This article has been accepted for publication in IEEE Transactions on Knowledge and Data Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2024.3419215

Ze-HFS: Zentropy-Based Uncertainty Measure for Heterogeneous Feature Selection and Knowledge Discovery

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

***Abstract*—Knowledge discovery of heterogeneous data is an active topic in knowledge engineering. Feature selection for heterogeneous data is an important part of effective data analysis. Although there have been many attempts to study the feature selection for heterogeneous data, there are still some challenges, such as the unbalanced problem between the stability and validity of the designed model. Hence, this paper focuses on how to design an effective and robust heterogeneous feature selection method, namely a zentropy-based uncertainty measure for heterogeneous feature selection(Ze-HFS). Different from other entropy-based uncertainty measures, the proposed method does not consider single-level information measures but systematically analyzes and integrates the information between different granular levels, which has an obvious advantage in the study of heterogeneous data knowledge discovery. Specifically, a heterogeneous distance metric is first introduced to construct heterogeneous neighborhood granules and heterogeneous neighborhood rough sets(HNRS). Then, the zentropy-based uncertainty measure is developed by analyzing the granular level structure in the HNRS model. Finally, two significant measures based on the above research are designed for heterogeneous feature selection. Compared with other state-of-the-art methods, the experimental results on 18 public datasets demonstrate the robustness and effectiveness of the proposed method.**

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

I. Introduction

BIG data, encompassing high dimensionality, heterogeneity, and incompleteness, is progressively pervasive in diverse fields, presenting abundant valuable information for data mining and knowledge discovery [1]–[4]. However, this abundance of data also brings forth challenges such as the curse of dimensionality, high storage costs, and increased computational complexity [5]–[7]. To address these challenges, feature selection has emerged as a vital processing technique that aims to reduce the dimensionality of data and enhance the interpretability of models [8]–[11]. By selecting the most relevant and informative subset of features, feature selection techniques enable us to extract valuable information from complex data, improving the efficiency and accuracy of data analysis. In practical applications, data are usually heterogeneous [12] that include categorical and numerical attributes. How to extract proper features from these heterogeneous data is of great importance in clustering and classification tasks.

Granular Computing (GrC), as a new paradigm in intelligence information processing, is useful to simulate the cognitive mechanism of humans dealing with complex problems and has been successfully applied to knowledge discovery [13], decision-making [14] and feature selection [15]. Many GrC models have been proposed to process various complex issues, such as fuzzy sets [16], rough sets [17], [18], three-way decisions [19]–[21], formal concept analysis [22], and concept-cognitive learning [23], [24]. In particular, rough set (RS) proposed by Pawlak [25] provides a formal framework to handle the incomplete and imprecise knowledge in information systems, which has been widely employed to target approximation and feature selection. Note that the classical rough set based on an equivalence relation is only suitable for the discrete data, and the numerical data needs discretization before processing [26], [27]. To avoid the information loss induced by discretization, the neighborhood rough set (NRS) model based on similarity relation was proposed for decision information systems (DIS) in [28]. This NRS model improves the tolerance of differences among objects by introducing a neighborhood parameter to distance metrics, enabling it to handle heterogeneous data simultaneously. Some distance metrics [28]–[30], such as heterogeneous Euclidean-overlap metric function (HEOM), heterogeneous value difference metric (HVDM), and interpolated value difference metric (IVDM), are designed to measure the differences in heterogeneous data. However, these metrics mainly focus on the maximum difference among objects and easily strengthen the importance of categorical features while ignoring the influence of other features in the neighborhood rough set. Therefore, developing a reasonable distance metric for heterogeneous data is one of the motivations of this paper.

How to extract useful information and knowledge from heterogeneous information systems is vital for human decision-making [31]. Considering the heterogeneity induced by various features, many uncertainty measures based on rough set theory are studied to characterize uncertain knowledge and successfully applied in feature selection [32]. Hu et al. [28] designed dependency degree for heterogeneous data to feature selection. As for fuzzy data, Wang et al. [33] proposed a rough set model

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

Witold Pedrycz is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6R 2V4, Canada(e-mail: wpedrycz@ualberta.ca).

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

Hongyun Zhang is with the Department of Computer Science and Technology, Tongji University, Shanghai 201804, China(e-mail: zhanghongyun@tongji.edu.cn).

combining neighborhood granules and designed a feature selection algorithm based on variable precision dependency. Qian et al. [34] developed a local reduction method based on approximation accuracy to improve computational efficiency. In addition, to better describe the classification error, Wang et al. [35] proposed an inner product dependency based on dependency degree. Moreover, the entropy theory is introduced to a rough set to construct uncertainty measures for characterizing the specific information of feature subsets. Zhang et al. [36] designed an incremental selection using a fuzzy-rough-set-based information entropy. Sang et al. [37] developed two incremental approaches using a conditional entropy for monotonic classification in a dynamic order information system. Moreover, Xu et al. [38] considered the decision distribution to define a composite entropy for feature evaluation. The above-referred uncertainty measures are primarily based on a single granule level, i.e., the approximation space or object granules, while ignoring the interaction between different information granule levels. This incomprehensive description of uncertainty could lead to poor performance of learning models. Therefore, developing a novel uncertainty measure is necessary to characterize uncertain knowledge.

Zentropy is a systematic thought that characterizes system chaos from multiple scales [39]. The larger scale provides a whole reflection of lower scales, and the lower scale is the refinement of large scales. In zentropy theory, system entropy is a whole reflection of current scales and all lower scales, where "z" derived from the German zusstandssumme expresses the summation of different scales. Consequently, it can accurately characterize entropy changes with system changes combining different scales, which have been successfully applied in negative thermal expansion of Ce and Fe_3Pt with temperature [39], [40]. This systematic thought is consistent with granular level structure in rough approximation. Inspired by the above analysis, this paper applies this naive zentropy thought to investigate a novel uncertainty measure with the granular level structure of heterogeneous data processing. The main contributions of this paper are as follows.

1) It proposes a robust and effective heterogeneous feature selection based on the zentropy measure and heterogeneous neighborhood model. Plenty of experiment results illustrate the effectiveness of the proposed method.
2) It designs an effective heterogeneous distance metric group (HDMG) to evaluate the difference between objects in heterogeneous decision information systems (HDIS).
3) It defines a robust heterogeneous neighborhood rough set model based on the designed HDMG measure, which could avoid the deviation induced by a single maximum distance in heterogeneous target approximation.
4) It provides a novel zentropy-based uncertainty measure for feature selection by analyzing the finer-coarse granular level in the approximation process. The experimental results on eighteen public datasets illustrate its good performance compared to others.

The paper is organized as follows. Section II reviews some basic concepts. A novel heterogeneous distance metric is proposed to construct heterogeneous neighborhood rough set in Section III. Moreover, Section IV presents a zentropy-based uncertainty measure and designs a corresponding feature selection algorithm based on it. The experiment analysis is shown in Section V. Finally, Section VI concludes this paper.

II. Preliminaries

This section formally reviews the basic notions about neighborhood rough set (NRS) [28] and information measures [34], [37] in DIS and analysis the limitations of this model.

A. Neighborhood Rough Set

A quaternion $\mathbb{DIS} =< \Omega, A, V, M >$ is known as a decision information system, where Ω and A are the universe and feature set, such that.

- $\Omega = \{x_1, x_2, \ldots, x_n\}$ is a non-empty finite set.
- $A = C \cup D$ is a feature set including conditional feature C and decision feature D.
- $V = \bigcup_{a \in A} V_a$ is the value domain of all objects on A.
- $M : \Omega \times A \longrightarrow V$ is a mapping function, where $M(x_i, a)$ represents the value of object $x_i \in \Omega$ in terms of feature a.

For any $x, y, z \in \Omega$ and $B \subseteq C$, a distance denoted as Δ_B is used to measure the difference between these objects, which satisfies the following conditions.

1) *Nonnegative*: $\Delta_B(x, y) \geq 0, \Delta_B(x, x) = 0$;
2) *Symmetry*: $\Delta_B(x, y) = \Delta_B(y, x)$;
3) *Triangle Inequality*: $\Delta_B(x, z) \leq \Delta_B(x, y) + \Delta_B(y, z)$.

Given $B \subseteq C$, the Euclidean distance d_B is usually adopted to evaluate the difference between objects and construct neighborhood granules in a $\mathbb{DIS}$, which is defined as follows.

$$d_B(x, y) = \sqrt{\sum_{b \in B} |M(x, b) - M(y, b)|^2}. \tag{1}$$

Then the neighborhood granule $N_B(x)$ of object $x \in \Omega$ induced by B is represented by:

$$N_B^\delta(x) = \{y \in \Omega | d_B(x, y) \leq \delta\}, \tag{2}$$

where δ is the neighborhood radius.

Definition 1. *[34] Given a $\mathbb{DIS} =< \Omega, A, V, M >$, where $\Omega/D = \{D_1, D_2, \ldots, D_s\}$. For $D_i \in \Omega/D$, $B \subseteq C$ and δ, the lower and upper approximations of D_i with respect to B are defined as follows.*

$$\begin{aligned} \underline{\mathcal{N}_B^\delta}(D_i) &= \{x \in \Omega | N_B^\delta(x) \subseteq D_i\}, \\ \overline{\mathcal{N}_B^\delta}(D_i) &= \{x \in \Omega | N_B^\delta(x) \cap D_i \neq \emptyset\}. \end{aligned} \tag{3}$$

Then, the pair $< \underline{\mathcal{N}_B^\delta}(D_i), \overline{\mathcal{N}_B^\delta}(D_i) >$ is called neighborhood rough set. The positive, negative and boundary regions of D_i are $Pos_B(D_i) = \underline{\mathcal{N}_B^\delta}(D_i)$, $Neg_B(D_i) = U - \overline{\mathcal{N}_B^\delta}(D_i)$ and $Bon_B(D_i) = \overline{\mathcal{N}_B^\delta}(D_i) - \underline{\mathcal{N}_B^\delta}(D_i)$. For convenience, approximation sets are referred to as the lower and upper approximations in this paper.

In order to characterize the precision of NRS, the accuracy measure of D_i on B is proposed as follows.

$$\Upsilon_B(D_i) = \frac{|\underline{\mathcal{N}_B^\delta}(D_i)|}{|D_i|}, \tag{4}$$

where $| \bullet |$ denotes the cardinality, and the accurate measure of D on B is $\Upsilon_B(D) = \frac{\sum_{i=1}^{s} |\underline{\mathcal{N}_B^\delta}(D_i)|}{|U|}$.

B. Entropy-based Uncertainty Measures

Information entropy as an important tool for processing uncertain knowledge has been widely applied to many scenarios. To describe the information amount induced by neighborhood granules, paper [28] defined neighborhood entropy as follows.

Definition 2. *[28] Given a $\mathbb{DIS} =< \Omega, A, V, M >$. For $B \subseteq C$ and δ, neighborhood entropy $NE(B)$ is denoted as follows.*

$$NE(B) = -\frac{1}{|\Omega|}\sum_{i=1}^{|\Omega|} log_2 \frac{|N_B^{\delta}(x_i)|}{|\Omega|}. \tag{5}$$

Similarly, neighborhood conditional entropy is also proposed based on the neighborhood conditional probability [28]. Moreover, Roffo et al. combined the mutual information between features and class to design normalized feature indicator.

Definition 3. *[41] Given a $\mathbb{DIS} =< \Omega, A, V, M >$. For $B \subseteq C$ and decision D, the normalized mutual information between B and D is defined as follows.*

$$M_B = \sum_{y\in D}\sum_{x\in B} p(x,y) log(\frac{p(x,y)}{p(x)p(y)}), \tag{6}$$

where $p(\cdot,\cdot)$ is the joint probability.

To effectively process heterogeneous information, Zhang et al. proposed a neighborhood combination entropy [30].

Definition 4. *[30] Given a $\mathbb{DIS} =< \Omega, A, V, M >$. For $B \subseteq C$ and δ, neighborhood combination entropy $NCE(B)$ is denoted as follows.*

$$NCE(B) = \frac{1}{n}\sum_{i=1}^{n} \frac{C_n^2 - C^2_{|N_B^{\delta}(x_i)|}}{C_n^2}. \tag{7}$$

Many entropy-based measures have been investigated and applied to uncertain information processing since Shannon entropy was proposed. Compared with the accuracy measure and roughness degree based on approximation space, these measures depict the average uncertainty from specific feature values, providing a finer method to analyze uncertainty.

C. Discussion

As an emerging computing paradigm, granular computing is used to address complex problems by simulating human thinking and recognition processes. It focuses on structuring multi-granularity representations for solved issues based on data granulation. Coarser granularity offers a general description, while finer granularity provides details. For instance, animals can be classified into seven main grades, including boundary, phylum, class, order, family, genus, and species, forming a progressively refined multi-granularity structure. GrC thought efficiently enables problem recognition and resolution by integrating information at multiple granular levels.

The rough set is an important GrC model that has been widely used for processing imprecise and uncertain knowledge. In RS, data granulation and granularity information processing are two important issues. As for heterogeneous data granulation, most existing heterogeneous methods evaluate objects from the same space while ignoring the essential differences induced by feature types. Some heterogeneous distance functions, such as heterogeneous distance [30] and heterogeneous Euclidean-overlap metric [28], mainly focus on the maximum distance among objects or set all the distances in the same computing space, ignoring the overall differences in heterogeneous features. These measures impact the precision of data granulation and further influence the accuracy of uncertainty evaluation. Hence, an appropriate distance metric to measure object difference is imperative for information processing in heterogeneous decision information systems.

Note that the approximation process is a multi-granularity structure in the RS theory. It includes target concepts, approximation sets, similarity classes, and further specific objects. The existing measure shown in equation (4) is used to evaluate models based on the lower approximation, which is an overall description of the approximation space. The most entropy-based uncertainty measures characterize systems' uncertainty from object granules, a relatively finer method but ignores the distribution of target decisions. They are all single granular level methods for uncertainty measure. The approximation process is multilevel from object to object granule and further to approximation space and target decision. Therefore, a systematic method combining granular levels is indispensable for the comprehensive characterization of uncertain knowledge.

III. Neighborhood Rough Set in HDIS

To flexible handle heterogeneous information, this section proposes a novel heterogeneous distance metric group (HDMG) to evaluate the difference among objects and then define a neighborhood rough set for HDIS. Compared with an Euclidean distance for numerical data or overlap metric for categorical data, this HDMG could produce a sound comparative analysis and comparison.

A. Heterogeneous Distance Metric

Given a $\mathbb{DIS} =< \Omega, C\cup D, V, M >$, if $C = C^{nu} \cup C^{ca}$ (C^{nu} and C^{ca} are the numerical and categorical feature sets), then the $\mathbb{HDIS} =< \Omega, C\cup D, V, M >$ is defined as a heterogeneous decision information system. Then, the HDMG function is defined for comparing the difference of objects in $\mathbb{HDIS}$.

Definition 5. *Given a $\mathbb{HDIS} =< \Omega, C \cup D, V, M >$, where $C = C^{nu} \cup C^{ca}$. For $B_1 \subseteq C^{nu}$, $B_2 \subseteq C^{ca}$, and $B = B_1 \cup B_2 \subseteq C$, the heterogeneous distance metric group $HDMG_B(x,y)$ of $x, y \in \Omega$ is defined as follows.*

$$HDMG_B(x,y) = (NE_{B_1}(x,y), CO_{B_2}(x,y)), \tag{8}$$

where

$$\begin{aligned} NE_{B_1}(x,y) &= \sqrt{\sum_{b\in B_1} |M(x,b) - M(y,b)|^2}, \\ CO_{B_2}(x,y) &= \sum_{b\in B_2} overlap_{\{b\}}(x,y)/|B_2|, \end{aligned} \tag{9}$$

The overlap degree is used to evaluate the difference between categorical features, which is

$$overlap_{\{b\}}(x,y) = \begin{cases} 1, & M(x,b) \neq M(y,b), \\ 0, & M(x,b) = M(y,b). \end{cases} \tag{10}$$

This article has been accepted for publication in IEEE Transactions on Knowledge and Data Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2024.3419215

In the categorical features without real numerical significance, the overlap degree focuses on the common or differences of categories rather than their specific values, thus helping mitigate the impact of outliers or errors in datasets. It transforms the class differences to specific distance values and can flexibly handle categorical data to construct neighborhood rough modeling by introducing neighborhood parameters. This HDMG, a mapping function from $\Omega \times \Omega$ to $R \times R$, is a 2-tuple used to evaluate the difference among objects in numerical and categorical data, providing an overall evaluation approach to construct heterogeneous neighborhood granules.

B. Heterogeneous Neighborhood Granule

According to Def. 5, the heterogeneous neighborhood granule of $x \in \Omega$ can be defined using the HDMG function and two neighborhood parameters. Initially, it is essential to note that the $(u, v) \leq (\alpha, \beta)$ holds if and only if $u \leq \alpha$ and $v \leq \beta$.

