25	82.38±0.18	73.81±0.26	70.71±0.25	77.14±0.12	**84.05**±0.13
	D5	77.20±0.07	85.75±0.05	86.03±0.06	71.79±0.04	86.04±0.06	87.75±0.05	90.03±0.03	85.08±0.05	91.46±0.06	86.05±0.07	89.47±0.05	**93.16**±0.04
	D6	63.39±0.14	**97.32**±0.06	83.04±0.21	69.64±0.22	97.32±0.06	84.82±0.18	65.54±0.13	91.61±0.10	89.11±0.11	87.68±0.10	90.54±0.09	95.71±0.07
	D7	79.19±0.13	79.52±0.15	78.90±0.12	68.95±0.12	70.76±0.12	81.14±0.12	79.76±0.11	79.71±0.05	79.95±0.13	75.10±0.11	80.62±0.14	**82.48**±0.09
	D8	88.65±0.01	88.69±0.01	88.78±0.01	33.57±0.01	88.43±0.01	88.77±0.01	88.67±0.01	67.83±0.02	88.83±0.01	88.83±0.01	88.83±0.01	**88.97**±0.02
	D9	81.03±0.08	81.95±0.09	81.95±0.07	84.03±0.09	77.95±0.05	83.18±0.09	84.13±0.08	86.11±0.10	76.37±0.09	82.11±0.09	85.11±0.07	**86.11**±0.09
	D10	70.00±0.03	71.70±0.04	71.60±0.03	69.70±0.03	68.20±0.04	71.70±0.04	69.10±0.04	70.40±0.05	71.00±0.05	67.80±0.04	70.40±0.03	**72.60**±0.03
	D11	79.43±0.04	79.39±0.07	80.09±0.10	71.60±0.08	79.46±0.09	80.10±0.06	73.73±0.06	74.94±0.11	75.24±0.07	79.00±0.09	79.06±0.09	**80.16**±0.07
	D12	9.23±0.07	53.85±0.10	49.23±0.16	40.00±0.18	35.38±0.08	33.08±0.13	20.00±0.10	63.08±0.17	63.08±0.13	65.38±0.14	62.31±0.12	**65.38**±0.17
	D13	91.75±0.04	92.44±0.03	92.26±0.03	91.92±0.02	82.60±0.03	95.25±0.03	94.02±0.03	94.75±0.02	95.43±0.02	95.78±0.02	95.43±0.03	**97.54**±0.02
	D14	53.41±0.01	53.81±0.01	53.90±0.01	35.77±0.02	63.12±0.01	53.67±0.01	71.59±0.00	34.58±0.01	55.52±0.01	55.52±0.01	55.52±0.01	55.52±0.01
	D15	47.21±0.12	58.63±0.17	58.54±0.11	51.13±0.16	55.33±0.10	65.46±0.15	67.29±0.12	45.64±0.13	62.88±0.13	66.00±0.13	39.04±0.10	**68.54**±0.18
	D16	80.35±0.04	80.90±0.01	80.85±0.03	76.75±0.03	79.15±0.05	78.95±0.02	80.85±0.03	59.59±0.03	74.25±0.03	78.35±0.04	80.70±0.03	**80.90**±0.03
	D17	95.43±0.03	95.80±0.04	95.05±0.05	89.59±0.03	94.70±0.02	95.79±0.02	95.07±0.03	90.48±0.04	95.25±0.02	95.43±0.02	95.43±0.02	**96.72**±0.02
	D18	29.33±0.04	27.98±0.03	26.97±0.04	31.83±0.04	27.71±0.04	26.97±0.03	28.39±0.04	32.42±0.04	31.63±0.05	31.63±0.05	31.63±0.05	**33.58**±0.03
	Ave	67.88±0.07	74.82±0.06	72.41±0.08	65.46±0.08	71.71±0.06	73.53±0.07	71.43±0.07	70.48±0.07	73.63±0.08	75.22±0.07	73.19±0.07	**82.69**±0.07
No.s	Datasets	FSDK	SFSG	HSFSG	IFSG	RRPC	SFSM	FSNCE	FGZE	Ab-Ent	Ab-Appr	Ab-Obj	Ours
DT	D1	50.67±0.08	51.78±0.07	53.33±0.07	49.56±0.08	57.56±0.11	55.56±0.08	59.78±0.08	53.45±0.08	65.22±0.08	60.89±0.07	65.00±0.05	**65.56**±0.07
	D2	80.43±0.03	84.71±0.02	84.95±0.03	63.12±0.04	73.90±0.02	84.81±0.04	83.96±0.04	79.94±0.03	85.14±0.03	84.99±0.02	80.91±0.03	**85.33**±0.02
	D3	87.78±0.04	91.30±0.04	91.30±0.04	85.00±0.04	85.00±0.06	92.41±0.03	90.56±0.03	90.12±0.04	90.93±0.03	90.19±0.04	91.30±0.03	**93.15**±0.04
	D4	50.00±0.20	88.81±0.16	80.95±0.18	75.48±0.16	**88.81**±0.11	65.95±0.23	64.05±0.19	77.38±0.18	79.76±0.15	87.62±0.12	88.33±0.14	**88.81**±0.10
	D5	80.60±0.06	86.04±0.04	88.05±0.06	84.04±0.05	89.75±0.06	88.87±0.06	91.75±0.05	87.28±0.05	90.90±0.05	90.03±0.03	90.03±0.05	**92.31**±0.04
	D6	63.93±0.11	95.71±0.10	84.64±0.15	92.86±0.08	91.79±0.07	76.61±0.13	58.21±0.17	96.00±0.08	90.00±0.10	98.57±0.05	91.43±0.07	**98.57**±0.05
	D7	76.52±0.13	72.81±0.13	68.24±0.09	73.67±0.08	68.81±0.14	72.33±0.13	74.95±0.07	75.10±0.09	73.62±0.10	73.62±0.18	75.00±0.11	**78.33**±0.10
	D8	92.06±0.01	94.81±0.00	**94.81**±0.01	33.90±0.01	90.56±0.01	92.12±0.01	92.05±0.01	67.83±0.02	92.11±0.01	92.13±0.01	92.11±0.01	92.01±0.01
	D9	86.58±0.09	87.26±0.06	85.08±0.08	85.55±0.08	84.68±0.10	86.63±0.06	86.79±0.07	88.33±0.08	87.13±0.09	84.63±0.08	88.68±0.06	**89.68**±0.09
	D10	69.80±0.04	**72.00**±0.03	71.90±0.05	68.90±0.05	69.20±0.05	71.70±0.04	71.00±0.05	68.20±0.04	70.40±0.04	67.70±0.05	64.90±0.05	**72.00**±0.06
	D11	72.24±0.09	73.35±0.10	71.14±0.09	74.89±0.08	73.03±0.09	74.19±0.08	77.95±0.05	72.83±0.08	72.27±0.06	74.96±0.07	72.71±0.10	**79.39**±0.09
	D12	22.31±0.11	59.23±0.13	50.00±0.13	53.85±0.12	41.54±0.14	26.92±0.19	14.62±0.10	54.62±0.11	67.69±0.15	55.38±0.10	50.77±0.17	**63.85**±0.09
	D13	92.45±0.03	90.15±0.04	91.21±0.04	90.34±0.02	78.04±0.06	94.02±0.04	93.84±0.03	93.20±0.02	93.15±0.02	91.75±0.02	92.97±0.03	**94.91**±0.02
	D14	59.92±0.00	60.44±0.01	60.70±0.01	51.29±0.00	55.84±0.00	59.92±0.00	60.02±0.01	33.99±0.01	60.72±0.01	60.72±0.01	60.72±0.01	**60.86**±0.01
	D15	57.96±0.17	60.96±0.11	61.75±0.14	54.96±0.19	71.13±0.16	69.00±0.13	66.13±0.11	57.73±0.08	72.29±0.12	73.00±0.06	69.71±0.10	**74.17**±0.07
	D16	81.70±0.03	82.35±0.03	82.00±0.02	79.20±0.03	81.35±0.04	82.50±0.02	82.15±0.03	59.43±0.02	80.95±0.02	81.80±0.03	82.00±0.02	**82.90**±0.03
	D17	93.98±0.03	94.70±0.03	93.98±0.04	87.94±0.04	95.07±0.02	94.53±0.03	95.07±0.02	87.25±0.03	95.07±0.02	95.07±0.02	95.25±0.02	**95.26**±0.02
	D18	29.00±0.04	30.54±0.03	30.08±0.04	28.73±0.03	27.78±0.03	27.38±0.04	29.26±0.03	29.09±0.05	28.05±0.05	28.05±0.05	28.05±0.05	**30.55**±0.04
	Ave	69.33±0.07	76.50±0.06	74.67±0.07	68.51±0.07	73.55±0.07	73.08±0.07	71.79±0.06	70.65±0.06	77.52±0.06	77.28±0.06	76.66±0.06	**84.15**±0.06

