Robust Semi-supervised Feature Selection with Multi-granularity Zentropy Modeling

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

***Abstract*—High-dimensional and weakly supervised (HiDWS) data present significant challenges for traditional machine learning and pattern recognition. Although semi-supervised feature selection has shown effectiveness in improving the quality of HiDWS data, existing methods remain sensitive and lack robustness due to the unreliability of unlabeled data learning and the uncertainty in modeling processes. Hence, this study focuses on a multi-granularity zentropy modeling (Ze-MGM) framework with model-agnostic for highly-accuracy and robust semi-supervised feature selection. Unlike existing methods, Ze-MGM does not rely on specific settings such as rough or fuzzy set assumptions and can effectively capture the granularity of information under high-dimensional and weakly supervised data scenarios. Specifically, we first introduce a strategic soft label ($S2-$Label) learning method that integrates object proximity and classification certainty to reduce uncertainty between features and labels. This method also enables the selection of compatible instances, thereby mitigating the negative impact of incompatible objects on label learning. Subsequently, a multi-granularity knowledge space and zentropy uncertainty measure are constructed by analyzing the hierarchical relationships among labels, decisions, and specific classes, which enables accurate multi-granularity knowledge representation and multi-granularity uncertainty characterization in HiDWS data modeling processing. Finally, two multi-granularity significance measures based on multi-granularity uncertainty are defined for feature evaluation and selection via a semi-supervised paradigm. Extensive experiments on multiple benchmark datasets demonstrate that the proposed Ze-MGM method achieves superior generalization performance and robustness compared to state-of-the-art methods.**

***Index Terms*—Granular computing; semi-supervised learning; feature selection; data mining; uncertainty**

I. Introduction

LOW-QUALITY data is increasingly becoming a commonplace challenge in data science and machine learning [1], [2]. Highly accurate extraction of reliable patterns and knowledge from this low-quality data is a hot and challenging problem in intelligent data analysis processing [3]–[5]. In practice, low-quality data often exhibit complex characteristics such as multimodal diversity, high-order correlations, and continuous evolution, which significantly increase uncertainty in data processing and model construction [6], [7]. Particularly, the phenomenon of missing labels, also called HiDWS data, further enhances learning challenges due to human errors in record-keeping, high acquisition costs, and sensing technology constraints [8]–[10]. This issue is particularly prevalent in high-dimensionality scenarios. Consequently, effectively exploiting the latent structure of unlabeled data is necessary for robust and generalizable modeling.

Semi-supervised learning (SSL) has shown strong potential by jointly utilizing labeled and unlabeled samples [11]–[13]. Within SSL, semi-supervised feature selection (SSFS) is particularly effective for high-dimensional data reduction, as it combines supervised guidance with structural insights from unlabeled data [14]–[17]. Despite its success, SSFS still faces two significant challenges: (1) **High sensitivity to the quality of unlabeled data**, as the effectiveness of model training is significantly affected by the unlabeled data structure and the pseudo-labels reliability; and (2) **Limited robustness in feature selection**, since the inherent uncertainty in both data structure and the modeling process makes it difficult to ensure stable and accurate assessments of feature evaluation, often resulting in suboptimal feature subsets. Therefore, developing an effective and robust approach to mitigate the inherent uncertainty while fully exploiting data structure in HiDWS becomes essential for advancing semi-supervised feature selection.

To address the first challenge, several studies exploit geometric and statistical structures resulting from unlabeled data [18]–[21]. For example, Sheikhpour et al. [20] enhanced structure modeling via Hessian-based graphs. Li et al. [21] extended the uncorrelated constraint framework based on ridge regression, where unlabeled samples provide geometric structural information. Others adopt metric-adaptive strategies using discernibility matrices or correlation measures [22], [23]. However, most fail to capture implicit relationships between labeled and unlabeled samples. In contrast, pseudo-labeling methods offer better integration. Techniques include spatial filtering [24], binary hashing [25], and fuzzy rough sets [26]. Nevertheless, these methods suffer from two main issues: (1) **Inter-class consistency assumption**: They primarily focus on intra-class similarity or membership degrees while neglecting inter-class variability and the underlying data distribution, potentially introducing bias in label inference; and (2) **Sample compatibility assumption**: The assumption that all samples

This paper was supported in part by the National Natural Science Foundation of China under Grant 624B2104, Grant 62576251, and Grant 62376198.*(Corresponding author: Duoqian Miao.)*

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

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

Witold Pedrycz is with the Silesian University of Technology (SUT), Department of Measurement and Control Systems, Gliwice, Akademicka 2, 44-100 Poland, Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2R3, Canada, also with the Research Center of Performance and Productivity Analysis, Istinye University, Istanbul, Türkiye(e-mail: wpedrycz@ualberta.ca).

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

are equally compatible and necessary for label propagation is inconsistent with real-world scenarios, which can degrade the accuracy of unlabeled data learning and model robustness. Therefore, a more effective label learning method that explicitly accounts for intra-class diversity and sample compatibility is crucial for enhancing learning effectiveness in partially labeled data.

To address the second challenge, recently efforts integrating uncertainty quantification have been proposed to enhance SSFS robustness [27]–[33]. For instance, Goffo et al. [28] designed a graph-based framework capturing feature relevance and redundancy via label mutual information. In unsupervised settings, Li et al. [33] leveraged Bayesian variational learning to model uncertainty. In particular, multi-granularity computing methods offer a new perspective on uncertainty measure by simulating human cognitive and reasoning mechanisms across multiple granularity levels when handling complex problems [34]–[39]. Zhang et al. [34] proposed combination entropy on similarity for heterogeneous data. Xu et al. [35] introduced a composite entropy tailored to distributions for feature evaluation. To support multi-scale data, Sang et al. [38] developed fuzzy rough combination entropy on a optimal scale for feature selection. Moreover, Yuan et al. [39] investigated a zentropy-based measure that integrates the granularity structure of rough sets for attribute reduction. Additionally, granular ball-based approaches [40], [41] have emerged, combining clustering and granularity for efficient selection. Although these uncertainty-aware methods have demonstrated promising results, two key issues remain: (1) **Assumption-specific modeling**: Existing uncertainty modeling strategies, such as approximation accuracy in rough set, belief and plausibility in Dempster-shafer theory, confidence interval in probabilistic graph, are often limited to specific modeling assumptions, which hinders their generalizability in diverse domains and learning paradigms; and (2) **Single granular level modeling**: Most approaches focus on modeling uncertainty at a single granularity level, overlooking the hierarchical structures and interactions inherent in both data and models. Therefore, developing a generalizable and robust multi-granularity uncertainty modeling framework is critical for high-dimensional and weakly supervised learning scenarios.

Inspired by the preceding discussion, this paper introduces a generalizable multi-granularity zentropy modeling framework focusing on the perspective of multi-granularity uncertainty. This approach enables robust and effective feature selection by capturing uncertainty across different granularities. The main contributions of the study are as follows:

1) **A generalizable multi-granularity modeling framework**: We develop a multi-granularity zentropy modeling framework with model-independent for knowledge acquisition quantitation in HiDWS data, which breaks the limitations of specific model-dependent in granular computing by the multi-granularity modeling in the intrinsic data structure.
2) **A novel soft label strategy learning method**: We provide a new $S2-$Label method based on strategic object selection and evaluation to enhance label inference accuracy in semi-supervised learning. Compared with the existing boolean and soft learning methods, it has a more precise and robust ability to label learning by jointly considering data multi-granularity distribution and uncertainty.
3) **An effective zentropy uncertainty measure mechanism**: We propose a zentropy uncertainty measure based on multi-granularity knowledge space, which can explicitly analyze and measure uncertainty in HiDWS data. It is a generalized and principled approach that significantly enhances knowledge quantification capability by considering the granularity of hierarchy and interaction.
4) **A robust semi-supervised feature selection strategy**: We design a robust feature selection strategy based on Ze-MGM, which effectively reduce the redundant features' impact and improve HiDWS data quality. Extensive experiments on benchmark datasets demonstrate its high-accuracy and robustness compared with other methods.

The remainder of this paper is organized as follows. Section II reviews and analyzes the related work. Section III introduces some basic concepts. The proposed Ze-MGM is presented in Section IV. Section V describes the multi-granularity reduction strategy for partially labeled data. Experimental results and analyses are provided in Section VI. Finally, Section VII concludes the paper and outlines future research directions.

II. Preliminaries

Considering that the proposed Ze-MGM leverages the multi-granularity zentropy modeling to address the challenge of partial label feature selection, this section provides a brief overview of the fundamental concepts related to HiDWS, information granulation, and zentropy modeling, with more detailed discussions available in [42]–[44].

A. Problem Definition

The high-dimensional and weakly supervised data can be represented as a triplet $HiDWS = (\mathcal{O}, \mathcal{A} \cup \mathcal{L})$, where:

- $\mathcal{O} = \mathcal{O}^L \cup \mathcal{O}^U$ represents the universe consisting of n objects, with $\mathcal{O}^L = \{x_i\}_{i=1}^{n_1}$ denoting the set of labeled objects and $\mathcal{O}^U = \{x_j\}_{j=1}^{n_2}$ representing the set of unlabeled objects.
- $\mathcal{A} = \{a_1, a_2, \ldots, a_m\}$ is the set of conditional features. For any $a \in \mathcal{A}$ and $x \in \mathcal{O}$, $a(x)$ denotes the feature value of object x with respect to feature a.
- $\mathcal{L} = \mathcal{L}^L \cup \mathcal{L}^U$ is the decision label set corresponding to $\mathcal{O}$, where $\mathcal{L}^L$ contains the labels for the objects in $\mathcal{O}^L$, and $\mathcal{L}^U$ contains the labels for the objects in $\mathcal{O}^U$. For any $x \in \mathcal{O}^U$, the decision label $\mathcal{L}^U(x)$ is marked as $*$.

With respect to the label information, the labeled objects are divided into classes as $\mathcal{O}^L/\mathcal{L}^L = \{Y_1, Y_2, \ldots, Y_s\}$. Without loss of generality, we assume that no new class label in $\mathcal{L}^U$ differs from those in $\mathcal{L}^L$. When the unlabeled object set $\mathcal{O}^U$ is empty, the partial decision system is called supervised data.

To improve model performance and reduce computational complexity in HiDWS data, feature selection has been widely studied as a crucial technique. From an information theoretic perspective, the objective is to identify a feature subset that minimizes uncertainty, which can be formulated as follows.

Definition 1. *[42] Given a $HiDWS = (\mathcal{O}, \mathcal{A} \cup \mathcal{L})$. Let $\mathcal{S}^*$ be a subset of $\mathcal{A}$. It is considered optimal if it satisfies:*

$$\mathcal{S}^* = \arg \min_{\mathcal{S} \subseteq \mathcal{A},\, |\mathcal{S}| \leq m} \mathcal{J}(\mathcal{S}), \tag{1}$$

where $\mathcal{J}$ is an uncertainty evaluate function.

In practice, exhaustively searching for the global optimal solution is highly time-consuming or infeasible. Therefore, it is common to seek a locally optimal solution $\mathcal{S}^*$ that achieves a better evaluation score than the initial candidates in $\mathcal{J}$.

B. Multi-granularity Knowledge Characterization

Information granules are the basic units that intuitively characterize complex problem constructs, which play a fundamental role in cognitive learning and decision activities. Among various granulation techniques, the fuzzy similarity relation provides an intuitive and practical approach and has been extensively utilized in supervised, semi-supervised, and unsupervised learning scenarios.

Definition 2. *[42] Given a $HiDWS = (\mathcal{O}, \mathcal{A} \cup \mathcal{L})$. For any $\mathcal{B} \subseteq \mathcal{A}$, the relation $\mathcal{R}_\mathcal{B} : \mathcal{O} \times \mathcal{O} \to [0, 1]$ is called a fuzzy similarity relation on $\mathcal{B}$ if it satisfies the following properties:*

1) ***Reflexivity**: $\forall x \in \mathcal{O}, \mathcal{R}_\mathcal{B}(x, x) = 1$;*
2) ***Symmetry**: $\forall x, y \in \mathcal{O}, \mathcal{R}_\mathcal{B}(x, y) = \mathcal{R}_\mathcal{B}(y, x)$.*

For any $x \in \mathcal{O}$, the fuzzy similarity object granule of x in terms of $\mathcal{B}$ is expressed as:

$$[x]_{\mathcal{R}_\mathcal{B}} = \frac{[x]_{\mathcal{R}_\mathcal{B}}(x_1)}{x_1} + \frac{[x]_{\mathcal{R}_\mathcal{B}}(x_2)}{x_2} + \cdots + \frac{[x]_{\mathcal{R}_\mathcal{B}}(x_n)}{x_n}, \tag{2}$$

where $[x]_{\mathcal{R}_\mathcal{B}}(x_i) = \mathcal{R}_\mathcal{B}(x, x_i)$ represents the membership degree of object x_i $(i = 1, 2, \ldots, n)$ to granule $[x]_{\mathcal{R}_\mathcal{B}}$. The cardinality of the granule, denoted as $|[x]_{\mathcal{R}_\mathcal{B}}|$, is computed as:

$$|[x]_{\mathcal{R}_\mathcal{B}}| = \sum_{i=1}^{n} [x]_{\mathcal{R}_\mathcal{B}}(x_i). \tag{3}$$

The fuzzy object granules of $\mathcal{O}$ induced by the similarity relation $\mathcal{R}_\mathcal{B}$ are collectively represented as:

$$\mathcal{G}^{\mathcal{O}}_{\mathcal{R}_\mathcal{B}} = \{[x_1]_{\mathcal{R}_\mathcal{B}}, [x_2]_{\mathcal{R}_\mathcal{B}}, \ldots, [x_n]_{\mathcal{R}_\mathcal{B}}\}. \tag{4}$$

In practical applications, fuzzy granulation methods based on distance functions are commonly employed due to their intuitive and computationally efficient characteristics. Let $\mathcal{B} \subseteq \mathcal{A}$ and for all $x, y \in \mathcal{O}$, consider the fuzzy similarity relation defined by $\mathcal{R}_\mathcal{B}(x, y) = e^{-d_\mathcal{B}(x,y)}$, where $d_\mathcal{B}(x, y)$ represents the distance between x and y induced by $\mathcal{B}$. For any subset $\mathcal{C} \subseteq \mathcal{B} \subseteq \mathcal{A}$, the fuzzy object granules exhibit variations in granulation thickness, such that $\mathcal{G}^{\mathcal{O}}_{\mathcal{R}_\mathcal{B}} \subseteq \mathcal{G}^{\mathcal{O}}_{\mathcal{R}_\mathcal{C}}$.

C. Zentropy in Rough Sets

Entropy [45], defined on a given probability distribution, is an effective tool for characterizing systems uncertainty. However, it operates under single-level conditions, which limits the capture of the multi-granularity structure inherent in knowledge acquisition and computational processes. The zentropy measure relying on rough sets was introduced to address this limitation, as illustrated in Fig. 1, extending traditional entropy concepts to handle multi-level structures [44], [46]. This approach has proven particularly successful in thermal negative expansion of Fe_3Pt and uncertainty measurement [46], [47]. In a decision information system $\mathcal{I} = (\mathcal{O}, \mathcal{A} \cup \mathcal{L})$, the zentropy measure is designed to characterize uncertainty under the rough set context.

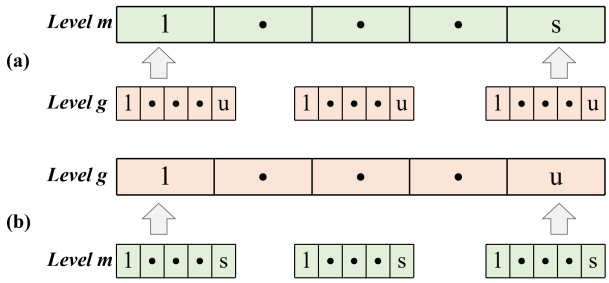


Fig. 1. Relationship among multiple levels. (a) At level m, there are s subconfigurations, where each subconfiguration can be decomposed into finer subconfigurations at level g. (b) Similarly, each subconfiguration at level g determines the overall performance of configurations at coarser level m.

Definition 3. *[44] Given a $\mathcal{I} = (\mathcal{O}, \mathcal{A} \cup \mathcal{L})$, where $\mathcal{O}/\mathcal{L} = \{Y_1, Y_2, \ldots, Y_s\}$. For any $\mathcal{B} \subseteq \mathcal{A}$ and $Y_i \in \mathcal{O}/\mathcal{L}$, the neighborhood lower and upper approximations of Y_i are defined as follows:*

$$\begin{aligned} \underline{\mathcal{R}_\mathcal{B}}(Y_i) &= \{x \in \mathcal{O} | N^\delta_\mathcal{B}(x) \subseteq Y_i\}, \\ \overline{\mathcal{R}_\mathcal{B}}(Y_i) &= \{x \in \mathcal{O} | N^\delta_\mathcal{B}(x) \cap Y_i \neq \emptyset\}. \end{aligned} \tag{5}$$

where $N^\delta_\mathcal{B}(x) = \{y \in \mathcal{O} | d_\mathcal{B}(x, y) \leq \delta\}$ is the neighborhood similarity class of x under $\mathcal{B}$. The pair $< \underline{\mathcal{R}_\mathcal{B}}(Y_i), \overline{\mathcal{R}_\mathcal{B}}(Y_i) >$ is called neighborhood approximation space.

As pointed out in [44], the rough approximation process is influenced by four granular levels, gradually refining from the target decision to the approximation set, similarity class, and specific objects. To accurately capture the information presented at different levels, the zentropy measure is proposed.

Definition 4. *[43] Given a $\mathcal{I} = (\mathcal{O}, \mathcal{A} \cup \mathcal{L})$, where $\mathcal{O}/\mathcal{L} = \{Y_1, Y_2, \ldots, Y_s\}$. For any $\mathcal{B} \subseteq \mathcal{A}$ and $Y_i \in \mathcal{O}/\mathcal{L}$, with $< \underline{\mathcal{R}_\mathcal{B}}(Y_i), \overline{\mathcal{R}_\mathcal{B}}(Y_i) >$ being the neighborhood approximation space, the zentropy measure is defined as follows:*

$$\mathcal{Z}_\mathcal{B}(\mathcal{O}) = -\sum_{i=1}^{s} p_i \log p_i + \sum_{i=1}^{s} p_k \mathcal{Z}_i, \tag{6}$$

where $-\sum_{i=1}^{s} p_i \log p_i$ is the external entropy, reflecting the current information. The $\mathcal{Z}_i$ denotes the internal entropy of the i-th granule, which can be further decomposed using the same formula as in Eq. (6) at finer levels.

In uncertainty measure and feature selection, this zentropy measure has shown strong robustness and effectiveness in multi-granularity data analysis [43], [44]. However, it is subject to two primary limitations: (1) it is inherently restricted to fully labeled datasets, which hinders its applicability in scenarios involving partial label learning; and (2) it is constrained by the specific rough approximation process, resulting in high computational complexity and a lack of generalizability, thus

restricting its broader applicability across various domains. Therefore, a universal framework that can accommodate partial label learning while enhancing computational efficiency and generalization across diversion is necessary.

III. Multi-granularity Zentropy Modeling

Considering the inherent multi-granularity characteristic in data processing and modeling, this section introduces a multi-granularity zentropy model for high-dimensional and weakly supervised data. Compared with existing methods, the proposed framework enhances robustness and effectiveness, offering a generalizable solution across diverse scenarios.

A. Strategic Soft Label Learning Method

This subsection introduces a label learning method for HiDWS data that involves two key steps: 1) a strategic selection mechanism for identifying data points compatible with the labeled data distribution and 2) a soft learning method for handling unlabeled data.

The first step selects a compatible subset of unlabeled data, $\mathcal{O}^U_{\text{sub}} \subseteq \mathcal{O}^U$, where compatibility refers to the lower proximity of objects to the existing decision classes in the feature space.

Definition 5. *Given a $HiDWS = (\mathcal{O}, \mathcal{A} \cup \mathcal{L})$, where $\mathcal{O} = \mathcal{O}^L \cup \mathcal{O}^U$ and $\mathcal{O}^L/\mathcal{L}^L = \{Y_1, Y_2, \ldots, Y_s\}$. The compatible objects can be formulated as follows:*

$$\mathcal{O}^U_{sub} = \left\{ x \in \mathcal{O}^U \middle| \|\mathcal{A}(x) - \overline{\mathcal{A}}(Y_j)\|_{\mathcal{A}} \leq \hat{\sigma}_{\mathcal{A}}(Y_j), \exists Y_j \in \mathcal{O}^L/\mathcal{L}^L \right\}, \tag{7}$$

where $\|\cdot\|_{\mathcal{A}}$ denotes the Euclidean distance on the feature set $\mathcal{A}$, $\overline{A}(Y_j)$ and $\hat{\sigma}_{\mathcal{A}}(Y_j)$ are the mean and standard deviation, respectively, of $j-$th class Y_j within the $\mathcal{A}$.