Definition 6. *Given a $\mathbb{HDIS} =< \Omega, C \cup D, V, M >$. For $B \subseteq C$, parameters α and β. The heterogeneous neighborhood granule $HN_B(x)$ of $x \in \Omega$ is defined as follows.*

$$HN_B^{\alpha,\beta}(x) = \{y \in \Omega | HDMG_B(x, y) \leq (\alpha, \beta)\}. \tag{11}$$

Since $(u, v) \leq (\alpha, \beta)$, the heterogeneous neighborhood granule can be decomposed into the neighborhood granules of x with respect to numerical and categorical features.

$$\begin{aligned} HN_B^{\alpha,\beta}(x) &= \{y \in \Omega | HDMG_B(x, y) \leq (\alpha, \beta)\} \\ &= \{y \in \Omega | NE_{B_1}(x, y) \leq \alpha \wedge CO_{B_2}(x, y) \leq \beta\} \\ &= N_{B_1}(x) \cap N_{B_2}(x). \end{aligned} \tag{12}$$

The heterogeneous neighborhood granule structure induced by feature subset B is represented as $HN_B^{\alpha,\beta} = \{HN_B^{\alpha,\beta}(x_1), HN_B^{\alpha,\beta}(x_2), \ldots, HN_B^{\alpha,\beta}(x_n))\}$. In addition, for $B \subseteq Q \subseteq C$, the $HN_Q^{\alpha,\beta}(x) \subseteq HN_B^{\alpha,\beta}(x)$ is easily obtained according to the Eq. 11.

C. Heterogeneous Neighborhood Rough Set

According to the heterogeneous neighborhood granule, the lower and upper approximations of $\mathbb{HDIS}$ can be obtained from the relationship between HN_B and target decisions.

Definition 7. *Given a $\mathbb{HDIS} =< \Omega, C \cup D, V, M >$. For $B \subseteq C$, $D_i \in U/D$, parameters α and β, then the lower and upper approximations of D_i under B are represented by:*

$$\begin{aligned} \underline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_i) &= \{x \in \Omega | HN_B^{\alpha,\beta}(x) \subseteq D_i\}, \\ \overline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_i) &= \{x \in \Omega | HN_B^{\alpha,\beta}(x) \cap D_i \neq \emptyset\}. \end{aligned} \tag{13}$$

Then, the pair $< \underline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_i), \overline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_i) >$ is a neighborhood rough set in $\mathbb{HDIS}$, which is also called a heterogeneous neighborhood rough set ($HNRS$). Obviously, the lower and upper approximations are affected by parameters α and β. Some properties can be easily obtained as follows.

Property 1. *Given a $\mathbb{HDIS} =< \Omega, C \cup D, V, M >$. For $B \subseteq C$, α and β, $\underline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_i)$ and $\overline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_i)$ are the approximations of $D_i \in U/D$, the following properties hold.*

1) *$\emptyset \subseteq \underline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_i) \subseteq D_i \subseteq \overline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_i) \subseteq \Omega$;*
2) *For $P \subseteq B$, $\underline{\mathcal{HN}_{\mathcal{P}}^{\alpha,\beta}}(D_i) \subseteq \underline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_i)$, $\overline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_i) \subseteq \overline{\mathcal{HN}_{\mathcal{P}}^{\alpha,\beta}}(D_i)$;*
3) *For $X \subseteq D_i$, $\underline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(X) \subseteq \underline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_i)$, $\overline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(X) \subseteq \overline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_i)$;*
4) *For $(\mu, \nu) \leq (\alpha, \beta)$, $\underline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_i) \subseteq \underline{\mathcal{HN}_{\mathcal{B}}^{\mu,\nu}}(D_i)$, $\overline{\mathcal{HN}_{\mathcal{B}}^{\mu,\nu}}(D_i) \subseteq \overline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_i)$.*

Proof:

1) It is obviously obtained from Def. 7.
2) From Def. 6, for any $x \in \Omega$, the heterogeneous neighborhood granules $HN_B^{\alpha,\beta}(x) \subseteq HN_P^{\alpha,\beta}(x)$, then $\underline{\mathcal{HN}_{\mathcal{P}}^{\alpha,\beta}}(D_i) \subseteq \underline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_i)$ and $\overline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_i) \subseteq \overline{\mathcal{HN}_{\mathcal{P}}^{\alpha,\beta}}(D_i)$ hold according to Def. 7.
3) For $X \subseteq D_i$ and $x \in \underline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(X)$, $HN_B^{\alpha,\beta}(x) \subseteq X \subseteq D_i$ holds, then $x \in \underline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_i)$. Similarly, $HN_B^{\alpha,\beta}(x) \cap X \neq \emptyset \Rightarrow HN_B^{\alpha,\beta}(x) \cap D_i \neq \emptyset$, thus $\overline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(X) \subseteq \overline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_i)$.
4) According to Def. 6, the $HN_B^{\mu,\nu}(x) \subseteq HN_B^{\alpha,\beta}(x)$ holds when $(\mu, \nu) \leq (\alpha, \beta)$. Similar to 3), the $\underline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_i) \subseteq \underline{\mathcal{HN}_{\mathcal{B}}^{\mu,\nu}}(D_i)$, $\overline{\mathcal{HN}_{\mathcal{B}}^{\mu,\nu}}(D_i) \subseteq \overline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_i)$ hold.

In the $\mathbb{HDIS}$, the lower and upper approximations of D under B can also be represented by:

$$\begin{aligned} \underline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D) &= \{\underline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_1), \underline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_2), \ldots, \underline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_s)\}, \\ \overline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D) &= \{\overline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_1), \overline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_2), \ldots, \overline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_s)\}. \end{aligned} \tag{14}$$

The positive, negative, and boundary regions of D under B in $\mathbb{HDIS}$ are $Pos_B(D) = \cup_{i=1}^{s} \underline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_i)$, $Neg_B(D) = \cup_{i=1}^{s}(U - \overline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_i))$ and $Bon_B(D) = \cup_{i=1}^{s}(\overline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_i) - \underline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_i))$.

Compared with the classical neighborhood rough set, the proposed $HNRS$ considers the difference in feature types, which is relatively comprehensive in describing the relationship between neighborhood granules and target decisions. Thereby, this model settles on the limitations of issue 1 in subsection II-C, and the robustness of this model relative to parameters is verified in the experiments section.

IV. Zentropy-Based Uncertainty Measure for Heterogeneous Data

As for the incomplete uncertainty description, this section investigates the granular structure in approximation process and proposes a novel zentropy-based uncertainty measure for information processing in heterogeneous data.

A. Granular Level Analysis to Approximation Process

The problem formulation and its solution is usually set in a specific conceptual framework composed of some generic and conceptually meaningful entities (i.e., information granules) [13]. The coarser granules are also decomposed into finer granules. Therefore, the final cognition is a systematic and

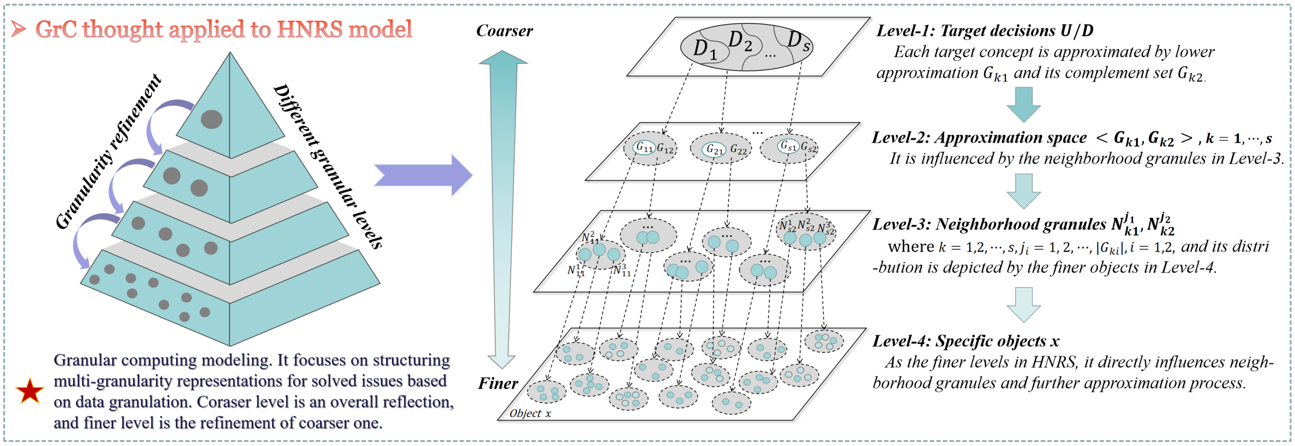


Fig. 1. Granular level structure in HNRS. It shows the granular computing thought in heterogeneous neighborhood rough set models. In the approximation process, there are four granular levels, including target decision, approximation space, neighborhood granules, and specific objects, which form a finer-coarser granular level structure. Each granular level reflects the information from different granules, and the finer level is the refinement of the coarser level.

comprehensive representation from multiple granule levels. As a universal phenomenon in massive data, uncertainty directly influences the learning models' performance. Comprehensively and accurately evaluating uncertainty becomes essential for information processing and knowledge discovery.

Neighborhood rough set as an effective granulation method is widely applied to the uncertainty measure in real-valued data, especially in heterogeneous data. In the approximation process of NRS, the data uncertainty is reflected from a universe to specific objects and quantized by the relationship between neighborhood granule and target concept. This approximation process can be described in Fig.1. Accuracy measure, roughness degree, and positive region are important measures based on approximation space to describe uncertainty from model precision, which can be reflected in the approximation space level (Level-2 in Fig.1). Neighborhood entropy and neighborhood conditional entropy depict the average uncertainty from neighborhood granules level, reflected in neighborhood granules level (Level-3 in Fig.1). These measures provide visions for uncertainty measures from different levels but ignore the interaction between granular levels.

It can be seen from Fig.1 that the approximation process is a finer-coarser granular level structure from the specific object to the whole universe. Each granular level reflects the uncertainty from different granules, and the coarser level is influenced by the finer level. The uncertainty measure on any single granular level is incomplete. Hence, to comprehensively depict the uncertainty in massive complex data, the uncertainty on different granular levels should all be considered and integrated to design an uncertainty measure. This granular level structure is consistent with the zentropy systematic thought. Therefore, developing a novel uncertainty measure combining zentropy thought and granular level is reasonable in HNRS.

B. Zentropy-Based Uncertainty Measure

According to the above analysis, this subsection proposes a novel zentropy-based uncertainty measure to systematically and comprehensively characterize the uncertainty in HDIS.

Definition 8. *Given a* $\mathbb{HDIS} =< \Omega, C \cup D, V, M >$. *For* $B \subseteq C$, *the zentropy-based uncertainty measure under* B *can be defined by:*

$$\mathbf{Z}_B(D) = -\sum_{k=1}^{s} P_k log P_k + \sum_{k=1}^{s} P_k \mathbf{Z}_k, \tag{15}$$

where $P_k = |D_k|/|\Omega|$ *is the probability of* $k-th$ *decision granule in the universe,* $\mathbf{Z}_k$ *is the internal entropy of* $k-th$ *decision granule and can be further decomposed into granules in the finer levels with the same formula as Eq.15.*

Specifically, the internal entropy $\mathbf{Z}_k$ in $level-1$ can be decomposed to the finer approximation layer as follows.

$$\mathbf{Z}_k = -\sum_{j=1}^{2} P_{kj} log P_{kj} + \sum_{j=1}^{2} P_{kj} \mathbf{Z}_{kj}, \tag{16}$$

where $P_{k1} = \frac{\underline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_k)}{|D_k|}$, $P_{k2} = \frac{D_k - \underline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_k)}{|D_k|}$, and $\mathbf{Z}_{kj}$ is the internal entropy of $j-th$ approximation granule in $level-2$, which can be investigated by the neighborhood granules in $level-3$.

$$\mathbf{Z}_{kj} = -\sum_{w=1}^{|G_{kj}|} P_{kjw} log P_{kjw} + \sum_{j=1}^{|G_{kj}|} P_{kjw} \mathbf{Z}_{kjw}, \tag{17}$$

where $G_{k1} = \underline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_k), G_{k2} = D_k - \underline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_k)$, and $P_{kjw} = 1/|G_{kj}|$. The $\mathbf{Z}_{kjw}$ reflects the uncertainty of the neighborhood granule, which can be described as follows.

$$\mathbf{Z}_{kjw} = -\sum_{q=1}^{2} P_{kjwq} log P_{kjwq}, \tag{18}$$

where $P_{kjw1} = \frac{|HN_B^{\alpha,\beta}(w) \cap D_k|}{|HN_B^{\alpha,\beta}(w)|}$, $P_{kjw2} = \frac{|HN_B^{\alpha,\beta}(w) \cap D_k^c|}{|HN_B^{\alpha,\beta}(w)|}$, and D_k^c is the complement set of D_k in Ω.

According to the above definition, this zentropy-based uncertainty measure has the following properties.

Property 2. *Given a* $\mathbb{HDIS} =< \Omega, C \cup D, V, M >$. *For* $B \subseteq C$, $\mathbf{Z}_B(D)$ *is the zentropy-based uncertainty measure of* B *with* D, *then the following properties hold.*

This article has been accepted for publication in IEEE Transactions on Knowledge and Data Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2024.3419215

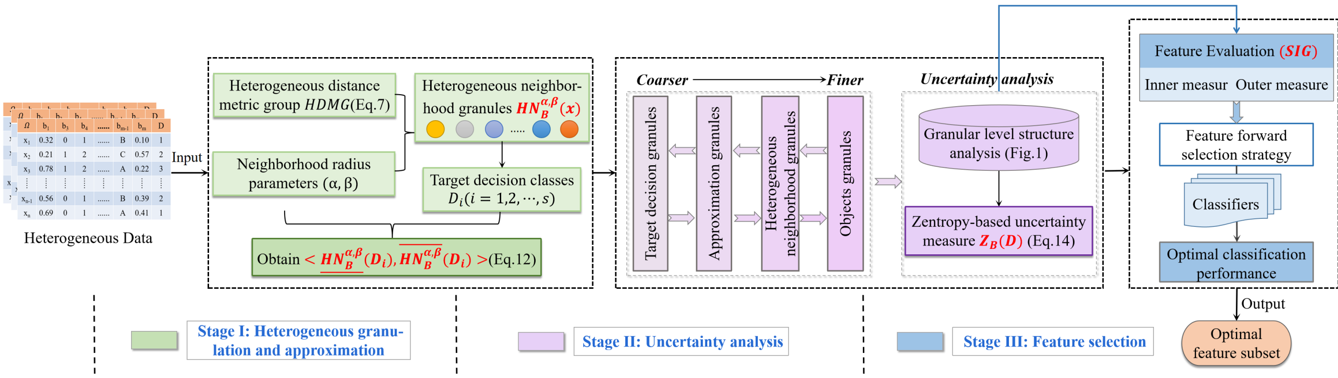


Fig. 2. Framework of the proposed Ze-HFS algorithm.

1) $\mathbf{Z}_B(D) \geq 0$;
2) For $D_k \in \Omega/D$, $\mathbf{Z}_{k1} = \sum_{w=1}^{|G_{k1}|} P_{k1w} log P_{k1w}$;
3) If $\underline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_k) = D_k$, $\mathbf{Z}_k = \mathbf{Z}_{k1}$;
4) If $\underline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_k) = \emptyset$, $\mathbf{Z}_k = \mathbf{Z}_{k2}$;
5) For $P \subseteq B$, $\mathbf{Z}_P(D)$ and $\mathbf{Z}_B(D)$ are indistinguishable.

Proof:

1) From the Defs. 6 and 7, all the probability at each granular level are non-negative, then the function about $-PlogP \geq 0$, thus $\mathbf{Z}_B(D) \geq 0$ for any $B \subseteq C$.
2) According to the Def. 7, for $D_k \in \Omega/D$ and $w \in \underline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_k)$, the neighborhood granule $HN_B^{\alpha,\beta}(w) \subseteq D_k$, thus $P_{kjw1} = 1$ and $\mathbf{Z}_{kjw} = 0$. Therefore, $\mathbf{Z}_{k1} = \sum_{w=1}^{|G_{k1}|} P_{k1w} log P_{k1w}$ holds for any $D_k \in \Omega/D$.
3) From equation (16), $P_{k1} = \frac{\underline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_k)}{|D_k|} = 1$ and $P_{k2} = 1 - P_{k1} = 0$. Thus, $\mathbf{Z}_k = \mathbf{Z}_{k1}$ when $\underline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_k) = D_k$.
4) Similarly to 3), $\mathbf{Z}_k = \mathbf{Z}_{k2}$ holds when $\underline{\mathcal{HN}_{\mathcal{B}}^{\alpha,\beta}}(D_k) = \emptyset$.
5) To prove this property, the size of $\mathbf{Z}_P(D) - \mathbf{Z}_B(D)$ should be calculated. Take the entropy of fouth granular level, for $P \subseteq B$ and $w \in \Omega$, the neighborhood granule $HN_B^{\alpha,\beta}(w) \subseteq HN_P^{\alpha,\beta}(w)$ but the size between $\frac{|HN_P^{\alpha,\beta}(w) \cap D_k|}{|HN_P^{\alpha,\beta}(w)|}$ and $\frac{|HN_B^{\alpha,\beta}(w) \cap D_k|}{|HN_B^{\alpha,\beta}(w)|}$ is indistinguishable, thus $\mathbf{Z}_P(D)$ and $\mathbf{Z}_B(D)$ are indistinguishable.

