Feature Selection Using Zentropy-Based Uncertainty Measure

Kehua Yuan, Duoqian Miao, Yiyu Yao, Hongyun Zhang, and Xuerong Zhao

***Abstract*—Feature selection and entropy theory are two efficacious data analysis tools for investigating uncertainty information processing in artificial intelligence. The fruitful marriage of the two has been an active research topic in knowledge discovery. Currently, most feature selection methods via entropy theory mainly focus on the information measures at a single granular level. However, it ignores the interaction between granular levels, which leads to the poor stability and accuracy of related methods. Hence, this article proposes a novel zentropy-based uncertainty measure to design a feature selection method by exploiting the granular level structure in knowledge space. Subsequently, by analyzing the granular level structure in decision data, the zentropy-based uncertainty measure and its properties are designed and analyzed to depict the uncertainty knowledge from whole and internal. Moreover, two importance measures are defined to evaluate features based on the designed uncertainty measure, and then a corresponding feature selection algorithm is developed. Finally, some experiments are carried out on public datasets to demonstrate that the proposed method can achieve state-of-the-art performance among methods, especially regarding stability and classification accuracy.**

***Index Terms*—Decision information system, feature selection, granular computing, rough set (RS), uncertainty measure.**

I. Introduction

FEATURE selection, a crucial issue of knowledge discovery in information systems, aims to select the key features (remove the irrelevant and redundant features with the target) from the original features to enhance the performance of task-specific learning. Undoubtedly, feature selection has become a hot topic in many fields, such as machine learning [14], pattern recognition [20], data mining [18], and approximate reasoning [40].

In recent years, big data have added new vigor and vitality to the development of artificial intelligence, in which data are a crucial resource. The processing of uncertainty for big data is related to the credibility and interpretability of artificial intelligence [7], [11], [22]. An influential theory to this processing is granular computing, which plays an important role in studying uncertain phenomena when humans use concepts, symbols, and models to describe the objective world [32]. From the perspective of granular computing, to characterize an uncertainty concept, several theoretical tools were applied to process this phenomenon from its representation, measurement, and reasoning methods, such as the three decision regions in three-way decision theory [5], [33], the boundary region in rough set (RS) [35], [37], and the pseudoconcept in concept cognitive learning [6], [29], [30]. Up to now, numerous uncertainty measures of granular computing have been proposed to meet different requirements of uncertainty processing [4], [8], [36]. In these theories, it is worth stressing that the fruitful marriage of entropy theory and granular computing has become a practical method for exploring uncertainty in artificial intelligence [15]. In particular, the RS theory provides a mathematical set representation of uncertainty [19]. In this theory, the vague concepts are described by two precision approximate sets, which provide an "accurate" method for dealing with imprecision knowledge. At present, the RS model has been widely used in knowledge acquisition [18], decision-making [9], [33], and artificial intelligence [20].

Note that the classical Pawlak RS can only handle discrete data and the discretization of data could lead to the lack of helpful information, thus, it is also extended to solve complex situations, including fuzzy data [24], [39], interval data [12], and ordered data [21]. In particular, neighborhood rough set (NRS) as a crucial RS model is usually adopted to characterize the uncertainty knowledge in real-valued data [26], [27]. In the NRS model, the vague concept is approximately characterized based on neighborhood similarity classes, which could avoid noise interference to a certain extent and improve the tolerance of differences between objects in continuous data [16], [17]. This neighborhood thought has been widely adopted to extend the RS model and successfully applied to concept approximation and feature selection. This is one reason for studying the uncertainty measure of real-valued data via the NRS in this article.