By removing out-of-distribution objects from $\mathcal{O}^U$, we construct a strategically selected universe $\mathcal{O}^{'} = \mathcal{O}^L \cup \mathcal{O}^U_{sub}$ as defined above. Here, out-of-distribution refers to objects that are located far from all decision classes in the feature space. In the second step, a soft label learning method is applied to these "friendly" objects, effectively utilizing information within the weakly supervised data.

Definition 6. *Given a $HiDWS = (\mathcal{O}^{'}, \mathcal{A} \cup \mathcal{L})$, where $\mathcal{O}^{'} = \mathcal{O}^L \cup \mathcal{O}^U_{sub}$ and $\mathcal{O}^L/\mathcal{L}^L = \{Y_1, Y_2, \ldots, Y_s\}$. For any $Y_j \in \mathcal{O}^L/\mathcal{L}^L$ and $x \in \mathcal{O}^U_{sub}$, the classify certainty of x to Y_j is computed as follows:*

$$\Phi^o_x(Y_j) = \frac{\sum_{y \in Y_j} e^{-\frac{\|a(x)-a(y)\|_{\mathcal{A}}}{1+\hat{\sigma}_{\mathcal{A}}(Y_j)}}}{|Y_j|}, \tag{8}$$

where $|Y_j|$ is the cardinality of objects in Y_j.

Then, the classify certainty of x to each decision class can be obtained as

$$\Phi^o_x = \{\Phi^o_x(Y_1), \Phi^o_x(Y_2), \ldots, \Phi^o_x(Y_s)\}. \tag{9}$$

From the above equations, this defined certainty measure considers the similarity between x and other objects in the same class Y_j and the intra-class distribution, thus providing a comprehensive method for handling unlabeled data. Moreover, considering the differences among different classes, the inter-class separation is further refined to classify certainty.

Definition 7. *Given a $HiDWS = (\mathcal{O}^{'}, \mathcal{A} \cup \mathcal{L})$, where $\mathcal{O}^{'} = \mathcal{O}^L \cup \mathcal{O}^U_{sub}$ and $\mathcal{O}^L/\mathcal{L}^L = \{Y_1, Y_2, \ldots, Y_s\}$. For any $Y_j \in \mathcal{O}^L/\mathcal{L}^L$ and $x \in \mathcal{O}^U_{sub}$, the refined classify certainty of x to Y_j is computed as follows:*

$$\Phi_x(Y_j) = \frac{\Phi^o_x(Y_j)}{\sum_{i \neq j} \Phi^o_x(Y_i)}. \tag{10}$$

Based on the above definitions, the soft label for object $x \in \mathcal{O}^U_{sub}$ can be computed as:

$$\tilde{Y_j}(x) = \frac{\Phi_x(Y_j)}{\sum_{i=1}^{s} \Phi_x(Y_i)}. \tag{11}$$

Thus, the soft label vector for $x \in \mathcal{O}^U_{sub}$ is defined as $\tilde{\mathcal{L}}^U(x) = \{\tilde{Y_1}(x), \tilde{Y_2}(x), \ldots, \tilde{Y_s}(x)\}$. Similarly, for $x \in \mathcal{O}^L$, the label vector $\tilde{\mathcal{L}}^L(x)$ will be a unit vector with s dimensions, where $\tilde{Y_i}(x) = 1$ if $\mathcal{L}^L(x) = Y_i$, and 0 otherwise. All the label vector of objects in $\mathcal{O}^{'}$ is recorded as $\tilde{\mathcal{L}}$.

Algorithm 1: $S2$-Label learning algorithm

Input : $HiDWS = (\mathcal{O}, \mathcal{A} \cup \mathcal{L})$.
Output : The strategically selected $\mathcal{O}^{'}$ and soft label $\tilde{\mathcal{L}}$.

Initialize $\mathcal{O}^U_{sub} = \emptyset$;
Step I: Strategic selection mechanism
for $Y_j \in \mathcal{O}^L/\mathcal{L}^L$ **do**
  Calculate class center $\overline{\mathcal{A}}(Y_j)$ and standard deviation $\hat{\sigma}_{\mathcal{A}}(Y_j)$;
end
for $x \in \mathcal{O}^U$ **do**
  if $\|\mathcal{A}(x) - \overline{\mathcal{A}}(Y_j)\|_{\mathcal{A}} \leq \hat{\sigma}_{\mathcal{A}}(Y_j)0$ **then**
    $\mathcal{O}^U_{sub} = \mathcal{O}^U_{sub} \cup \{x\}$;
  end
end
$\mathcal{O}^{'} = \mathcal{O}^L \cup \mathcal{O}^U_{sub}$;
Step II: Soft label learning approach
for $x \in \mathcal{O}^{'}$ **do**
  if $x \in \mathcal{O}^U$ **then**
    for $Y_j \in \mathcal{O}^L/\mathcal{L}^L$ **do**
      Calculate $\Phi_x(Y_j)$ according to Eq. (10);
    end
    Calculate $\tilde{Y_j}(x)$ according to Eq. (11);
  end
  if $x \in \mathcal{O}^L$ **then**
    for $Y_j \in \mathcal{O}^L/\mathcal{L}^L$ **do**
      $\tilde{Y_i}(x) = 1$ when $\mathcal{L}^L(x) = Y_i$;
      $\tilde{Y_i}(x) = 0$ when $\mathcal{L}^L(x) \neq Y_i$;
    end
  end
  $\tilde{\mathcal{L}}(x) = \{\tilde{Y_1}(x), \tilde{Y_2}(x), \ldots, \tilde{Y_s}(x)\}$.
end

The S2-Label learning approach provides a more informative and nuanced alternative to Boolean labels. By considering the class standard deviation and comparing the relative classification certainty of the current class with that of other classes, it effectively captures both the intra-class distribution and inter-class separation. The detailed process is illustrated in Fig. 2 and Algorithm 1. In step I, this strategic selection reduces semi-supervised learning errors by excluding out-of-distribution or "unfriendly". The time complexity of this step is $O(n_1 m + n_2 s m)$, where n_1 is the number of labeled samples, n_2 is the number of unlabeled samples, m is the feature dimensionality, s is the number of class centers. Meanwhile, in step II, soft labels are assigned to the n_2' selected unlabeled

instances by jointly considering intra-class similarity and inter-class separation across s classes, which has a complexity of $O(n_2' sm)$. Therefore, the whole time complexity of S2-Label learning algorithm algorithm is $O(n_1 m + (n_2 + n_2^{'})sm)$.

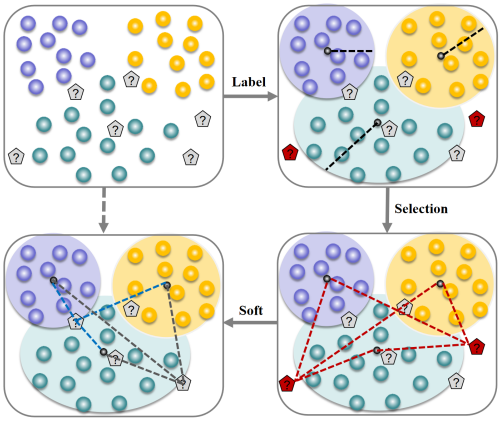


Fig. 2. S2-label learning process. The labeled data in $\mathcal{O}^L$ is first partitioned into s granules. Subsequently, compatible objects are identified by difference criteria defined by Eq. (7). Finally, the soft labels are derived based on the intra-class distribution in Eq. (8) and inter-class separation in Eq. (10).

B. Multi-granularity Knowledge Space Construction

Multi-granularity computing has emerged as a method that collaborates across multiple levels of granularity to discover the latent knowledge within data. To comprehensively capture the granularity information in HiDWS data, this subsection presents the multi-granularity knowledge space construction through relation granulation.

Definition 8. *Given a HiDWS* $= (\mathcal{O}^{'}, \mathcal{A} \cup \mathcal{L})$*, where* $\mathcal{O}^{'} = \mathcal{O}^L \cup \mathcal{O}^U_{sub}$ *and* $\mathcal{L} = \mathcal{L}^L \cup \mathcal{L}^U$*. The universe* $\mathcal{O}^{'}$ *can be observed at the first granularity level based on label information, which induces the following granulation:*

$$\mathcal{G}^1(\mathcal{L}) = \{\mathcal{G}_1^1(\mathcal{L}), \mathcal{G}_2^1(\mathcal{L})\}, \tag{12}$$

where $\mathcal{G}_1^1(\mathcal{L}) = \{x \in \mathcal{O}^{'} \mid \mathcal{L}(x) \in \mathcal{L}^L\}$ *and* $\mathcal{G}_2^1(\mathcal{L}) = \{x \in \mathcal{O}^{'} \mid \mathcal{L}(x) \notin \mathcal{L}^L\}$.

Specifically, for each granule $\mathcal{G}_i^1(\mathcal{L}) \in \mathcal{G}^1(\mathcal{L})$, where $i = 1, 2$, the granule is further refined as follows:

$$\mathcal{G}_i^1(\mathcal{L}) = \{\mathcal{G}_{i1}^2(Y_1), \mathcal{G}_{i2}^2(Y_2), \ldots, \mathcal{G}_{is}^2(Y_s)\}, \tag{13}$$

where $\mathcal{G}_{ij}^2(Y_j) = \frac{\tilde{Y}_j(x_1)}{x_1} + \frac{\tilde{Y}_j(x_2)}{x_2} + \cdots + \frac{\tilde{Y}_j(x_n^{'})}{x_n^{'}}$ represents the j-th decision granule at the second granularity level. Here, $\tilde{Y}_j(x_r)$ is the soft label of $x_r \in \mathcal{O}^{'}$ associated with Y_j, as defined in Def. 7.

Furthermore, by examining the relationships among the objects, each decision granule can be further refined.

$$\mathcal{G}_{ij}^2(Y_j) = \{\mathcal{G}_{ij1}^3(x_1), \mathcal{G}_{ij2}^3(x_2), \ldots, \mathcal{G}_{ijw}^3(x_w)\}, \tag{14}$$

where $j = 1, 2, \ldots s$, w being the object number in the granule $\mathcal{G}_{ij}^2(Y_j)$, $\mathcal{G}_{ijr}^3(x_r) = [x_r]_{\mathcal{R}_\mathcal{A}^\lambda} = \frac{\mathcal{R}_\mathcal{A}^\lambda(x_r, x_1)}{x_1} + \frac{\mathcal{R}_\mathcal{A}^\lambda(x_r, x_2)}{x_2} + \cdots + \frac{\mathcal{R}_\mathcal{A}^\lambda(x_r, x_{|\mathcal{G}_i^1(\mathcal{L})|})}{x_{|\mathcal{G}_i^1(\mathcal{L})|}}$, and $\mathcal{R}_\mathcal{A}^\lambda(x_r, x_j) = 0$ when it lower than λ. Here, the binary relation $\mathcal{R}_\mathcal{A}$ is defined as follows.

$$\mathcal{R}_\mathcal{A}^\lambda(x, y) = e^{-\frac{\|a(x) - a(y)\|_\mathcal{A}^2}{\delta^2}}, \tag{15}$$

where δ^2 is an adjustment parameter that controls the sensitivity of the similarity measure.

According to [48], setting δ^2 as the 20th percentile of $\|a(x) - a(y)\|_\mathcal{A}^2$ computed over all object pairs $(x, y) \in \mathcal{G}_{ij}^2(Y_j)$ provides robustness against outliers and ensures local adaptability. Following this strategy, the resulting object granules remain stable and well-structured, thus enhancing the reliability of the multi-level granule construction process.

The above Def. 8 outlines the multi-granularity knowledge space construction, including the label, decision, and object hierarchical levels. Unlike existing approaches, the proposed framework avoids reliance on specific models and exhibits greater generalizability when applied to HiDWS data. Several theoretical properties are established to investigate further how the multi-granularity knowledge space evolves with respect to the feature space and the parameter λ as follows.

Property 1. *Given a HiDWS* $= (\mathcal{O}^{'}, \mathcal{A} \cup \mathcal{L})$*, where* $\mathcal{O}^{'} = \mathcal{O}^L \cup \mathcal{O}^U_{sub}$ *and* $\mathcal{O}^L/\mathcal{L}^L = \{Y_1, Y_2, \ldots, Y_s\}$*. Let the object-level granule induced by* $\mathcal{B}$ *and* λ *be denoted as* $\mathcal{G}_{ijw}^{3,\mathcal{B},\lambda}$*, where* $i = 1, 2$*,* $j = 1, 2, \ldots, s$*, and* w *is the number of objects in decision granule* $\mathcal{G}_{ij}^2(Y_j)$*. Then, the following properties hold:*

1) ***Monotonicity with features****: For* $\mathcal{C} \subseteq \mathcal{B}$*,* $\mathcal{G}_{ijw}^{3,\mathcal{B},\lambda} \subseteq \mathcal{G}_{ijw}^{3,\mathcal{C},\lambda}$*;*
2) ***Monotonicity with*** λ*: For* $\lambda_1 \leq \lambda_2$*,* $\mathcal{G}_{ijw}^{3,\mathcal{B},\lambda_1} \supseteq \mathcal{G}_{ijw}^{3,\mathcal{B},\lambda_2}$*.*

Proof: These properties follow directly from Def 8.

- For 1), from Eq. (14), we have $\mathcal{G}_{ijw}^{3,\mathcal{B},\lambda} = \sum_{r=1}^{|\mathcal{G}_i^1(\mathcal{L})|} \frac{\mathcal{R}_\mathcal{B}^\lambda(x_r, x_r')}{x_r'}$, where the binary relation $\mathcal{R}_\mathcal{B}^\lambda$ is defined in Eq. (15). When $\mathcal{C} \subseteq \mathcal{B}$, the similarity measure becomes less discriminative, i.e., $\mathcal{R}_\mathcal{B}^\lambda \subseteq \mathcal{R}_\mathcal{C}^\lambda$, which implies that $\mathcal{G}_{ijw}^{3,\mathcal{B},\lambda} \subseteq \mathcal{G}_{ijw}^{3,\mathcal{C},\lambda}$.
- For 2), similarly, for $\lambda_1 \leq \lambda_2$, the corresponding similarity relation satisfies $\mathcal{R}_\mathcal{B}^{\lambda_1} \supseteq \mathcal{R}_\mathcal{B}^{\lambda_2}$, leading to larger granules when using λ_1. Thus, $\mathcal{G}_{ijw}^{3,\mathcal{B},\lambda_1} \supseteq \mathcal{G}_{ijw}^{3,\mathcal{B},\lambda_2}$.

As demonstrated in Property 1, the object granules in multi-granularity knowledge space exhibit monotonic behavior with respect to the changes in feature subset and λ. In contrast, the granules at label and decision levels are determined by the inherent data structure and S2-Label learning method. Across the entire multi-granularity space, granule numbers increase monotonically from coarser to finer levels. This hierarchical relationship can be formally stated as follows.

Theorem 1. *Let* $\mathcal{S} \subseteq \mathcal{B}$*, the* $\mathcal{G}^1(\mathcal{L}), \mathcal{G}_i^1(\mathcal{L}), \mathcal{G}_{ij}^2(Y_j)$ *for* $i = 1, 2$*,* $j = 1, 2, \ldots, s$ *are the granules in the induced multi-granularity knowledge space. Then, the following holds.*

$$|\mathcal{G}^1(\mathcal{L})| \leq \sum_{i=1}^{2} |\mathcal{G}_i^1(\mathcal{L})| \leq \sum_{i=1}^{2} \sum_{j=1}^{s} |\mathcal{G}_{ij}^2(Y_j)|. \tag{16}$$

Proof: The theorem is proved as follows.

1) According to Eqs. (12) and (13), the $|\mathcal{G}^1(\mathcal{L})| = 2$ and $\sum_{i=1}^{2} |\mathcal{G}_i^1(\mathcal{L})| = 2s$. Since the number of decision classes $s \geq 1$, it is evident that $|\mathcal{G}^1(\mathcal{L})| \leq \sum_{i=1}^{2} |\mathcal{G}_i^1(\mathcal{L})|$.

2) According to Eq. (14), $|\mathcal{G}_{ij}^2(Y_j)| = w$, where each $\mathcal{G}_{ij}^2(Y_j)$ contains at least one object x_w. Therefore, $\sum_{i=1}^{2}\sum_{j=1}^{s}|\mathcal{G}_{ij}^2(Y_j)| \geq 2s$. This implies that $\sum_{i=1}^{2}|\mathcal{G}_i^1(\mathcal{L})| \leq \sum_{i=1}^{2}\sum_{j=1}^{s}|\mathcal{G}_{ij}^2(Y_j)|$.

By combining the above 1) and 2), the theorem is proved.

The above granulation process results in a multi-granularity structure with three levels, including: label granularity, decision granularity, and object granularity. Particularly, with the increase of granule number at different levels, the granularity information is refined accordingly, as illustrated in Fig. 3. Consequently, the information at each granularity must be considered within the entire dataset. This granulation approach aligns with the systematic thought of zentropy method. Therefore, an effective uncertainty measure with zentropy modeling is essential for managing uncertainty in the HiDWS data.

C. Multi-granularity Zentropy Uncertainty Measure

According to the multi-granularity knowledge space analysis introduced in subsection III-B, this subsection develops a novel multi-granularity zentropy uncertainty measure mechanism to accurately handle uncertainty information in partially labeled data learning. Unlike existing zentropy-based approaches [27], [43], [44], which are often constrained by specific model assumptions or the requirement of fully labeled data, the proposed method introduces a unified and flexible framework tailored for HiDWS scenarios. Notably, it can naturally degenerate into a supervision scenario, demonstrating its generality and compatibility with existing paradigms.

Definition 9. *Given a $HiDWS = (\mathcal{O}^{'}, \mathcal{A} \cup \mathcal{L})$, where $\mathcal{O}^{'} = \mathcal{O}^L \cup \mathcal{O}_{sub}^U$ and $\mathcal{O}^L/\mathcal{L}^L = \{Y_1, Y_2, \ldots, Y_s\}$. For $\mathcal{B} \subseteq \mathcal{A}$, the multi-granularity zentropy uncertainty measure of $\mathcal{O}^{'}$ on $\mathcal{B}$ is defined as:*

$$\mathbf{M}_{\mathcal{B}}^{\mathbf{Z}}(\mathcal{O}^{'}) = -\sum_{i=1}^{2} p_i log p_i + \sum_{i=1}^{2} p_i \mathbf{M}_i^{\mathbf{Z}}, \tag{17}$$

where $p_1 = \frac{|\mathcal{G}_1^1(\mathcal{L})|}{|\mathcal{G}^1(\mathcal{L})|} = \frac{|\mathcal{L}^L|}{|\mathcal{L}|}$ and $p_2 = \frac{|\mathcal{G}_2^1(\mathcal{L})|}{|\mathcal{G}^1(\mathcal{L})|} = \frac{|\mathcal{L}^U|}{|\mathcal{L}|}$ are the probability induced by Eq. (12). The $\mathbf{M}_i^{\mathbf{Z}}$, internal entropy of the i-th granule, can be decomposed into as follows.

$$\mathbf{M}_i^{\mathbf{Z}} = -\sum_{j=1}^{s} p_{ij} log p_{ij} + \sum_{j=1}^{s} p_{ij} \mathbf{M}_{ij}^{\mathbf{Z}}, \tag{18}$$

where $p_{ij} = \frac{|\mathcal{G}_{ij}^2(Y_j)|}{|\mathcal{G}_i^1(\mathcal{L})|} = \frac{\sum_{x \in \mathcal{G}_i^1(\mathcal{L})} |\tilde{Y}_j(x)|}{\sum_{j=1}^{s}\sum_{x \in \mathcal{G}_i^1(\mathcal{L})} |\tilde{Y}_j(x)|}$ is the probability of the j-th decision granule in $\mathcal{G}_i^1(\mathcal{L})$, with finer similarity granules providing additional refinement.

$$\mathbf{M}_{ij}^{\mathbf{Z}} = -\sum_{w=1}^{n_{ij}} p_{ijw} log p_{ijw}, \tag{19}$$

where n_{ij} is the object number in $\mathcal{G}_i^1(\mathcal{L})$ classed to Y_j via the soft label maximum principle, and $p_{ijw} = \frac{|[x_w]_{\mathcal{R}_\mathcal{B}} \cap \tilde{Y}_j|}{|[x_w]_{\mathcal{R}_\mathcal{B}}|}$ quantifies the certainty of x_w to Y_j.

The proposed multi-granularity zentropy uncertainty measure accurately captures both the granularity structure and the interaction information embedded in the multi-granularity knowledge space illustrated in Fig. 3. Unlike traditional approaches limited to a rough approximation structure, this measure provides a generalizable granulation-based framework for uncertainty quantification. As the decision system evolves, the measure adapts accordingly. Based on this formulation, several related properties can be derived as follows.