C. Zentropy Measure to Feature Evaluation

In this subsection, two feature importance measures combining the proposed zentropy-based measure are developed and employed to design a heterogeneous feature selection in HDIS.

Given a $\mathbb{HDIS} =< \Omega, C \cup D, V, M >$. For $\forall c \in C$, the inner importance of c relative C is defined as follows.

$$\mathcal{SIG}_{inner}(c, C, D) = \mathbf{Z}_{C-c}(D) - \mathbf{Z}_C(D). \quad (19)$$

This inner importance of features selects the essential features relative to the original feature set C. When $\mathcal{SIG}_{inner}(c, C, D) > 0$, the reducing of c increases the zentropy values(i.e., increases the system uncertainty), indicating the feature c is essential in $HDIS$. Then, the outer importance of features is designed to select the other features.

Given a $\mathbb{HDIS} =< \Omega, C \cup D, V, M >$. For $B \subseteq C$, the outer importance of $c \in C - B$ is defined as follows.

$$\mathcal{SIG}_{outer}(c, B, D) = \mathbf{Z}_B(D) - \mathbf{Z}_{B \cup \{c\}}(D). \quad (20)$$

In each cycle, the feature with the maximum value of $\mathcal{SIG}_{outer}(c, B, D)$ is selected in feature selection process.

Algorithm 1 Zentropy-based uncertainty measure for heterogeneous feature selection (Ze-HFS).

Input: $\mathbb{HDIS} =< \Omega, C \cup D, V, M >, \alpha, \beta$.
Output: The feature reduction $\mathcal{R}$.

```
1: Initialize R ← ∅, start = 1;
2: Compute Z_C(D) according to Def. 8;
3: for c ∈ C do /* Select the inner important features */
4:     Calculate SIG_inner(c, C, D) from Eq. 19;
5:     if SIG_inner(c, C, D) > 0 then R_0 ← c;
6:     end if
7: end for
8: if R_0 ≠ ∅ then
9:     Compute Z_{R_0}(D) according to Def. 8;
10:    if Z_{R_0}(D) ≤ Z_C(D) then start = 0;
11:    end if
12: end if
13: R = R_0;
14: while start = 1 do
15:    for c ∈ C − R do
16:        Calculate SIG_outer(c, R, D) according to Eq. 20;
17:    end for
18:    Obtain c_0 = argmax_{c∈C−R} SIG_outer(c, R, D);
19:    R ← c_0; /* Select the outer important features */
20:    Compute the zentropy-based measure Z_R(D);
21:    if Z_R(D) > Z_C(D) then start = 1;
22:    else start = 0;
23:    end if
24: end while
25: for r ∈ R do /* Delete the redundant features */
26:    Compute Z_{R−r}(D);
27:    if Z_{R−r}(D) < Z_R(D) then R ← R − r;
28:    end if
29: end for
30: return R.
```

Definition 9. *Given a $\mathbb{HDIS} =< \Omega, C \cup D, V, M >$, the $\mathcal{R}$ is a feature reduct of C when the following properties satisfy:*
1) $\mathbf{Z}_{\mathcal{R}}(D) \leq \mathbf{Z}_B(D)$;
2) For any $r \in \mathcal{R}$, $\mathbf{Z}_{\mathcal{R}-r}(D) \geq \mathbf{Z}_{\mathcal{R}}(D)$.

According to the important measures, a feature selection algorithm considering granular level structure is designed in algorithm 1. In this algorithm, step 2 computes $\mathbf{Z}_C(D)$ with time complexity $O(mn^2)$ due to the computation between n objects under m features. Steps 3-7 are used to select the inner important features and obtain the origin reduct set $\mathcal{R}_0$, whose time complexity is $O(m(m-1)n^2)$. When $\mathcal{R}_0 \neq \emptyset$, steps 8-12 needs to compute $\mathbf{Z}_{\mathcal{R}_0}(D)$, whose time complexity is $O(|\mathcal{R}_0|n^2)$. Similarly, if there are l selected features in steps 13-22, the time complexity is $O(\sum_{i=1}^{l}(|C-\mathcal{R}_0|-i-1)(|\mathcal{R}_0|+i)n^2)$. Finally, the redundant features are removed in steps 23-27 with the time complexity $O(|\mathcal{R}_0|(|\mathcal{R}_0|-1)n^2)$. Hence, the whole complexity of this algorithm is $O(m^2n^2)$. The detailed process is shown in Fig. 2.

V. Experimental Analysis

In this section, a series of experiments are carried out to demonstrate the robustness and effectiveness of the proposed method. All the experiments are run on a computer with OS: Microsoft WIN10; Processor: Intel(R) Core(TM) i7-6800K CPU @ 3.4GHz×12; Memory: 62.7 GB; Programming language: MATLAB R2020b.

A. Experimental Design

Eighteen datasets from UCI Machine Learning Repository [1], ASU feature selection datasets[2], and SEER Program[3] are selected for experiments, where the detailed information is shown in Table I. All the numerical features are first transformed to interval $[0, 1]$ by the Max-min normalization.

TABLE I
Dataset Description

No.s	Datasets	Objects	Features	Classes	Types
D1	Heart disease	270	14	2	Mixed
D2	Seer breast cancer	4024	15	2	Mixed
D3	Segmentation	2310	20	7	Numerical
D4	South german credit	1000	21	2	Numerical
D5	Soybean large	250	36	13	Categorical
D6	Spectf heart	267	45	2	Categorical
D7	Thyroid	7200	22	3	Mixed
D8	Arrhythmia	450	250	3	Mixed
D9	Colon	62	2001	2	Numerical
D10	Glioma	50	4435	4	Numerical
D11	Leukemia	72	7071	2	Numerical
D12	Lung	203	3312	5	Numerical
D13	Lymphography	96	4027	9	Categorical
D14	Warppie10p	210	2421	10	Numerical
D15	German	1000	21	2	Mixed
D16	Nci9	60	9173	9	Categorical
D17	Sports	1000	60	2	Categorical
D18	Warpar10p	130	2400	10	Numerical

In the experimental evaluation, the robustness of the proposed heterogeneous distance metric group (HDMG) is illustrated by comparing it with the classical heterogeneous distance (CHD) [30] and heterogeneous Euclidean-overlap metric function (HEOM) [29] in approximation accuracy under different parameters. In addition, to illustrate the superiority of proposed Ze-HFS, twelve other representative feature selection algorithms, including feature selection with Gaussian kernelized fuzzy rough sets (FGR) [42], fuzzy rough algorithm with minimum misclassification rate (FRMR) [35], rough set based online streaming feature selection (ROFS) [43], rough set based on the relative stability of local redundancy (RLR) [44], mutual infinite latent feature selection (MI) [45], Laplacian score for feature selection (LS) [46], feature selection with neighborhood combination entropy (FCE) [30], wrapper set-based integer-coded fuzzy granular evolutionary (WSFE) algorithm [47], infinite feature selection (IFS) [41] with filtering framework, evolutionary-filter approach neighborhood component analysis (ENCA) [48], principal component analysis (PCA) [49], and sparse feature selection via fast embedding spectral analysis (SFSE) [50] are compared.

Four classifiers, including K-nearest neighbor (KNN, K=3), naive Bayesian (NB, kernel function), decision tree (DT, Gini index), and feedforward neural network (FNN, Tanh function, and ten neurons) are adopted to evaluate the classification performance of different methods. The 10-fold cross-validation method is employed to make comparisons of classification performance in different feature selection algorithms. All the average results and standard deviation are abbreviated as Ave and St, and the best results are in bold.

It is clear that α and β are two critical parameters that directly influence the selected features of Ze-HFS by influencing neighborhood granules. To investigate the influence of these parameters on selected features, the parameters α and β are set from 0 to 0.25 with a step of 0.025 to find the optimal feature subset. The parameters of other compared algorithms are set consistent with their corresponding references.

TABLE II
Accuracy measure comparison under different parameters

No.s	Metric	0.05	0.15	0.25	0.35	0.45	0.55	Ave$\pm St$
D2	HDM	1.00	1.00	1.00	1.00	1.00	1.00	**1.00±0.00**
	HEOM	0.94	0.62	0.44	0.03	0.00	0.00	0.34±0.39
	HDMG	1.00	1.00	0.98	0.97	0.56	0.43	0.82±0.26
D3	HDM	0.98	0.65	0.35	0.22	0.05	0.00	0.37±0.38
	HEOM	0.76	0.25	0.01	0.00	0.00	0.00	0.17±0.30
	HDMG	1.00	0.88	0.68	0.41	0.35	0.32	**0.61±0.29**
D4	HDM	1.00	1.00	1.00	0.76	0.72	0.34	0.80±0.26
	HEOM	1.00	1.00	1.00	1.00	1.00	1.00	**1.00±0.00**
	HDMG	1.00	1.00	1.00	0.99	0.98	0.93	0.98±0.03
D6	HDM	1.00	1.00	1.00	1.00	1.00	1.00	**1.00±0.00**
	HEOM	1.00	1.00	1.00	1.00	1.00	1.00	**1.00±0.00**
	HDMG	1.00	1.00	1.00	1.00	1.00	1.00	**1.00±0.00**
D8	HDM	1.00	1.00	1.00	0.87	0.72	0.63	0.87±0.16
	HEOM	1.00	0.19	0.01	0.00	0.00	0.00	0.20±0.40
	HDMG	1.00	1.00	1.00	1.00	1.00	1.00	**1.00±0.00**
D9	HDM	0.97	0.32	0.02	0.02	0.02	0.02	0.22±0.38
	HEOM	0.02	0.02	0.02	0.02	0.02	0.02	0.02±0.00
	HDMG	1.00	0.84	0.30	0.06	0.02	0.02	**0.37±0.44**
D13	HDM	1.00	1.00	1.00	1.00	1.00	1.00	**1.00±0.00**
	HEOM	1.00	1.00	1.00	0.99	0.81	0.13	0.82±0.35
	HDMG	1.00	0.99	0.41	0.05	0.01	0.00	0.41±0.48
D14	HDM	1.00	1.00	1.00	1.00	0.98	0.86	0.97±0.06
	HEOM	1.00	0.46	0.00	0.00	0.00	0.00	0.24±0.41
	HDMG	1.00	1.00	1.00	1.00	1.00	1.00	**1.00±0.00**

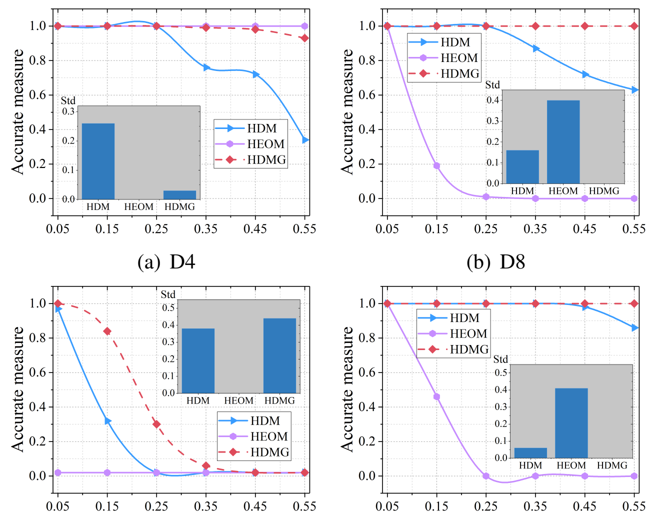
B. Robustness Evaluation of HNRS Model

This subsection illustrates the robustness of the proposed heterogeneous distance metric by analyzing the fluctuation of accurate measures with parameter variation. Note that the accurate measure is calculated by equation (4). The parameters of compared heterogeneous distance metric are all set from 0.05 to 0.55 with a step of 0.1.

[1]http://archive.ics.uci.edu/ml/datasets.php
[2]http://featureselection.asu.edu/datasets.php
[3]https://seer.cancer.gov/data-software/

This article has been accepted for publication in IEEE Transactions on Knowledge and Data Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2024.3419215

Under different parameters, the accurate measure values of three compared distance metrics on eight randomly selected datasets are recorded in Table II. The last column displays the average values, and the large values denote a higher accuracy of corresponding models. From this table, the accurate measures for HDM, HEOM, and HDMG decline as the parameters increase. That is because the radius increase will enlarge the neighborhood particles, causing the inclusion relation to not hold. The HDMG achieves five maximum values and four minimum standard errors in eight datasets, indicating the excellent approximation accuracy and robustness of the proposed metric related to parameters. The detailed comparison of four datasets is shown in Fig.3.



(a) D4 (b) D8

(c) D9 (d) D14

Fig. 3. Accurate measure values of different heterogeneous functions under various neighborhood parameters.

C. Classification Performance Evaluation

This subsection mainly evaluates the proposed Ze-HFS method by comparing it with twelve other representative feature selection methods in classification performance. All the experimental results of eighteen public datasets are analyzed from three aspects: running time and feature number, classification accuracy, and statistical test.

*1) **Running time and feature number** :* This part compares the running time and selected feature number of different methods on 18 public datasets. Notably, MI and LS are feature ranking methods, thus ensuring consistency in the selected feature number with Ze-HFS for equitable comparison. The running time of different methods is recorded in Table III, where the unit is second (s). From this table, the RLR, MIFS, LS, IFS, ENCA, and PCA achieve the minimum running time on 4, 1, 9, 3, 1, and 1 datasets. For the average times, the Ze-HFS can be better than the FGR, ROFS, FCE, WSFE, ENCA, and SFSE methods. Moreover, for selected feature numbers, Ze-HFS achieves minimum numbers three times on KNN and NB classifiers, and two times on DT and FNN classifiers from Tables IV and V. In comparison, ENCA obtains the minimum number seven times on four classifiers while

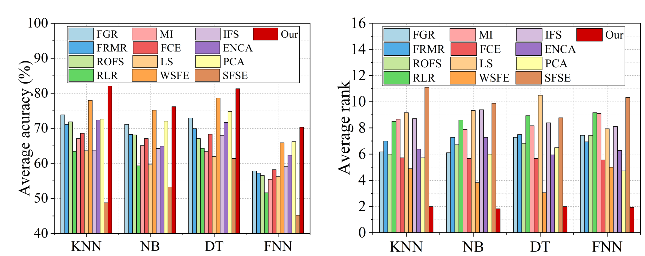
TABLE III
RUNNING TIME OF DIFFERENT METHODS (S)

No.s	FGR	FRMR	ROFS	RLR	MI	FCE	LS	WSFE	IFS	ENCA	PCA	SFSE	Our
D1	2.6	2.0	6.7	**0.4**	1.8	15.3	1.8	14.2	1.0	2.4	17.2	19.1	1.1
D2	523	28.7	1741	9.7	4.0	63.6	4.0	8.6	**2.5**	2.7	63.3	103	377
D3	1056	84.2	287	**7.7**	22.2	195	12.8	94.7	17.4	7.9	259	220	62.9
D4	65.5	3.9	44.1	1.6	6.0	52.3	1.9	10.8	1.5	**0.9**	36.5	46.3	64.1
D5	26.9	8.7	18.1	2.6	2.2	60.1	**2.0**	131	5.4	7.3	160	318	4.8
D6	9.2	2.7	21.3	**0.7**	10.6	96.0	8.6	9.6	2.2	1.9	27.9	103	15.8
D7	2165	202	8017	29.8	**0.8**	675	**0.8**	99.2	5.1	17.0	603	508	1616
D8	349	261	719	335	26.6	1384	**9.3**	3420	10.0	20.0	204	12588	241
D9	107	1.1	276	9.7	1.7	78.7	1.3	41.0	28.0	685	**1.0**	129	24.9
D10	39.0	7.0	494	21.0	1.2	166	**1.1**	5298	168	1086	57.0	147	52.8
D11	70.2	15.1	969	35.8	1.0	536	**0.8**	157	311	2099	143	248	94.7
D12	525	257	1026	50.9	1.7	566	**1.3**	19481	156	862	1218	3169	217
D13	1.7	3.0	4.5	**0.4**	1.2	23.2	1.1	14.0	2.0	1.0	31.0	35.0	1.5
D14	761	83.7	1216	128	5.7	4413	**5.1**	7408	145	738	46.0	5582	608
D15	71.0	16.9	30.3	1.9	12.9	715	7.6	15.0	**1.0**	2.0	30.0	36.0	126
D16	368	99.2	284	909	9.3	152	**7.5**	8778	2448	23957	355	449	520
D17	605	42.2	80.1	21.3	2.8	5568	3.6	12.0	**2.0**	3.0	40.0	252	123
D18	221	70.9	176	247	9.0	45.1	**8.5**	6222	136	984	119	2306	158
Ave	387.0	66.1	856.1	100.7	6.7	822.4	**4.4**	2845.2	191.2	1693.2	189.5	1458.8	239.4

selecting many features on other high-dimensional datasets, such as D8, D9, and D11. This method is inefficient for selecting the appropriate feature subset in feature selection, which significantly influences the classification performance. Similarly, it can be seen from the FCE method. On different classifiers, the proposed method obtains the minimum average number and rank compared with other methods.

*2) **Classification accuracy**:* This part evaluates the classification accuracy of different methods on public datasets. From Tables VI-IX, the proposed Ze-HFS method achieves better classification accuracy in most datasets on different classifiers.