D12, the unsupervised IFSG exhibits poor performance due to selecting excessive redundant features.

With the increase in missing labels, the superiority of the proposed method remains evident. At the 60% setting, Ce-MGK achieves competitive results on 16, 16, and 14 datasets for three classifiers. In contrast, the HSFSG method achieves better results in only a few cases, such as D12 on NB, and D8 and D12 on DT. That's mainly because it ignores multi-level structural information in feature spaces. The comparison of average accuracies in Fig. 7 further confirms the superiority of Ce-MGK in classification tasks.

In particular, the Ce-MGK generally outperforms the existing fuzzy granule zentropy-based FGZE method in different conditions, except on D6 at 60% with DT. This further validates the effectiveness of the proposed multi-level zentropy in enhancing fuzzy rough approximations. Overall, the analysis above demonstrates that Ce-MGK can reliably identify appropriate feature subsets and achieve excellent classification performance across different levels of label incompleteness.

3) Comparison Evaluation on Imbalanced Datasets: To verify the generality of Ce-MGK, four imbalanced datasets are employed, and their details are recorded in Table VI.

As reported in Tables IV–V, the proposed Ce-MGK consistently achieves the best performance across the four imbalanced datasets under the 30% missing label setting. A similar trend is observed at the 60% missing ratio, except for

Authorized licensed use limited to: TONGJI UNIVERSITY. Downloaded on November 11,2026 at 13:57:24 UTC from IEEE Xplore. Restrictions apply.

TABLE V
CLASSIFICATION ACCURACY OF DIFFERENT METHODS UNDER 60% MISSING PROPORTIONS (%)