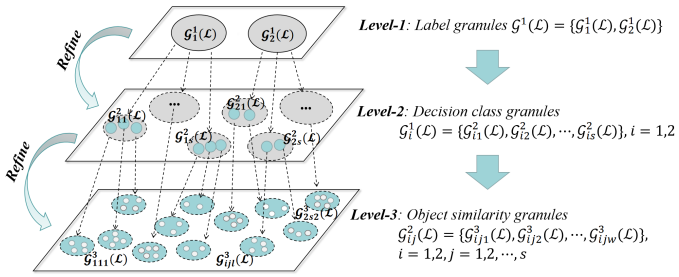


Fig. 3. The multi-granularity knowledge space construction in HiDWS data.

Property 2. *Given a $HiDWS = (\mathcal{O}^{'}, \mathcal{A} \cup \mathcal{L})$, where $\mathcal{O}^{'} = \mathcal{O}^L \cup \mathcal{O}_{sub}^U$ and $\mathcal{O}^L/\mathcal{L}^L = \{Y_1, Y_2, \ldots, Y_s\}$. For any $\mathcal{B} \subseteq \mathcal{A}$, $\mathbf{M}_{\mathcal{B}}^{\mathbf{Z}}(\mathcal{O}^{'})$ is the multi-granularity zentropy uncertainty measure of $\mathcal{O}^{'}$ on $\mathcal{B}$. Then, the following properties hold.*

1) ***Non-negativity**: $\mathbf{M}_{\mathcal{B}}^{\mathbf{Z}}(\mathcal{O}^{'}) \geq 0$;*
2) ***Degenerate case**: If $\mathcal{O}_{sub}^U = \emptyset$, $\mathbf{M}_{\mathcal{B}}^{\mathbf{Z}}(\mathcal{O}^{'}) = -\sum_{j=1}^{s} p_{1j} log p_{1j} + \sum_{j=1}^{s} p_{1j} \mathbf{M}_{ij}^{\mathbf{Z}}$;*
3) ***Unit vector case**: If $[x]_{\mathcal{R}_\mathcal{B}}$ is a unit vector for $\forall x \in \mathcal{O}^{'}$, $\mathbf{M}_1^{\mathbf{Z}} = -\sum_{j=1}^{s} p_{1j} log p_{1j}$, $\mathbf{M}_2^{\mathbf{Z}} = -\sum_{j=1}^{s} p_{2j} log p_{2j} - \sum_{j=1}^{s} p_{2j} \sum_{w=1}^{n_{2j}} \tilde{Y}_j(x_w) log \tilde{Y}_j(x_w)$;*
4) ***Maximum change**: For $\mathcal{S} \subsetneq \mathcal{B}$, the change amount $\triangle = |\mathbf{M}_{\mathcal{S}}^{\mathbf{Z}}(\mathcal{O}^{'}) - \mathbf{M}_{\mathcal{B}}^{\mathbf{B}}(\mathcal{O}^{'})| \leq \frac{n}{2s} log 2$.*

Proof: This property can be derived from Defs. 8 and 9.

- For 1), all the probability in this measure are larger or equal to 0 according to Def. 9, thus $\mathbf{M}_{\mathcal{B}}^{\mathbf{Z}}(\mathcal{O}^{'}) \geq 0$ holds.
- For 2), when $\mathcal{O}^U = \emptyset$, it follows that $p_2 = \frac{|\mathcal{L}^U|}{|\mathcal{L}|} = 0$. In this case, the system entropy will be determined by the decision granularity and object granularity. Thus, the entropy measure $\mathbf{M}_{\mathcal{B}}^{\mathbf{Z}}(\mathcal{O}^{'})$ simplifies to $\mathbf{M}_{\mathcal{B}}^{\mathbf{Z}}(\mathcal{O}^{'}) = \sum_{j=1}^{s} p_{1j} log p_{1j} + \sum_{j=1}^{s} p_{1j} \mathbf{M}_{ij}^{\mathbf{Z}}$.
- For 3), when $[x]_{\mathcal{R}_\mathcal{B}}$ is a unit vector, the probability $p_{ijw} = \frac{|[x]_{\mathcal{R}_\mathcal{B}} \cap \tilde{Y}_j|}{|[x]_{\mathcal{R}_\mathcal{B}}|}$ equals 1 for $x \in \mathcal{G}_{1j}^2(Y_j)$ and $\tilde{Y}_j(x)$ for $x \in \mathcal{G}_{2j}^2(Y_j)$, according to Def. 9. As a result, $\mathbf{M}_{1j}^{\mathbf{Z}} = 0$ and $\mathbf{M}_{2j}^{\mathbf{Z}} = \sum_{w=1}^{n_{2j}} \tilde{Y}_j(x_w) log \tilde{Y}_j(x_w)$ for $j = 1, 2, \ldots, s$. Therefore, the entropy measure simplifies to $\mathbf{M}_1^{\mathbf{Z}} = -\sum_{j=1}^{s} p_{1j} log p_{1j}$ and $\mathbf{M}_2^{\mathbf{Z}} = -\sum_{j=1}^{s} p_{2j} log p_{2j} - \sum_{j=1}^{s} p_{2j} \sum_{w=1}^{n_{2j}} \tilde{Y}_j(x_w) log \tilde{Y}_j(x_w)$.
- For 4), according to Def. 9, the variation in multi-granularity zentropy measure depends on the changes in entropy at the object granules, as defined in Eq. (19). Consider the case where the neighborhood granule $[x_w]_{\mathcal{R}_\mathcal{B}}$ contains only the object x_w, i.e., $[x_w]_{\mathcal{R}_\mathcal{B}}(x_w) = 1$ and $[x_w]_{\mathcal{R}_\mathcal{B}}(x_j) = 0$ for all $x_j \neq x_w$. In this situation, $p_{ijw} = \frac{|[x_w]_{\mathcal{R}_\mathcal{B}} \cap \tilde{Y}_j|}{|[x_w]_{\mathcal{R}_\mathcal{B}}|} = 1$, and the corresponding entropy component $\mathbf{M}_{ij}^{\mathbf{Z}}$ reaches its **minimum value**, i.e., $\mathbf{M}_{ij}^{\mathbf{Z}} = 0$. On the other hand, when the label distribution within the granule is most uncertain-specifically when

$p_{ijw} = \frac{|[x_w]_{\mathcal{R}_\mathcal{B}} \cap \tilde{Y}_j|}{|[x_w]_{\mathcal{R}_\mathcal{B}}|} = \frac{1}{2}$, the entropy $\mathbf{M}_{ij}^{\mathbf{Z}}$ achieves its **maximum**, with $\mathbf{M}_{ij}^{\mathbf{Z}} = \frac{n_{ij}}{2}\log 2$.

Currently, the full zentropy at object level i is given by $\mathbf{M}_i^{\mathbf{Z}} = -\sum_{j=1}^{s} p_{ij}\log p_{ij} + \sum_{j=1}^{s} p_{ij} \cdot \frac{n_{ij}}{2}\log 2$, where the second summation term satisfies the following inequality: $\sum_{j=1}^{s} p_{ij} \cdot \frac{n_{ij}}{2}\log 2 \leq \sum_{j=1}^{s} \frac{n_{ij}}{sn_i} \cdot \frac{n_{ij}}{2}\log 2$. Therefore, the total **maximum change** in the zentropy across all objects is:

$$\triangle = \left(n_2^{'} \sum_{j=1}^{s} n_{1j}^2 + n_1 \sum_{j=1}^{s} n_{2j}^2 \right) \cdot \frac{\log 2}{2sn_1 n_2^{'}}.$$

By applying the **Cauchy-Schwarz inequality**, we obtain an upper bound:

$$\triangle \leq \left(n_2^{'} n_1^2 + n_1 n_2^{'2} \right) \cdot \frac{\log 2}{2sn_1 n_2^{'}} = \frac{n^{'}}{2s}\log 2,$$

where $n^{'} = n_1 + n_2^{'}$.

In Property 2, key characteristics such as non-negativity, maximal variation, degeneracy, and the unit vector case are discussed. Among them, monotonicity with respect to changes in the feature space is particularly critical in the feature selection process, as it directly impacts the design of selection strategies in specific algorithms. In the following, we formally present and prove the monotonicity of the proposed measure.

Theorem 2. *Given a $HiDWS = (\mathcal{O}^{'}, \mathcal{A} \cup \mathcal{L})$, where $\mathcal{O}^{'} = \mathcal{O}^L \cup \mathcal{O}_{sub}^U$ and $\mathcal{O}^L/\mathcal{L}^L = \{Y_1, Y_2, \ldots, Y_s\}$. For $\mathcal{S} \subseteq \mathcal{B}$, $\mathbf{M}_\mathcal{S}^{\mathbf{Z}}(\mathcal{O}^{'})$ and $\mathbf{M}_\mathcal{B}^{\mathbf{Z}}(\mathcal{O}^{'})$ are indistinguishable.*

Proof: When $\mathcal{S} \subseteq \mathcal{B}$, it follows from Eq. (15) that $[x]_{\mathcal{R}_\mathcal{B}} \subseteq [x]_{\mathcal{R}_\mathcal{S}}$. However, the quantities $\frac{|[x]_{\mathcal{R}_\mathcal{S}} \cap \tilde{Y}_j|}{|[x]_{\mathcal{R}_\mathcal{S}}|}$ and $\frac{|[x]_{\mathcal{R}_\mathcal{B}} \cap \tilde{Y}_j|}{|[x]_{\mathcal{R}_\mathcal{B}}|}$ are not directly comparable for $j = 1, 2, \ldots, s$. Consequently, $\mathbf{M}_{ij}^{\mathbf{Z}}$ exhibits non-monotonic behavior. As a result, $\mathbf{M}_\mathcal{S}^{\mathbf{Z}}(\mathcal{O}^{'})$ and $\mathbf{M}_\mathcal{B}^{\mathbf{Z}}(\mathcal{O}^{'})$ become indistinguishable.

The above theorem indicates that there is no monotonicity in $\mathbf{M}_\mathcal{S}^{\mathbf{Z}}(\mathcal{O}^{'})$ with the variation of feature subsets, which inevitable includes some redundant features when heuristic strategies are employed. Consequently, redundant feature elimination is necessary in the feature selection process.

IV. Semi-supervised Feature Selection Based on Ze-MGM

By integrating the proposed multi-granularity zentropy modeling in section III, this section investigates a semi-supervised feature selection strategy grounded in the multi-granularity information in HiDWS, which effectively captures the intrinsic granular structure of partially labeled data. In particular, we also summarize the overall procedure of the proposed multi-granularity zentropy modeling framework.

A. Multi-granularity Reduction Guarantee

The ultimate objective of feature selection is to identify a subset of features that enhances the learning process while preserving the integrity of the information. Accordingly, feature reduction is formalized in the following definition.

Definition 10. *Given a $HiDWS = (\mathcal{O}, \mathcal{A} \cup \mathcal{L})$, where $\mathcal{O}^{'} = \mathcal{O}^L \cup \mathcal{O}_{sub}^U$ is the universe after strategic selection method. For any $\mathcal{H} \subseteq \mathcal{A}$, $\mathcal{H}$ is called a feature reduction of $\mathcal{A}$ if it has the following properties:*

1) $\mathbf{M}_\mathcal{H}^{\mathbf{Z}}(\mathcal{O}^{'}) \leq \mathbf{M}_\mathcal{A}^{\mathbf{Z}}(\mathcal{O}^{'})$;

2) For $\forall h \in \mathcal{H}$, $\mathbf{M}_\mathcal{H}^{\mathbf{Z}}(\mathcal{O}^{'}) > \mathbf{M}_{\mathcal{H}-\{h\}}^{\mathbf{Z}}(\mathcal{O}^{'})$.

According to the above two conditions in Def. 10, the final selected subset retains the original information content while each feature is indispensable. Furthermore, we introduce inner and outer critical measures to evaluate and select the most informative features during the selection process.

Algorithm 2: The proposed SSFS via Ze-MGM

Input : $HiDWS = (\mathcal{O}^{'}, \mathcal{A} \cup \mathcal{L})$ and parameter λ.
Output : The reduction subset $\mathcal{H}$.

Initialize $\mathcal{H} \leftarrow \emptyset, Core_\mathcal{A} \leftarrow \emptyset$;
Step I: Inner critical feature analysis
Obtain $\mathcal{O}^{'}$ and $\tilde{\mathcal{L}}$ according to Algorithm 1;
Calculate $\mathbf{M}_\mathcal{A}^{\mathbf{Z}}(\mathcal{O}^{'})$ by Def. 9;
forall $a \in \mathcal{A}$ **do**
  Calculate $\mathcal{I}nner_\mathcal{A}(a, \mathcal{A})$ by Eq. (20);
  if $\mathcal{I}nner(a, \mathcal{A}) > 0$ **then**
    $Core_\mathcal{A} \leftarrow Core_\mathcal{A} \cup \{a\}$;
  end
end
Step II: Outer critical feature analysis
Let $\mathcal{H} \leftarrow Core_\mathcal{A}$ and calculate $\mathbf{M}_\mathcal{H}^{\mathbf{Z}}(\mathcal{O}^{'})$;
while $\mathbf{M}_\mathcal{H}^{\mathbf{Z}}(\mathcal{O}^{'}) > \mathbf{M}_\mathcal{A}^{\mathbf{Z}}(\mathcal{O}^{'})$ **do**
  forall $b \in \mathcal{A} - \mathcal{H}$ **do**
    Calculate $\mathcal{O}uter(b, \mathcal{H})$ by Eq. (21);
  end
  Select $b_o = argmax_{b \in \mathcal{A}-\mathcal{H}} \mathcal{O}uter(b, \mathcal{H})$;
  $\mathcal{H} \leftarrow \mathcal{H} \cup \{b_o\}$;
end
Step III: Redundant feature deletion
for $h \in \mathcal{H}$ **do**
  Calculate $\mathbf{M}_{\mathcal{H}-\{h\}}^{\mathbf{Z}}(\mathcal{O}^{'})$;
  if $\mathbf{M}_{\mathcal{H}-\{h\}}^{\mathbf{Z}}(\mathcal{O}^{'}) \leq \mathbf{M}_\mathcal{H}^{\mathbf{Z}}(\mathcal{O}^{'})$ **then**
    $\mathcal{H} \leftarrow \mathcal{H} - \{h\}$;
  end
end

Definition 11. *Given a $HiDWS = (\mathcal{O}, \mathcal{A} \cup \mathcal{L})$, where $\mathcal{O}^{'} = \mathcal{O}^L \cup \mathcal{O}_{sub}^U$ is the universe. For $\forall a \in \mathcal{A}$, the inner critical measure of a relative to $\mathcal{A}$ is defined as follows:*

$$\mathcal{I}nner(a, \mathcal{A}) = \mathbf{M}_{\mathcal{A}-\{a\}}^{\mathbf{Z}}(\mathcal{O}^{'}) - \mathbf{M}_\mathcal{A}^{\mathbf{Z}}(\mathcal{O}^{'}). \tag{20}$$

Similarly, for $\mathcal{B} \subseteq \mathcal{A}$ and $b \in \mathcal{A} - \mathcal{B}$, the outer critical measure of b relative to $\mathcal{B}$ is defined as follows:

$$\mathcal{O}uter(b, \mathcal{B}) = \mathbf{M}_{\mathcal{B}-\{b\}}^{\mathbf{Z}}(\mathcal{O}^{'}) - \mathbf{M}_\mathcal{B}^{\mathbf{Z}}(\mathcal{O}^{'}). \tag{21}$$

As the aforementioned critical measures in Def. 11, if $\mathcal{I}nner_\mathcal{A}(a, \mathcal{A}) > 0$, the feature a is considered critical relative to $\mathcal{A}$ and is referred to as a core feature. Similarly, if $\mathcal{O}uter(b, \mathcal{B}) > 0$, the feature b is considered critical relative to $\mathcal{B}$, with a larger value indicating a greater degree of criticality.

B. Algorithm Design and Analysis

Based on the multi-granularity reduction guarantee in subsection IV-A, a feature selection considering S2-Label learning

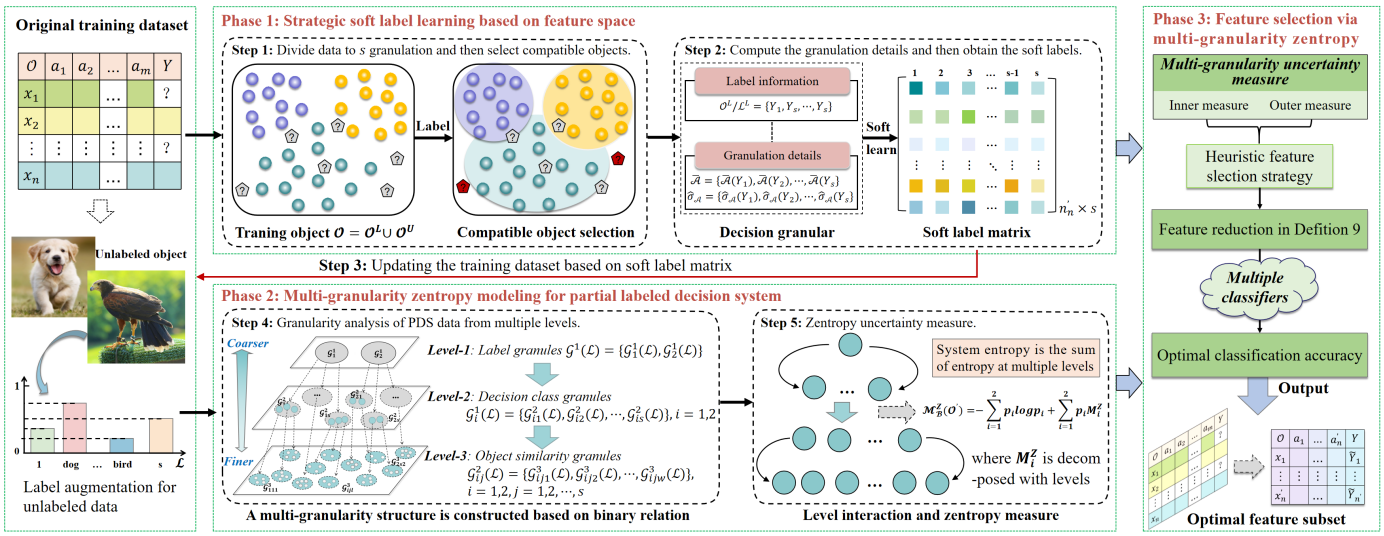


Fig. 4. The multi-granularity zentropy modeling framework for semi-supervised feature selection.

and multi-granularity zentropy uncertainty measure is presented to achieve the goal described in Def. 10. The number of selected features is determined automatically according to the two conditions specified in this definition. The algorithm consists of three key steps: (1) inner critical feature analysis, (2) outer important feature analysis, and (3) redundant feature deletion. Step I identifies the inner critical features using Eq. (20). In Step II, the outer critical features are sequentially chosen relative to the previously selected inner ones. Finally, Step III involves deleting redundant features to ensure that only the essential features are retained.

Algorithm 2 shows the entire feature selection process of $HiDWS$. In step I, the time complexity is $O(n_1 m + (n_2 + n_2^{'})sm)$. Additionally, to compute $\mathbf{M}_{\mathcal{A}}^{\mathbf{Z}}(\mathcal{O}^{'})$, it must calculate the similarity between n_1 objects in $\mathcal{O}^L$ and $n_2^{'}$ objects in $\mathcal{O}_{sub}^U$ under m features. Therefore, the time complexity of multi-granularity zentropy computation is $O((n_1 + n_2^{'})^2 m)$. Similarly, the time complexity of calculating the inner critical feature is also $O((n_1 + n_2^{'})^2 m)$. In step II, suppose l_0 features are selected initially, the time complexity of step II is $O\left(\sum_{h=l_0}^{l}((n_1 + n_2^{'})^2(h+1)(m-h)\right)$ when the total number of features after selection is l. Finally, the time complexity of feature reduction in step III is $O((n_1+n_2^{'})^2(h-1)h)$, where h is the number of final selected features. Thus, the overall time complexity of this algorithm is $O((n_1 + n_2^{'})^2 m^2)$.

C. Overall Procedure

In this subsection, we summarize the overall procedure of multi-granularity zentropy modeling for HiDWS feature selection in Fig. 4, which consists of three main phases:

• **Phase 1: Strategic soft label learning based on feature space**. To mitigate the negative impact of "unfriendly" samples, compatible unlabeled data $\mathcal{O}_{sub}^U$ are first selected based on distributional proximity to labeled instances. Subsequently, soft labels $\tilde{\mathcal{L}}_{n^{'}\times s}$ are assigned by integrating class similarity and separability, enabling more accurate modeling of HiDWS data and laying the foundation for the subsequent phase.

• **Phase 2: Multi-granularity zentropy modeling for partialed label decision system**. In this phase, an explainable framework is constructed based on binary granulation, which hierarchically evolves from label granules $\mathcal{G}_i^1(\mathcal{L})$ to decision granules $\mathcal{G}_{ij}^2(Y_j)$ and finally to object granules $\mathcal{G}_{ijw}^3(x_w)$. This design captures information dynamics across multiple granular levels, enhancing interpretability and robustness.