Specifically, Ze-HFS obtains the best accuracy in 13, 13, 13, and 14 datasets on KNN, NB, DT, and FNN classifiers, respectively. Moreover, it could obtain the following conclusions by analyzing different data types. On the KNN classifier, Ze-HFS achieves maximum accuracy in five mixed, five categorical, and eight numerical datasets, with 2, 5, and 6 datasets, respectively. On the NB classifier, Ze-HFS performs better in 5, 3, and 5 datasets within the respective categories. When compared with DT classifier, Ze-HFS achieves the maximum accuracy 2, 4, and 7 times, and 4, 4, and 6 times in selected mixed, categorical, and numerical datasets on FNN classifier. Additionally, it exhibits superior average classification accuracy and ranking across all tested classifiers, as shown in Fig. 4.



(a) Accuracy (b) Rank

Fig. 4. The average accuracy and rank of different algorithms on 18 datasets.

According to the above analysis, the proposed Ze-HFS could achieve better classification accuracy with fewer features, illustrating its superiority in feature selection.

This article has been accepted for publication in IEEE Transactions on Knowledge and Data Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2024.3419215

TABLE IV
THE NUMBER OF SELECTED FEATURES BY THIRTEEN COMPARED METHODS ON KNN AND NB CLASSIFIERS

No.s	FGR		FRMR		ROFS		RLR		MI		FCE		LS		WSFE		IFS		ENCA		PCA		SFSE		Our	
	KNN	NB	KNN	NB	KNN	NB	KNN	NB	KNN	NB	KNN	NB	KNN	NB	KNN	NB	KNN	NB	KNN	NB	KNN	NB	KNN	NB	KNN	NB
D1	7	7	13	13	1	1	4	4	6	5	12	11	6	5	3	3	2	2	1	1	8	2	10	1	6	5
D2	11	11	11	11	2	2	11	11	6	5	1	1	6	5	1	1	2	2	1	1	8	1	10	3	6	5
D3	11	11	17	17	9	9	9	9	13	14	14	15	13	14	5	5	4	4	1	1	9	9	2	2	13	14
D4	10	10	20	20	2	2	6	6	6	7	7	3	6	7	1	3	3	3	1	1	12	10	4	1	6	7
D5	11	11	11	11	1	1	23	23	28	20	15	14	28	20	16	10	7	7	9	9	14	18	5	1	28	20
D6	8	8	31	31	3	3	4	4	3	1	5	3	3	1	1	1	8	8	4	4	5	5	34	1	3	1
D7	20	20	18	18	10	10	20	20	3	3	16	20	3	3	3	4	3	3	1	1	17	4	3	3	3	3
D8	8	8	107	107	31	31	28	28	6	9	38	66	6	9	139	127	27	27	49	49	27	6	206	1	6	9
D9	24	24	1	1	15	15	7	7	2	16	1	1	2	16	2	3	218	218	675	675	1	1	62	28	2	16
D10	3	3	9	9	13	13	1	1	2	3	1	1	2	3	77	9	732	732	1755	1755	27	9	17	44	2	3
D11	3	3	20	20	10	10	11	11	2	2	1	1	2	2	34	37	36	36	36	36	56	5	34	72	2	2
D12	4	4	58	58	59	59	1	1	7	7	1	1	7	7	368	355	407	407	126	126	41	14	136	202	7	7
D13	6	6	16	16	4	4	9	9	12	13	10	15	12	13	5	3	3	3	1	1	4	12	2	1	12	13
D14	5	5	12	12	62	62	4	4	19	20	128	192	19	20	176	213	388	388	107	107	10	11	43	209	19	20
D15	10	10	20	20	5	5	11	11	9	2	6	7	9	2	3	2	2	2	1	1	14	10	4	2	9	2
D16	10	10	6	6	1	1	12	12	8	1	1	1	8	1	151	157	1506	1506	4475	4475	15	5	4	3	8	1
D17	21	21	38	1	1	1	5	5	1	1	35	9	1	1	3	2	6	6	6	6	16	3	16	1	1	1
D18	5	5	6	6	15	15	6	6	5	8	162	191	5	8	137	166	386	386	138	138	13	15	15	129	5	8
Ave	9.83	9.83	23.00	20.94	13.56	13.56	9.56	9.56	7.67	7.61	25.22	30.67	7.67	7.61	62.50	61.17	207.78	207.78	410.39	410.39	16.50	7.78	33.72	39.11	7.67	7.61
Rank	7.39	7.44	9.17	8.72	5.11	5.61	6.61	7.00	4.72	4.89	6.89	6.94	4.72	4.89	6.50	6.78	7.33	7.72	6.39	6.50	7.89	6.28	7.33	5.11	4.72	4.89

TABLE V
THE NUMBER OF SELECTED FEATURES BY THIRTEEN COMPARED METHODS ON DT AND FNN CLASSIFIERS

No.s	FGR		FRMR		ROFS		RLR		MI		FCE		LS		WSFE		IFS		ENCA		PCA		SFSE		Our	
	DT	FNN	DT	FNN	DT	FNN	DT	FNN	DT	FNN	DT	FNN	DT	FNN	DT	FNN	DT	FNN	DT	FNN	DT	FNNN	DT	FNN	DT	FNN
D1	7	7	13	13	1	1	4	4	3	10	12	11	3	10	3	3	2	2	1	1	5	8	11	2	3	10
D2	11	11	11	11	2	2	11	11	3	6	1	1	3	6	1	1	2	2	1	1	9	9	5	8	3	6
D3	11	11	17	17	9	9	9	9	11	13	15	14	11	13	5	5	4	4	1	1	8	8	10	16	11	13
D4	10	10	20	20	2	2	6	6	13	8	10	18	13	8	3	2	3	3	1	1	6	10	5	17	13	8
D5	11	11	11	11	1	1	23	23	28	30	15	15	28	30	17	16	7	7	9	9	6	18	32	13	28	30
D6	8	8	1	31	3	3	4	4	1	3	1	41	1	3	1	1	8	8	4	4	2	5	2	36	1	3
D7	20	20	1	18	10	10	20	20	20	19	20	19	20	19	4	4	3	3	1	1	17	16	3	3	20	19
D8	8	8	107	107	31	31	28	28	1	1	80	68	1	1	13	129	27	27	49	49	11	9	10	5	1	1
D9	24	24	1	1	15	15	7	7	3	1	1	1	3	1	21	8	218	218	675	675	1	1	56	2	3	1
D10	3	3	9	1	13	13	1	1	2	2	1	1	2	2	241	170	732	732	1755	1755	12	9	48	5	2	2
D11	3	3	20	1	10	10	11	11	2	2	1	1	2	2	34	1	36	36	36	36	33	56	70	2	2	2
D12	4	4	58	58	59	59	1	1	3	4	1	1	3	4	164	360	407	407	126	126	32	9	203	203	3	4
D13	6	6	16	1	4	4	9	9	8	6	12	8	8	6	5	5	3	3	1	1	8	3	7	14	8	6
D14	5	5	12	12	62	62	4	4	20	19	49	139	20	19	223	127	388	388	107	107	11	10	210	209	20	19
D15	10	10	1	20	5	5	11	11	3	6	10	10	3	6	3	3	2	2	1	1	11	13	5	18	3	6
D16	10	10	6	6	1	1	12	12	10	8	1	1	10	8	182	222	1506	1506	4475	4475	21	13	57	3	10	8
D17	21	21	1	38	1	1	5	5	2	2	40	33	2	2	3	3	6	6	6	6	7	12	8	9	2	2
D18	5	5	6	6	15	15	6	6	7	7	146	145	7	7	207	189	386	386	138	138	23	19	130	127	7	7
Ave	9.83	9.83	17.28	20.67	13.56	13.56	9.56	9.56	7.78	8.17	23.11	29.28	7.78	8.17	62.78	69.39	207.78	207.78	410.39	410.39	12.39	12.67	48.44	38.44	7.78	8.17
Rank	7.06	7.22	6.56	7.33	5.50	5.17	6.83	6.83	4.83	5.06	6.67	7.39	4.83	5.06	6.78	6.50	7.61	7.28	6.67	6.22	6.94	7.50	8.94	8.17	4.83	5.06

TABLE VI
CLASSIFICATION ACCURACY OF THIRTEEN COMPARED METHODS WITH KNN CLASSIFIER(%)

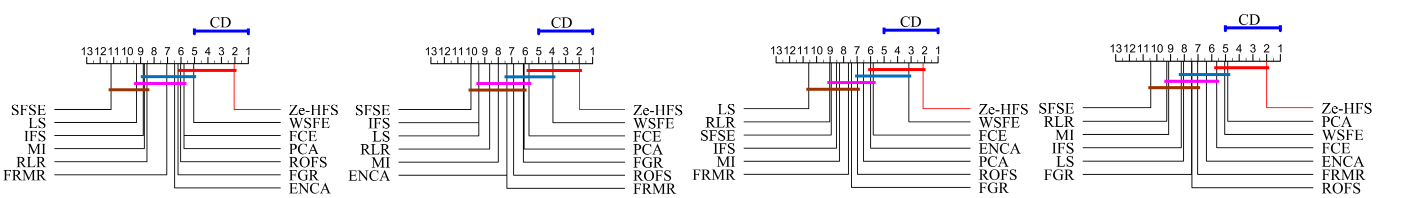
No.s	FGR	FRMR	ROFS	RLR	MI	FCE	LS	WSFE	IFS	ENCA	PCA	SFSE	Our
D1	69.26$_{\pm 0.07}$	78.15$_{\pm 0.06}$	57.78$_{\pm 0.11}$	65.56$_{\pm 0.07}$	76.30$_{\pm 0.11}$	80.37$_{\pm 0.09}$	72.96$_{\pm 0.06}$	78.15$_{\pm 0.10}$	65.19$_{\pm 0.14}$	67.41$_{\pm 0.08}$	81.48$_{\pm 0.10}$	48.89$_{\pm 0.12}$	**83.33**$_{\pm 0.07}$
D2	87.48$_{\pm 0.01}$	87.18$_{\pm 0.01}$	**89.04**$_{\pm 0.02}$	87.37$_{\pm 0.02}$	87.82$_{\pm 0.02}$	88.37$_{\pm 0.02}$	46.47$_{\pm 0.10}$	87.85$_{\pm 0.02}$	19.93$_{\pm 0.02}$	87.67$_{\pm 0.02}$	84.37$_{\pm 0.03}$	71.15$_{\pm 0.29}$	88.07$_{\pm 0.02}$
D3	96.49$_{\pm 0.01}$	95.97$_{\pm 0.01}$	96.28$_{\pm 0.01}$	89.74$_{\pm 0.02}$	96.15$_{\pm 0.01}$	96.32$_{\pm 0.02}$	90.95$_{\pm 0.02}$	95.02$_{\pm 0.01}$	63.03$_{\pm 0.04}$	49.65$_{\pm 0.02}$	88.74$_{\pm 0.02}$	14.42$_{\pm 0.02}$	**96.84**$_{\pm 0.01}$
D4	71.20$_{\pm 0.05}$	72.00$_{\pm 0.05}$	69.00$_{\pm 0.05}$	73.50$_{\pm 0.06}$	69.30$_{\pm 0.06}$	74.60$_{\pm 0.03}$	71.70$_{\pm 0.04}$	70.00$_{\pm 0.05}$	69.80$_{\pm 0.05}$	70.00$_{\pm 0.05}$	73.40$_{\pm 0.05}$	70.20$_{\pm 0.04}$	**75.50**$_{\pm 0.05}$
D5	66.40$_{\pm 0.08}$	60.00$_{\pm 0.07}$	28.00$_{\pm 0.07}$	80.00$_{\pm 0.06}$	84.80$_{\pm 0.11}$	60.40$_{\pm 0.07}$	81.20$_{\pm 0.08}$	**86.00**$_{\pm 0.05}$	47.20$_{\pm 0.11}$	85.20$_{\pm 0.07}$	82.00$_{\pm 0.10}$	16.80$_{\pm 0.07}$	**86.00**$_{\pm 0.07}$
D6	75.98$_{\pm 0.12}$	74.89$_{\pm 0.09}$	75.97$_{\pm 0.05}$	73.75$_{\pm 0.09}$	73.03$_{\pm 0.09}$	76.93$_{\pm 0.06}$	74.54$_{\pm 0.10}$	69.69$_{\pm 0.08}$	76.15$_{\pm 0.09}$	75.36$_{\pm 0.07}$	76.47$_{\pm 0.10}$	71.14$_{\pm 0.10}$	**77.79**$_{\pm 0.09}$
D7	93.79$_{\pm 0.01}$	92.86$_{\pm 0.01}$	98.00$_{\pm 0.01}$	94.00$_{\pm 0.01}$	96.78$_{\pm 0.01}$	95.40$_{\pm 0.01}$	92.58$_{\pm 0.01}$	**98.03**$_{\pm 0.01}$	92.58$_{\pm 0.01}$	95.56$_{\pm 0.01}$	97.51$_{\pm 0.01}$	92.58$_{\pm 0.01}$	96.88$_{\pm 0.01}$
D8	53.11$_{\pm 0.08}$	56.67$_{\pm 0.07}$	58.00$_{\pm 0.07}$	57.11$_{\pm 0.06}$	20.67$_{\pm 0.24}$	59.56$_{\pm 0.08}$	51.11$_{\pm 0.08}$	59.11$_{\pm 0.07}$	53.11$_{\pm 0.06}$	**61.56**$_{\pm 0.09}$	61.11$_{\pm 0.07}$	54.22$_{\pm 0.09}$	59.78$_{\pm 0.05}$
D9	59.52$_{\pm 0.24}$	50.24$_{\pm 0.18}$	80.71$_{\pm 0.15}$	54.76$_{\pm 0.20}$	54.29$_{\pm 0.26}$	59.76$_{\pm 0.13}$	59.05$_{\pm 0.23}$	66.19$_{\pm 0.21}$	57.62$_{\pm 0.18}$	52.38$_{\pm 0.21}$	53.57$_{\pm 0.20}$	65.71$_{\pm 0.18}$	**90.24**$_{\pm 0.12}$
D10	58.00$_{\pm 0.26}$	62.00$_{\pm 0.24}$	72.00$_{\pm 0.17}$	28.00$_{\pm 0.23}$	60.00$_{\pm 0.25}$	48.00$_{\pm 0.17}$	44.00$_{\pm 0.25}$	84.00$_{\pm 0.13}$	64.00$_{\pm 0.23}$	74.00$_{\pm 0.14}$	76.00$_{\pm 0.18}$	36.00$_{\pm 0.23}$	**86.00**$_{\pm 0.14}$
D11	96.75$_{\pm 0.04}$	86.07$_{\pm 0.15}$	80.89$_{\pm 0.13}$	72.14$_{\pm 0.17}$	65.18$_{\pm 0.15}$	65.89$_{\pm 0.24}$	65.18$_{\pm 0.18}$	93.04$_{\pm 0.07}$	83.39$_{\pm 0.11}$	87.50$_{\pm 0.12}$	91.43$_{\pm 0.12}$	61.43$_{\pm 0.15}$	**97.32**$_{\pm 0.06}$
D12	80.79$_{\pm 0.06}$	94.07$_{\pm 0.06}$	**95.67**$_{\pm 0.06}$	64.43$_{\pm 0.12}$	78.29$_{\pm 0.09}$	64.50$_{\pm 0.12}$	79.24$_{\pm 0.12}$	95.52$_{\pm 0.06}$	92.67$_{\pm 0.07}$	95.60$_{\pm 0.03}$	86.10$_{\pm 0.03}$	68.50$_{\pm 0.11}$	89.76$_{\pm 0.09}$
D13	75.05$_{\pm 0.11}$	76.29$_{\pm 0.11}$	79.14$_{\pm 0.11}$	75.71$_{\pm 0.06}$	72.95$_{\pm 0.17}$	79.76$_{\pm 0.11}$	78.43$_{\pm 0.16}$	77.81$_{\pm 0.12}$	54.57$_{\pm 0.09}$	67.57$_{\pm 0.10}$	79.05$_{\pm 0.10}$	47.24$_{\pm 0.16}$	**82.48**$_{\pm 0.09}$
D14	87.14$_{\pm 0.05}$	80.00$_{\pm 0.10}$	95.24$_{\pm 0.04}$	29.23$_{\pm 0.09}$	83.81$_{\pm 0.09}$	**99.05**$_{\pm 0.02}$	63.33$_{\pm 0.08}$	92.38$_{\pm 0.05}$	95.19$_{\pm 0.06}$	93.81$_{\pm 0.08}$	51.90$_{\pm 0.12}$	19.52$_{\pm 0.08}$	95.24$_{\pm 0.05}$
D15	71.40$_{\pm 0.05}$	72.90$_{\pm 0.06}$	70.90$_{\pm 0.06}$	71.40$_{\pm 0.04}$	69.90$_{\pm 0.03}$	**76.00**$_{\pm 0.05}$	67.80$_{\pm 0.04}$	67.80$_{\pm 0.06}$	54.60$_{\pm 0.07}$	64.80$_{\pm 0.11}$	71.70$_{\pm 0.03}$	30.70$_{\pm 0.05}$	**73.50**$_{\pm 0.05}$
D16	46.67$_{\pm 0.22}$	21.67$_{\pm 0.16}$	13.33$_{\pm 0.15}$	18.33$_{\pm 0.12}$	15.00$_{\pm 0.12}$	15.00$_{\pm 0.12}$	11.67$_{\pm 0.11}$	43.33$_{\pm 0.26}$	53.33$_{\pm 0.26}$	43.33$_{\pm 0.18}$	43.33$_{\pm 0.21}$	21.67$_{\pm 0.16}$	**55.00**$_{\pm 0.21}$
D17	79.30$_{\pm 0.03}$	78.80$_{\pm 0.02}$	64.20$_{\pm 0.06}$	79.90$_{\pm 0.05}$	63.50$_{\pm 0.03}$	81.10$_{\pm 0.04}$	66.70$_{\pm 0.03}$	78.20$_{\pm 0.04}$	62.50$_{\pm 0.05}$	79.20$_{\pm 0.04}$	71.80$_{\pm 0.02}$	63.60$_{\pm 0.05}$	**74.00**$_{\pm 0.02}$
D18	60.77$_{\pm 0.12}$	40.77$_{\pm 0.15}$	69.23$_{\pm 0.14}$	26.92$_{\pm 0.09}$	40.00$_{\pm 0.14}$	27.69$_{\pm 0.16}$	27.69$_{\pm 0.09}$	61.54$_{\pm 0.08}$	43.08$_{\pm 0.14}$	52.31$_{\pm 0.07}$	37.69$_{\pm 0.11}$	23.85$_{\pm 0.12}$	**69.58**$_{\pm 0.09}$
Ave$_{\pm St}$	73.84$_{\pm 0.09}$	71.14$_{\pm 0.09}$	71.85$_{\pm 0.08}$	63.44$_{\pm 0.09}$	67.10$_{\pm 0.11}$	68.59$_{\pm 0.09}$	63.59$_{\pm 0.10}$	77.98$_{\pm 0.08}$	63.77$_{\pm 0.10}$	72.38$_{\pm 0.08}$	72.65$_{\pm 0.09}$	48.76$_{\pm 0.11}$	**82.07**$_{\pm 0.07}$
Rank	6.17	7.00	6.00	8.50	8.67	5.72	9.17	4.89	8.72	6.39	5.72	11.11	**2.00**