No.s	Datasets	FSDK	SFSG	HSFSG	IFSG	RRPC	SFSM	FSNCE	FGZE	Ab-Ent	Ab-Appr	Ab-Obj	Ours
KNN	D1	54.00±0.07	65.11±0.07	**65.78**±0.07	52.67±0.08	62.67±0.09	65.11±0.06	56.89±0.06	41.21±0.09	60.22±0.07	58±0.09	57.56±0.08	57.56±0.07
	D2	77.23±0.04	85.70±0.02	85.70±0.02	60.86±0.02	76.01±0.03	**87.86**±0.02	80.48±0.02	79.52±0.02	78.13±0.02	81.09±0.02	79.82±0.03	83.93±0.02
	D3	90.56±0.04	92.59±0.03	92.96±0.03	89.81±0.03	90.56±0.05	92.59±0.03	91.30±0.04	88.75±0.05	93.52±0.04	88.70±0.05	91.11±0.05	**93.70**±0.03
	D4	42.86±0.25	80.24±0.15	75.95±0.22	80.95±0.16	83.81±0.22	77.14±0.24	57.62±0.18	82.62±0.11	55.48±0.25	83.57±0.11	76.90±0.12	**84.52**±0.14
	D5	81.79±0.08	87.75±0.05	88.02±0.04	84.07±0.05	88.62±0.06	90.58±0.07	90.29±0.05	88.08±0.05	87.76±0.05	88.90±0.04	90.33±0.05	**91.48**±0.04
	D6	58.75±0.17	91.61±0.10	**94.29**±0.10	82.68±0.17	84.29±0.11	91.61±0.07	64.64±0.20	81.25±0.16	82.32±0.16	86.43±0.13	77.68±0.12	**94.29**±0.10
	D7	78.38±0.12	76.48±0.09	76.29±0.11	56.29±0.14	73.62±0.16	76.38±0.08	62.76±0.13	81.55±0.15	79.76±0.06	77.81±0.12	79.10±0.07	**84.48**±0.10
	D8	91.08±0.01	88.71±0.01	89.04±0.01	19.61±0.13	89.58±0.01	91.03±0.01	90.79±0.01	91.05±0.01	91.06±0.01	91.06±0.01	91.06±0.01	**91.15**±0.01
	D9	96.42±0.04	93.89±0.06	93.34±0.06	85.08±0.05	93.32±0.03	93.39±0.06	93.39±0.06	95.32±0.07	88.24±0.06	88.71±0.05	95.37±0.04	**96.92**±0.03
	D10	66.10±0.05	71.50±0.02	71.80±0.05	70.00±0.04	69.60±0.04	72.10±0.04	72.00±0.04	67.90±0.05	71.50±0.04	70.00±0.05	69.30±0.04	**72.30**±0.03
	D11	74.89±0.07	75.63±0.07	74.96±0.10	77.85±0.11	72.64±0.06	76.10±0.11	73.11±0.09	71.96±0.06	75.64±0.06	71.91±0.06	73.79±0.07	**78.60**±0.08
	D12	26.92±0.15	57.69±0.10	47.69±0.11	41.54±0.09	29.23±0.20	48.46±0.10	18.46±0.09	60.00±0.10	45.38±0.13	59.23±0.14	36.92±0.15	**60.00**±0.17
	D13	95.07±0.03	94.73±0.03	95.08±0.03	92.09±0.02	94.91±0.03	96.48±0.01	97.19±0.01	94.10±0.02	87.45±0.02	97.19±0.02	95.43±0.02	**97.36**±0.02
	D14	62.46±0.01	43.42±0.02	44.91±0.01	48.89±0.01	42.62±0.01	59.89±0.00	61.76±0.01	17.95±0.01	51.37±0.01	51.37±0.01	51.37±0.01	**63.33**±0.01
	D15	57.21±0.11	65.46±0.18	64.71±0.12	47.88±0.15	66.79±0.14	71.63±0.11	70.42±0.10	8.63±0.14	64.08±0.10	69.13±0.13	42.75±0.12	**76.13**±0.12
	D16	86.65±0.02	87.30±0.02	87.20±0.02	82.15±0.03	87.10±0.03	86.45±0.02	86.00±0.04	43.13±0.04	86.65±0.02	87.20±0.02	87.25±0.02	**87.30**±0.02
	D17	93.80±0.04	93.98±0.02	93.97±0.03	89.04±0.03	93.26±0.04	94.15±0.01	94.16±0.03	87.28±0.06	94.16±0.02	93.80±0.02	94.16±0.02	**94.34**±0.03
	D18	30.08±0.04	30.07±0.03	30.34±0.04	25.49±0.04	30.95±0.04	30.34±0.03	30.42±0.04	19.72±0.04	30.41±0.04	30.41±0.04	30.41±0.04	**31.49**±0.03
	Ave	70.24±0.07	76.77±0.06	76.22±0.06	65.94±0.08	73.86±0.07	77.85±0.06	71.76±0.07	66.67±0.07	73.51±0.06	76.36±0.06	73.35±0.06	**79.94**±0.06
No.s	Datasets	FSDK	SFSG	HSFSG	IFSG	RRPC	SFSM	FSNCE	FGZE	Ab-Ent	Ab-Appr	Ab-Obj	Ours
NB	D1	59.78±0.09	60.67±0.07	58.89±0.07	58.89±0.06	64±0.08	**64.89**±0.07	58.89±0.09	41.73±0.08	66.00±0.04	67.78±0.05	63.56±0.07	59.56±0.04
	D2	65.43±0.03	65.24±0.02	65.19±0.03	61.48±0.02	65.42±0.05	65.99±0.03	67.54±0.03	64.45±0.04	42.80±0.12	69.16±0.04	68.49±0.03	**69.19**±0.02
	D3	91.30±0.03	93.52±0.03	93.52±0.03	91.48±0.04	91.48±0.03	93.33±0.04	91.48±0.04	90.63±0.06	93.52±0.03	91.48±0.03	91.48±0.03	**93.70**±0.03
	D4	64.76±0.24	80.48±0.15	79.29±0.15	82.38±0.17	75.71±0.15	74.05±0.18	59.52±0.28	81.19±0.16	64.05±0.21	64.05±0.27	60.95±0.20	**83.10**±0.16
	D5	77.19±0.07	82.92±0.06	82.32±0.07	71.23±0.09	91.15±0.07	90.60±0.05	90.30±0.07	89.65±0.04	85.20±0.08	80.34±0.07	81.77±0.07	**91.44**±0.06
	D6	56.96±0.19	90.00±0.15	87.50±0.11	75.71±0.21	78.93±0.12	**91.61**±0.10	65.00±0.25	82.32±0.16	75.18±0.19	81.96±0.15	82.50±0.17	**91.61**±0.14
	D7	74.81±0.16	77.71±0.11	77.76±0.08	68.90±0.11	70.33±0.10	72.90±0.09	67.00±0.13	77.67±0.08	78.24±0.13	75.14±0.13	81.19±0.12	**82.52**±0.15
	D8	88.65±0.01	88.69±0.01	88.78±0.01	33.57±0.01	88.43±0.01	88.77±0.01	88.87±0.01	87.42±0.01	88.83±0.01	88.83±0.01	88.83±0.01	**88.87**±0.01
	D9	81.42±0.09	82.63±0.10	84.74±0.10	82.55±0.04	80.58±0.11	85.08±0.07	82.24±0.09	82.53±0.11	80.55±0.13	83.08±0.11	85.24±0.10	**85.74**±0.06