• **Phase 3: Feature selection via multi-granularity zentropy**. A Ze-MGM method based on the multi-granularity zentropy measure is developed for HiDWS data. Guided by the principle of maximizing information preservation, we define inner and outer criticality measures to identify informative features. The final subset is determined to retain the essential information content while reducing redundancy.

TABLE I
BASIC DESCRIPTION OF SIXTEEN BENCHMARK DATASETS

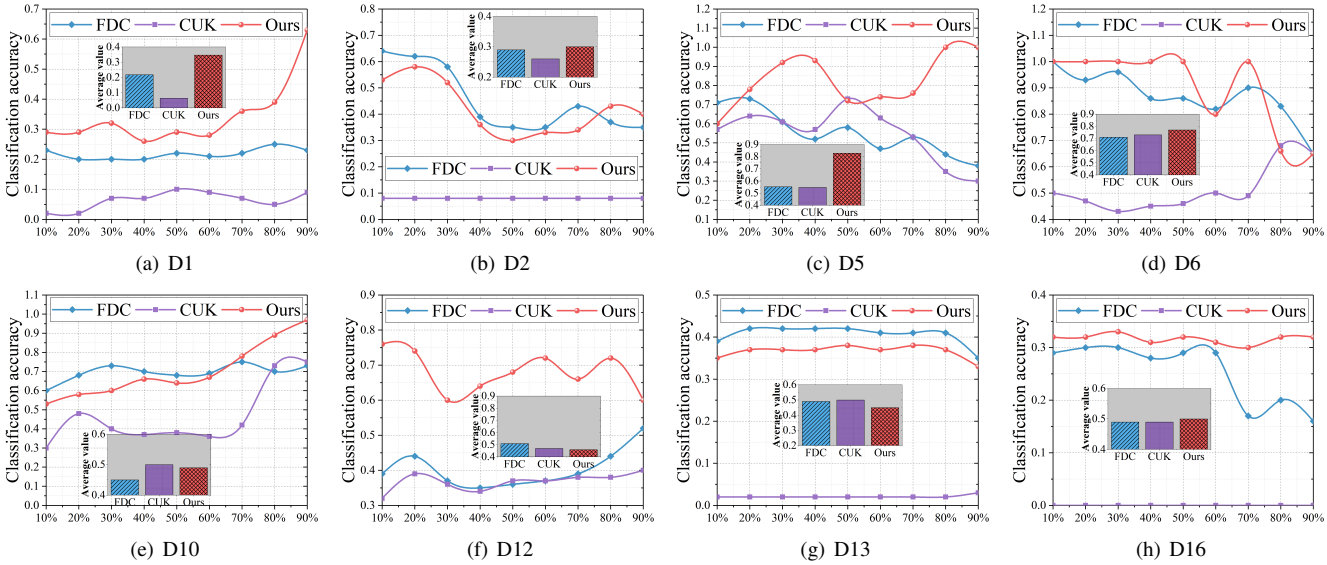
No.s	Datasets	Abbreviation	Objects	Features	Classes
D1	Arrhythmia	Arrh	450	277	13
D2	Cardiotocography	Card	2126	21	3
D3	Climate Model Simulation Crashes	Cmsc	504	18	2
D4	Colonok	Colo	62	2001	2
D5	Glioma	Glio	50	4435	4
D6	Leukemia	Leuk	72	7071	2
D7	Lymphography	Lymp	96	4027	9
D8	Mice Protein Expression	Mipe	1077	68	8
D9	Nursery	Nurs	12960	9	5
D10	Parkinsons	Park	197	23	2
D11	South German Credit	Sogc	1000	21	2
D12	Spectfheart	Spec	267	45	2
D13	Thyroid	Thyr	7200	22	3
D14	WarpAR10P	Warp	130	2400	10
D15	Wdbc	Wdbc	569	31	2
D16	Wine Quality White	Wiqw	4898	12	7

V. Experiment Analysis

Numerical experiments were conducted to demonstrate the effectiveness and robustness of proposed method for seme-supervised learning in partial decision systems. All the experiments were performed on a computer with OS: Microsoft

TABLE II
CLASSIFICATION ACCURACY OF DIFFERENT LABEL LEARNING METHODS UNDER DIFFERENT PROPORTIONS OF MISSING LABELS

No.s	Methods	10%	20%	30%	40%	50%	60%	70%	80%	90%	$Ave \pm Std$	No.s	Methods	10%	20%	30%	40%	50%	60%	70%	80%	90%	$Ave \pm Std$
	FDC	0.23	0.20	0.20	0.20	0.22	0.21	0.22	0.25	0.23	0.22±0.02		FDC	0.62	0.61	0.56	0.51	0.44	0.45	0.49	0.53	0.50	0.52±0.06
D1	CUK	0.02	0.02	0.07	0.07	0.10	0.09	0.07	0.05	0.09	0.06±0.03	D9	CUK	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01±0.00
	S2-Label	0.29	0.29	0.32	0.26	0.29	0.28	0.36	0.39	0.63	**0.35**±0.12		S2-Label	0.71	0.76	0.76	0.75	0.73	0.76	0.73	0.51	0.62	**0.70**±0.08
	FDC	0.64	0.62	0.58	0.39	0.35	0.35	0.43	0.37	0.35	**0.45**±0.12		FDC	0.60	0.68	0.73	0.70	0.68	0.69	0.75	0.70	0.73	**0.70**±0.04
D2	CUK	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08±0.00	D10	CUK	0.30	0.48	0.40	0.37	0.38	0.36	0.42	0.73	0.75	0.46±0.16
	S2-Label	0.53	0.58	0.52	0.36	0.30	0.33	0.34	0.43	0.40	0.42±0.10		S2-Label	0.53	0.58	0.60	0.66	0.64	0.67	0.78	0.89	0.97	**0.70**±0.15
	FDC	0.75	0.72	0.75	0.67	0.67	0.69	0.78	0.80	0.74	0.73±0.05		FDC	0.60	0.49	0.54	0.59	0.58	0.62	0.63	0.65	0.51	0.58±0.05
D3	CUK	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09±0.00	D11	CUK	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30±0.00
	S2-Label	0.83	0.85	0.82	0.86	0.84	0.87	0.83	0.86	0.88	**0.85**±0.02		S2-Label	0.61	0.52	0.59	0.66	0.65	0.65	0.68	0.67	0.62	**0.63**±0.05
	FDC	0.57	0.54	0.68	0.60	0.71	0.82	0.55	0.69	0.51	0.63±0.10		FDC	0.39	0.44	0.37	0.35	0.36	0.37	0.39	0.44	0.52	0.40±0.06
D4	CUK	0.43	0.38	0.37	0.60	0.52	0.47	0.41	0.76	0.54	0.50±0.13	D12	CUK	0.32	0.39	0.36	0.34	0.37	0.37	0.38	0.38	0.40	0.37±0.03
	S2-Label	0.80	0.56	0.50	0.63	0.73	1.00	0.75	0.78	0.56	**0.70**±0.16		S2-Label	0.76	0.74	0.60	0.64	0.68	0.72	0.66	0.72	0.60	**0.68**±0.06
	FDC	0.71	0.73	0.61	0.52	0.58	0.47	0.53	0.44	0.38	0.55±0.12		FDC	0.39	0.42	0.42	0.42	0.42	0.41	0.41	0.41	0.35	**0.41**±0.02
D5	CUK	0.57	0.64	0.61	0.57	0.73	0.63	0.53	0.35	0.30	0.55±0.14	D13	CUK	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.03	0.02±0.00
	S2-Label	0.60	0.78	0.92	0.93	0.72	0.74	0.76	1.00	1.00	**0.83**±0.14		S2-Label	0.35	0.37	0.37	0.37	0.38	0.37	0.38	0.37	0.33	0.36±0.02
	FDC	1.00	0.93	0.96	0.86	0.86	0.82	0.90	0.83	0.65	0.87±0.10		FDC	0.45	0.23	0.38	0.48	0.44	0.44	0.44	0.50	0.58	**0.44**±0.09
D6	CUK	0.50	0.47	0.43	0.45	0.46	0.50	0.49	0.68	0.65	0.51±0.09	D14	CUK	0.70	0.27	0.45	0.40	0.43	0.40	0.48	0.54	0.12	0.42±0.16
	S2-Label	1.00	1.00	1.00	1.00	1.00	0.80	1.00	0.66	0.65	**0.90**±0.15		S2-Label	0.45	0.27	0.40	0.35	0.39	0.38	0.48	0.54	0.58	0.42±0.10
	FDC	0.67	0.63	0.70	0.66	0.68	0.65	0.71	0.72	0.75	0.68±0.04		FDC	0.95	0.96	0.97	0.96	0.96	0.96	0.95	0.94	0.93	0.95±0.01
D7	CUK	0.06	0.03	0.04	0.03	0.03	0.66	0.67	0.67	0.64	0.31±0.33	D15	CUK	0.38	0.37	0.37	0.37	0.37	0.37	0.38	0.38	0.38	0.38±0.00
	S2-Label	0.82	0.90	0.86	0.89	0.85	0.92	0.88	0.86	0.92	**0.88**±0.03		S2-Label	0.95	0.95	0.97	0.95	0.96	0.96	0.96	0.95	0.95	**0.96**±0.01
	FDC	0.38	0.29	0.24	0.26	0.30	0.29	0.30	0.32	0.31	0.30±0.04		FDC	0.29	0.30	0.30	0.28	0.29	0.29	0.17	0.20	0.16	0.25±0.06
D8	CUK	0.32	0.35	0.25	0.27	0.27	0.28	0.29	0.22	0.33	0.29±0.04	D16	CUK	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00±0.00
	S2-Label	0.38	0.33	0.37	0.40	0.40	0.47	0.45	0.44	1.00	**0.47**±0.20		S2-Label	0.32	0.32	0.33	0.31	0.32	0.31	0.30	0.32	0.32	**0.32**±0.01



(a) D1 (b) D2 (c) D5 (d) D6

(e) D10 (f) D12 (g) D13 (h) D16

Fig. 5. The classification accuracy of different label learning methods on some datasets under nine proportions of missing labels.

WIN10; Processor: Intel Core i7-6800K CPU @ 3.4GHz×12; Memory: 62.7 GB; Programming language: MATLAB 2020a.

A. General Setting

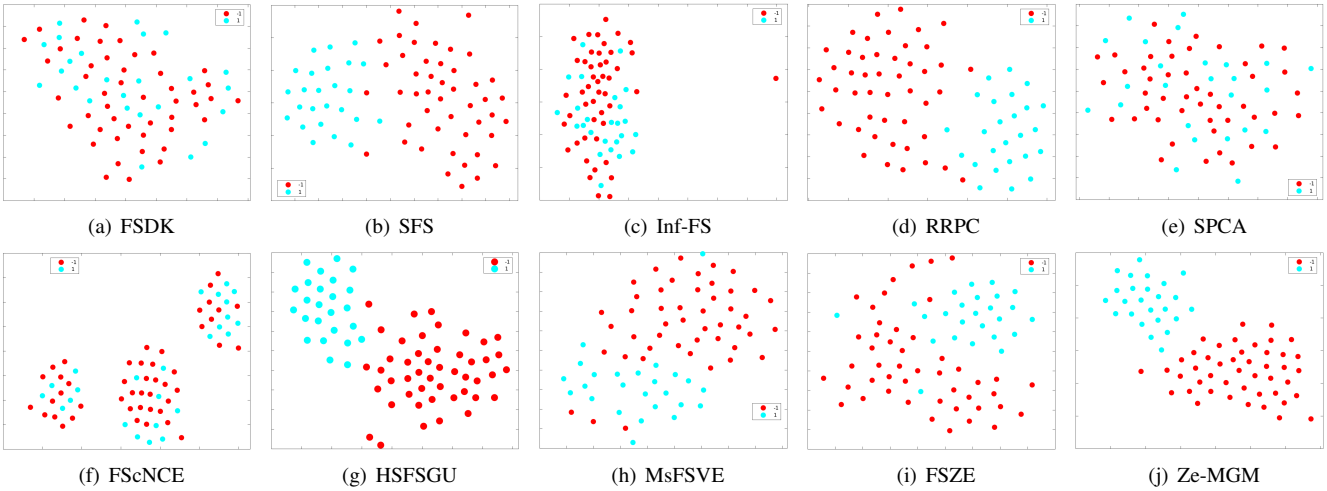
This subsection mainly outlines the general experimental settings, including the datasets utilized, the selected baseline methods for comparison, and the experimental designs employed for comparative analysis.

*1) **Datasets**:* To evaluate the performance of the proposed method, sixteen benchmark datasets from the UCI repository and a public website were selected, whose detailed information is shown in Tbale I. To ensure consistency and comparability, all datasets were normalized using the min-max scaling method, defined as follows:

$$\widehat{a}_j(x_i) = \frac{a_j(x_i) - min(V(a_j))}{max(V(a_j)) - min(V(a_j))}, \forall a_j \in \mathcal{A} \qquad (22)$$

where $a_j(x_i)$ is the specific value of x_i with respect to conditional feature $a_j \in \mathcal{A}$, the $max(V(a_j))$ and $min(V(a_j))$ are the maximum and minimum values across all objects in the datasets under feature a_j, respectively.

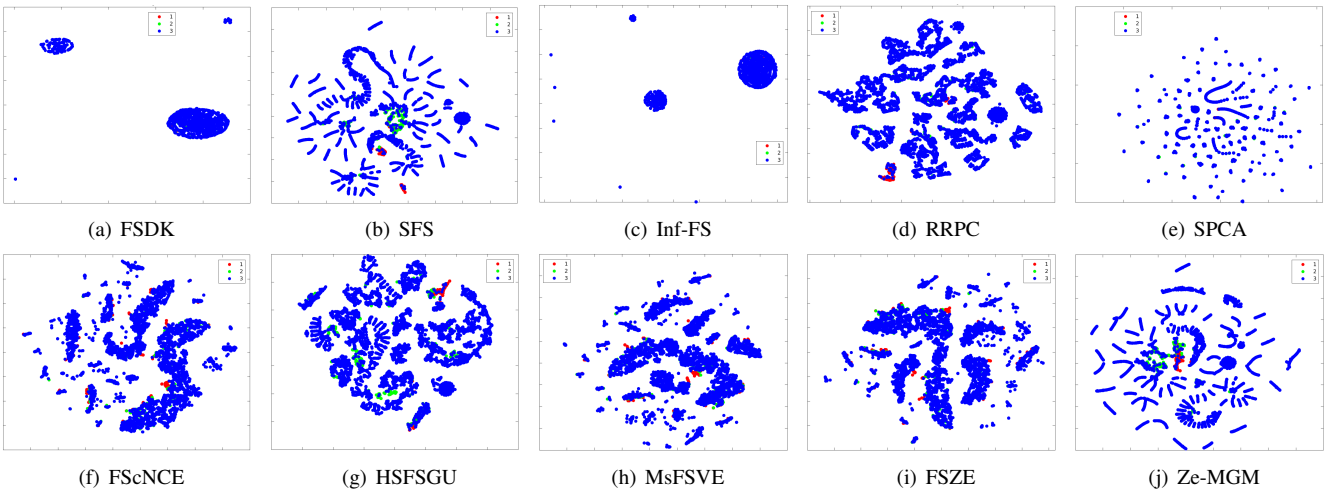
*2) **Baselines**:* To enhance the reliability and comprehensiveness of the experimental comparisons for the proposed method, we evaluate its performance in both label learning and feature selection scenarios. Specifically, two baseline label learning approaches, including classification uncertainty based



(a) FSDK (b) SFS (c) Inf-FS (d) RRPC (e) SPCA

(f) FScNCE (g) HSFSGU (h) MsFSVE (i) FSZE (j) Ze-MGM

Fig. 6. t-SNE visualization of compared methods on the Leuk dataset indcued by the first classifier and missing label proportion.



(a) FSDK (b) SFS (c) Inf-FS (d) RRPC (e) SPCA

(f) FScNCE (g) HSFSGU (h) MsFSVE (i) FSZE (j) Ze-MGM

Fig. 7. t-SNE visualization of compared methods on the Thyr dataset indcued by the first classifier and missing label proportion.

on KNN granules (CUK) [26] and the fuzzy decision classification (FDC) [42], are utilized to benchmark the proposed strategic label learning framework detailed in Algorithm 1. Additionally, seven representative feature selection methods, including fast sparse discriminative K-means for feature selection (FSDK) [49], semi-supervised feature selection based on the generalized uncorrelated constraint (SFS) [21], infinite feature selection based on graph theory (Inf-FS) [28], semi-supervised feature selection with relevance and redundancy criteria based on Pearson's correlation (RRPC) [22], hessian-based semi-supervised feature selection method based on generalized ucorrelated constraint (HSFSGU) [20], feature selection based on conditional neighborhood combination entropy (FScNCE) [34], sparse principal component analysis for feature selection (SPCA) [50], feature selection by using fuzzy rough combinatorial entropy on consistency optimal scale (MsFSVE) [38], and feature selection based on zentropy measure (FSZE) [44], are employed to validate the effectiveness of the multi-granularity zentropy measure for feature evaluation.

*3) **Experimental designs**:* In all experiments, two widely used classifiers, including K-nearest neighbor (KNN, $K = 3$) and decision tree (DT, Gini index), are employed as baseline models to evaluate the classification performance of the feature selection methods under investigation. For convenience, we refer to them as C1 and C2, respectively. To ensure a fair comparison, a ten-fold cross-validation procedure is adopted. Notably, the parameter δ plays a critical role, as it directly influences the construction of neighborhood granules, thereby affecting the feature selection process. To examine its impact comprehensively, δ is varied from 0 to 1 with increments of 0.05. Other parameter settings for the compared methods are configured according to their respective references to maintain consistency and reproducibility. In addition, to further demonstrate the effectiveness of the proposed S2-label learning approach, datasets with missing label ratios from 10% to 90% (in 10% increments) are generated by randomly removing a proportion of labels, where each ratio denotes the percentage of missing labels in the dataset.