TABLE VII
CLASSIFICATION ACCURACY OF THIRTEEN COMPARED METHODS WITH NB CLASSIFIER(%)

No.s	FGR	FRMR	ROFS	RLR	MI	FCE	LS	WSFE	IFS	ENCA	PCA	SFSE	Our
D1	72.96$_{\pm0.07}$	79.63$_{\pm0.07}$	60.37$_{\pm0.11}$	66.30$_{\pm0.10}$	82.22$_{\pm0.08}$	80.00$_{\pm0.07}$	77.41$_{\pm0.11}$	78.52$_{\pm0.10}$	65.19$_{\pm0.14}$	67.41$_{\pm0.09}$	81.11$_{\pm0.05}$	48.52$_{\pm0.11}$	**82.22**$_{\pm0.06}$
D2	88.30$_{\pm0.02}$	88.32$_{\pm0.02}$	**89.69**$_{\pm0.02}$	87.85$_{\pm0.02}$	88.74$_{\pm0.01}$	89.31$_{\pm0.02}$	84.86$_{\pm0.03}$	89.31$_{\pm0.02}$	84.69$_{\pm0.01}$	89.36$_{\pm0.01}$	85.31$_{\pm0.01}$	23.39$_{\pm0.23}$	**89.69**$_{\pm0.02}$
D3	92.12$_{\pm0.01}$	90.65$_{\pm0.02}$	90.65$_{\pm0.01}$	86.62$_{\pm0.02}$	92.42$_{\pm0.01}$	91.82$_{\pm0.02}$	72.94$_{\pm0.03}$	**92.55**$_{\pm0.01}$	64.20$_{\pm0.03}$	50.26$_{\pm0.03}$	82.90$_{\pm0.02}$	15.32$_{\pm0.04}$	92.38$_{\pm0.01}$
D4	70.90$_{\pm0.03}$	70.50$_{\pm0.05}$	70.00$_{\pm0.05}$	71.40$_{\pm0.03}$	69.60$_{\pm0.06}$	**73.00**$_{\pm0.04}$	70.00$_{\pm0.06}$	72.20$_{\pm0.05}$	69.20$_{\pm0.05}$	70.00$_{\pm0.05}$	71.40$_{\pm0.05}$	69.90$_{\pm0.07}$	72.50$_{\pm0.05}$
D5	43.60$_{\pm0.10}$	40.00$_{\pm0.11}$	10.80$_{\pm0.04}$	40.80$_{\pm0.07}$	42.80$_{\pm0.07}$	39.60$_{\pm0.12}$	39.20$_{\pm0.06}$	46.80$_{\pm0.13}$	33.60$_{\pm0.11}$	36.40$_{\pm0.07}$	43.60$_{\pm0.06}$	16.80$_{\pm0.11}$	**54.40**$_{\pm0.08}$
D6	77.55$_{\pm0.07}$	69.62$_{\pm0.08}$	79.67$_{\pm0.08}$	78.97$_{\pm0.07}$	79.36$_{\pm0.12}$	79.54$_{\pm0.11}$	79.39$_{\pm0.10}$	79.44$_{\pm0.11}$	71.44$_{\pm0.09}$	77.54$_{\pm0.09}$	72.54$_{\pm0.05}$	79.44$_{\pm0.06}$	79.44$_{\pm0.09}$
D7	94.03$_{\pm0.01}$	92.56$_{\pm0.02}$	94.07$_{\pm0.01}$	94.04$_{\pm0.01}$	96.67$_{\pm0.01}$	94.10$_{\pm0.01}$	92.58$_{\pm0.01}$	95.44$_{\pm0.01}$	92.58$_{\pm0.01}$	95.79$_{\pm0.01}$	94.47$_{\pm0.01}$	92.58$_{\pm0.01}$	**97.44**$_{\pm0.01}$
D8	59.11$_{\pm0.09}$	61.33$_{\pm0.06}$	61.11$_{\pm0.06}$	59.33$_{\pm0.09}$	54.22$_{\pm0.06}$	**62.67**$_{\pm0.07}$	56.00$_{\pm0.05}$	62.00$_{\pm0.08}$	59.56$_{\pm0.06}$	62.33$_{\pm0.07}$	56.44$_{\pm0.08}$	33.78$_{\pm0.07}$	**62.67**$_{\pm0.06}$
D9	64.29$_{\pm0.22}$	65.00$_{\pm0.26}$	54.52$_{\pm0.31}$	30.00$_{\pm0.26}$	58.81$_{\pm0.15}$	65.71$_{\pm0.22}$	40.95$_{\pm0.26}$	65.48$_{\pm0.22}$	31.43$_{\pm0.17}$	36.90$_{\pm0.22}$	62.62$_{\pm0.22}$	63.33$_{\pm0.14}$	**66.90**$_{\pm0.19}$
D10	66.00$_{\pm0.23}$	58.00$_{\pm0.15}$	70.00$_{\pm0.24}$	34.00$_{\pm0.19}$	58.00$_{\pm0.20}$	50.00$_{\pm0.22}$	48.00$_{\pm0.23}$	72.00$_{\pm0.23}$	24.00$_{\pm0.23}$	36.00$_{\pm0.31}$	76.00$_{\pm0.23}$	52.00$_{\pm0.32}$	**84.00**$_{\pm0.16}$
D11	92.86$_{\pm0.18}$	65.54$_{\pm0.17}$	76.96$_{\pm0.17}$	68.04$_{\pm0.12}$	59.46$_{\pm0.21}$	65.89$_{\pm0.24}$	63.75$_{\pm0.14}$	91.79$_{\pm0.10}$	77.86$_{\pm0.17}$	73.75$_{\pm0.17}$	88.75$_{\pm0.13}$	70.18$_{\pm0.22}$	**94.64**$_{\pm0.09}$
D12	82.79$_{\pm0.10}$	84.26$_{\pm0.09}$	90.21$_{\pm0.06}$	68.29$_{\pm0.14}$	79.26$_{\pm0.09}$	68.45$_{\pm0.11}$	72.93$_{\pm0.15}$	86.71$_{\pm0.07}$	87.21$_{\pm0.04}$	78.43$_{\pm0.14}$	91.12$_{\pm0.06}$	78.83$_{\pm0.06}$	**91.19**$_{\pm0.05}$
D13	75.71$_{\pm0.10}$	79.10$_{\pm0.13}$	72.57$_{\pm0.14}$	76.33$_{\pm0.10}$	75.48$_{\pm0.15}$	80.29$_{\pm0.12}$	78.24$_{\pm0.15}$	71.10$_{\pm0.11}$	69.86$_{\pm0.12}$	55.38$_{\pm0.11}$	75.52$_{\pm0.13}$	43.24$_{\pm0.07}$	**80.52**$_{\pm0.10}$
D14	78.57$_{\pm0.09}$	67.14$_{\pm0.14}$	90.95$_{\pm0.09}$	21.54$_{\pm0.10}$	56.19$_{\pm0.11}$	90.95$_{\pm0.07}$	24.29$_{\pm0.11}$	92.86$_{\pm0.03}$	88.57$_{\pm0.07}$	94.29$_{\pm0.06}$	74.76$_{\pm0.03}$	78.10$_{\pm0.07}$	80.95$_{\pm0.10}$
D15	71.40$_{\pm0.04}$	70.80$_{\pm0.06}$	69.90$_{\pm0.05}$	71.20$_{\pm0.05}$	70.00$_{\pm0.04}$	72.10$_{\pm0.04}$	69.20$_{\pm0.04}$	72.00$_{\pm0.04}$	69.80$_{\pm0.04}$	70.00$_{\pm0.05}$	70.90$_{\pm0.03}$	65.00$_{\pm0.15}$	**72.20**$_{\pm0.06}$
D16	20.00$_{\pm0.17}$	21.67$_{\pm0.18}$	10.00$_{\pm0.14}$	16.67$_{\pm0.16}$	6.67$_{\pm0.09}$	5.00$_{\pm0.08}$	11.67$_{\pm0.16}$	45.00$_{\pm0.18}$	51.67$_{\pm0.27}$	41.67$_{\pm0.23}$	35.00$_{\pm0.22}$	25.00$_{\pm0.12}$	23.33$_{\pm0.14}$
D17	74.90$_{\pm0.04}$	74.30$_{\pm0.03}$	63.80$_{\pm0.08}$	75.20$_{\pm0.04}$	63.50$_{\pm0.04}$	74.30$_{\pm0.05}$	75.70$_{\pm0.05}$	71.20$_{\pm0.04}$	62.90$_{\pm0.05}$	74.70$_{\pm0.04}$	65.80$_{\pm0.02}$	63.50$_{\pm0.03}$	**75.90**$_{\pm0.03}$
D18	55.38$_{\pm0.07}$	50.00$_{\pm0.09}$	70.77$_{\pm0.13}$	20.77$_{\pm0.10}$	37.69$_{\pm0.15}$	24.62$_{\pm0.16}$	15.38$_{\pm0.09}$	69.23$_{\pm0.15}$	52.31$_{\pm0.09}$	58.46$_{\pm0.15}$	69.23$_{\pm0.16}$	40.00$_{\pm0.13}$	**70.79**$_{\pm0.16}$
Ave$_{\pm St}$	71.14$_{\pm0.09}$	68.24$_{\pm0.10}$	68.11$_{\pm0.10}$	59.30$_{\pm0.09}$	65.06$_{\pm0.09}$	67.08$_{\pm0.10}$	59.58$_{\pm0.10}$	75.20$_{\pm0.09}$	64.23$_{\pm0.10}$	64.93$_{\pm0.10}$	72.08$_{\pm0.09}$	53.27$_{\pm0.11}$	**76.18**$_{\pm0.08}$
Rank	6.11	7.28	6.72	8.61	7.89	5.67	9.33	3.83	9.39	7.28	6.00	9.89	**1.83**

TABLE VIII
CLASSIFICATION ACCURACY OF THIRTEEN COMPARED METHODS WITH DT CLASSIFIER(%)

No.s	FGR	FRMR	ROFS	RLR	MI	FCE	LS	WSFE	IFS	ENCA	PCA	SFSE	Our
D1	64.07$_{\pm0.10}$	74.44$_{\pm0.09}$	67.78$_{\pm0.06}$	66.67$_{\pm0.09}$	77.41$_{\pm0.09}$	79.63$_{\pm0.07}$	75.19$_{\pm0.10}$	84.81$_{\pm0.08}$	64.81$_{\pm0.14}$	74.07$_{\pm0.07}$	78.52$_{\pm0.09}$	55.56$_{\pm0.07}$	**86.30**$_{\pm0.06}$
D2	86.31$_{\pm0.01}$	86.43$_{\pm0.02}$	**89.16**$_{\pm0.01}$	85.56$_{\pm0.02}$	87.55$_{\pm0.02}$	88.19$_{\pm0.02}$	84.67$_{\pm0.03}$	89.14$_{\pm0.01}$	84.69$_{\pm0.01}$	88.56$_{\pm0.01}$	82.41$_{\pm0.02}$	84.69$_{\pm0.02}$	88.67$_{\pm0.01}$
D3	95.80$_{\pm0.01}$	95.45$_{\pm0.01}$	95.80$_{\pm0.01}$	90.17$_{\pm0.01}$	13.38$_{\pm0.02}$	96.06$_{\pm0.01}$	68.87$_{\pm0.02}$	96.23$_{\pm0.01}$	67.88$_{\pm0.04}$	51.08$_{\pm0.03}$	88.14$_{\pm0.01}$	12.16$_{\pm0.01}$	**96.32**$_{\pm0.01}$
D4	70.70$_{\pm0.06}$	69.10$_{\pm0.05}$	65.70$_{\pm0.06}$	65.90$_{\pm0.03}$	70.80$_{\pm0.03}$	73.10$_{\pm0.05}$	67.50$_{\pm0.04}$	73.10$_{\pm0.05}$	68.80$_{\pm0.05}$	71.70$_{\pm0.03}$	70.60$_{\pm0.04}$	70.30$_{\pm0.05}$	**73.40**$_{\pm0.04}$
D5	76.80$_{\pm0.10}$	66.00$_{\pm0.05}$	21.60$_{\pm0.09}$	82.80$_{\pm0.09}$	82.80$_{\pm0.08}$	54.80$_{\pm0.13}$	78.40$_{\pm0.08}$	86.40$_{\pm0.05}$	54.40$_{\pm0.14}$	86.40$_{\pm0.08}$	81.20$_{\pm0.08}$	14.40$_{\pm0.04}$	86.00$_{\pm0.08}$
D6	74.22$_{\pm0.08}$	75.24$_{\pm0.07}$	75.66$_{\pm0.08}$	71.51$_{\pm0.07}$	75.56$_{\pm0.11}$	77.37$_{\pm0.06}$	70.67$_{\pm0.13}$	77.45$_{\pm0.09}$	75.33$_{\pm0.07}$	73.02$_{\pm0.07}$	73.03$_{\pm0.06}$	74.57$_{\pm0.08}$	**77.48**$_{\pm0.07}$
D7	**99.64**$_{\pm0.00}$	92.58$_{\pm0.01}$	99.17$_{\pm0.00}$	99.32$_{\pm0.00}$	97.31$_{\pm0.01}$	97.61$_{\pm0.00}$	92.58$_{\pm0.01}$	98.28$_{\pm0.01}$	92.58$_{\pm0.01}$	96.03$_{\pm0.01}$	96.79$_{\pm0.01}$	92.28$_{\pm0.01}$	98.07$_{\pm0.00}$
D8	53.56$_{\pm0.07}$	56.00$_{\pm0.06}$	62.67$_{\pm0.05}$	58.22$_{\pm0.06}$	53.33$_{\pm0.03}$	**64.22**$_{\pm0.09}$	45.78$_{\pm0.08}$	48.89$_{\pm0.06}$	49.33$_{\pm0.06}$	58.44$_{\pm0.06}$	55.11$_{\pm0.05}$	47.11$_{\pm0.06}$	50.67$_{\pm0.05}$
D9	70.24$_{\pm0.19}$	60.24$_{\pm0.08}$	82.38$_{\pm0.19}$	64.29$_{\pm0.26}$	60.48$_{\pm0.15}$	65.00$_{\pm0.16}$	61.67$_{\pm0.24}$	84.76$_{\pm0.14}$	78.10$_{\pm0.19}$	71.43$_{\pm0.17}$	52.38$_{\pm0.18}$	65.24$_{\pm0.16}$	**90.71**$_{\pm0.13}$
D10	50.00$_{\pm0.25}$	64.00$_{\pm0.26}$	50.00$_{\pm0.24}$	32.00$_{\pm0.22}$	50.00$_{\pm0.25}$	50.00$_{\pm0.14}$	44.00$_{\pm0.21}$	60.00$_{\pm0.13}$	48.00$_{\pm0.17}$	64.00$_{\pm0.21}$	68.00$_{\pm0.19}$	56.00$_{\pm0.21}$	**78.00**$_{\pm0.1}$
D11	97.14$_{\pm0.06}$	91.43$_{\pm0.07}$	83.21$_{\pm0.13}$	76.43$_{\pm0.18}$	65.18$_{\pm0.19}$	61.61$_{\pm0.21}$	57.14$_{\pm0.16}$	**100.00**$_{\pm0.00}$	93.04$_{\pm0.07}$	94.46$_{\pm0.10}$	90.54$_{\pm0.11}$	77.68$_{\pm0.10}$	**100.00**$_{\pm0.00}$
D12	78.88$_{\pm0.06}$	86.21$_{\pm0.09}$	83.19$_{\pm0.10}$	61.67$_{\pm0.13}$	77.29$_{\pm0.11}$	58.17$_{\pm0.14}$	76.71$_{\pm0.09}$	90.69$_{\pm0.08}$	87.74$_{\pm0.09}$	87.19$_{\pm0.10}$	88.74$_{\pm0.08}$	85.67$_{\pm0.10}$	**91.12**$_{\pm0.05}$
D13	72.19$_{\pm0.16}$	74.24$_{\pm0.12}$	79.71$_{\pm0.11}$	78.52$_{\pm0.11}$	78.29$_{\pm0.10}$	82.48$_{\pm0.05}$	72.90$_{\pm0.09}$	80.38$_{\pm0.09}$	73.86$_{\pm0.10}$	74.38$_{\pm0.09}$	76.33$_{\pm0.10}$	54.90$_{\pm0.16}$	**83.86**$_{\pm0.10}$
D14	72.38$_{\pm0.14}$	50.00$_{\pm0.08}$	60.95$_{\pm0.12}$	36.67$_{\pm0.10}$	73.33$_{\pm0.10}$	84.76$_{\pm0.11}$	39.52$_{\pm0.10}$	78.57$_{\pm0.05}$	80.48$_{\pm0.08}$	71.90$_{\pm0.12}$	76.52$_{\pm0.10}$	**86.67**$_{\pm0.09}$	80.48$_{\pm0.07}$
D15	67.50$_{\pm0.07}$	70.00$_{\pm0.05}$	70.2$_{\pm0.05}$	67.70$_{\pm0.06}$	69.8$_{\pm0.04}$	73.40$_{\pm0.05}$	68.90$_{\pm0.03}$	72.20$_{\pm0.04}$	68.00$_{\pm0.05}$	70.90$_{\pm0.04}$	70.00$_{\pm0.02}$	70.10$_{\pm0.05}$	**73.70**$_{\pm0.03}$
D16	60.00$_{\pm0.20}$	28.33$_{\pm0.16}$	5.00$_{\pm0.11}$	20.00$_{\pm0.13}$	6.67$_{\pm0.12}$	18.33$_{\pm0.12}$	13.33$_{\pm0.07}$	48.33$_{\pm0.27}$	20.00$_{\pm0.13}$	25.00$_{\pm0.16}$	55.00$_{\pm0.16}$	38.33$_{\pm0.12}$	**61.67**$_{\pm0.18}$
D17	72.90$_{\pm0.04}$	75.60$_{\pm0.03}$	64.10$_{\pm0.08}$	72.90$_{\pm0.06}$	63.50$_{\pm0.02}$	**79.20**$_{\pm0.02}$	77.10$_{\pm0.05}$	78.90$_{\pm0.04}$	64.10$_{\pm0.06}$	73.30$_{\pm0.05}$	77.40$_{\pm0.06}$	63.50$_{\pm0.05}$	**79.20**$_{\pm0.03}$
D18	50.77$_{\pm0.10}$	43.08$_{\pm0.15}$	51.54$_{\pm0.19}$	26.92$_{\pm0.10}$	37.69$_{\pm0.18}$	26.15$_{\pm0.08}$	20.00$_{\pm0.15}$	**67.69**$_{\pm0.09}$	53.08$_{\pm0.15}$	58.46$_{\pm0.15}$	66.15$_{\pm0.17}$	55.38$_{\pm0.17}$	**67.69**$_{\pm0.11}$
Ave$_{\pm St}$	72.95$_{\pm0.09}$	69.91$_{\pm0.08}$	67.10$_{\pm0.09}$	64.29$_{\pm0.10}$	63.35$_{\pm0.09}$	68.34$_{\pm0.08}$	61.94$_{\pm0.09}$	78.66$_{\pm0.07}$	68.01$_{\pm0.09}$	71.69$_{\pm0.09}$	74.83$_{\pm0.08}$	61.38$_{\pm0.09}$	**81.30**$_{\pm0.07}$
Rank	7.28	7.50	6.83	8.94	8.17	5.67	10.50	3.06	8.39	5.94	6.50	8.78	**2.00**