	D10	71.00±0.05	70.60±0.04	71.40±0.04	69.50±0.04	72.40±0.05	72.80±0.03	68.90±0.04	71.30±0.06	70.20±0.04	67.80±0.04	71.10±0.04	**73.20**±0.05
	D11	79.39±0.09	79.05±0.10	78.63±0.05	71.51±0.12	77.19±0.09	77.08±0.06	79.42±0.06	78.70±0.07	79.37±0.05	79.37±0.05	79.37±0.05	**79.43**±0.06
	D12	13.85±0.13	60.77±0.14	**63.08**±0.17	46.15±0.14	26.15±0.12	47.69±0.09	20.77±0.12	56.00±0.21	66.92±0.06	53.08±0.13	53.08±0.11	58.46±0.15
	D13	91.92±0.03	92.44±0.03	93.32±0.03	91.57±0.03	92.26±0.03	94.19±0.04	94.03±0.03	94.32±0.03	93.84±0.04	94.20±0.04	95.08±0.02	**96.13**±0.03
	D14	53.41±0.01	53.81±0.01	53.90±0.01	35.77±0.02	51.09±0.01	53.67±0.01	55.00±0.01	34.58±0.01	55.52±0.01	55.52±0.01	55.52±0.01	**55.52**±0.01
	D15	48.46±0.09	52.83±0.15	55.92±0.09	51.13±0.16	52.83±0.14	55.46±0.14	64.21±0.10	37.09±0.11	64.21±0.07	64.83±0.15	64.38±0.18	**65.92**±0.13
	D16	79.45±0.03	66.40±0.02	66.60±0.03	76.75±0.03	74.50±0.04	78.20±0.03	80.75±0.04	71.20±0.02	81.15±0.04	66.75±0.04	47.00±0.03	**81.15**±0.02
	D17	95.98±0.03	95.62±0.03	94.86±0.05	89.59±0.03	93.43±0.02	96.89±0.02	95.26±0.04	87.48±0.03	95.06±0.02	96.51±0.02	95.06±0.02	**97.27**±0.03
	D18	33.64±0.03	27.98±0.03	26.97±0.04	31.83±0.04	27.71±0.04	27.99±0.02	28.39±0.03	32.27±0.05	31.63±0.05	31.63±0.05	31.63±0.05	**34.25**±0.02
	Ave	68.19±0.08	73.41±0.07	73.48±0.06	66.11±0.08	70.76±0.07	73.95±0.06	69.86±0.08	70.03±0.07	72.90±0.07	72.86±0.08	72.01±0.07	**77.06**±0.07
No.s	Datasets	FSDK	SFSG	HSFSG	IFSG	RRPC	SFSM	FSNCE	FGZE	Ab-Ent	Ab-Appr	Ab-Obj	Ours
DT	D1	51.11±0.08	57.56±0.07	58.22±0.09	49.56±0.06	56.22±0.07	58.44±0.09	54.67±0.06	43.29±0.09	56.67±0.06	58.89±0.08	**60.00**±0.06	58.89±0.09
	D2	76.76±0.03	85.37±0.03	86.31±0.04	63.08±0.03	73.94±0.04	86.17±0.02	82.08±0.02	75.60±0.04	80.71±0.02	84.24±0.03	80.25±0.03	**86.53**±0.05
	D3	87.78±0.04	91.30±0.04	90.93±0.04	85.19±0.05	85.56±0.05	92.22±0.04	87.78±0.05	89.03±0.06	92.78±0.04	90.74±0.05	89.81±0.04	**93.89**±0.04
	D4	65.00±0.20	80.48±0.15	75.71±0.24	82.38±0.13	**84.52**±0.16	77.38±0.08	65.00±0.13	84.29±0.14	76.19±0.15	84.05±0.13	84.29±0.14	**84.52**±0.14
	D5	81.75±0.05	85.47±0.03	88.03±0.05	82.33±0.06	88.90±0.06	90.31±0.04	90.87±0.04	89.02±0.05	89.48±0.05	87.19±0.04	87.75±0.04	**91.75**±0.03
	D6	53.75±0.19	84.82±0.18	95.71±0.07	89.29±0.12	87.68±0.07	**98.57**±0.05	56.79±0.18	90.00±0.16	83.39±0.09	84.64±0.13	79.46±0.15	87.68±0.10
	D7	64.90±0.09	70.24±0.10	71.67±0.14	73.67±0.11	67.90±0.15	66.71±0.14	68.95±0.10	77.05±0.10	73.00±0.11	70.86±0.17	76.52±0.13	**83.14**±0.09
	D8	92.06±0.01	94.81±0.00	**94.81**±0.01	33.90±0.01	90.56±0.01	92.12±0.01	91.96±0.01	75.54±0.14	92.11±0.01	92.13±0.01	92.11±0.01	92.28±0.01
	D9	84.76±0.12	83.63±0.09	83.05±0.10	83.08±0.08	83.11±0.10	86.61±0.11	81.05±0.11	89.04±0.13	83.61±0.06	75.97±0.08	83.58±0.10	**90.84**±0.07
	D10	70.60±0.04	70.30±0.04	69.50±0.04	69.40±0.05	65.90±0.05	70.00±0.05	68.90±0.05	69.40±0.04	71.00±0.04	70.60±0.03	68.40±0.04	**72.00**±0.04
	D11	76.03±0.10	73.45±0.06	74.53±0.04	75.66±0.05	70.43±0.08	73.77±0.10	77.17±0.08	72.79±0.08	72.29±0.10	75.70±0.08	73.42±0.08	**78.25**±0.06
	D12	25.38±0.12	56.92±0.13	**58.46**±0.10	56.15±0.08	37.69±0.11	50.77±0.09	16.92±0.09	38.00±0.11	50.77±0.14	56.92±0.13	28.46±0.15	**58.46**±0.16
	D13	91.57±0.04	92.09±0.04	92.79±0.03	90.86±0.05	92.98±0.02	91.74±0.04	92.98±0.03	93.05±0.03	92.62±0.03	92.63±0.02	93.33±0.03	**95.08**±0.02
	D14	60.38±0.01	59.44±0.01	53.70±0.01	51.29±0.00	57.69±0.00	59.92±0.00	59.41±0.00	33.99±0.01	60.72±0.01	60.72±0.01	60.72±0.01	**61.02**±0.01
	D15	59.21±0.15	68.38±0.12	72.92±0.09	54.96±0.19	72.50±0.11	68.63±0.13	73.67±0.09	37.75±0.12	71.79±0.17	70.38±0.10	72.92±0.07	**74.17**±0.11
	D16	79.80±0.03	80.30±0.03	80.25±0.03	79.20±0.03	76.60±0.05	79.95±0.02	82.15±0.04	71.93±0.05	83.10±0.04	82.60±0.03	81.95±0.02	82.75±0.03
	D17	94.54±0.03	94.52±0.03	94.35±0.04	87.94±0.04	94.89±0.02	93.99±0.04	93.61±0.04	87.09±0.05	95.07±0.02	95.44±0.02	95.44±0.02	**95.99**±0.03
	D18	29.54±0.04	31.22±0.03	31.22±0.01	28.73±0.03	29.94±0.03	30.44±0.03	30.48±0.04	30.42±0.03	29.13±0.03	31.21±0.05	29.19±0.04	**31.23**±0.04
	Ave	69.16±0.08	75.57±0.07	76.23±0.06	68.70±0.07	73.17±0.07	75.99±0.06	70.80±0.07	69.29±0.08	75.25±0.06	75.83±0.07	74.31±0.06	**78.80**±0.07