B. Strategic Label Learning Comparison

To demonstrate the effectiveness of the proposed strategic label learning method, we compare its performance against

TABLE III
AVERAGE COMPUTATIONAL TIME (SECONDS) AND SELECTED FEATURE NUMBER UNDER THREE MISSING LABEL PROPORTIONS

Proportion	Datasets	FSDK	SFS	Inf-FS	RRPC	SPCA	FScNCE	HSFSGU	MsFSVE	FSZE	Ze-MGM
20%	D1	40.00(13.00)	30.00(13.00)	10.18(27.00)	63.46(13.00)	100.49(13.00)	133.61(61.50)	67.80(13.00)	191.17(276.00)	225.88(11.50)	119.07(13.00)
	D2	26.23(8.00)	27.19(8.00)	5.42(3.00)	52.63(8.00)	57.59(8.00)	60.76(20.00)	24.20(8.00)	1656.75(20.00)	66.92(20.00)	295.04(8.00)
	D3	6.99(8.00)	7.14(8.00)	0.65(3.00)	7.09(8.00)	16.26(8.00)	8.94(16.50)	3.74(8.00)	26.87(17.00)	4.20(4.00)	7.07(8.00)
	D4	11.00(2.50)	80.00(2.50)	12.70(335.0)	3.96(2.50)	268.64(2.50)	114.23(1.00)	430.30(2.50)	7.07(14.00)	45.61(4.00)	84.42(2.50)
	D5	10.00(6.00)	340.00(6.00)	55.00(732.0)	13.45(6.00)	752.35(6.00)	256.80(1.00)	3716.50(6.00)	63.80(50.00)	168.67(8.50)	73.12(6.00)
	D6	90.00(4.00)	16923(4.00)	56.54(1115)	7.67(4.00)	1349.70(4.00)	1127.73(1.00)	4332.50(4.00)	28.66(10.00)	200.8(3.00)	79.88(4.00)
	D7	13.00(10.00)	12.00(10.00)	0.84(3.00)	10.94(10.00)	15.50(10.00)	5.10(12.00)	15.00(10.00)	1.76(15.00)	3.05(7.50)	101.73(10.00)
	D8	110.31(24.00)	116.12(24.00)	7.27(12.00)	98.97(24.00)	295.27(24.00)	90.92(63.00)	48.69(24.00)	896.68(45.00)	67.99(62.50)	8.10(24.00)
	D9	303.52(7.00)	240.38(7.00)	47.61(1.00)	228.42(7.00)	1812.50(7.00)	425.39(7.00)	149.83(7.00)	13250.75(8.00)	340.45(8.00)	1971.00(7.00)
	D10	4.91(5.00)	4.32(5.00)	2.52(3.00)	4.01(5.00)	16.63(5.00)	6.81(12.50)	5.28(5.00)	5.91(14.00)	4.32(22.00)	23.27(5.00)
	D11	12.11(10.50)	11.50(10.50)	1.13(3.00)	10.82(10.50)	28.31(10.50)	20.29(10.50)	4.96(10.50)	40.97(20.00)	4.20(14.00)	5.83(10.50)
	D12	7.61(3.00)	8.42(3.00)	1.03(8.00)	8.24(3.00)	15.98(3.00)	12.24(21.00)	5.04(3.00)	12.44(29.00)	8.49(6.50)	7.55(3.00)
	D13	69.43(3.00)	59.65(3.00)	16.99(3.00)	56.27(3.00)	144.60(3.00)	482.44(20.00)	30.29(3.00)	25602(21.00)	1616.35(21.00)	3032.65(3.00)
	D14	96.00(14.50)	4611(14.50)	72.05(386)	83.80(14.50)	414.60(14.50)	508.00(1.00)	937.70(14.50)	283.22(75.00)	168.83(2.00)	1491.73(14.50)
	D15	10.41(9.50)	12.13(9.50)	0.93(6.00)	10.15(9.50)	22.14(9.50)	17.36(26.00)	6.46(9.50)	56.91(18.00)	19.31(24.00)	12.46(9.50)
	D16	97.36(11.00)	107.83(11.00)	6.11(2.00)	308.50(11.00)	460.28(11.00)	150.30(10.00)	88.40(11.00)	24976.75(11.00)	237.88(11.00)	432.63(11.00)
	Ave	56.81(8.69)	1411.92(8.69)	18.56(165.13)	60.52(8.69)	360.68(8.69)	213.81(17.75)	616.67(8.69)	4193.86(40.13)	198.93(14.34)	484.10(8.69)
Proportion	Datasets	FSDK	SFS	Inf-FS	RRPC	SPCA	FScNCE	HSFSGU	MsFSVE	FSZE	Ze-MGM
40%	D1	40.00(15.00)	40.00(15.00)	10.18(27.00)	60.70(15.00)	104.68(15.00)	95.08(74.00)	62.00(15.00)	201.43(76.00)	163.04(11.00)	111.79(15.00)
	D2	15.99(8.50)	15.55(8.50)	5.42(3.00)	39.90(8.50)	47.24(8.50)	28.37(18.50)	17.7(8.50)	728.50(18.50)	72.14(19.00)	273.65(8.50)
	D3	4.72(4.50)	5.88(4.50)	0.86(3.00)	4.76(4.50)	8.37(4.50)	7.16(17.00)	2.30(4.50)	13.67(17.00)	2.45(3.00)	13.28(4.50)
	D4	14.00(2.00)	40.00(2.00)	12.70(335.0)	3.02(2.00)	271.66(2.00)	66.09(1.00)	316.00(2.00)	4.89(9.00)	26.53(2.00)	76.31(2.00)
	D5	20.00(7.50)	490(7.50)	55.00(732.0)	20.85(7.50)	715.77(7.50)	147.06(1.00)	3859.20(7.50)	26.53(3.00)	110.51(4.50)	50.11(7.50)
	D6	131.00(2.50)	17697(2.50)	56.54(1115)	6.96(2.50)	1360.6(2.50)	484.75(1.00)	4072.00(2.50)	14.89(9.00)	161.71(3.00)	47.5(2.50)
	D7	16.00(12.50)	18.00(12.50)	0.84(3.00)	13.29(12.50)	16.20(12.50)	3.10(8.00)	15.20(12.50)	1.55(11.50)	4.60(4.50)	84.58(12.50)
	D8	53.52(23.00)	65.55(23.00)	7.02(12.00)	52.74(23.00)	118.14(23.00)	66.93(62.00)	24.34(23.00)	409.58(18.00)	33.62(51.50)	12.03(23.00)
	D9	360.24(7.00)	385.94(7.00)	52.76(1.00)	339.57(7.00)	1006.90(7.00)	225.63(7.00)	146.55(7.00)	13250.75(7.00)	334.04(7.00)	1839.28(7.00)
	D10	3.15(6.00)	2.51(6.00)	2.17(3.00)	3.12(6.00)	14.94(6.00)	3.95(13.00)	4.29(6.00)	3.14(13.00)	3.29(18.00)	14.57(6.00)
	D11	8.92(8.50)	14.20(8.50)	1.10(3.00)	9.67(8.50)	23.79(8.50)	14.12(11.50)	4.10(8.50)	19.91(20.00)	3.24(10.50)	9.50(8.50)
	D12	4.68(10.00)	5.58(10.00)	1.02(8.00)	5.03(10.00)	9.20(10.00)	7.55(19.50)	2.48(10.00)	6.91(29.00)	5.09(5.00)	7.33(10.00)
	D13	107.18(4.00)	108.50(4.00)	21.52(3.00)	75.74(4.00)	148.00(4.00)	279.88(20.00)	43.89(4.00)	25602(21.00)	522.51(20.00)	2674.20(4.00)
	D14	103.00(16.00)	4462.0(16.00)	72.05(386.0)	61.40(16.00)	413.50(16.00)	322.25(1.00)	1694.00(16.00)	131.46(76.00)	779.00(27.50)	2027.58(16.00)
	D15	6.77(10.50)	8.07(10.50)	1.19(6.00)	6.91(10.50)	11.75(10.50)	11.24(26.00)	3.19(10.50)	27.61(19.00)	12.43(30.00)	10.07(10.50)
	D16	99.85(11.00)	105.13(11.00)	6.11(3.00)	317.01(11.00)	472.63(11.00)	101.21(10.50)	83.10(11.00)	14521.50(11.00)	181.67(11.00)	184.69(11.00)
	Ave	61.81(9.28)	1466.49(9.28)	19.15(165.19)	63.79(9.28)	296.46(9.28)	116.52(18.19)	646.90(9.28)	3435.27(22.38)	150.99(14.22)	464.78(9.28)
Proportion	Datasets	FSDK	SFS	Inf-FS	RRPC	SPCA	FScNCE	HSFSGU	MsFSVE	FSZE	Ze-MGM
60%	D1	40.00(14.00)	40.00(14.00)	10.18(27.00)	67.09(14.00)	102.34(14.00)	49.78(21.00)	58.40(14.00)	78.50(65.00)	49.78(10.00)	71.29(14.00)
	D2	22.03(12.50)	24.22(12.50)	5.42(3.00)	83.88(12.50)	86.38(12.50)	17.20(18.00)	23.30(12.50)	238.10(20.00)	17.20(19.00)	302.82(12.50)
	D3	2.72(6.00)	2.71(6.00)	0.45(3.00)	2.62(6.00)	8.08(56.00)	3.81(14.00)	1.74(6.00)	6.05(17.00)	3.81(4.00)	20.56(6.00)
	D4	13.00(3.00)	100.00(3.00)	12.70(335.00)	5.71(3.00)	255.74(3.00)	48.91(1.00)	224.40(3.00)	3.02(6.00)	48.91(2.00)	40.66(3.00)
	D5	10.00(5.00)	500.00(5.00)	55(732.00)	11.73(5.00)	727.04(5.00)	106.23(1.00)	3374.60(5.00)	19.85(11.00)	106.23(2.00)	11.91(5.00)
	D6	120.00(2.50)	16827(2.50)	56.54(386.0)	7.13(2.50)	1359.5(2.50)	173.65(1.00)	5059.50(2.50)	8.93(6.00)	173.65(2.00)	45.50(2.50)
	D7	11.00(5.50)	10.00(5.50)	0.84(3.00)	11.99(5.50)	14.70(5.50)	2.55(1.00)	10.90(5.50)	0.97(13.00)	2.55(4.00)	44.13(5.50)
	D8	25.34(16.50)	25.90(16.50)	5.03(12.00)	25.74(16.50)	84.36(16.50)	41.02(62.50)	19.17(16.50)	136.09(22.50)	41.02(40.50)	13.04(16.50)
	D9	181.10(7.00)	175.55(7.00)	29.38(1.00)	140.81(7.00)	1867.60(7.00)	94.27(7.00)	158.13(7.00)	6117.03(7.00)	94.27(7.00)	1339.38(7.00)
	D10	2.29(4.50)	2.55(4.50)	2.17(3.00)	2.39(4.50)	14.82(4.50)	3.52(8.00)	3.65(4.50)	2.86(16.00)	3.52(2.00)	14.94(4.50)
	D11	6.02(10.00)	7.13(10.00)	0.66(3.00)	6.03(10.00)	24.05(10.00)	7.76(11.00)	5.32(10.00)	6.72(20.00)	7.76(7.00)	59.87(10.00)
	D12	4.77(14.00)	4.86(14.00)	0.66(8.00)	5.67(14.00)	14.40(14.00)	5.84(29.00)	3.19(14.00)	3.41(30.00)	5.84(5.00)	24.12(14.00)
	D13	49.41(5.00)	56.07(5.00)	9.34(3.00)	34.96(5.00)	269.20(5.00)	115.67(20.00)	58.13(5.00)	11323(21.00)	115.67(20.00)	910.93(5.00)
	D14	103.00(11.00)	4539(11.00)	72.05(386.0)	51.59(11.00)	411.20(11.00)	123.33(1.00)	618.20(11.00)	56.24(34.00)	123.33(11.00)	996.10(11.00)
	D15	4.08(11.50)	4.84(11.50)	0.65(6.00)	4.28(11.50)	14.05(11.50)	7.46(28.00)	3.03(11.50)	10.61(17.00)	7.46(14.50)	12.31(11.50)
	D16	103.75(11.00)	121.15(11.00)	6.11(3.00)	396.04(11.00)	585.87(11.00)	36.35(11.00)	83.80(11.00)	3766.40(11.00)	36.35(11.00)	180.97(11.00)
	Ave	43.66(8.69)	1402.56(8.69)	16.70(119.63)	53.60(8.69)	364.96(8.69)	52.33(14.66)	606.59(8.69)	1361.11(19.78)	52.33(10.06)	255.53(8.69)

two baseline approaches, CUK and FDC, under varying proportions of missing labels ranging from 10% to 90%. The classification accuracies of all three methods are summarized in Table II. In this table, $Ave \pm St$ denotes the average performance across the nine missing label proportions, with the best results highlighted in bold.

As shown in the table, the proposed S2-Label achieves superior classification accuracy in most scenarios. Specifically, it outperforms the baseline methods on 12, 11, 11, 11, 11, 11, 14, 13, and 15 datasets under nine different missing label proportions, respectively. Notably, the performance of S2-Label improves as the proportion of missing labels increases. This enhancement can be attributed to S2-Label's ability to consider the overall distribution of each decision class and strategically exclude unfavorable samples during label learning compared with FDC and CUK. Additionally, S2-Label achieves the highest average values on 13 occasions across 16 datasets. These results collectively validate the effectiveness of S2-Label in addressing label learning challenges.

A more vivid comparison of some datasets is presented in Fig. 5. Under different proportions of missing labels, the S2-Label curve consistently outperforms the other two methods in most scenarios, particularly on the D5, D10, and D12 datasets. Additionally, the S2-Label bar in this figure is higher than the other methods in most cases, further reinforcing its effectiveness in label learning.

C. Performance Comparison and Analysis

This subsection presents a comparative analysis of the feature selection performance of nine methods across three

TABLE IV
CLASSIFICATION ACCURACY OF COMPARED FEATURE SELECTION METHODS UNDER UNDER 20% MISSING LABEL PROPORTION

No.s	Dataset	FSDK	SFS	Inf-FS	RRPC	SPCA	FScNCE	HSFSGU	MsFSVE	FSZE	Ze-MGM
C1	D1	50.00±0.10	67.56±0.05	52.67±0.08	62.89±0.1	56.67±0.08	55.78±0.07	66.89±0.06	57.33±0.06	60.22±0.07	**70.67**±0.05
	D2	74.93±0.02	**87.40**±0.02	60.25±0.02	75.68±0.03	76.39±0.02	81.28±0.02	87.20±0.02	81.04±0.02	80.58±0.03	86.83±0.02
	D3	91.85±0.03	91.48±0.02	89.44±0.04	91.85±0.04	92.41±0.04	92.04±0.04	91.85±0.03	92.59±0.04	93.52±0.03	**94.44**±0.03
	D4	56.9±0.27	85.71±0.14	78.57±0.14	82.38±0.19	52.86±0.23	46.90±0.12	**91.90**±0.12	82.62±0.17	83.57±0.16	89.05±0.10
	D5	46.00±0.19	78.00±0.20	66.00±0.27	32.00±0.17	46.00±0.19	44.00±0.23	78.00±0.20	42.00±0.18	56.00±0.23	**86.00**±0.16
	D6	62.68±0.16	92.86±0.10	80.36±0.12	94.46±0.07	62.68±0.14	65.00±0.19	94.29±0.10	85.00±0.15	94.29±0.10	**94.82**±0.09
	D7	72.24±0.08	79.71±0.09	58.00±0.12	72.29±0.06	78.48±0.13	75.62±0.09	81.24±0.11	77.76±0.07	77.62±0.15	**82.57**±0.11
	D8	99.44±0.01	99.44±0.01	99.54±0.01	99.07±0.01	99.54±0.01	99.53±0.01	99.07±0.01	99.07±0.01	99.17±0.01	**99.72**±0.00
	D9	91.1±0.010	85.25±0.01	17.72±0.09	89.51±0.01	90.95±0.01	91.10±0.01	85.29±0.01	90.83±0.01	90.82±0.01	**91.10**±0.01
	D10	92.37±0.05	95.87±0.04	86.61±0.07	91.74±0.10	93.29±0.06	93.37±0.07	96.32±0.06	91.24±0.08	93.34±0.03	**96.45**±0.04
	D11	71.90±0.05	72.70±0.04	69.80±0.05	72.20±0.05	72.70±0.04	71.60±0.02	72.00±0.04	72.50±0.03	71.70±0.02	**75.20**±0.03
	D12	73.85±0.08	71.95±0.07	78.28±0.07	68.97±0.07	72.74±0.07	70.43±0.10	72.64±0.08	71.14±0.09	77.44±0.11	**79.74**±0.09
	D13	92.58±0.01	96.82±0.01	92.58±0.01	92.46±0.01	90.71±0.01	94.10±0.01	96.76±0.00	93.90±0.01	93.92±0.01	**96.97**±0.01
	D14	31.54±0.08	72.31±0.11	42.31±0.17	38.46±0.11	40.00±0.13	16.92±0.08	74.62±0.11	55.38±0.11	28.46±0.10	**74.62**±0.13
	D15	96.13±0.02	96.66±0.02	91.74±0.02	96.49±0.02	96.66±0.03	96.49±0.03	96.84±0.03	96.83±0.02	96.13±0.03	**97.01**±0.02
	D16	56.55±0.01	57.06±0.02	45.20±0.03	57.02±0.01	56.98±0.02	56.92±0.03	57.00±0.02	57.19±0.03	57.55±0.01	**57.55**±0.01
	Ave±st	72.50±0.07	83.17±0.06	69.32±0.08	76.09±0.07	73.69±0.08	71.94±0.07	83.87±0.06	77.90±0.07	78.39±0.07	**85.80**±0.06
	Rank	7.06	4.00	7.69	6.94	5.81	6.56	3.94	5.56	5.25	**1.19**
	Dataset	FSDK	SFS	Inf-FS	RRPC	SPCA	FScNCE	HSFSGU	MsFSVE	FSZE	Ze-MGM
C2	D1	48.22±0.06	61.78±0.10	48.89±0.06	53.11±0.11	43.78±0.06	65.78±0.06	62.00±0.09	57.11±0.09	55.11±0.06	**71.11**±0.04
	D2	78.98±0.03	84.99±0.03	63.12±0.04	72.77±0.04	75.63±0.04	84.95±0.02	85.75±0.02	84.10±0.02	82.60±0.02	**85.94**±0.01
	D3	91.85±0.05	90.19±0.04	86.30±0.06	88.52±0.05	91.67±0.04	91.48±0.04	91.85±0.04	91.85±0.04	89.26±0.04	**92.04**±0.04
	D4	64.52±0.10	70.71±0.11	73.81±0.20	83.81±0.14	65.00±0.18	64.05±0.18	90.24±0.12	81.19±0.19	79.05±0.13	**90.48**±0.11
	D5	42.00±0.24	64.00±0.21	40.00±0.16	32.00±0.22	50.00±0.24	36.00±0.16	78.00±0.18	62.00±0.15	66.00±0.21	**80.00**±0.19
	D6	56.79±0.13	95.71±0.10	87.50±0.14	94.29±0.07	71.96±0.23	59.29±0.16	95.71±0.10	81.96±0.14	93.04±0.10	**97.32**±0.06
	D7	70.19±0.09	51.90±0.06	73.57±0.10	63.57±0.09	71.62±0.11	77.62±0.12	52.62±0.10	77.81±0.11	76.38±0.12	**79.14**±0.11
	D8	84.86±0.03	85.15±0.03	80.13±0.03	85.52±0.03	84.86±0.04	85.34±0.02	85.70±0.04	84.68±0.04	85.42±0.04	**86.17**±0.03
	D9	92.15±0.00	88.88±0.01	34.20±0.02	90.56±0.01	91.51±0.01	**92.22**±0.01	88.99±0.01	92.13±0.01	92.08±0.00	**92.22**±0.01
	D10	86.76±0.09	85.18±0.06	84.00±0.11	84.13±0.10	85.79±0.07	86.74±0.08	87.11±0.05	87.21±0.11	86.29±0.15	**88.26**±0.04
	D11	67.90±0.05	70.50±0.07	69.00±0.04	68.40±0.06	62.60±0.05	68.40±0.04	69.50±0.04	73.00±0.06	70.70±0.05	**73.20**±0.03
	D12	76.35±0.09	73.29±0.09	77.52±0.09	79.43±0.09	71.85±0.10	79.43±0.09	75.64±0.09	72.64±0.09	73.48±0.08	**80.46**±0.05
	D13	92.58±0.01	98.00±0.00	92.58±0.01	92.15±0.01	89.19±0.01	**99.60**±0.00	98.08±0.00	**99.60**±0.00	99.53±0.00	98.06±0.01
	D14	31.54±0.16	65.38±0.13	60.77±0.11	33.85±0.15	29.23±0.12	19.23±0.10	**71.54**±0.09	43.85±0.12	30.77±0.09	63.85±0.13
	D15	93.67±0.03	93.50±0.03	91.40±0.03	90.34±0.05	91.92±0.02	92.28±0.04	93.31±0.03	94.21±0.03	92.61±0.02	**95.08**±0.03
	D16	56.96±0.02	57.76±0.02	50.45±0.02	57.94±0.02	58.49±0.01	57.19±0.02	57.78±0.02	58.47±0.03	58.11±0.02	**58.62**±0.03
	Ave±st	70.96±0.07	77.31±0.07	69.58±0.08	73.15±0.08	70.94±0.08	72.47±0.07	80.24±0.06	77.61±0.08	76.90±0.07	**83.25**±0.06
	Rank	6.56	5.69	7.69	6.75	7.56	5.56	3.88	4.19	5.19	**1.38**

missing label proportions: 20%, 40%, and 60%. The evaluation metrics include the number of selected features, computational time, and classification accuracy on four adopted classifiers.

*1) **Selected Feature Number**:* The number of selected features under different missing label proportions is recorded in Table III, where the reported values represent the average number of selected features across four adopted classifiers for each method. For a fair comparison, the number of selected features for ranking-based methods is consistent with that of the proposed Ze-MGM method.

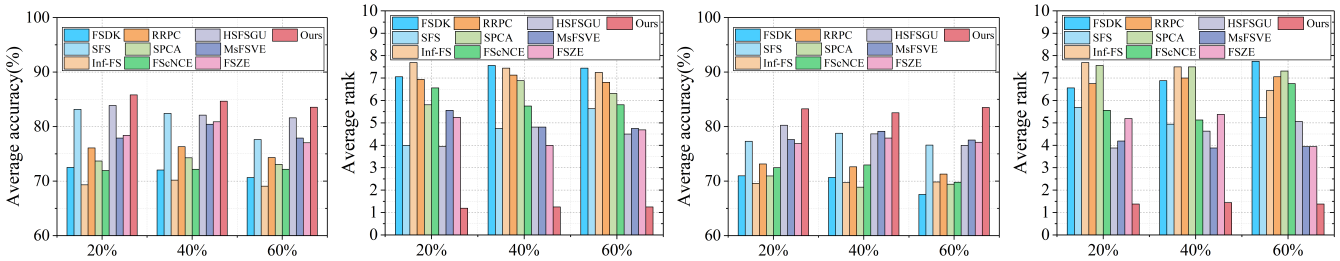
At a 20% missing label proportion, the SFS, Inf-FS, RRPC, FScNCE, FSZE, and Ze-MGM methods achieve the minimum number of selected features on 1, 9, 1, 4, 2 and 1 datasets, respectively. As shown in this table, when the missing label proportion increases to 40%, the Inf-FS, FScNCE, and FSZE achieve the minimum feature number on 10, 4, and 2 datasets out of the total 16 datasets. Similarly, under a 60% missing label proportion, these four methods including Inf-FS, SPCA, FScNCE, and FSZE select the fewest features on 7, 2, 5, and 2 datasets, respectively. It is worth noting that FScNCE occasionally selects only a single dominant feature, which may be insufficient for reliable classification. In contrast, Inf-FS exhibits instability in feature selection: a large number of features on datasets D4, D5, and D6, but only one on D9. Such inconsistency may lead to degraded classification performance. In addition, Ze-MGM consistently yields a lower average number of selected features across all datasets and missing label proportions compared to the other competing methods. This observation demonstrates the effectiveness of Ze-MGM in identifying the most informative features, particularly under challenging conditions with incomplete supervision.