(a) KNN (b) NB (c) DT (d) FNN

Fig. 5. Nemenyi test of Ze-HFS and twelve other methods. (a)-(c) show the test results of classification accuracy on KNN, NB, DT, and FNN classifiers.

*3) **Statistical test:*** To test whether Ze-HFS is significantly different from other compared methods, the Friedman test [51] with significance $P = 0.1$ is first performed on 18 datasets. The null hypothesis of Friedman test is that all the compared methods are equivalent, and it is rejected when the test p-value is smaller than significance level P. The testing p-values of Friedman test on four classifiers are 1.68×10^{-11}, 3.75×10^{-11}, 5.52×10^{-12} and 1.43×10^{-10} that all far less than 0.1, indicating there exist statistical differences among all compared methods. To further evaluate the difference between any two methods, Nemenyi's post hoc test [51] is conducted. The null hypothesis is that the compared methods are the same, which can be rejected when the rank difference is greater than the CD critical value, computed as follows.

$$CD = q_\alpha \sqrt{\frac{k(k+1)}{6N}}, \tag{21}$$

where $q_{\alpha=0.1} = 1.9076$ when $k = 18$, and $N = 13$.

The results of Nemenyi's post hoc are shown in Fig. 5. The classification accuracy of Ze-HFS ranks first and is significantly different from other 9, 10, 9, and 9 methods on KNN, NB, DT, and FNN classifiers. All results show the best performance of the proposed feature selection in classification.

TABLE IX
CLASSIFICATION ACCURACY OF THIRTEEN COMPARED METHODS WITH FNN CLASSIFIER(%)

No.s	FGR	FRMR	ROFS	RLR	MI	FCE	LS	WSFE	IFS	ENCA	PCA	SFSE	Our
D1	71.85$_{\pm 0.06}$	77.41$_{\pm 0.10}$	68.15$_{\pm 0.07}$	67.78$_{\pm 0.09}$	79.26$_{\pm 0.09}$	81.85$_{\pm 0.10}$	78.89$_{\pm 0.06}$	81.11$_{\pm 0.09}$	65.93$_{\pm 0.14}$	74.07$_{\pm 0.07}$	83.33$_{\pm 0.08}$	55.19$_{\pm 0.12}$	**83.70**$_{\pm 0.04}$
D2	89.81$_{\pm 0.02}$	89.89$_{\pm 0.01}$	89.69$_{\pm 0.01}$	89.39$_{\pm 0.02}$	88.99$_{\pm 0.02}$	89.81$_{\pm 0.02}$	85.19$_{\pm 0.02}$	89.84$_{\pm 0.02}$	84.57$_{\pm 0.01}$	89.89$_{\pm 0.01}$	89.14$_{\pm 0.02}$	84.69$_{\pm 0.02}$	**90.43**$_{\pm 0.02}$
D3	90.35$_{\pm 0.06}$	85.84$_{\pm 0.09}$	94.94$_{\pm 0.02}$	87.79$_{\pm 0.02}$	77.06$_{\pm 0.07}$	95.45$_{\pm 0.01}$	88.05$_{\pm 0.03}$	94.11$_{\pm 0.02}$	61.04$_{\pm 0.04}$	50.13$_{\pm 0.03}$	85.41$_{\pm 0.02}$	14.33$_{\pm 0.02}$	**95.84**$_{\pm 0.01}$
D4	8.80$_{\pm 0.03}$	8.10$_{\pm 0.02}$	1.30$_{\pm 0.02}$	7.90$_{\pm 0.04}$	4.00$_{\pm 0.02}$	10.50$_{\pm 0.04}$	7.60$_{\pm 0.03}$	9.30$_{\pm 0.04}$	9.00$_{\pm 0.03}$	2.80$_{\pm 0.02}$	9.30$_{\pm 0.04}$	0.80$_{\pm 0.01}$	**11.10**$_{\pm 0.07}$
D5	70.40$_{\pm 0.08}$	56.40$_{\pm 0.09}$	25.20$_{\pm 0.08}$	72.80$_{\pm 0.11}$	76.80$_{\pm 0.04}$	46.80$_{\pm 0.12}$	75.20$_{\pm 0.10}$	80.4$_{\pm 0.07}$	49.60$_{\pm 0.09}$	80.00$_{\pm 0.07}$	78.80$_{\pm 0.11}$	17.60$_{\pm 0.07}$	**85.20**$_{\pm 0.07}$
D6	2.96$_{\pm 0.06}$	5.63$_{\pm 0.05}$	4.91$_{\pm 0.06}$	2.98$_{\pm 0.06}$	0.38$_{\pm 0.01}$	**10.90**$_{\pm 0.09}$	4.15$_{\pm 0.08}$	2.25$_{\pm 0.03}$	4.97$_{\pm 0.08}$	2.98$_{\pm 0.04}$	5.24$_{\pm 0.06}$	5.57$_{\pm 0.08}$	6.01$_{\pm 0.08}$
D7	93.53$_{\pm 0.01}$	92.94$_{\pm 0.01}$	93.21$_{\pm 0.01}$	94.25$_{\pm 0.01}$	95.44$_{\pm 0.01}$	96.81$_{\pm 0.01}$	92.58$_{\pm 0.01}$	95.40$_{\pm 0.02}$	92.58$_{\pm 0.01}$	94.64$_{\pm 0.01}$	96.49$_{\pm 0.01}$	92.58$_{\pm 0.01}$	**96.85**$_{\pm 0.01}$
D8	59.78$_{\pm 0.08}$	60.89$_{\pm 0.11}$	67.78$_{\pm 0.07}$	57.11$_{\pm 0.08}$	55.78$_{\pm 0.05}$	64.89$_{\pm 0.07}$	56.67$_{\pm 0.08}$	52.22$_{\pm 0.20}$	56.22$_{\pm 0.03}$	**68.67**$_{\pm 0.06}$	62.89$_{\pm 0.05}$	54.22$_{\pm 0.09}$	55.78$_{\pm 0.09}$
D9	19.29$_{\pm 0.13}$	35.48$_{\pm 0.15}$	34.05$_{\pm 0.23}$	29.05$_{\pm 0.23}$	31.90$_{\pm 0.11}$	35.00$_{\pm 0.16}$	34.52$_{\pm 0.20}$	34.29$_{\pm 0.31}$	30.00$_{\pm 0.17}$	25.24$_{\pm 0.15}$	34.52$_{\pm 0.24}$	52.62$_{\pm 0.17}$	**56.67**$_{\pm 0.06}$
D10	46.00$_{\pm 0.28}$	52.00$_{\pm 0.17}$	64.00$_{\pm 0.23}$	36.00$_{\pm 0.08}$	42.00$_{\pm 0.18}$	50.00$_{\pm 0.22}$	34.00$_{\pm 0.16}$	74.00$_{\pm 0.21}$	52.00$_{\pm 0.22}$	70.00$_{\pm 0.25}$	80.00$_{\pm 0.25}$	38.00$_{\pm 0.15}$	**86.00**$_{\pm 0.16}$
D11	1.43$_{\pm 0.05}$	27.86$_{\pm 0.10}$	18.04$_{\pm 0.13}$	16.79$_{\pm 0.19}$	33.39$_{\pm 0.20}$	31.96$_{\pm 0.15}$	30.36$_{\pm 0.18}$	58.04$_{\pm 0.25}$	69.46$_{\pm 0.21}$	59.82$_{\pm 0.22}$	61.00$_{\pm 0.00}$	**71.43**$_{\pm 0.21}$	63.57$_{\pm 0.14}$
D12	81.74$_{\pm 0.10}$	80.36$_{\pm 0.07}$	89.11$_{\pm 0.07}$	69.05$_{\pm 0.07}$	76.93$_{\pm 0.09}$	67.07$_{\pm 0.11}$	79.33$_{\pm 0.07}$	84.50$_{\pm 0.25}$	82.78$_{\pm 0.22}$	89.06$_{\pm 0.06}$	84.62$_{\pm 0.05}$	77.33$_{\pm 0.06}$	**89.12**$_{\pm 0.10}$
D13	79.10$_{\pm 0.09}$	70.95$_{\pm 0.08}$	79.76$_{\pm 0.10}$	71.57$_{\pm 0.11}$	75.67$_{\pm 0.10}$	80.43$_{\pm 0.08}$	77.00$_{\pm 0.10}$	74.48$_{\pm 0.12}$	73.19$_{\pm 0.10}$	74.38$_{\pm 0.09}$	81.05$_{\pm 0.14}$	54.19$_{\pm 0.09}$	**82.48**$_{\pm 0.12}$
D14	74.29$_{\pm 0.08}$	79.05$_{\pm 0.09}$	81.90$_{\pm 0.07}$	46.67$_{\pm 0.09}$	84.76$_{\pm 0.07}$	100.00$_{\pm 0.00}$	87.62$_{\pm 0.08}$	**100.00**$_{\pm 0.00}$	98.57$_{\pm 0.02}$	99.52$_{\pm 0.02}$	99.52$_{\pm 0.02}$	28.10$_{\pm 0.12}$	94.76$_{\pm 0.04}$
D15	71.70$_{\pm 0.05}$	71.90$_{\pm 0.05}$	72.70$_{\pm 0.05}$	73.00$_{\pm 0.04}$	70.90$_{\pm 0.04}$	70.30$_{\pm 0.05}$	73.00$_{\pm 0.04}$	73.20$_{\pm 0.04}$	68.30$_{\pm 0.05}$	71.70$_{\pm 0.06}$	71.10$_{\pm 0.03}$	70.10$_{\pm 0.03}$	**75.30**$_{\pm 0.05}$
D16	51.67$_{\pm 0.27}$	20.00$_{\pm 0.11}$	15.00$_{\pm 0.12}$	16.67$_{\pm 0.18}$	8.33$_{\pm 0.12}$	20.00$_{\pm 0.15}$	11.67$_{\pm 0.11}$	46.67$_{\pm 0.17}$	40.00$_{\pm 0.31}$	40.00$_{\pm 0.21}$	41.67$_{\pm 0.16}$	16.67$_{\pm 0.16}$	**55.00**$_{\pm 0.08}$
D17	77.20$_{\pm 0.05}$	76.10$_{\pm 0.08}$	64.70$_{\pm 0.06}$	73.10$_{\pm 0.20}$	63.50$_{\pm 0.05}$	73.50$_{\pm 0.05}$	77.10$_{\pm 0.04}$	75.70$_{\pm 0.16}$	65.10$_{\pm 0.06}$	71.30$_{\pm 0.15}$	71.30$_{\pm 0.16}$	63.60$_{\pm 0.04}$	**77.20**$_{\pm 0.04}$
D18	50.77$_{\pm 0.13}$	39.23$_{\pm 0.13}$	53.08$_{\pm 0.27}$	16.92$_{\pm 0.08}$	33.08$_{\pm 0.15}$	22.31$_{\pm 0.06}$	19.23$_{\pm 0.09}$	60.08$_{\pm 0.21}$	60.15$_{\pm 0.10}$	58.46$_{\pm 0.17}$	56.15$_{\pm 0.18}$	16.92$_{\pm 0.08}$	**60.77**$_{\pm 0.21}$
Ave$_{\pm St}$	57.81$_{\pm 0.09}$	57.22$_{\pm 0.08}$	56.53$_{\pm 0.09}$	51.60$_{\pm 0.09}$	55.45$_{\pm 0.08}$	58.20$_{\pm 0.08}$	56.23$_{\pm 0.08}$	65.87$_{\pm 0.12}$	59.08$_{\pm 0.10}$	62.37$_{\pm 0.09}$	66.20$_{\pm 0.0}$9	45.22$_{\pm 0.09}$	**70.32**$_{\pm 0.08}$
Rank	7.44	6.94	7.44	9.17	9.11	5.56	7.94	5.00	8.11	6.28	4.72	10.33	**1.94**

TABLE X
AVERAGE NUMBER OF SELECTED FEATURES BY THIRTEEN COMPARED METHODS UNDER SIX NOISE LEVELS ON NB AND FNN CLASSIFIERS