TABLE VI
BASIC DESCRIPTION OF FOUR IMBALANCED DATASETS

No.s	Datasets	Objects	Features	Classes
D15	Autos	159	25	48||46||29||20||13||3
D16	Landsat	2000	36	470||461||397||237||224||211
D17	Page	548	11	492||33||12||8||3
D18	Yeast	1483	9	243||429||463||44||35||51||163||30||20||5

the Page dataset on the DT classifier. These results demonstrate the effectiveness of Ce-MGK in handling imbalanced scenarios. Despite these promising outcomes, investing in a dedicated learning framework for extremely imbalanced settings remains an important direction for future research.

4) Ablation Study: To further validate the effectiveness of the consistent object selection mechanism and the multi-level uncertainty analysis in the Ce-MGK framework, three ablation variants are constructed: 1) Ce-MGK without consistent object selection (Ab-Obj), and 2) Ce-MGK without multi-level uncertainty analysis, including only fuzzy approximations (Ab-Appr) or fuzzy entropy of objects (Ab-Ent).

The ablation study results are reported in Tables IV–V. As observed from these tables, the proposed Ce-MGK consistently achieves superior results under different missing label proportions. Specifically, at the 30% ratio, Ce-MGK obtains the highest accuracy in 18, 17, and 16 cases across the three classifiers, respectively, whereas Ab-Appr and Ab-Ent outperform the other

Authorized licensed use limited to: TONGJI UNIVERSITY. Downloaded on November 11,2026 at 13:57:24 UTC from IEEE Xplore. Restrictions apply.

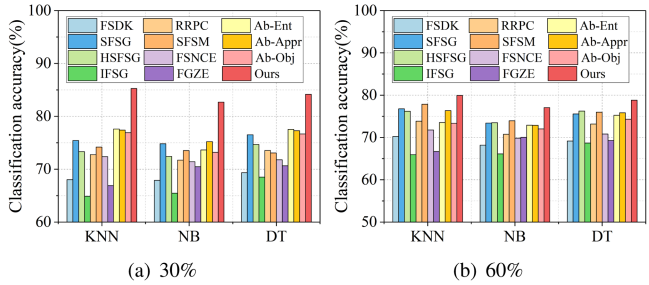


Fig. 7. Average accuracy of Ce-MGK under 30% and 60% proportions.

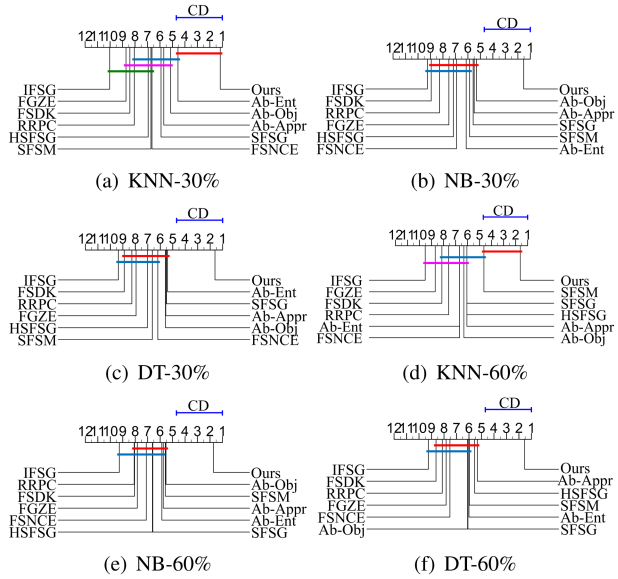


Fig. 8. The Nemenyi post-hoc test results on three classifiers under 30% and 60% missing label proportions.