Moreover, the t-SNE [51] method is adopted to visualize the feature space induced by different methods. Considering the space limitation, the Leuk and Thyr are two examples in Figs. 6 and 7. As illustrated, the Ze-MGM method yields feature spaces with reduced class overlap compared to other methods, demonstrating the superior discriminative capability achieved through the multi-granularity uncertainty framework.

*2) **Computational Time**:* The computational time of the compared methods under three different missing label proportions is reported in Tables III. Overall, the proposed Ze-MGM method demonstrates superior efficiency in the feature selection process compared to several baselines. Specifically, at 20% missing label proportion, Ze-MGM achieves competi-

TABLE V
CLASSIFICATION ACCURACY OF COMPARED FEATURE SELECTION METHODS UNDER UNDER 40% MISSING LABEL PROPORTION

No.s	Dataset	FSDK	SFS	Inf-FS	RRPC	SPCA	FScNCE	HSFSGU	MsFSVE	FSZE	Ze-MGM
C1	D1	50.89±0.06	62.44±0.09	53.11±0.08	60.00±0.06	52.89±0.09	56.89±0.08	62.44±0.07	58.22±0.06	60.67±0.05	**68.89**±0.04
	D2	77.51±0.02	85.99±0.02	60.86±0.02	78.03±0.02	82.08±0.03	79.92±0.04	86.12±0.03	80.91±0.04	80.66±0.03	**87.11**±0.02
	D3	89.81±0.04	93.15±0.04	89.81±0.04	89.07±0.03	91.11±0.05	91.85±0.04	92.59±0.02	91.67±0.05	92.59±0.04	**93.70**±0.03
	D4	39.52±0.24	77.86±0.17	85.71±0.14	84.52±0.20	59.29±0.26	48.33±0.22	77.38±0.08	78.81±0.16	87.14±0.13	**88.81**±0.13
	D5	44.00±0.26	74.00±0.23	68.00±0.19	36.00±0.16	50.00±0.27	40.00±0.21	**76.00**±0.21	64.00±0.21	70.00±0.24	70.00±0.17
	D6	66.79±0.14	95.71±0.07	80.36±0.12	95.89±0.07	62.68±0.14	65.18±0.19	97.14±0.06	94.29±0.07	91.61±0.18	**97.32**±0.06
	D7	76.48±0.11	75.81±0.10	57.38±0.12	77.05±0.13	80.43±0.07	79.00±0.09	74.43±0.11	78.48±0.10	75.00±0.03	**83.62**±0.12
	D8	98.24±0.01	99.25±0.01	99.44±0.00	99.35±0.01	98.42±0.01	99.72±0.01	98.79±0.01	99.63±0.01	99.26±0.01	**100.00**±0.00
	D9	90.94±0.01	85.25±0.01	17.72±0.09	89.51±0.01	90.95±0.01	91.13±0.01	85.29±0.01	91.08±0.00	91.13±0.01	**91.13**±0.01
	D10	94.87±0.06	92.74±0.08	86.61±0.07	87.05±0.08	94.32±0.05	92.39±0.05	92.71±0.07	94.45±0.05	94.89±0.03	**95.42**±0.04
	D11	73.40±0.05	71.80±0.05	70.00±0.05	70.00±0.04	68.90±0.03	73.70±0.05	71.50±0.03	72.40±0.05	73.90±0.04	**74.10**±0.03
	D12	74.53±0.09	78.33±0.07	79.39±0.08	69.29±0.07	79.47±0.06	71.62±0.10	77.14±0.06	73.79±0.11	75.24±0.08	**80.11**±0.08
	D13	92.58±0.01	96.15±0.01	92.58±0.01	92.46±0.01	91.42±0.01	93.92±0.01	95.94±0.01	93.93±0.01	95.07±0.01	**96.85**±0.01
	D14	31.54±0.08	**77.69**±0.09	44.62±0.09	45.38±0.18	33.85±0.14	16.92±0.11	73.08±0.15	61.54±0.19	53.08±0.13	72.31±0.15
	D15	94.73±0.04	95.95±0.02	91.91±0.03	90.15±0.03	95.78±0.03	96.49±0.02	95.96±0.03	96.31±0.03	96.66±0.02	**97.53**±0.02
	D16	56.80±0.02	56.55±0.01	45.20±0.03	57.02±0.01	56.98±0.02	57.57±0.03	57.00±0.02	57.19±0.03	57.08±0.02	**57.57**±0.03
	Ave±st	72.04±0.08	82.42±0.07	70.17±0.07	76.30±0.07	74.28±0.08	72.17±0.08	82.10±0.06	80.42±0.07	80.87±0.07	**84.66**±0.06
	Rank	7.56	4.75	7.44	7.13	6.88	5.75	4.81	4.81	4.00	**1.25**
	Dataset	FSDK	SFS	Inf-FS	RRPC	SPCA	FScNCE	HSFSGU	MsFSVE	FSZE	Ze-MGM
C2	D1	46.89±0.06	66.00±0.10	48.44±0.11	53.56±0.07	44.89±0.05	62.89±0.04	64.44±0.06	60.67±0.04	55.78±0.05	**68.67**±0.05
	D2	78.98±0.03	84.48±0.02	63.08±0.03	72.77±0.03	78.04±0.03	84.33±0.03	84.86±0.02	84.15±0.03	82.98±0.02	**85.61**±0.01
	D3	83.89±0.07	91.67±0.04	87.22±0.06	83.89±0.07	87.22±0.05	89.81±0.06	92.59±0.03	91.67±0.03	87.78±0.05	**93.33**±0.04
	D4	56.43±0.23	69.76±0.18	76.19±0.17	84.05±0.10	65.95±0.18	64.76±0.16	65.95±0.22	81.67±0.21	86.90±0.13	**88.33**±0.11
	D5	38.00±0.15	**72.00**±0.22	44.00±0.16	24.00±0.16	44.00±0.16	52.00±0.23	64.00±0.28	48.00±0.23	60.00±0.21	70.00±0.17
	D6	65.18±0.18	94.29±0.10	87.50±0.14	94.46±0.07	52.32±0.25	56.96±0.15	94.29±0.10	**97.14**±0.06	93.39±0.12	94.64±0.09
	D7	79.05±0.07	71.05±0.15	73.57±0.09	76.90±0.12	66.90±0.09	76.43±0.13	72.95±0.09	79.67±0.11	74.95±0.09	**81.57**±0.13
	D8	83.85±0.05	84.68±0.03	79.94±0.03	81.33±0.05	77.34±0.04	84.22±0.04	85.42±0.04	84.21±0.04	**82.54**±0.05	**85.42**±0.03
	D9	92.15±0.00	88.88±0.01	34.20±0.02	90.56±0.01	91.51±0.01	92.16±0.01	88.99±0.01	92.04±0.01	**92.22**±0.01	**92.22**±0.01
	D10	85.66±0.10	85.66±0.07	84.00±0.11	79.89±0.12	88.79±0.07	87.11±0.08	87.61±0.06	85.50±0.09	83.08±0.08	**89.66**±0.10
	D11	69.00±0.04	71.20±0.04	68.90±0.04	73.20±0.04	66.90±0.05	71.70±0.02	71.20±0.04	71.70±0.03	70.40±0.03	**73.40**±0.05
	D12	75.63±0.07	74.02±0.10	76.00±0.07	68.59±0.06	72.32±0.13	77.92±0.07	73.83±0.09	73.45±0.07	75.27±0.10	**79.42**±0.08
	D13	92.57±0.01	99.01±0.00	92.58±0.01	92.15±0.01	88.99±0.01	99.60±0.00	99.06±0.00	**99.64**±0.00	99.60±0.00	98.18±0.01
	D14	33.85±0.17	58.46±0.14	58.46±0.12	40.00±0.17	25.38±0.15	17.69±0.08	63.08±0.15	64.62±0.18	51.54±0.15	**66.15**±0.10
	D15	92.27±0.01	92.26±0.03	91.04±0.04	88.59±0.04	92.98±0.05	92.79±0.04	92.25±0.04	93.85±0.02	92.97±0.04	**95.08**±0.04
	D16	57.57±0.03	56.96±0.02	50.45±0.02	57.96±0.02	58.49±0.01	56.84±0.02	57.84±0.02	58.04±0.02	56.72±0.02	**58.62**±0.03
	Ave±st	70.68±0.08	78.77±0.08	69.72±0.08	72.62±0.07	68.88±0.08	72.95±0.07	78.65±0.08	79.13±0.07	77.88±0.07	**82.52**±0.06
	Rank	6.88	4.94	7.50	7.00	7.50	5.13	4.63	3.88	5.38	**1.44**



(a) C1Accuracy (b) C1 Rank (c) C2 Accuracy (d) C2 Rank

Fig. 8. The average accuracy and rank of different methods on two classifiers under three missing label proportions.

tive efficiency, outperforming methods such as SFS, HSFSGU, and MsFSVE in most cases. Notably, compared with FSZE, which also leverages multi-granularity information, Ze-MGM shows a computational advantage in most cases, as it avoids the complexity induced by rough approximation mechanisms.

At 40% and 60% missing proportions, Ze-MGM continues to show lower computational time than the other three methods (i.e., SFS, HSFSGU, and MsFSVE) and four methods (i.e., SFS, SPCA, HSFSGU, and MsFSVE), maintaining its efficiency advantage. Interestingly, as the missing label proportion increases, the computational time of supervised methods such as FScNCE, MsFSVE, and FSZE tends to decrease. This is because these methods rely solely on labeled data, which becomes sparser under higher missingness, thereby reducing their time complexity.

Moreover, the time gap between FSZE and Ze-MGM narrows as the missing label proportion increases. This trend further highlights the effectiveness of the proposed explainability-driven multi-granularity construction. Compared to traditional multi-granularity modeling based on rough approximation, Ze-

TABLE VI
CLASSIFICATION ACCURACY OF COMPARED FEATURE SELECTION METHODS UNDER 60% MISSING LABEL PROPORTION

No.s	Dataset	FSDK	SFS	Inf-FS	RRPC	SPCA	FScNCE	HSFSGU	MsFSVE	FSZE	Ze-MGM
C1	D1	48.00±0.09	58.67±0.06	52.00±0.07	60.89±0.11	53.78±0.09	55.56±0.04	58.00±0.06	57.33±0.06	60.44±0.08	**68.22**±0.06
	D2	79.07±0.02	83.96±0.02	60.86±0.02	76.57±0.02	81.23±0.02	80.48±0.02	84.10±0.02	81.04±0.02	80.86±0.02	**84.48**±0.01
	D3	89.07±0.04	92.96±0.04	89.63±0.03	89.07±0.05	90.00±0.05	90.56±0.05	93.89±0.02	92.59±0.04	92.41±0.03	**94.44**±0.04
	D4	48.33±0.20	68.33±0.15	75.71±0.21	80.95±0.18	54.52±0.28	54.52±0.28	83.81±0.14	82.62±0.17	78.81±0.14	**85.48**±0.15
	D5	40.00±0.23	74.00±0.21	66.00±0.23	32.00±0.30	44.00±0.30	48.00±0.23	76.00±0.16	42.00±0.18	44.00±0.21	**78.00**±0.20
	D6	53.21±0.15	55.36±0.18	80.36±0.12	87.32±0.16	67.68±0.27	65.54±0.13	**92.86**±0.10	85.00±0.15	80.54±0.10	88.93±0.18
	D7	81.76±0.11	76.24±0.10	56.14±0.12	70.24±0.12	79.76±0.08	62.90±0.15	76.90±0.10	77.76±0.07	72.10±0.11	**82.48**±0.05
	D8	98.33±0.01	98.42±0.01	99.72±0.00	98.42±0.02	94.89±0.02	99.63±0.01	98.61±0.01	99.07±0.01	99.35±0.01	**99.91**±0.00
	D9	90.76±0.01	89.07±0.01	14.54±0.09	89.75±0.01	90.95±0.01	90.86±0.01	88.80±0.01	90.83±0.01	**91.15**±0.01	**91.15**±0.01
	D10	87.16±0.07	89.21±0.07	87.74±0.08	88.68±0.05	**94.32**±0.06	93.84±0.04	88.68±0.04	91.24±0.08	84.58±0.05	92.76±0.08
	D11	72.40±0.05	70.50±0.03	70.00±0.05	71.30±0.05	68.60±0.03	72.10±0.05	70.00±0.04	72.50±0.03	72.80±0.06	**74.10**±0.03
	D12	77.83±0.09	71.18±0.08	77.88±0.08	69.57±0.09	76.42±0.08	73.80±0.05	72.28±0.10	71.14±0.09	72.99±0.08	**78.29**±0.06
	D13	92.58±0.01	95.88±0.01	92.58±0.01	92.86±0.00	91.64±0.02	93.99±0.01	95.76±0.01	93.90±0.01	94.97±0.00	**96.01**±0.01
	D14	30.77±0.07	66.15±0.15	44.62±0.09	29.23±0.14	28.46±0.11	19.23±0.10	**73.08**±0.13	55.38±0.11	53.85±0.16	67.69±0.20
	D15	95.95±0.02	95.78±0.02	92.09±0.03	95.26±0.04	95.96±0.03	96.83±0.03	95.60±0.02	96.83±0.02	96.66±0.02	**97.53**±0.03
	D16	45.20±0.03	56.43±0.02	45.20±0.03	57.02±0.01	56.98±0.02	56.80±0.02	57.00±0.02	57.19±0.03	57.21±0.03	**57.33**±0.02
	Ave±st	70.65±0.08	77.63±0.07	69.07±0.08	74.32±0.08	73.07±0.09	72.17±0.08	81.59±0.06	77.90±0.07	77.04±0.07	**83.55**±0.07
	Rank	7.44	5.63	7.25	6.81	6.31	5.81	4.50	4.75	4.69	**1.25**
	Dataset	FSDK	SFS	Inf-FS	RRPC	SPCA	FScNCE	HSFSGU	MsFSVE	FSZE	Ze-MGM
C2	D1	44.22±0.07	59.56±0.05	49.56±0.06	54.89±0.09	47.11±0.05	46.00±0.07	59.11±0.08	57.11±0.09	62.89±0.08	**68.22**±0.06
	D2	84.24±0.03	84.62±0.02	63.08±0.03	74.75±0.04	77.38±0.03	82.08±0.02	83.54±0.03	84.10±0.02	84.34±0.02	**85.47**±0.02
	D3	85.74±0.05	91.48±0.02	87.59±0.04	85.93±0.04	83.52±0.04	86.48±0.04	92.78±0.03	91.85±0.04	92.41±0.03	**93.15**±0.03
	D4	48.33±0.27	64.05±0.19	72.62±0.22	82.38±0.16	67.62±0.15	64.76±0.14	76.19±0.18	81.19±0.19	81.90±0.10	**85.00**±0.15
	D5	40.00±0.16	76.00±0.18	48.00±0.19	28.00±0.17	38.00±0.18	34.00±0.21	52.00±0.23	62.00±0.15	50.00±0.19	**94.00**±0.10
	D6	67.86±0.18	66.43±0.18	87.50±0.14	85.89±0.22	76.25±0.17	59.64±0.11	87.14±0.20	81.96±0.14	84.82±0.13	**87.68**±0.13
	D7	60.10±0.09	70.81±0.14	73.62±0.09	70.95±0.09	68.38±0.08	70.95±0.09	70.90±0.12	77.81±0.11	78.95±0.13	**83.10**±0.10
	D8	84.32±0.03	82.92±0.03	80.50±0.03	83.10±0.03	64.81±0.07	84.03±0.04	82.53±0.04	84.68±0.04	84.50±0.04	**85.70**±0.03
	D9	92.08±0.00	92.19±0.01	34.20±0.01	90.60±0.01	91.51±0.01	92.10±0.00	**94.81**±0.01	92.13±0.01	92.05±0.01	92.22±0.01
	D10	81.53±0.09	84.53±0.07	83.16±0.05	82.53±0.09	89.26±0.07	83.11±0.05	83.63±0.07	87.21±0.11	87.18±0.09	**89.74**±0.05
	D11	59.70±0.04	71.70±0.04	69.00±0.05	67.80±0.03	70.20±0.05	68.80±0.03	71.80±0.04	71.50±0.03	71.10±0.03	**72.40**±0.06
	D12	74.59±0.08	72.93±0.12	74.53±0.10	72.98±0.11	70.78±0.10	74.91±0.05	69.69±0.10	72.64±0.09	71.57±0.04	**79.36**±0.07
	D13	92.57±0.01	99.11±0.00	92.58±0.01	92.01±0.01	88.15±0.01	99.58±0.00	99.19±0.00	99.60±0.00	**99.63**±0.00	98.40±0.00
	D14	23.08±0.11	60.00±0.11	58.46±0.12	36.15±0.17	25.38±0.13	19.23±0.10	53.08±0.12	43.85±0.12	40.00±0.18	**68.46**±0.11
	D15	92.10±0.04	91.56±0.04	92.44±0.03	74.69±0.07	93.85±0.03	94.20±0.02	90.66±0.05	94.21±0.03	94.38±0.03	**94.73**±0.03
	D16	50.45±0.02	57.68±0.03	50.45±0.02	57.90±0.02	58.49±0.01	56.92±0.02	57.84±0.02	58.47±0.03	58.39±0.02	**58.49**±0.02
	Ave±st	67.56±0.08	76.60±0.08	69.83±0.07	71.28±0.09	69.42±0.07	69.80±0.06	76.56±0.08	77.52±0.07	77.13±0.07	**83.51**±0.06
	Rank	7.75	5.25	6.44	7.06	7.31	6.75	5.06	3.94	3.94	**1.38**

MGM alleviates the associated computational complexity, as discussed in subsection II-C, and provides a more scalable solution for feature selection under incomplete supervision.

*3) **Classification Accuracy**:* To evaluate the classification performance of the compared methods in partially labeled decision systems, this subsection analyzes the classification accuracy under different missing label proportions. The detailed results are presented in Table IV-VI, where the best-performing results are highlighted in bold.

Classification accuracy is measured across four adopted classifiers at three levels of missing label proportions (20%, 40%, and 60%). Overall, the proposed Ze-MGM method consistently achieves superior classification performance compared to other baselines. At the 20% missing label proportion, Ze-MGM obtains the highest classification accuracy on 14 and 14 datasets across the two classifiers, respectively. Moreover, Ze-MGM reaches the best average accuracy performance of 85.80% and 83.52%, further illustrating its excellence in classification task.

Similarly, Ze-MGM maintains its performance advantage at 40% missing proportion, achieving the highest accuracy on 14 and 13 datasets for the two classifiers, respectively. The only exceptions occur on a few datasets, such as D5 and D14 on C1, and D5, D6, and D13 on C2. When the missing label proportion increases to 60%, Ze-MGM still outperforms other methods in most cases, achieving the highest accuracy on 13 and 14 datasets, respectively. Although FSDK and RRPC occasionally yield competitive results, Ze-MGM remains consistently effective across classifiers and datasets. In addition, the average performance is presented in Fig. 8, where the average accuracy and rank of the proposed method all achieve the best results compared with others.