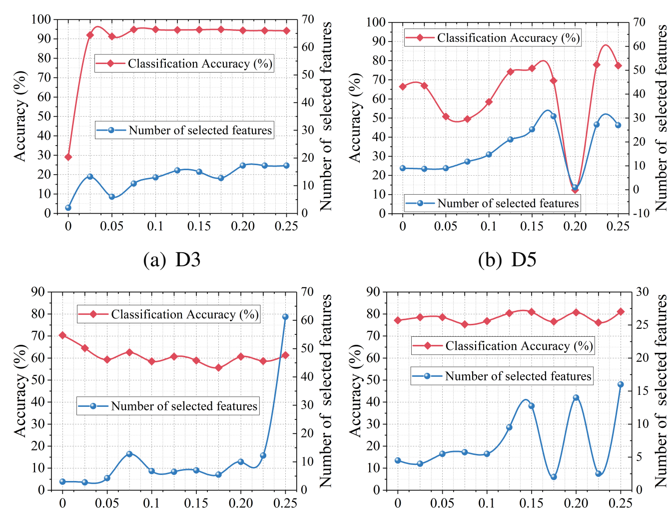
No.s	FGR		FRMR		ROFS		RLR		MI		FCE		LS		WSFE		IFS		ENCA		PCA		SFSE		Our	
	NB	FNN	NB	FNN	NB	FNN	NB	FNN	NB	FNN	NB	FNN	NB	FNN	NB	FNN	NB	FNN	NB	FNN	NB	FNN	NB	FNN	NB	FNN
D1	7.0	7.0	9.0	7.0	6.8	6.8	4.8	4.8	6.8	6.2	9.3	12.2	6.8	6.2	3.8	4.5	2.3	2.3	1.0	1.0	7.7	4.5	8.2	4.0	6.8	6.2
D2	11.5	11.5	3.2	11.8	6.0	6.0	11.3	11.3	2.0	5.0	1.0	9.7	2.0	5.0	1.2	1.2	2.2	2.2	1.0	1.0	7.0	8.0	6.5	6.8	2.0	5.0
D3	10.8	10.8	16.0	16.0	7.7	7.7	8.2	8.2	11.2	12.0	15.7	17.8	11.2	12.0	5.5	5.5	1.7	1.7	1.0	1.0	11.7	12.7	8.0	10.0	11.2	12.0
D4	9.7	9.7	1.0	20.0	1.5	1.5	6.3	6.3	3.0	8.0	4.7	16.3	3.0	8.0	1.3	1.5	2.8	2.8	2.3	2.3	8.3	11.8	7.3	7.5	3.0	8.0
D5	8.2	8.2	10.0	10.0	1.5	1.5	16.2	16.2	11.0	9.7	11.2	7.8	11.0	9.7	12.0	11.3	5.8	5.8	11.2	11.2	16.2	15.2	22.2	18.7	11.0	9.7
D6	7.0	7.0	1.0	40.7	9.2	9.2	3.5	3.5	1.3	2.8	8.8	27.2	1.3	2.8	1.2	2.3	6.2	6.2	6.0	6.0	15.2	17.3	19.0	33.8	1.3	2.8
D7	20.0	20.0	17.7	17.7	5.0	5.0	13.2	13.2	16.8	8.8	19.7	14.7	16.8	8.8	3.0	2.5	2.5	2.5	1.2	1.2	13.2	13.2	7.3	4.8	16.8	8.8
D8	8.5	8.5	17.0	18.3	15.2	15.2	38.0	38.0	1.2	2.8	30.7	20.5	1.2	2.8	15.3	9.8	30.2	30.2	78.0	78.0	21.7	8.8	37.5	5.5	1.2	2.8
D9	22.0	22.0	2.2	1.0	7.8	7.8	4.3	4.3	82.0	140.3	2.7	1.2	82.0	140.3	17.5	78.3	312.0	312.0	888.2	888.2	12.7	8.3	41.7	7.8	22.0	14.3
D10	3.0	3.0	4.7	4.7	6.8	6.8	2.7	2.7	3.0	3.5	1.0	1.0	3.0	3.5	448.3	372.0	640.3	640.3	1830.8	1830.8	8.8	15.8	29.8	28.0	3.0	3.5
D11	3.0	3.0	12.0	3.5	13.8	13.8	3.3	3.3	2.5	2.2	1.0	1.0	2.5	2.2	465.2	1.2	48.3	48.3	40.5	40.5	8.2	37.0	60.8	7.3	2.5	2.2
D12	4.7	4.7	35.8	35.0	14.8	14.8	3.2	3.2	4.8	4.8	1.0	1.0	4.8	4.8	32.0	20.8	321.7	321.7	137.3	137.3	17.5	6.7	194.8	38.3	4.8	4.8
D13	6.2	6.2	16.7	14.0	4.5	4.5	9.3	9.3	3.0	3.3	8.0	9.5	3.0	3.3	3.8	3.5	2.5	2.5	1.0	1.0	7.5	6.8	7.2	7.5	3.0	3.3
D14	5.8	5.8	10.5	9.2	19.7	19.7	4.7	4.7	8.7	9.7	80.8	97.3	8.7	9.7	214.0	279.0	336.2	336.2	180.0	180.0	56.0	36.5	187.8	196.0	8.7	9.7
D15	10.5	10.5	4.2	10.5	7.3	7.3	8.0	8.0	1.5	7.3	5.5	10.2	1.5	7.3	1.5	1.3	2.5	2.5	1.7	1.7	9.0	10.5	10.7	13.8	1.5	7.3
D16	6.5	6.5	5.5	4.7	1.8	8.8	21.5	21.5	6.8	7.2	1.0	1.0	6.8	7.2	150.8	231.8	1430.8	1430.8	4492.8	4492.8	32.2	23.5	48.0	48.5	6.8	7.2
D17	17.0	17.0	30.7	36.0	1.0	5.0	13.2	13.2	3.3	4.0	14.5	20.3	3.3	4.0	2.8	3.0	7.7	7.7	18.5	18.5	27.5	8.8	20.8	37.5	3.3	4.0
D18	5.0	5.0	6.0	5.2	3.3	6.3	3.7	3.7	7.7	7.5	1.8	1.8	7.7	7.5	275.3	301.3	351.5	351.5	126.2	126.2	30.5	43.2	115.8	107.3	7.7	7.5
Average	9.24	9.24	11.28	14.73	7.44	8.21	9.74	9.74	9.81	13.62	12.13	15.03	9.81	13.62	91.92	73.94	194.85	194.85	434.37	434.37	17.26	16.03	46.30	32.40	6.48	6.62

D. Parameter Analysis

This section investigates the sensitivity of Ze-HFS to parameters α, β by observing the variation of classification performance under different parameters. When dealing with heterogeneous data, the α, β are set from 0 to 0.25 in steps of 0.025. The β is set to 0 when the data is numerical, and the α is set to 0 when the data is categorical.

Fig. 6 describes the selected feature number and classification accuracy with parameter variation on D3, D5, D9, and D13 datasets. From this figure, the selected feature number increases with parameters increase, and the accuracy becomes higher and then decreases after reaching the highest. In addition, the classification results of heterogeneous datasets D2 and D15 are shown in Fig. 7, where the first and last two subfigures describe the feature number and accuracy with parameters, respectively. From this figure, the selected features and accuracy fluctuate violently with different parameters, and the optimal parameters differ on various datasets.

From the above analysis, the parameter directly influences the neighborhood modeling and feature selection process. Therefore, the neighborhood parameters can be adjusted to



(a) D3 (b) D5

(c) D9 (d) D13

Fig. 6. The classification performance of Ze-HFS under different parameters.

TABLE XI
AVERAGE CLASSIFICATION ACCURACY OF THIRTEEN COMPARED METHODS UNDER DIFFERENT NOISE LEVELS WITH NB CLASSIFIER(%)

No.s	FGR	FRMR	ROFS	RLR	MI	FCE	LS	WSFE	IFS	ENCA	PCA	SFSE	Our
D1	64.88±0.08	66.60±0.09	63.52±0.09	61.17±0.12	52.22±0.07	70.00±0.10	49.26±0.09	65.12±0.10	57.65±0.12	60.74±0.12	71.06±0.09	53.77±0.09	**71.42**±0.09
D2	83.75±0.04	85.34±0.02	85.15±0.02	81.95±0.04	82.98±0.02	85.74±0.02	81.98±0.04	81.40±0.07	75.31±0.11	85.48±0.02	84.68±0.02	83.49±0.05	**85.88**±0.02
D3	63.48±0.03	62.42±0.04	60.81±0.03	57.54±0.04	41.34±0.02	63.93±0.03	21.77±0.04	63.52±0.03	16.95±0.02	36.72±0.03	61.13±0.03	15.21±0.02	**63.82**±0.03
D4	65.90±0.06	68.82±0.06	69.45±0.03	68.10±0.06	70.00±0.03	70.98±0.04	67.00±0.09	70.00±0.04	66.32±0.08	69.77±0.04	70.87±0.04	69.80±0.05	**71.52**±0.04
D5	24.27±0.08	21.53±0.08	13.20±0.07	23.67±0.07	15.60±0.07	26.20±0.08	11.20±0.05	27.80±0.10	21.73±0.08	25.47±0.10	**33.8**±0.09	10.87±0.06	24.00±0.09
D6	72.24±0.08	79.41±0.09	74.41±0.09	77.73±0.07	79.39±0.04	79.47±0.08	79.43±0.06	79.41±0.08	72.95±0.09	74.63±0.07	79.27±0.08	78.91±0.09	**79.81**±0.09
D7	92.31±0.01	92.58±0.01	92.61±0.01	92.11±0.01	91.47±0.01	92.60±0.01	92.58±0.01	93.34±0.01	92.58±0.01	92.59±0.01	92.56±0.01	79.55±0.05	**93.46**±0.01
D8	47.96±0.08	53.70±0.06	37.56±0.09	39.22±0.09	54.22±0.07	55.22±0.07	46.44±0.08	53.26±0.07	13.48±0.06	26.00±0.09	53.89±0.08	41.19±0.08	54.78±0.08
D9	63.45±0.20	66.59±0.15	59.52±0.19	60.28±0.20	64.52±0.27	67.14±0.16	52.38±0.13	72.42±0.15	63.65±0.19	64.01±0.22	63.25±0.20	64.68±0.22	**74.88**±0.18
D10	50.67±0.23	44.67±0.24	51.33±0.22	27.67±0.16	48.00±0.22	37.33±0.21	22.00±0.18	57.00±0.23	21.00±0.18	26.00±0.22	56.33±0.20	37.33±0.23	**60.33**±0.24
D11	79.26±0.17	75.98±0.15	73.01±0.16	79.29±0.14	60.89±0.21	65.77±0.17	64.11±0.17	79.61±0.16	67.74±0.16	71.40±0.17	84.58±0.15	68.18±0.15	**89.49**±0.11
D12	73.56±0.08	72.13±0.08	74.87±0.09	65.06±0.11	62.10±0.08	68.73±0.10	31.76±0.19	77.92±0.09	64.62±0.13	75.23±0.11	**74.90**±0.08	71.10±0.10	**78.75**±0.08
D13	65.24±0.12	63.27±0.14	58.88±0.11	56.67±0.15	49.05±0.19	67.79±0.11	49.29±0.13	61.25±0.12	55.21±0.16	60.98±0.11	63.06±0.09	53.36±0.12	**70.13**±0.11
D14	41.98±0.10	35.16±0.09	42.78±0.08	30.63±0.08	27.62±0.10	57.62±0.09	22.38±0.07	53.25±0.10	**55.00**±0.13	62.22±0.10	44.76±0.10	41.11±0.09	44.92±0.10
D15	67.65±0.07	70.13±0.05	68.05±0.06	66.97±0.06	46.90±0.15	70.85±0.04	52.70±0.18	69.98±0.04	67.57±0.07	68.18±0.07	70.72±0.05	69.72±0.05	**71.13**±0.04
D16	11.39±0.13	12.22±0.14	11.39±0.14	14.17±0.13	13.33±0.13	7.22±0.10	8.33±0.12	15.00±0.15	16.94±0.16	14.72±0.14	**32.22**±0.16	22.22±0.16	21.67±0.15
D17	**71.02**±0.05	68.20±0.04	56.35±0.05	58.18±0.07	69.30±0.05	71.01±0.04	57.6±0.06	68.53±0.04	54.23±0.06	68.75±0.04	70.17±0.05	63.92±0.04	**71.02**±0.04
D18	30.64±0.12	20.64±0.14	17.05±0.09	19.62±0.12	19.23±0.09	15.38±0.09	7.69±0.07	**50.90**±0.14	11.67±0.09	12.95±0.08	48.90±0.14	20.90±0.11	33.59±0.14
Ave±St	59.42±0.09	58.86±0.09	56.11±0.09	54.45±0.09	52.68±0.10	59.61±0.09	45.44±0.1	63.32±0.10	49.70±0.10	55.32±0.10	**64.23**±0.09	52.52±0.10	**64.48**±0.09
Rank	6.94	6.22	9.11	8.06	4.50	8.94	10.94	4.17	7.06	10.00	4.33	8.50	**1.83**

TABLE XII
AVERAGE CLASSIFICATION ACCURACY OF THIRTEEN COMPARED METHODS UNDER DIFFERENT NOISE LEVELS WITH FNN CLASSIFIER(%)

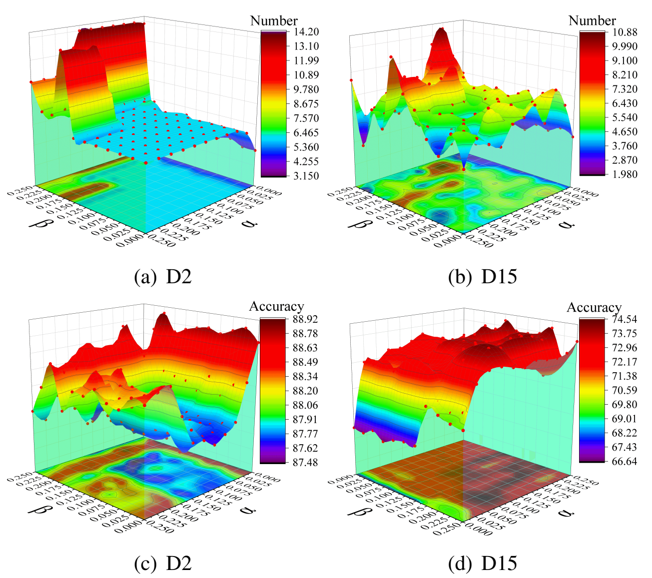
No.s	FGR	FRMR	ROFS	RLR	MI	FCE	LS	WSFE	IFS	ENCA	PCA	SFSE	Our
D1	68.09±0.08	70.12±0.07	67.59±0.08	67.72±0.09	61.48±0.09	71.85±0.09	62.22±0.10	**74.20**±0.07	61.42±0.09	68.83±0.08	**71.93**±0.08	56.36±0.08	**74.89**±0.08
D2	85.94±0.02	85.57±0.02	85.84±0.02	85.82±0.02	85.01±0.01	86.00±0.02	84.69±0.01	85.80±0.02	84.63±0.02	85.51±0.02	80.55±0.02	84.69±0.02	**86.51**±0.02
D3	64.35±0.03	65.17±0.04	62.40±0.03	61.09±0.03	42.99±0.04	66.45±0.03	37.36±0.02	62.45±0.04	16.76±0.03	37.97±0.04	63.53±0.04	14.87±0.02	**67.65**±0.03
D4	7.02±0.04	6.72±0.04	1.07±0.01	5.72±0.04	5.90±0.05	8.92±0.04	3.6±0.03	3.37±0.02	5.03±0.03	2.62±0.02	7.45±0.05	0.75±0.01	**9.40**±0.05
D5	41.20±0.09	32.73±0.09	19.80±0.08	40.80±0.08	16.80±0.10	30.47±0.10	24.00±0.11	47.20±0.10	30.27±0.09	45.87±0.09	47.53±0.10	18.47±0.08	**50.67**±0.10
D6	4.56±0.06	7.56±0.08	3.54±0.05	3.13±0.04	2.24±0.05	13.45±0.11	3.79±0.06	2.12±0.02	6.36±0.07	3.07±0.04	6.81±0.09	4.13±0.04	6.36±0.07
D7	93.47±0.01	93.3±0.01	94.25±0.01	93.57±0.01	92.58±0.01	94.18±0.01	92.58±0.01	94.40±0.01	92.58±0.01	93.18±0.01	93.62±0.01	92.58±0.01	**94.32**±0.01
D8	52.37±0.07	53.56±0.07	53.89±0.08	50.81±0.06	54.00±0.06	54.63±0.07	54.00±0.11	55.63±0.08	51.67±0.06	51.11±0.08	55.93±0.07	54.11±0.06	**57.48**±0.08
D9	22.94±0.19	32.78±0.23	31.19±0.19	29.72±0.19	22.86±0.14	35.12±0.20	33.33±0.24	33.45±0.22	22.94±0.19	25.91±0.20	34.44±0.19	34.37±0.20	**36.09**±0.20
D10	45.00±0.22	48.00±0.22	43.00±0.20	41.67±0.20	32.00±0.23	43.67±0.24	18.00±0.11	64.00±0.26	44.00±0.21	41.67±0.22	56.00±0.22	38.00±0.21	**61.33**±0.22
D11	8.57±0.10	25.45±0.18	13.66±0.13	6.82±0.11	33.39±0.19	33.15±0.17	22.50±0.17	20.54±0.18	14.02±0.15	10.30±0.12	19.79±0.18	**34.70**±0.16	12.38±0.13
D12	69.12±0.09	72.71±0.10	71.83±0.09	68.48±0.08	67.07±0.08	68.16±0.07	68.07±0.10	75.98±0.10	72.77±0.09	75.92±0.10	**77.84**±0.08	69.78±0.11	76.97±0.08
D13	61.42±0.16	64.65±0.10	63.87±0.13	63.02±0.13	53.33±0.09	70.12±0.11	50.67±0.13	68.54±0.12	59.17±0.14	64.07±0.12	73.99±0.11	55.57±0.13	**70.25**±0.12
D14	36.90±0.11	31.83±0.12	41.11±0.10	27.38±0.10	25.71±0.10	63.02±0.09	13.81±0.05	32.29±0.11	**67.86**±0.09	69.60±0.08	40.62±0.10	16.75±0.08	39.84±0.11
D15	69.03±0.04	70.35±0.04	69.77±0.05	68.43±0.05	68.80±0.04	71.05±0.04	68.60±0.05	69.92±0.05	69.33±0.05	69.48±0.04	71.35±0.04	70.13±0.04	**71.85**±0.04
D16	19.17±0.17	16.67±0.15	13.06±0.14	32.50±0.17	15.00±0.12	19.72±0.16	6.67±0.09	19.44±0.16	24.44±0.18	20.00±0.16	35.28±0.20	18.61±0.13	**35.28**±0.19
D17	70.17±0.04	70.03±0.06	67.22±0.05	67.65±0.06	66.00±0.05	74.28±0.04	61.90±0.06	72.25±0.04	63.33±0.05	69.15±0.04	73.75±0.05	63.90±0.05	**74.33**±0.04
D18	22.56±0.12	19.49±0.11	18.33±0.09	20.77±0.09	14.62±0.08	16.79±0.10	5.38±0.06	36.28±0.13	46.79±0.12	53.72±0.13	41.67±0.14	13.21±0.08	31.41±0.12
Ave±St	46.77±0.09	48.15±0.10	45.63±0.09	46.39±0.09	42.21±0.09	51.17±0.09	39.51±0.08	50.99±0.10	46.30±0.09	49.33±0.09	52.89±0.10	41.17±0.08	**53.17**±0.09
Rank	7.11	6.06	8.50	7.94	4.28	10.11	10.39	5.11	7.39	8.22	3.67	9.28	**2.39**

select an optimal feature subset.