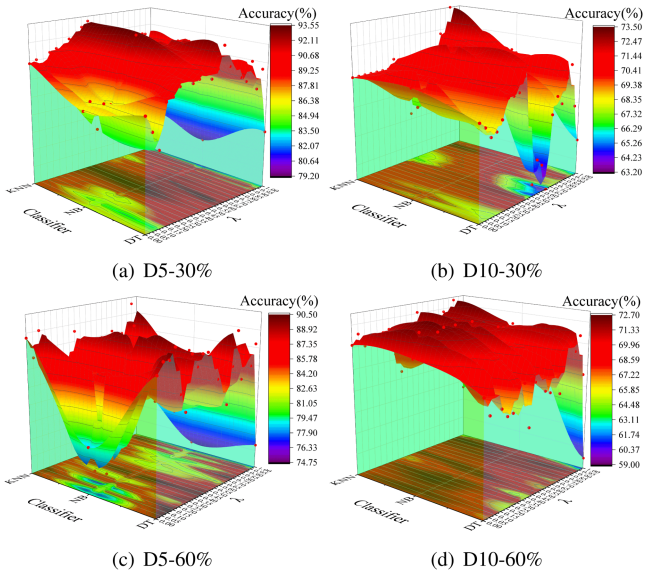


Fig. 9. Accuracy of Ce-MGK under different missing proportions.

methods in only 2 and 1 cases. Moreover, at the 60% ratio, Ce-MGK achieves the best performance in 17, 16, and 16 cases, while the ablation variants only occasionally obtain excellent results. These results illustrate both the effectiveness of consistent object and multi-level uncertainty analysis in semi-supervised feature selection.

D. Statistical Significance

To further assess the significant differences among these compared methods, the Friedman test [52] is first conducted at a significance level of $\alpha = 0.1$. The null hypothesis assumes that all methods perform equally, and it is rejected when the p-value is smaller than α. In this study, the obtained p-values are 6.58×10^{-13}, 8.92×10^{-10}, and 6.97×10^{-10} under the 30% setting, and 1.69×10^{-9}, 4.00×10^{-8}, and 2.72×10^{-9} under the 60% setting. These p-values are far smaller than the significance threshold, indicating the statistically significant differences among these methods.

Based on this finding, the Nemenyi post-hoc test [52] is further employed to examine pairwise differences between methods. The null hypothesis here supposes the equality of two methods, which is rejected when their average rank distance exceeds the critical distance. All the Nemenyi-hoc test results are presented in Fig. 8. From this figure, the proposed Ce-MGK consistently ranks first across all three classifiers under both missing label proportions. Specifically, Ce-MGK shows statistical significance over 10, 11, and 11 methods at the 30% setting, and over 10, 11, and 11 methods at the 60% setting. These results clearly validate the superiority of Ce-MGK in classification tasks from a statistical perspective.

E. Parameter Sensitivity Analysis

This section investigates the sensitivity of the proposed Ce-MGK method to the neighborhood parameter λ by evaluating its impact on classification accuracy. Representative results on datasets D5 and D10 with 30% and 60% missing label proportions are shown in Fig. 9.

As presented in Fig. 9, the accuracy fluctuates significantly with changes in λ, demonstrating that Ce-MGK is sensitive to this parameter. This behavior arises because λ directly affects the approximation process and the zentropy-based measure, which determines the effectiveness of semi-supervised feature selection. Moreover, the optimal λ value differs across datasets and missing label proportions. Therefore, properly adjusting λ is essential for selecting appropriate feature subsets and ensuring reliable classification performance.

VI. Conclusion

Ce-MGK is an effective multi-granularity knowledge modeling method for semi-supervised feature selection. Overall, it adopts threefold ideas: 1) a consistency-enhanced strategy to enhance the robustness of knowledge modeling, 2) multi-granularity knowledge representation for improving the semi-supervised data mining effectiveness, and 3) zentropy-based multi-level uncertainty reasoning and measurement. Theoretical

Authorized licensed use limited to: TONGJI UNIVERSITY. Downloaded on November 11,2026 at 13:57:24 UTC from IEEE Xplore. Restrictions apply.

and empirical studies show that Ce-MGK could achieve superior unlabeling resistance and classification performance.

The current study investigates a novel semi-supervised feature selection method by developing consistency-enhanced multi-granularity knowledge modeling. However, some limitations have not been considered, such as the computational cost of consistency-driven multi-granularity structures, especially in large-scale or streaming scenarios. Exploring acceleration strategies, such as matrix partitioning and distributed parallel computing, as well as incremental update mechanisms for fuzzy approximations and entropy measures, is an important research direction. Although the proposed Ce-MGK can significantly enhance the robustness and classification performance of the SFS, it is still unsuitable for open-world scenarios. Therefore, how to integrate evolutionary computing and sustainable learning theory into robust knowledge modeling and semi-supervised learning also deserves to be explored.

REFERENCES

[1] H. Gomes, J. Read, M. Grzenda, B. Pfahringer, and A. Bifet, "SLEADE: Disagreement-based semi-supervised learning for sparsely labeled evolving data streams," *IEEE Trans. Knowl. Data Eng.*, vol. 38, no. 3, pp. 1973–1985, Mar. 2026.

[2] G. Li, Z. Yu, K. Yang, C. L. P. Chen, and X. Li, "Ensemble-enhanced semi-supervised learning with optimized graph construction for high-dimensional data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 47, no. 2, pp. 1103–1119, Feb. 2025.

[3] Q. Liu, S. Liu, X. Liu, and J. Dai, "Latent semantics and anchor graph multi-layer learning for multi-view unsupervised feature selection," *IEEE Trans. Knowl. Data Eng.*, vol. 37, no. 10, pp. 6032–6045, Oct. 2025.

[4] M. Eftekhari et al., *How Fuzzy Concepts Contribute to Machine Learning*. Cham, Switzerland: Springer, 2022.

[5] D. Guo et al., "M-FCCL: Memory-based concept-cognitive learning for dynamic fuzzy data classification and knowledge fusion," *Inf. Fusion*, vol. 100, 2023, Art. no. 101962.

[6] K. Yuan, D. Miao, W. Ding, W. Pedrycz, and Y. Yao, "Robust semi-supervised feture selection with multi-granularity zentropy modeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 48, no. 4, pp. 4587–4604, Apr. 2026.

[7] K. Benabdeslem, D. E. K. Mansouri, and R. Makkhongkaew, "SCOs: Semi-supervised co-selection by a similarity preserving approach," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 6, pp. 2899–2911, Jun. 2022.