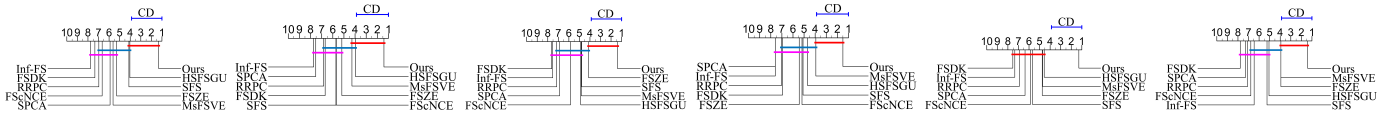
TABLE VII
BASIC DESCRIPTION OF FOUR LONG-TAILED DATASETS

No.s	Datasets	Abbreviation	Objects	Features	Classes
LD1	Autos	Auto	159	25	48||46||29||20||13||3
LD2	Landsat	Land	2000	36	470||461||397||237||224||211
LD3	Lungsmall	Luns	73	325	21||16||13||7||6||5||5||
LD4	Page	Page	548	11	492||33||12||8||3

Moreover, in comparison with the existing multi-granularity method FSZE, Ze-MGM demonstrates clear and significant improvements in classification accuracy. At 20% missing proportion, Ze-MGM outperforms FSZE on nearly all datasets

TABLE VIII
CLASSIFICATION PERFORMANCE OF COMPARED METHODS ON FOUR LONG-TAILED DATASETS UNDER DIFFERENT MISSING LABEL PROPORTIONS

Proportion	No.s	Datasets	FSDK	SFS	Inf-FS	RRPC	SPCA	FScNCE	HSFSGU	MsFSVE	FSZE	Ze-MGM
20%	C1	LD1	58.46$_{\pm0.12}$	59.21$_{\pm0.13}$	51.00$_{\pm0.14}$	66.67$_{\pm0.17}$	57.88$_{\pm0.11}$	72.38$_{\pm0.13}$	54.67$_{\pm0.12}$	58.04$_{\pm0.14}$	71.17$_{\pm0.15}$	**76.67**$_{\pm0.12}$
		LD2	87.40$_{\pm0.03}$	87.85$_{\pm0.02}$	81.95$_{\pm0.02}$	87.95$_{\pm0.04}$	88.60$_{\pm0.03}$	88.35$_{\pm0.03}$	88.20$_{\pm0.02}$	87.75$_{\pm0.02}$	87.80$_{\pm0.02}$	**88.80**$_{\pm0.02}$
		LD3	64.64$_{\pm0.18}$	69.29$_{\pm0.17}$	83.93$_{\pm0.12}$	61.61$_{\pm0.17}$	65.00$_{\pm0.16}$	34.82$_{\pm0.18}$	64.82$_{\pm0.16}$	**87.86**$_{\pm0.14}$	60.89$_{\pm0.18}$	78.57$_{\pm0.12}$
		LD4	93.80$_{\pm0.03}$	93.61$_{\pm0.03}$	89.42$_{\pm0.03}$	93.98$_{\pm0.03}$	93.98$_{\pm0.03}$	93.62$_{\pm0.03}$	93.98$_{\pm0.03}$	93.99$_{\pm0.03}$	94.18$_{\pm0.04}$	**94.90**$_{\pm0.02}$
		Ave±st	76.08$_{\pm0.09}$	77.49$_{\pm0.08}$	76.57$_{\pm0.08}$	77.55$_{\pm0.10}$	76.36$_{\pm0.08}$	72.29$_{\pm0.09}$	75.42$_{\pm0.08}$	81.91$_{\pm0.08}$	78.51$_{\pm0.09}$	**84.73**$_{\pm0.07}$
	C2	LD1	67.92$_{\pm0.11}$	62.92$_{\pm0.09}$	59.21$_{\pm0.14}$	74.88$_{\pm0.12}$	57.25$_{\pm0.08}$	76.71$_{\pm0.17}$	61.00$_{\pm0.11}$	76.75$_{\pm0.10}$	79.88$_{\pm0.11}$	**81.08**$_{\pm0.08}$
		LD2	83.05$_{\pm0.03}$	83.50$_{\pm0.02}$	79.50$_{\pm0.03}$	82.15$_{\pm0.02}$	83.10$_{\pm0.03}$	82.35$_{\pm0.02}$	82.25$_{\pm0.03}$	81.70$_{\pm0.03}$	80.70$_{\pm0.02}$	**83.75**$_{\pm0.02}$
		LD3	57.14$_{\pm0.20}$	51.96$_{\pm0.25}$	52.68$_{\pm0.24}$	46.25$_{\pm0.21}$	58.39$_{\pm0.23}$	32.68$_{\pm0.14}$	58.93$_{\pm0.20}$	64.46$_{\pm0.17}$	60.71$_{\pm0.14}$	**71.43**$_{\pm0.21}$
		LD4	93.61$_{\pm0.04}$	94.33$_{\pm0.03}$	87.96$_{\pm0.03}$	95.08$_{\pm0.03}$	94.33$_{\pm0.04}$	94.53$_{\pm0.02}$	95.07$_{\pm0.04}$	94.16$_{\pm0.04}$	95.07$_{\pm0.03}$	**95.62**$_{\pm0.03}$
		Ave±st	75.43$_{\pm0.10}$	73.18$_{\pm0.10}$	69.84$_{\pm0.11}$	74.59$_{\pm0.09}$	73.27$_{\pm0.09}$	71.57$_{\pm0.09}$	74.31$_{\pm0.09}$	79.27$_{\pm0.08}$	79.09$_{\pm0.08}$	**82.97**$_{\pm0.08}$
40%	C1	LD1	55.29$_{\pm0.16}$	68.00$_{\pm0.13}$	51.00$_{\pm0.14}$	67.29$_{\pm0.12}$	56.67$_{\pm0.16}$	71.83$_{\pm0.14}$	68.00$_{\pm0.07}$	63.54$_{\pm0.09}$	67.25$_{\pm0.07}$	**75.50**$_{\pm0.10}$
		LD2	87.90$_{\pm0.02}$	87.40$_{\pm0.01}$	81.95$_{\pm0.02}$	86.90$_{\pm0.03}$	82.30$_{\pm0.03}$	88.20$_{\pm0.01}$	87.70$_{\pm0.02}$	87.65$_{\pm0.02}$	88.15$_{\pm0.01}$	**88.55**$_{\pm0.01}$
		LD3	59.11$_{\pm0.15}$	54.46$_{\pm0.19}$	83.93$_{\pm0.12}$	60.18$_{\pm0.16}$	61.07$_{\pm0.16}$	36.25$_{\pm0.22}$	71.61$_{\pm0.18}$	**84.64**$_{\pm0.11}$	55.71$_{\pm0.33}$	75.54$_{\pm0.14}$
		LD4	93.99$_{\pm0.03}$	93.43$_{\pm0.03}$	89.42$_{\pm0.03}$	94.72$_{\pm0.02}$	93.97$_{\pm0.03}$	94.34$_{\pm0.03}$	93.97$_{\pm0.04}$	94.16$_{\pm0.04}$	94.15$_{\pm0.03}$	**95.07**$_{\pm0.03}$
		Ave±st	74.07$_{\pm0.09}$	75.82$_{\pm0.09}$	76.57$_{\pm0.08}$	77.27$_{\pm0.08}$	73.50$_{\pm0.09}$	72.66$_{\pm0.10}$	80.32$_{\pm0.08}$	82.50$_{\pm0.06}$	76.32$_{\pm0.11}$	**83.66**$_{\pm0.07}$
	C2	LD1	68.54$_{\pm0.14}$	60.33$_{\pm0.13}$	59.21$_{\pm0.14}$	75.50$_{\pm0.12}$	60.38$_{\pm0.10}$	76.08$_{\pm0.10}$	57.92$_{\pm0.15}$	77.38$_{\pm0.11}$	79.88$_{\pm0.10}$	**81.88**$_{\pm0.12}$
		LD2	81.95$_{\pm0.03}$	83.00$_{\pm0.03}$	79.50$_{\pm0.03}$	79.80$_{\pm0.02}$	83.50$_{\pm0.04}$	82.85$_{\pm0.02}$	82.65$_{\pm0.02}$	82.00$_{\pm0.02}$	81.85$_{\pm0.03}$	**83.70**$_{\pm0.02}$
		LD3	**72.68**$_{\pm0.09}$	51.07$_{\pm0.17}$	52.68$_{\pm0.24}$	39.82$_{\pm0.24}$	44.29$_{\pm0.19}$	34.29$_{\pm0.12}$	59.46$_{\pm0.22}$	62.50$_{\pm0.18}$	59.64$_{\pm0.21}$	63.21$_{\pm0.18}$
		LD4	93.98$_{\pm0.03}$	93.98$_{\pm0.03}$	87.96$_{\pm0.03}$	95.08$_{\pm0.03}$	93.59$_{\pm0.04}$	94.34$_{\pm0.02}$	93.43$_{\pm0.04}$	94.71$_{\pm0.03}$	94.53$_{\pm0.03}$	**95.43**$_{\pm0.03}$
		Ave±st	79.29$_{\pm0.07}$	72.10$_{\pm0.09}$	69.84$_{\pm0.11}$	72.55$_{\pm0.10}$	70.44$_{\pm0.09}$	71.89$_{\pm0.07}$	73.37$_{\pm0.11}$	79.15$_{\pm0.08}$	78.97$_{\pm0.09}$	**81.06**$_{\pm0.09}$
60%	C1	LD1	55.29$_{\pm0.16}$	76.71$_{\pm0.08}$	51.00$_{\pm0.14}$	63.54$_{\pm0.14}$	52.25$_{\pm0.12}$	70.42$_{\pm0.10}$	75.50$_{\pm0.07}$	68.54$_{\pm0.09}$	66.00$_{\pm0.12}$	**78.00**$_{\pm0.08}$
		LD2	87.80$_{\pm0.02}$	87.35$_{\pm0.02}$	81.95$_{\pm0.02}$	87.45$_{\pm0.04}$	82.50$_{\pm0.02}$	87.95$_{\pm0.02}$	87.05$_{\pm0.02}$	87.55$_{\pm0.02}$	88.05$_{\pm0.03}$	**88.45**$_{\pm0.02}$
		LD3	59.11$_{\pm0.15}$	68.75$_{\pm0.12}$	79.29$_{\pm0.16}$	61.61$_{\pm0.17}$	61.07$_{\pm0.16}$	34.64$_{\pm0.16}$	70.00$_{\pm0.18}$	64.29$_{\pm0.20}$	61.61$_{\pm0.13}$	**79.46**$_{\pm0.10}$
		LD4	93.99$_{\pm0.03}$	93.43$_{\pm0.03}$	89.42$_{\pm0.03}$	94.72$_{\pm0.02}$	93.97$_{\pm0.03}$	94.16$_{\pm0.03}$	93.97$_{\pm0.04}$	94.52$_{\pm0.03}$	94.34$_{\pm0.03}$	**95.07**$_{\pm0.03}$
		Ave±st	74.05$_{\pm0.09}$	81.56$_{\pm0.06}$	75.41$_{\pm0.09}$	76.83$_{\pm0.09}$	72.45$_{\pm0.08}$	71.79$_{\pm0.07}$	81.63$_{\pm0.08}$	78.72$_{\pm0.09}$	77.50$_{\pm0.07}$	**85.25**$_{\pm0.06}$
	C2	LD1	67.92$_{\pm0.11}$	69.71$_{\pm0.12}$	59.21$_{\pm0.14}$	74.92$_{\pm0.13}$	62.13$_{\pm0.14}$	73.67$_{\pm0.09}$	69.17$_{\pm0.12}$	76.17$_{\pm0.10}$	71.75$_{\pm0.11}$	**78.67**$_{\pm0.12}$
		LD2	81.95$_{\pm0.03}$	82.90$_{\pm0.02}$	79.50$_{\pm0.03}$	79.70$_{\pm0.02}$	82.20$_{\pm0.03}$	82.15$_{\pm0.04}$	83.45$_{\pm0.03}$	82.85$_{\pm0.03}$	82.50$_{\pm0.03}$	**83.45**$_{\pm0.04}$
		LD3	**72.68**$_{\pm0.09}$	63.39$_{\pm0.16}$	52.68$_{\pm0.24}$	35.71$_{\pm0.15}$	44.29$_{\pm0.19}$	30.36$_{\pm0.15}$	60.36$_{\pm0.18}$	57.50$_{\pm0.18}$	45.54$_{\pm0.13}$	63.21$_{\pm0.18}$
		LD4	93.98$_{\pm0.03}$	95.43$_{\pm0.03}$	87.96$_{\pm0.03}$	95.08$_{\pm0.03}$	93.59$_{\pm0.04}$	93.61$_{\pm0.04}$	94.88$_{\pm0.04}$	94.72$_{\pm0.03}$	94.72$_{\pm0.04}$	**95.43**$_{\pm0.03}$
		Ave±st	79.13$_{\pm0.07}$	77.86$_{\pm0.08}$	69.84$_{\pm0.11}$	71.35$_{\pm0.08}$	70.55$_{\pm0.1}$	69.95$_{\pm0.08}$	76.96$_{\pm0.09}$	77.81$_{\pm0.08}$	73.63$_{\pm0.08}$	**80.19**$_{\pm0.09}$



(a) C1-20% (b) C2-20% (c) C1-40% (d) C2-40% (e) C1-60% (f) C2-60%

Fig. 9. The Nemenyi post-hoc test results of different methods on two classifiers under three missing label proportions.

and classifiers, except D13 on C2. At 40%, Ze-MGM outperforms FSZE on 16 and 15 datasets across the two classifiers, respectively. A similar trend is observed at 60% missing proportion. Furthermore, Ze-MGM consistently achieves higher average classification accuracy than FSZE under all missing proportions, underscoring the effectiveness of the proposed multi-granularity construction in selecting informative features for classification tasks.

These above results demonstrate the robustness and adaptability of Ze-MGM in handling incomplete supervision scenarios and highlight its capability to enhance classification performance through effective feature selection.

4) Statistical Tests: This subsection presents the statistical analysis among compared methods. The Friedman test [52] is first performed across different classifiers and missing proportions. At a significance level of $\beta = 0.1$, the null hypothesis, i.e., all algorithms have equivalent classification performance, is rejected when the p-value is smaller than β. The computed p-values for the compared methods under the three proportions are 1.67×10^{-9}, 2.06×10^{-9}, and 2.85×10^{-8} on classifier C1, and 3.62×10^{-9}, 4.86×10^{-9}, and 5.90×10^{-10} on C2 classifier. All p-values are substantially below the threshold, confirming the significant differences among the methods.

Given this result, the Nemenyi post-hoc test [52] is further applied to conduct pairwise comparisons. The equality null hypothesis will be rejected when the observed rank distance exceeds the critical difference (CD) value, computed as:

$$CD_\beta = q_\beta \sqrt{\frac{M(M+1)}{6D}}, \tag{23}$$

where $CD_\beta = 3.1257$ for $M = 10$, $D = 16$, and $\beta = 0.1$.

The results of the Nemenyi post-hoc test are illustrated in Fig. 9. As shown, Ze-MGM consistently ranks first across different conditions, highlighting its superior performance. Specifically, Ze-MGM shows statistically significant improvements over 7 and 7 methods on the two classifiers under the 20% level; 8 and 8 methods under the 40% level; and 9 and 7 methods under the 60% level. These findings further validate the effectiveness of Ze-MGM in enhancing classification performance under incomplete supervision.

TABLE IX
CLASSIFICATION ACCURACY OF ABLATION EXPERIMENTS AND ZE-MGM

Proportion	No.s	Datasets	Ab-ZeM	Ab-OOD	Ze-MGM	No.s	Ab-ZeM	Ab-OOD	Ze-MGM
		D1	66.67±0.06	69.78±0.09	**70.67**±0.05		67.56±0.07	67.11±0.11	**71.11**±0.04
		D2	86.26±0.02	**86.83**±0.02	**86.83**±0.02		85.14±0.03	85.37±0.01	**85.94**±0.01
		D3	92.96±0.02	93.33±0.01	**94.44**±0.03		91.48±0.04	91.30±0.03	**92.04**±0.04
		D4	88.81±0.13	**89.05**±0.16	**89.05**±0.10		89.29±0.14	90.48±0.11	**90.48**±0.11
		D5	82.00±0.15	86.00±0.14	**86.00**±0.16		74.00±0.19	80.00±0.16	**80.00**±0.19
		D6	93.21±0.10	94.46±0.07	**94.82**±0.09		94.29±0.14	93.04±0.07	**97.32**±0.06
		D7	81.81±0.07	81.19±0.07	**82.57**±0.11		77.71±0.10	79.05±0.11	**79.14**±0.11
		D8	**99.72**±0.01	**99.72**±0.00	**99.72**±0.00		84.50±0.03	85.34±0.06	**86.17**±0.03
		D9	91.00±0.01	91.00±0.01	**91.10**±0.01		92.12±0.01	92.12±0.01	**92.22**±0.01
20%	C1	D10	96.34±0.05	96.37±0.05	**96.45**±0.04	C2	88.26±0.07	86.71±0.08	**88.26**±0.04
		D11	74.70±0.05	74.40±0.04	**75.20**±0.03		72.00±0.04	72.60±0.04	**73.20**±0.03
		D12	76.35±0.10	78.68±0.11	**79.74**±0.09		79.36±0.08	79.36±0.08	**80.46**±0.05
		D13	**97.26**±0.00	96.82±0.01	96.97±0.01		**99.43**±0.00	97.89±0.00	98.06±0.01
		D14	70.00±0.12	66.15±0.12	**74.62**±0.13		60.77±0.16	60.77±0.16	**63.85**±0.13
		D15	96.66±0.02	96.66±0.02	**97.01**±0.02		94.55±0.04	94.55±0.04	**95.08**±0.03
		D16	57.35±0.03	57.35±0.03	**57.55**±0.01		57.98±0.02	58.00±0.02	**58.62**±0.03
		LD1	76.17±0.11	74.38±0.14	**76.67**±0.12		79.92±0.11	79.83±0.09	**81.08**±0.08
		LD2	88.70±0.03	**88.80**±0.02	**88.80**±0.02		83.20±0.03	82.70±0.03	**83.75**±0.02
		LD3	78.04±0.16	76.61±0.10	**78.57**±0.10		70.71±0.18	69.46±0.19	**71.43**±0.21
		LD4	94.53±0.02	94.71±0.02	**94.90**±0.02		94.89±0.02	95.26±0.02	**95.62**±0.03
		Ave±st	84.43±0.06	84.61±0.06	**85.58**±0.00		81.86±0.08	82.05±0.07	**83.19**±0.06
Proportion	No.s	Datasets	Ab-ZeM	Ab-OOD	Ze-MGM	No.s	Ab-ZeM	Ab-OOD	Ze-MGM
		D1	64.89±0.08	67.78±0.09	**68.89**±0.04		62.67±0.09	68.22±0.09	**68.67**±0.05
		D2	85.80±0.02	85.37±0.02	**87.11**±0.02		85.00±0.02	84.95±0.02	**85.61**±0.01
		D3	92.78±0.03	93.52±0.05	**93.70**±0.03		91.85±0.02	92.04±0.04	**93.33**±0.04
		D4	83.57±0.16	87.14±0.12	**88.81**±0.13		82.38±0.18	**90.48**±0.11	88.33±0.11
		D5	**82.00**±0.15	60.00±0.19	70.00±0.17		**70.00**±0.17	62.00±0.22	**70.00**±0.17
		D6	91.79±0.07	94.46±0.07	**97.32**±0.06		94.46±0.1	94.29±0.14	**94.64**±0.09
		D7	81.14±0.09	83.14±0.07	**83.62**±0.12		78.95±0.12	79.05±0.11	**81.57**±0.13
		D8	99.81±0.01	99.81±0.01	**100.00**±0.00		84.59±0.02	85.00±0.20	**85.42**±0.03
		D9	91.00±0.01	91.00±0.01	**91.13**±0.01		92.12±0.01	92.12±0.01	**92.22**±0.01
40%	C1	D10	93.39±0.06	95.39±0.04	**95.42**±0.04	C2	86.63±0.06	86.53±0.11	**89.66**±0.10
		D11	72.40±0.04	73.60±0.04	**74.10**±0.03		72.20±0.05	72.50±0.07	**73.40**±0.05
		D12	78.02±0.09	77.14±0.07	**80.11**±0.08		79.36±0.08	75.58±0.11	**79.42**±0.08
		D13	**97.35**±0.01	96.74±0.01	96.85±0.01		**99.63**±0.00	98.11±0.00	98.18±0.01
		D14	**72.31**±0.08	68.46±0.20	**72.31**±0.15		63.85±0.11	60.77±0.11	**66.15**±0.10
		D15	96.83±0.02	97.37±0.03	**97.53**±0.02		94.91±0.03	94.03±0.03	**95.08**±0.04
		D16	57.35±0.03	57.35±0.03	**57.57**±0.03		57.92±0.02	57.88±0.02	**58.62**±0.03
		LD1	74.17±0.13	74.88±0.14	**75.50**±0.10		77.38±0.17	77.38±0.10	**81.88**±0.12
		LD2	88.60±0.03	**88.75**±0.02	88.55±0.01		83.30±0.02	82.95±0.03	**83.70**±0.02
		LD3	**75.54**±0.13	71.07±0.19	**75.54**±0.14		**63.21**±0.18	61.43±0.12	**63.21**±0.18
		LD4	94.52±0.02	94.34±0.02	**95.07**±0.03		93.97±0.03	94.17±0.03	**95.43**±0.03
		Ave±st	83.66±0.06	82.87±0.07	**84.46**±0.06		80.72±0.08	80.47±0.08	**82.23**±0.12
Proportion	No.s	Datasets	Ab-ZeM	Ab-OOD	Ze-MGM	No.s	Ab-ZeM	Ab-OOD	Ze-MGM
		D1	64.89±0.08	67.11±0.06	**68.22**±0.06		65.56±0.08	67.11±0.04	**68.22**±0.06
		D2	82.88±0.03	82.69±0.03	**84.48**±0.01		85.23±0.03	84.95±0.02	**85.47**±0.02
		D3	93.15±0.03	94.26±0.03	**94.44**±0.04		92.04±0.04	92.41±0.03	**93.15**±0.03
		D4	84.76±0.21	85.24±0.17	**85.48**±0.15		82.62±0.14	81.90±0.18	**85.00**±0.15
		D5	78.00±0.20	74.00±0.28	**78.00**±0.20		74.00±0.10	70.00±0.25	**94.00**±0.10
		D6	**88.93**±0.13	80.89±0.13	**88.93**±0.18		86.25±0.11	85.18±0.13	**87.68**±0.13
		D7	79.19±0.07	77.10±0.08	**82.48**±0.05		75.00±0.10	78.43±0.09	**83.10**±0.10
		D8	**99.91**±0.00	99.63±0.01	**99.91**±0.00		84.68±0.02	84.68±0.04	**85.70**±0.03
		D9	91.00±0.01	91.00±0.01	**91.15**±0.01		92.12±0.01	92.12±0.01	**92.22**±0.01
60%	C1	D10	**94.79**±0.06	90.26±0.04	92.76±0.08	C2	85.08±0.08	86.71±0.07	**89.74**±0.05
		D11	71.90±0.04	71.90±0.04	**74.10**±0.03		71.90±0.04	70.10±0.04	**72.40**±0.06
		D12	77.92±0.09	75.27±0.07	**78.29**±0.06		77.41±0.12	74.16±0.06	**79.36**±0.07
		D13	**97.50**±0.01	95.83±0.01	96.01±0.01		99.58±0.00	**99.63**±0.00	98.40±0.00
		D14	64.62±0.14	56.92±0.18	**67.69**±0.20		62.31±0.17	53.08±0.09	**68.46**±0.11
		D15	96.84±0.03	96.84±0.03	**97.53**±0.03		94.38±0.04	93.50±0.04	**94.73**±0.03
		D16	56.82±0.03	56.94±0.02	**57.33**±0.02		58.35±0.03	58.23±0.03	**58.49**±0.02
		LD1	74.17±0.11	76.08±0.10	**78.00**±0.08		77.38±0.07	78.08±0.09	**78.67**±0.12
		LD2	87.85±0.03	88.35±0.02	**88.45**±0.02		82.80±0.03	82.90±0.02	**83.45**±0.04
		LD3	77.86±0.18	73.04±0.18	**79.46**±0.10		62.68±0.26	51.79±0.22	**63.21**±0.18
		LD4	94.52±0.02	**95.44**±0.03	95.07±0.03		95.26±0.02	**95.43**±0.02	**95.43**±0.03
		Ave±st	82.87±0.07	81.44±0.08	**83.89**±0.07		80.23±0.07	79.02±0.07	**82.84**±0.07

D. Comparison Evaluation on Long-tailed Datasets

To further illustrate the generality of the proposed method in long-tailed situations, four datasets from the KEEL Data-Mining Software Tool are employed to make comparisons, where the detailed information is summarized in Table VII.