E. Robustness Evaluation of Ze-HFS Method

This subsection evaluates the robustness of Ze-HFS in noise environments. For each dataset, 10%, 20%, 30%, 40%, 50%, and 60% samples are randomly added with Gaussian noise. Then, the selected features and accuracy of compared methods are obtained at different noise levels. To streamline presentation, NB and FNN classifiers are adopted for evaluation.

The average number of selected features for each method under six noise levels is represented in Table X. From this table, Ze-HFS has the minimum average feature number 3 and 2 times, and also has the minimum average values 6.48 and 6.62 on NB and FNN classifiers, respectively. Furthermore, the average classification accuracy of thirteen methods under six noise levels is shown in Tables XI-XII. Compared with other methods, Ze-HFS outperforms other methods on 13 datasets for both employed classifiers. Moreover, its average accuracy across 18 datasets surpasses that of its counterparts, indicating the superior robustness of Ze-HFS in noisy environments.



(a) D2 (b) D15

(c) D2 (d) D15

Fig. 7. The classification performance of Ze-HFS under different parameters on D2 and D15 dataset. (a)-(b) and (c)-(d) show the average feature number and classification accuracy with parameter varies.

VI. Conclusions

This paper proposes a novel feature selection method for heterogeneous data combining the granular level structure. First, a heterogeneous distance metric is defined to construct the neighborhood granules and approximation space, which is adequate for handling high-dimensional data by considering the differences among multiple features. Then, the granular level structure in the approximation process is analyzed, and a novel zentropy-based uncertainty measure is proposed to describe uncertain knowledge. Finally, two importance measures based on zentropy are designed to evaluate features, and a feature selection algorithm is further designed. Extensive experiments on public datasets illustrate that the proposed method is robust with parameter changes and good in classification performance. Notably, the computational complexity of zentropy-based metrics within heterogeneous datasets is high owing to the complex computation of information across various levels, especially in dynamic environments. Consequently, a more efficient computational framework combining the approximation process needs further investigation. Meanwhile, parameters α and β are two important parameters that significantly influence the algorithm's complexity and performance. An adaptive mechanism for neighborhood models combining data characteristics plays an important role in improving algorithm efficiency. Furthermore, the exploration of zentropy thought for uncertain information processing in diverse application domains also deserves to be explored.

Acknowledgments

This paper was supported in part by the National Key Research and Development Program of China "Key Special Project on Cyberspace Security Governance" under Grant 2022YFB3104702, in part by the National Natural Science Foundation of China under Grant 62376198.

References

[1] M. Wang, W.J. Fu, X.N. He, et al., "A Survey on Large-scale Machine Learning," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, pp. 2574-2594, 2020.

[2] H.S. Yan, J.F. Zhang, N. Gang, et al., "CIFS: Improving Adversarial Robustness of CNNs via Channel-wise Importance-based Feature Selection," in *Processing of 21th International Conference on Machine Learning (ICML)*, pp. 11693-11703, 2021.

[3] Y. Xue, X. Han, Z. Wang, "Self-Adaptive Weight based on Dual-Attention for Differentiable Neural Architecture Search," in *IEEE Transactions on Industrial Informatics*, vol. 20, no. 4, pp. 6394-6403, 2024.

[4] K.H. Yuan, D.Q. Miao, Y.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.

[5] X.S. Zhao, K.P. Liu, W. Fan, et al., "Simplifying Reinforced Feature Selection via Restructured Choice Strategy of Single Agent," in *Processing of 2020 IEEE International Conference on Data Mining (ICDM)*, pp. 871-880, 2020.

[6] Y. Zhou, J.H. Zhang, S. Kwong, et al., "An Evolutionary Multi-objective Optimization Framework of Discretization-based Feature Selection for Classification," in *Swarm and Evolutionary Computation*, vol. 60, pp. 100770, 2021.

[7] Y. Xue, C. Lu, F. Neri, et al., "Improved Differentiable Architecture Search With Multi-Stage Progressive Partial Channel Connections," in *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 8, no. 1, pp. 32-43, 2024.

[8] D. Cai, C.Y. Zhang, X.F. He, "Unsupervised Feature Selection for Multi-Cluster Data," in *Processing of 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* pp. 333-342, 2010.

[9] Y.J. Lin, H.Y. Liu, H. Zhao, et al., "Hierarchical Feature Selection based on Label Distribution Learning," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 6, pp. 5964-5976, 2023.

[10] J.H. Wan, H.M. Chen, T.R. Li, et al., "Interactive and Complementary Feature Selection via Fuzzy Multigranularity Uncertainty Measures," in *IEEE Transactions on Cybernetics*, vol. 53, no. 2, pp. 1208-1221, 2023.

[11] Y. Xue, C. Zhang, F. Neri, et al., "An Externa Attention-oased Feature Ranker Forlarge-scale Featureselection," in *Knowledge-Based Systems*, vol.281, pp. 111084, 2023.

[12] L. Chen, L. Lu, K.R. Feng, et al., "Multiple Classifier Integration for the Prediction of Protein Structural Classes," in *Journal of Computational Chemistry*, vol. 30, pp. 2248-2254, 2009.

[13] J.T. Yao, A.V. Vasilakos, W. Pedrycz, "Granular Computing: Perspectives and Challenges," in *IEEE Transactions on Cybernetics*, vol. 43, no. 6, pp. 1977-1989, 2013.

[14] J. Yang, G.Y. Wang, Q.H. Zhang, et al., "Optimal Granularity Selection based on Cost-sensitive Sequential Three-way Decisions with Rough Fuzzy Sets," in *Knowledge-Based Systems*, vol. 163, pp. 131-144, 2019.

[15] D.D. Guo, C.M. Jiang, R.X. Sheng, et al., "A Novel Outcome Evaluation Model of Three-Way Decision: A Change Viewpoint," in *Information Sciences*, vol. 607, pp. 1089-1110, 2022.

[16] H. Bustince, E. Barrenechea, M. Pagola, et al., "A Historical Account of Types of Fuzzy Sets and Their Relationships," in *IEEE Transactions on Fuzzy Systems*, vol. 24, pp. 179-194, 2016.

[17] X.W. Wei, B. Pang, J.S. Mi, "Axiomatic Characterizations of L-valued Rough Sets using a Single Axiom," in *Information Sciences*, vol. 283-310, pp. 283-310, 2021.

[18] X.D. Yue, J. Zhou, Y.Y. Yao, et al., "Shadowed Neighborhoods based on Fuzzy Rough Transformation for Three-Way Classification," in *IEEE Transactions on Fuzzy Systems*, vol. 28, pp. 978-991, 2020.

[19] W.H. Xu, D.D. Guo, J. Mi, et al., "Two-way Concept-cognitive Learning via Concept Movement Viewpoint," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 10, pp. 6798-6812, 2023.

[20] Y.Y. Yao, "Three-way Granular Computing, Rough Sets, and Formal Concept Analysis," in *International Journal of Approximate Reasoning*, vol. 116, pp. 106-125, 2020.

[21] D.D. Guo, W.H. Xu, Y.H. Qian, et al., "Fuzzy-granular Concept-cognitive Learning via Three-way Decision: Performance Evaluation on Dynamic Knowledge Discovery," in *IEEE Transactions on Fuzzy Systems*, vol. 32, no. 3, pp. 1409-1423, 2024.

[22] W.H. Xu, D.D. Guo, Y.H. Qian, et al., "Two-way Concept-cognitive Learning Method: a Fuzzy-based Progressive Learning," in *IEEE Transactions on Fuzzy Systems*, vol. 31, no. 6, pp. 1885-1899, 2023.

[23] D.D. Guo, W.H. Xu, W.P. Ding, et al., "Concept-cognitive Learning Survey: Mining and Fusing Knowledge from Data," in *Information Fusion*, vol. 109, pp. 102426, 2024.

[24] D.D. Guo, W.H. Xu, Y.H. 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, 2024.

[25] Z. Pawlak, "Rough Sets," in *International Journal of Computer and Information Sciences*, vol. 11, pp. 341-356, 1982.

[26] W.H. Xu, K.H. Yuan, W.T. Li, "Dynamic Updating Approximations of Local Generalized Multigranulation Neighborhood Rough Set," in *Applied Intelligence,* vol. 52, no. 8, pp. 9148-9173, 2022.

[27] K.H. Yuan, W.H. Xu, D.Q. Miao, "A Local Rough Set Method for Feature Selection by Variable Precision Composite Measure," in *Applied Soft Computing*, vol. 155, pp. 111450, 2024.

[28] Q.H. Hu, D.R. Yu, J.F. Liu, et al., "Neighborhood Rough Set based Heterogeneous Feature Subset Selection," in *Information Sciences*, vol. 178, no. 18, pp. 3577-3594, 2008.

[29] H. Wang, "Nearest Neighbors by Neighborhood Counting," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 6, pp. 942-953, 2006.

[30] P.F. Zhang, T.R. 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.

[31] Y. Xue, J. Qin, "Partial Connection based on Channel Attention for Differentiable Neural Architecture Search," in *IEEE Transactions on Industrial Informatics*, vol. 19, no. 5, pp. 6804-6813, 2023.

[32] W.T. Li, H.X. Zhou, W.H. Xu, et al., "Interval Dominance-based Feature Selection for Interval-Valued Ordered Data," in *IEEE Transactions on*

Neural Networks and Learning Systems, vol. 34, no. 10, pp. 6898-6912, 2023.

[33] C.Z. Wang, M.W. Shao, Q. He, et al., "Feature Subset Selection based on Fuzzy Neighborhood Rough Sets," in *Knowledge-Based Systems*, vol. 111, pp. 173-179, 2016.

[34] Q. Wang, Y.H. Qian, X.Y. Liang, et al., "Local Neighborhood Rough Set, " in *Knowledge-Based Systems*, vol. 153, pp. 53-64, 2018.

[35] C.Z. Wang, Y.H. Qian, W.P. Ding, et al., "Feature Selection with Fuzzy-Rough Minimum Classification Error Criterion," in *IEEE Transactions on Fuzzy systems*, vol. 30, no. 8, pp. 2930-2942, 2022.

[36] X. Zhang, C.L. Mei, D.G. Chen, et al., "Active Incremental Feature Selection using a Fuzzy-Rough-Set-based Information Entropy," in *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 5, pp. 901-915, 2020.

[37] B.B. Sang, H.M. Chen, L. Yang, et al., "Incremental Feature Selection using a Conditional Entropy based on Fuzzy Dominance Neighborhood Rough Sets," in *IEEE Transactions on Fuzzy Systems*, vol. 30, no. 6, pp. 1683-1697, 2022.

[38] W.H. Xu, K.H. Yuan, W.T. 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, pp. 76-88, 2023.

[39] Z.K. Liu, Y. Wang, and S.L. Shang, "Zentropy Theory for Positive and Negative Thermal Expansion," in *Journal of Phase Equilibria and Diffusion*, vol. 43, pp. 598-605, 2022.

[40] S.L. Shang, Y. Wang, Z.K. Liu, "Quantifying the Degree of Disorder and Associated Phenomena in Materials Through Zentropy: Illustrated with Invar Fe_3Pt," in *Scripta Materialia*, vol. 225, pp. 115164, 2023.

[41] 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.

[42] Q.H. Hu, D.R. Yu, W. Pedrycz, et al., "Kernelized Fuzzy Rough Sets and Their Applications," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 11, pp. 1649-1667, 2011.

[43] P. Zhou, P.P. Li, S. Zhao, et al., "Online Early Terminated Streaming Feature Selection based on Rough Set Theory," in *Applied Soft Computing*, vol. 113, pp. 107993, 2021.

[44] S.Y. Xia, X.Y. Bai, G.Y. Wang, "An Efficient and Accurate Rough Set for Feature Selection, Classification, and Knowledge Representation," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 8, pp. 7724-7735, 2023.

[45] G. Roffo, S. Melzi, U. Castellani, et al., "Infinite Latent Feature Selection: A Probabilistic Latent Graph-based Ranking Approach." in *Processing of the IEEE International Conference on Computer Vision*, pp. 1398-1406, 2017.

[46] X. He, D. Cai and P. Niyogi. "Laplacian Score for Feature Selection," in *Proceeding of International Conference on Neural Information Processing Systems*, pp. 507-514, 2005.

[47] H. Saadatmand, M.R. Akbarzadeh, "Set-based Integer-coded Fuzzy Granular Evolutionary Algorithms for High-dimensional Feature Selection," in *Applied Soft Computing*, vol. 142, pp. 110240, 2023.

[48] E. Hancer, B. Xue, M.J. Zhang, "An Evolutionary Filter Approach to Feature Selection in Classification for Both Single-and Multi-objective Scenarios," in *Knowledge-Based Systems*, vol. 280, pp. 111008, 2023.

[49] A. Malhi, R. X. Gao, "PCA-based Feature Selection Scheme for Machine Defect Classification," in *IEEE Transactions on Instrumentation and Measurement*, vol. 53, no. 6, pp. 1517-1525, 2004.

[50] J.Y. Wang, H.M. Wang, F.P. Nie, et al., "Sparse Feature Selection via Fast Embedding Spectral Analysis," in *Pattern Recognition*, vol. 139, pp. 109472, 2023.

[51] J. Demsar, "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 College of Electronics and Information Engineering, Tongji University, Shanghai, China. Her current research interests include granular computing, uncertainty analysis, and machine learning.

Duoqian Miao is a Professor with the School of Electronics and Information Engineering at Tongji University, Shanghai, China. He has published more than 180 academic papers, which include prestigious journals and international conferences, and the hindex is 39. 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.

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.

Weiping Ding (M'16-SM'19) received the Ph.D. degree in Computer Science, Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2013. He is a Full Professor with the School of Information Science and Technology, Nantong University, Nantong, China, and also the supervisor of Ph.D postgraduate by the Faculty of Data Science at City University of Macau, China. His main research directions involve deep neural networks, multimodal machine learning, and medical images analysis. He has published over 250 articles, including over 110 IEEE Transactions papers. He serves as an Associate Editor/Editorial Board member of IEEE Transactions on Neural Networks and Learning Systems, IEEE Transactions on Fuzzy Systems, IEEE/CAA Journal of Automatica Sinica, IEEE Transactions on Intelligent Transportation Systems, IEEE Transactions on Intelligent Vehicles, IEEE Transactions on Emerging Topics in Computational Intelligence, IEEE Transactions on Artificial Intelligence, Information Fusion, Information Sciences, Neurocomputing, Applied Soft Computing. He is the Leading Guest Editor of Special Issues in several prestigious journals, including IEEE Transactions on Evolutionary Computation, IEEE Transactions on Fuzzy Systems, and Information Fusion. He is the Co-Editor-in-Chief of both Journal of Artificial Intelligence and Systems and Journal of Artificial Intelligence Advances.

Hongyun Zhang received the Ph.D. degree in pattern recognition and intelligence system from Tongji University, Shanghai, China, in 2005. She is currently an Associate Professor at Tongji University. She is the author or coauthor of nearly 60 journal papers and conference proceedings in principal curves, pattern recognition, machine learning, granular computing, and rough set. Her current research interests include principal curves, pattern recognition, data mining, image retrieval, and granular computing.