[8] D. Shi, L. Zhu, J. Li, Z. Cheng, and Z. Liu, "Binary label learning for semi-supervised feature selection," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 3, pp. 2299–2312, Mar. 2023.

[9] Q. Liu, S. Liu, J. Dai, X. Zhu, and X. Liu, "Linear complexity multi-view unsupervised feature selection via anchor-based feature relationship construction," *IEEE Trans. Image Process.*, vol. 34, pp. 7889–7902, 2025.

[10] K. Yuan, D. Miao, W. Pedrycz, and Y. Yao, "Zentropy-enhanced multi-granularity knowledge modeling for robust feature selection," *IEEE Trans. Cybern.*, vol. 56, no. 7, pp. 4096–4107, Jul. 2026.

[11] Z. Bao, Z. Tan, J. Wan, X. Ma, G. Guo, and Z. Lei, "Divergence-driven consistency training for semi-supervised facial age estimation," *IEEE Trans. Inf. Forensics Secur.*, vol. 18, pp. 221–232, 2023.

[12] N. Gaw, J. Li, and H. Yoon, "A novel semi-supervised learning model for smartphone-based health telemonitoring," *IEEE Trans. Automat. Sci. Eng.*, vol. 21, no. 1, pp. 428–441, Jan. 2024.

[13] F. Nie, Z. Wang, R. Wang, and X. Li, "Submanifold-preserving discriminant analysis with an auto-optimized graph," *IEEE Trans. Cybern.*, vol. 50, no. 8, pp. 3682–3695, Aug. 2020.

[14] R. Sheikhpour, K. Berahmand, and S. Forouzandeh, "Hessian-based semi-supervised feature selection using generalized uncorrelated constraint," *Knowl.-Based Syst.*, vol. 269, 2023, Art. no. 110521.

[15] K. Sechidis and G. Brown, "Simple strategies for semi-supervised feature selection," *Mach. Learn.*, vol. 107, no. 2, pp. 357–395, 2018.

[16] Z. Qiu et al., "A-SFS: Semi-supervised feature selection based on multi-task self-supervision," *Knowl.-Based Syst.*, vol. 252, 2024, Art. no. 109449.

[17] J. Xu, B. Tang, H. He, and H. Man, "Semisupervised feature selection based on relevance and redundancy criteria," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 9, pp. 1974–1984, Sep. 2017.

[18] R. Sheikhpour et al., "Semi-supervised feature selection with concept factorization and robust label learning," in *Pattern Recognit.*, vol. 172, 2026, Art. no. 112317.

[19] J. Tan, Z. Qiu, and N. Gui, "Graph representation learning enhanced semi-supervised feature selection," *ACM Trans. Knowl. Discov. Data*, vol. 18, no. 9, pp. 1–20, 2024.

[20] W. Shi, Y. Gong, B. Chen, and X. Hei, "Transductive semisupervised deep hashing," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 8, pp. 3713–3726, Aug. 2022.

[21] J. Tan, N. Gui, and Z. Qiu, "GAEFS: Self-supervised graph auto-encoder enhanced feature selection," *Knowl.-Based Syst.*, vol. 290, 2024, Art. no 111523.

[22] J. Wang et al., "Deep semi-supervised learning method based on sample adaptive weights and discriminative feature learning," *J. Vis. Commun. Image Representation*, vol. 115, 2026, Art. no 104689.

[23] K. Liu et al., "A survey on rough feature selection: Recent advances and challenges," *IEEE/CAA J. Automatica Sinica*, vol. 13, no. 3, pp. 521–542, Mar. 2026.

[24] X. Li, Y. Zhang, and R. Zhang, "Semisupervised feature selection via generalized uncorrelated constraint and manifold embedding," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 9, pp. 5070–5079, Sep. 2022.

[25] 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.

[26] Z. Huang, J. Li, and C. Wang, "Robust feature selection using multi-granulation variable-precision distinguishing indicators for fuzzy covering decision systems," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 54, no. 2, pp. 903–914, Feb. 2024.

[27] 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.

[28] 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.

[29] Y. Ding, W. Xu, W. Ding, and Y. Qian, "IFCRL: Interval-intent fuzzy concept re-cognition learning model," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 6, pp. 3581–3593, Jun. 2024.

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

[31] 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.

[32] S. Liu, J. Wang, R. Jiao, J. Zhou, and K. Lu, "Ensemble workload prediction with fluctuation division control in the computing power network," *IEEE Trans. Cloud Comput.*, vol. 14, no. 2, pp. 475–489, Second Quarter 2026.

[33] C. Luo, S. Wang, T. Li, H. Chen, J. Lv, and Z. Yi, "Large-scale meta-heuristic feature selection based on BPSO assisted rough hypercuboid approach," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 12, pp. 10889–10903, Dec. 2023.

[34] 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.

[35] D. Guo, W. Xu, S. Xia, W. Ding, Y. Qian, and K. Yuan, "Adaptive granular-ball concept-cognitive learning for efficient and robust fuzzy knowledge representation in classification tasks," *IEEE Trans. Fuzzy Syst.*, vol. 34, no. 5, pp. 1523–1535, May 2026.

[36] P. Liang et al., "Feature selection based on robust fuzzy rough sets using kernel-based similarity and relative classification uncertainty measures," *Knowl.-Based Syst.*, vol. 255, 2022, Art. no 109795.

[37] 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.

[38] N. Zhou, S. Liao, H. Chen, W. Ding, and Y. Lu, "Semisupervised feature selection with multiscale fuzzy information fusion: From both global and local perspectives," *IEEE Trans. Fuzzy Syst.*, vol. 33, no. 6, pp. 1825–1839, Jun. 2025.

Authorized licensed use limited to: TONGJI UNIVERSITY. Downloaded on November 11,2026 at 13:57:24 UTC from IEEE Xplore. Restrictions apply.