As shown in Table VIII, the proposed Ze-MGM achieves the highest classification accuracy in most cases, except for LD3 on the C1 classifier at 20% and 40% proportions and the C2 classifier at 60% proportion. Moreover, the average results obtained on both classifiers are significantly better than those of other methods. These results indicate that the proposed Ze-MGM maintains excellent performance even in long-tailed scenarios. Although it exhibits outstanding results, developing a specific learning mechanism to handle extreme class imbalance remains a promising direction for future work.

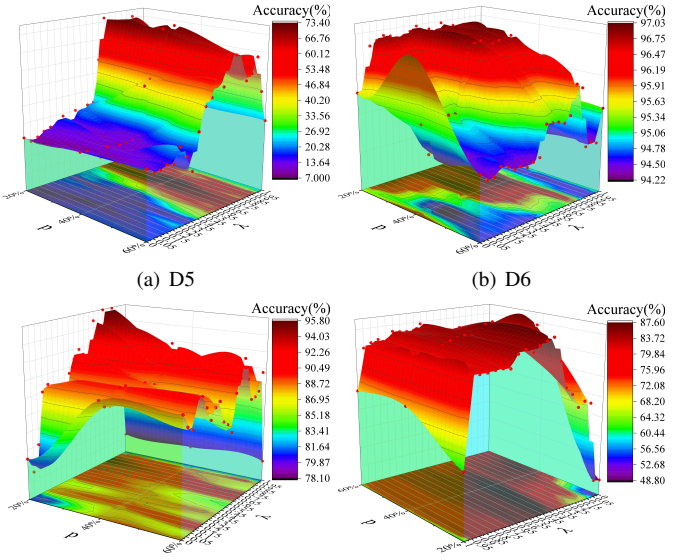
E. Ablation Study

To verify the effectiveness of the two core modules, the Ze-MGM without out-of-distribution removal (Ab-OOD) and Ze-MGM without multi-granularity uncertainty measure (Ab-ZeM) are two ablation experiments for further validation. All the classification results are summarized in Table IX.

From this table, the proposed Ze-MGM achieves superior classification performance under different missing label levels. On the two classifiers, it attains the highest accuracy 19 and 19 times under the 20% level, 17 and 18 times under the 40% level, and 17 and 19 times under the 60% level, respectively. In contrast, both Ab-OOD and Ab-ZeM achieve the best results only a few times. Furthermore, the average accuracy of Ze-MGM consistently surpasses that of the two ablated variants. These results clearly demonstrate the effectiveness of both the out-of-distribution removal and multi-granularity uncertainty modules in enhancing the semi-supervised feature selection capability of the Ze-MGM method.

F. Parameter Sensitivity Analysis

This subsection investigates the impact of the parameter λ on Ze-MGM under three different missing label proportions. As discussed in subsection III-B, the λ directly influences the granularity structure, which plays a key role in the multi-granularity zentropy modeling for HiDWS data.



(a) D5 (b) D6

(c) D10 (d) D11

Fig. 10. Classification accuracy of Ze-MGM on KNN with λ and p.

To examine how λ affects classification performance, the classification accuracy of Ze-MGM using the KNN classifier is presented in Fig. 10, where λ denotes the granularity parameter, and p is the missing label proportion. As observed, the accuracy fluctuates with variations in λ, particularly on the D5 and D11 datasets, indicating the sensitivity of Ze-MGM to this parameter. Furthermore, the optimal value of λ induced by the best accuracy differs across various missing label proportions due to the dynamic changes in granularity structure induced by unlabeled instances. Therefore, adjusting

λ appropriately is necessary to achieve optimal performance under different missing proportions.

VI. Conclusions

This paper develops a generalizable multi-granularity modeling framework to address the challenges of high sensitivity and limited robustness in high-dimensional and sparsely labeled scenarios. In contrast to existing methods, the approach uniquely incorporates object compatibility for informed sample selection and enhances label learning by integrating object proximity with classification certainty. Moreover, the designed interpretable multi-granularity structure leverages zentropy modeling to accurately capture uncertainty across various granularities. Furthermore, an SSFS approach tailored for partially labeled data is proposed by introducing strategic label augmentation combined with zentropy modeling. Extensive theoretical analysis and experiments demonstrate the effectiveness of the proposed Ze-MGM method compared to state-of-the-art techniques. While the proposed approach mitigates the limitations of conventional uncertainty measures that rely on fixed modeling assumptions, the computational overhead of granularity optimization warrants further study. Moreover, practical systems often encounter continuously evolving data streams and open scenarios. Hence, incorporating incremental mechanisms to handle such dynamic environments is a promising yet nontrivial direction for future work.

References

[1] Á. Troncoso-Garca, M. Martínez-Ballesteros, F. Martnez-Álvarez, et al., "A New Metric Based on Association Rules to Assess Feature-Attribution Explainability Techniques for Time Series Forecasting," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 47, no. 5, pp. 4140-4155, 2025.

[2] Y. Yang, N. Jiang, Y. Xu, et al., "Robust Semi-Supervised Learning by Wisely Leveraging Open-Set Data," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 12, pp. 8334-8347, 2024.

[3] W. Li, B. Yang, W. Pedrycz, et al., "Adaptive Hyper-box Granulation with Justifiable Granularity for Feature Selection," in *IEEE Transactions on Knowledge and Data Engineering*, doi: 10.1109/TKDE.2025.3617583, 2025.
T. Yang, Y. Deng, B. Yu, et al., "Local Feature Selection for Large-Scale Data Sets With Limited Labels," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 7, pp. 7152-7163, 2023.

[4] S. Niu, L. Lin, J. Huang, et al., "OwMatch: Conditional Self-Labeling with Consistency for Open-World Semi-Supervised Learning," in *Proceedings of the Neural Information Processing Systems*, vol. 37, pp. 99836-99866, 2024.

[5] H. Nguyen, D. Cheng, X. Wang, et al., "Data-Efficient Deep Learning for Printed Circuit Board Defect Detection Using X-Ray Images," in *IEEE Transactions on Instrumentation and Measurement*, doi: 10.1109/TIM.2025.3529089, 2025.

[6] X. Liang, Y. Qian, Q. Guo, et al., "AF: An Association-Based Fusion Method for Multi-Modal Classification," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 12, pp. 9236-9254.

[7] Y. Yoshikawa, T. Iwata, "Gaussian Process Regression With Interpretable Sample-Wise Feature Weights," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 9, pp. 5789-5803, 2023.

[8] D. Guo, W. Xu, W. Ding, et al., "Concept-cognitive learning survey: Mining and fusing knowledge from data," in *Information Fusion*, vol. 109, pp. 102426, 2024.

[9] H. Saadatmand, M. Akbarzadeh-T, "Many-Objective Jaccard-Based Evolutionary Feature Selection for High-Dimensional Imbalanced Data Classification," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 12, pp. 8820-8835, 2024.

[10] P. Hao, K. Liu, W. Gao, "Double-layer Hybrid-label Identification Feature Selection for Multi-view Multi-label Learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 11, pp. 12295-12303, 2024.

[11] Z. Xiao, H. Tong, R. Qu, et al., "CapMatch: Semi-Supervised Contrastive Transformer Capsule with Feature-Based Knowledge Distillation for Human Activity Recognition," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 2, pp. 2690-2704, 2025.

[12] P. Mallapragada, R. Jin, A. Jain, et al., "SemiBoost: Boosting for Semi-Supervised Learning," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 2000-2014, 2009.

[13] F. Nie, L. Tian, R. Wang, et al., "Multiview Semi-Supervised Learning Model for Image Classification," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 12, pp. 2389-2400, 2020.

[14] T. Yin, M. Chen, Z. Yuan, et al., "A Robust Multilabel Feature Selection Approach Based on Graph Structure Considering Fuzzy Dependency and Feature Interaction," in *IEEE Transactions on Fuzzy Systems*, vol. 31, no. 12, pp. 4516-4528, 2023.

[15] G. Li, Z. Yu, K. Yang, et al., "Ensemble-Enhanced Semi-Supervised Learning With Optimized Graph Construction for High-Dimensional Data," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 47, no. 2, pp. 1103-1119, 2025.

[16] R. Sheikhpour, K. Berahmand, M. Mohammadi, et al., "Sparse Feature Selection Using Hypergraph Laplacian-based Semi-supervised Discriminant Analysis," in *Pattern Recognition*, vol. 157, pp. 110882, 2025.

[17] R. Xu, D. Wu, X. Luo, "Recursion-and-Fuzziness Reinforced Online Sparse Streaming Feature Selection," in *IEEE Transactions on Fuzzy Systems*, doi: 10.1109/TFUZZ.2025.3569272, 2025.

[18] B. Jiang, X. Wu, X. Zhou, et al., "Semi-Supervised Multiview Feature Selection With Adaptive Graph Learning," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 3, pp. 3615-3629, 2024.

[19] C. Sevilla-Salcedo, V. Gmez-Verdejo, P. Olmos, "Sparse Semi-supervised Heterogeneous Interbattery Bayesian Analysis," in *Pattern Recognition*, vol. 120, pp. 108141, 2021.

[20] R. Sheikhpour, K. Berahmand, S. Forouzandeh, "Hessian-based Semi-supervised Feature Selection Using Generalized Uncorrelated Constraint", in *Knowledge-Based Systems*, vol. 269, pp. 110521, 2023.

[21] X. Li, Y. Zhang, R. Zhang, "Semisupervised Feature Selection via Generalized Uncorrelated Constraint and Manifold Embedding," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 9, pp. 5070-5079, 2022.

[22] J. Xu, B. Tang, H. He, et al., "Semisupervised Feature Selection Based on Relevance and Redundancy Criteria," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 9, pp. 1974-1984, 2017.

[23] J. Dai, Q. Hu, J. Zhang, et al., "Attribute Selection for Partially Labeled Categorical Data By Rough Set Approach," in *IEEE Transactions on Cybernetics*, vol. 47, no. 9, pp. 2460-2471, 2017.

[24] X. Cao, C. Wei, Y. Ge, et al., "Semi-Supervised Hyperspectral Band Selection Based on Dynamic Classifier Selection," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 4, pp. 1289-1298, 2019.

[25] D. Shi, L. Zhu, J. Li, et al., "Binary Label Learning for Semi-Supervised Feature Selection," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 3, pp. 2299-2312, 2023.

[26] S. An, M. Zhang, C. Wang, et al., "Robust Fuzzy Rough Approximations With kNN Granules for Semi-supervised Feature Selection," in *Fuzzy Sets and Systems*, vol. 461, pp. 108476, 2023.

[27] K. Yuan, D. Miao, H. Zhang, et al., "An Efficient and Robust Feature Selection Approach Based on Zentropy Measure and Neighborhood-Aware Model," in *IEEE Transactions on Neural Networks and Learning Systems*, doi: 10.1109/TNNLS.2025.3565320, 2025.

[28] G. Roffo, S. Melzi, U. Castellani, et al., "Infinite Feature Selection: A Graph-based Feature Filtering Approach," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 12, pp. 4396-4410, 2021.

[29] C. Hou, R. Fan, L. Zeng, ET AL., "Adaptive Feature Selection With Augmented Attributes," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 8, pp. 9306-9324, 2023.

[30] D. Guo, W. Xu, Y. Qian, et al., "M-FCCL: Memory-Based Concept-Cognitive Learning for Dynamic Fuzzy Data Classification and Knowledge Fusion," in *Information Fusion*, vol. 100, pp. 101962, 2023.

[31] A. Mencattini, P. Casti, M. D'Orazio, et al., "Uncertainty-Based Feature Selection for Improved Adequacy of Dermoscopic Image Classification," in *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1-9, no. 4505909, 2023.

[32] R. Xu, D. Wu, R. Wang, et al., "A Highly-Accurate Three-Way Decision-Incorporated Online Sparse Streaming Features Selection Model," in *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 55, no. 6, pp. 4258-4272, 2025.

[33] Y. Li, M. Dong, J. Hua, "Simultaneous Localized Feature Selection and Model Detection for Gaussian Mixtures," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 5, pp. 953-960, 2009.
[34] P. Zhang, T. Li, Z. Yuan, et al., "Heterogeneous Feature Selection Based on Neighborhood Combination Entropy," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 3, pp. 3514-3527, 2024.
[35] W. Xu, K. Yuan, W. Li, et al., "An Emerging Fuzzy Feature Selection Method Using Composite Entropy-Based Uncertainty Measure and Data Distribution," in *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 7, no. 1, pp. 76-88, 2023.
[36] Q. Liu, M. Cai, Q. Li, "Low-Dimensional Representation-Driven TSK Fuzzy System for Feature Selection," in *IEEE Transactions on Fuzzy Systems*, vol. 33, no. 11, pp. 4044-4054, 2025.
[37] D. Guo, W. Xu, Y. Qian, et al., "Fuzzy-granular Cconcept-Cognitive Learning via Three-way Decision: Performance Evaluation on Dynamic Knowledge Discovery," in *IEEE Transactions on Fuzzy Systems*, vol. 32, no. 2, pp. 1409-1423, 2023.
[38] B. Sang, L. Yang, W. Xu, et al., "VCOS: Multi-scale Information Fusion to Feature Selection Using Fuzzy Rough Combination Entropy," in *Information Fusion*, vol. 117, pp. 102901, 2025.
[39] K. Yuan, D. Miao, W. Pedrycz, et al., "Multi-granularity Data Analysis with Zentropy Uncertainty Measure for Efficient and Robust Feature Selection," in *IEEE Transactions on Cybernetics*, vol. 55, no. 2, pp. 740-752., 2025.
[40] J. Xie, X. Xiang, S. Xia, et al., "MGNR: A Multi-Granularity Neighbor Relationship and Its Application in KNN Classification and Clustering Methods," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 12, pp. 7956-7972, 2024.
[41] W. Li, L. Wei, W. Pedrycz, et al., "Granular-ball Regeneration Clustering with Principle of Justifiable Granularity," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 10, pp. 18173-18187, 2025.
[42] K. Liu, T. Li, X. Yang, et al., "SemiFREE: Semisupervised Feature Selection With Fuzzy Relevance and Redundancy," in *IEEE Transactions on Fuzzy Systems*, vol. 31, no. 10, pp. 3384-3396, 2023.
[43] K. Yuan, D. Miao, W. Pedrycz, et al., "Ze-HFS: Zentropy-Based Uncertainty Measure for Heterogeneous Feature Selection and Knowledge Discovery," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 11, pp. 7326-7339, 2024.
[44] K. Yuan, D. Miao, Y. Yao, et al., "Feature Selection Using Zentropy-Based Uncertainty Measure," in *IEEE Transactions on Fuzzy Systems*, vol. 32, no. 4, pp. 2246-2260, 2024.
[45] C. Shannon, "A Mathematical Theory of Communication," in *The Bell system technical journal*, vol. 27, no.3, pp. 379-423, 1948.
[46] Z. Liu, Y. Wang, S. Shang, "Zentropy Theory for Positive and Negative Thermal Expansion," in *Journal of Phase Equilibria and Diffusion*, vol. 43, pp. 598-605, 2022.
[47] Z. Liu, "Fundamentals on Dependence of Volume on Pressure and Temperature," in *Physics and Chemistry of Minerals*, vol. 52, no. 3, doi: https://doi.org/10.1007/s00269-024-01305-5, 2025.
[48] D. Berthelot, N. Carlini, I. Goodfellow, et al., "MixMatch: A holistic approach to semi-supervised learning," in *International Conference on Neural Information Processing Systems(NeurIPS)*, pp. 5049-5059, 2019.
[49] F. Nie, Z. Ma, J. Wang, et al., "Fast Sparse Discriminative K-Means for Unsupervised Feature Selection," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 7, pp. 9943-9957, 2024.
[50] Z. Li, F. Nie, J. Bian, et al., "Sparse PCA via $\ell_{2,p}$-Norm Regularization for Unsupervised Feature Selection," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 4, pp. 5322-5328, 2023.
[51] V. Laurens, G. Hinton, "Visualizing Data Using t-SNE," in *Journal of machine learning research*, vol. 9, no. 11, pp. 2579-2605, 2008.
[52] J. Demšar, "Statistical Comparisons of Classifiers Over Multiple Data Sets," in *Journal of Machine learning research*, vol. 7, pp. 1-30, 2006.

Kehua Yuan received the M.Sc. degree from the College of Artificial Intelligence, Southwest University, Chongqing, China, in 2022. She is currently pursuing the Ph.D. degree from the School of Computer Science and Technology, Tongji University, Shanghai, China. Her current research interests include granular computing, uncertainty analysis, and machine learning. She has widely published on these subjects at high-level international journals such as *IEEE Transactions on Knowledge and Data Engineering*, *IEEE Transactions on Cybernetics*, *IEEE Transactions on Fuzzy Systems*, *IEEE Transactions on Neural Networks and Learning Systems*.

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

WeipingDing (M'16-SM'19) received the Ph.D. degree in Computer Science, Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2013. From 2014 to 2015, he was a Postdoctoral Researcher at the Brain Research Center, National Chiao Tung University, Hsinchu, Taiwan, China. In 2016, he was a Visiting Scholar at National University of Singapore, Singapore. From 2017 to 2018, he was a Visiting Professor at University of Technology Sydney, Australia. Now he is the Full Professor of Nantong University. His main research directions involve granular data mining and multimodal machine learning. He has published over 380 articles, including over 190 IEEE Transactions papers. His twenty authored/co-authored 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 Transactions on Intelligent Transportation Systems, Information Fusion, Neurocomputing, Applied Soft Computing, et al.* He was the Leading Guest Editor of Special Issues in several prestigious journals, including *IEEE Transactions on Evolutionary Computation, IEEE Transactions on Fuzzy Systems, et al.*

Witold Pedrycz (Life Fellow, IEEE) received the M.Sc., Ph.D., and D.Sc. degrees from the Silesian University of Technology, Gliwice, Poland, in 1977, 1980, and 1984, respectively. He is a Professor with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada. He is also with the Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland, where he was elected as a Foreign Member in 2009. He also holds an appointment as a Special Professor with the School of Computer Science, University of Nottingham, Nottingham, U.K. His current research interests include computational intelligence, fuzzy modeling and granular computing, knowledge discovery and data mining, fuzzy control, pattern recognition, knowledge-based neural networks, relational computing, and software engineering.

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

Yiyu Yao received the B.E. degree in computer science from Xi'an Jiaotong University, Xi'an, China, in 1983, and the M.Sc. and Ph.D. degrees in computer science from 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 published over 400 papers. He was a highly cited researcher from 2015 to 2019. He is currently the Associate Editors of *Information Sciences* and *International Journal of Approximate Reasoning*, and so on.