[39] T. Yang, Y. Deng, B. Yu, Y. Qian, and J. Dai,, "Local feature selection for large-scale data sets with limited labels," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 7, pp. 7152–7163, Jul. 2023.
[40] Z. Yan, S. He, J. Yu, and C. Wu, "Adaptive granules-based semi-supervised feature selection for hybrid data," *IEEE Trans. Knowl. Data Eng.*, vol. 38, no. 5, pp. 2561–2577, May 2026.
[41] S. An et al., "Robust fuzzy rough approximations with kNN granules for semi-supervised feature selection," *Fuzzy Sets Syst.*, vol. 461, 2023, Art. no 108476.
[42] K. Liu, T. Li, X. Yang, H. Chen, J. Wang, and Z. Deng, "SemiFREE: Semisupervised feature selection with fuzzy relevance and redundancy," *IEEE Trans. Fuzzy Syst.*, vol. 31, no. 10, pp. 3384–3396, Oct. 2023.
[43] C. Madhu and S. M. S., "An interpretable fuzzy graph learning for label propagation assisting data classification," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 3, pp. 1331–1345, Mar. 2024.
[44] R. Xu, D. Wu, R. Wang, and X. Luo, "A highly-accurate three-way decision-incorporated online sparse streaming features selection model," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 55, pp. 6, pp. 4258–4272, Jun. 2025.
[45] 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.
[46] D. Berthelot et al., "MixMatch: A holistic approach to semi-supervised learning," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, 2019, pp. 5049–5059.
[47] X. Zhang, C. Mei, D. Chen, Y. Yang, and J. Li, "Active incremental feature selection using a fuzzy-rough-set-based information entropy," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 5, pp. 901–915, May 2020.
[48] X. Yang et al., "A noise-aware fuzzy rough set approach for feature selection," *Knowl.-Based Syst.*, vol. 250, 2022, Art. no 109092.
[49] F. Nie, Z. Ma, J. Wang, and X. Li,, "Fast sparse discriminative K-Means for unsupervised feature selection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 7, pp. 9943–9957, Jul. 2024.
[50] G. Roffo, S. Melzi, U. Castellani, A. Vinciarelli, and M. Cristani, "Infinite feature selection: A graph-based feature filtering approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 12, pp. 4396–4410, Dec. 2021.
[51] V. Laurens and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, 2008.
[52] J. Demsar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, 2006.



Kehua Yuan received the MSc degree from the College of Artificial Intelligence, Southwest University, Chongqing, China, in 2022. She is currently working toward the PhD degree with the School of Computer Science and Technology, Tongji University, Shanghai, China. She has authored or coauthored on these subjects at international journals such as *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *IEEE Transactions on Knowledge and Data Engineering*, *IEEE Transactions on Cybernetics*, *IEEE Transactions on Fuzzy Systems*, *IEEE Transactions on Neural Networks and Learning Systems*. Her research interests include granular computing, uncertainty analysis, and machine learning.



Duoqian Miao is currently a professor with the School of computer science and technology, Tongji University, Shanghai, China. He has authored or coauthored more than 200 academic papers, which include prestigious journals and international conferences. His research interests include rough sets, soft computing, machine learning and intelligent systems. He is a fellow of International Rough Set Society and a Director of Chinese Association for Artificial Intelligence (CAAI). He is currently the associate editors of *Information Sciences*, *CAAI Transactions on Intelligence Technology*, and *International Journal of Approximate Reasoning*.



Weiping Ding (Senior Member, IEEE) received the PhD degree in computer science, from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2013. From 2014 to 2015, he was a postdoctoral researcher with the Brain Research Center, National Chiao Tung University, Hsinchu, Taiwan, China. In 2016, he was a visiting scholar with the National University of Singapore, Singapore. From 2017 to 2018, he was a visiting professor with the University of Technology Sydney, Australia. He is currently a full professor with Nantong University. His research directions involve granular data mining and multimodal machine learning. He has authored or coauthored more than 400 articles, including more than 220 IEEE Transactions papers, twenty papers have been selected as ESI Highly Cited Papers. He serves as an associate editor/area editor/editorial board member of more than 10 international prestigious journals, such as *IEEE Transactions on Neural Networks and Learning Systems, IEEE Transactions on Fuzzy Systems, IEEE/CAA Journal of Automatica Sinica, IEEE Transactions on Emerging Topics in Computational Intelligence, IEEE Transactions on Intelligent Transportation Systems, Information Fusion, Neurocomputing,* and*Applied Soft Computing*. He was the leading guest editor of special issues in several prestigious journals, including *IEEE Transactions on Evolutionary Computation, IEEE Transactions on Fuzzy Systems,* and*Information Fusion*.



Witold Pedrycz (Life Fellow, IEEE) received the MSc, PhD, and DSc degrees from the Silesian University of Technology, Gliwice, Poland, in 1977, 1980, and 1984, respectively. He is currently 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. He also holds an appointment as a special professor with the School of Computer Science, University of Nottingham, Nottingham, U.K. His 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. He 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 *Information Sciences*, *WIREs Data Mining and Knowledge Discovery*, and *Granular Computing*.



Hongyun Zhang received the PhD degree in pattern recognition and intelligence system from Tongji University, Shanghai, China, in 2005. She is currently an associate professor with Tongji University. She has the authored or coauthored of more than 60 journal papers and conference proceedings in Principal curves, pattern recognition, machine learning, granular computing, and rough set. Her research interests include Principal curves, pattern recognition, data mining, image retrieval, and granular computing.



Yiyu Yao received the BE degree in computer science from Xi'an Jiaotong University, Xi'an, China, in 1983, and the MSc and PhD 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, Canada. His research interests include three-way decision, granular computing, rough sets, formal concept analysis, information retrieval, data mining, and Web intelligence. He proposed a theory of three-way decision, a decision-theoretic rough set model, and a triarchic theory of granular computing. He has authored or coauthored more than 400 papers. He was a highly cited researcher from 2015 to 2019.

Authorized licensed use limited to: TONGJI UNIVERSITY. Downloaded on November 11,2026 at 13:57:24 UTC from IEEE Xplore. Restrictions apply.