Meanwhile, inspired by entropy-based uncertainty measures, some researchers have successfully applied these methods to knowledge reduction and rule extraction in decision information system [23], [31]. Liang [13] investigated the information entropy (IE) and rough entropy from knowledge granulation. Dai et al. [1], [2] studied two kinds of conditional IE for an incomplete decision system. Wang et al. [25] proposed four kinds of uncertainty measures to describe the uncertainty in NRS based on self-information. Gao et al. [3] introduced a monotonic uncertainty measure to describe the knowledge and applied it to attribute reduction in the decision-theoretic RS model. Moreover, considering the distribution structure of the

Manuscript received 31 March 2023; revised 12 September 2023 and 21 November 2023; accepted 20 December 2023. Date of publication 28 December 2023; date of current version 2 April 2024. This work was supported in part by the National Key Research and Development Program of China "Key Special Project on Cyberspace Security Governance" under Grant 2022YFB3104700 and in part by the National Natural Science Foundation of China under Grant 61976158, Grant 62376198, Grant 62076182, and Grant 62006172. Recommended by Associate Editor C. Mencar. *(Corresponding author: Duoqian Miao.)*

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

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

Digital Object Identifier 10.1109/TFUZZ.2023.3347757

decision, Xu et al. [31] defined a composite entropy combined with data distribution and designed the corresponding feature selection algorithm. Nevertheless, these methods mainly focus on handling uncertainty on a single granular level so that the interaction between granular leavers is ignored, leading to the poor stability and accuracy of the related method. Hence, another motivation in this article is to point out this issue and solve it by analyzing the granular level structure.

IE is an important measure to describe the uncertainty within a probability distribution. Nevertheless, the existing entropy-based measures mostly focus on the information at a single level while ignoring the granular level structure in information systems, which is an incomplete description and could lead to the poor performance of learning models. Zentropy is an emerging entropy to comprehensively characterize the chaos of system confusion from multiple levels [15]. In particular, zentropy shows an excellent performance in predicting the physical properties of the phase, such as the variation of volume with temperature. In material properties, phase refers to different states of substances, such as solids, liquids, and gases, each having specific physical properties that can be reflected as melting point, heat capacity, density, and bulk modulus. By comprehensively considering the relationship between different states, one can better understand and describe their behavior and characteristics. In the zentropy theory, system entropy is the overall reflection of each internal entropy on different scales related to the cognitive mechanism of granular computing. This article applies the naive idea of zentropy to the uncertainty measure method in information system. In this method, this article draws systematic thought to explore the expression of uncertainty to achieve a comprehensive and stable cognition of uncertainty between different granular levels. Therefore, the third motivation of this article is how to systematically integrate this idea to measure the uncertainty of granular structure.

Inspired by the abovementioned issues, this article proposes a zentropy-based uncertainty measure to characterize the granular level structure from particle to whole and explores its application. The main contributions of this article are as follows.

1) It offers a new thought for uncertainty measures based on entropy theory, focusing on the uncertainty at different granular levels to design feature evaluation functions. The core idea of this thought is no longer to focus on a single granular level but to present a more comprehensive analysis for an understanding of uncertainty.
2) It proposes a novel zentropy-based uncertainty measure by analyzing the relationship between the defined granular levels. Compared with other existing entropy-based measures, it has a more comprehensive and robust ability to depict uncertainty in the decision information system.
3) It applies the proposed zentropy-based measure to feature selection for selecting valuable features. The experiments on public datasets demonstrate that it can achieve state-of-the-art performance compared with others, especially in terms of stability and classification accuracy.

The rest of this article is organized as follows. Section II briefly reviews the related works and introduces the motivation for the study. Section III investigates the construction of the zentropy-based uncertainty measure. Moreover, Section IV represents the designed feature selection algorithm, and Section V analyzes the experimental results on public datasets. Finally, Section VI concludes this article.

II. Several Measures of Uncertainty

This section briefly reviews some basic notions about the NRS, uncertainty measures, and IE to facilitate the subsequent discussions.

A. NRS in Decision Information System

Let $\text{DIS} = (U, C \cup D, V, f)$ be a decision information system, where $U = \{x_1, x_2, \ldots, x_n\}$ is the universe, C and D are, respectively, the conditional attribute set and decision attribute set, $U/D = \{D_1, D_2, \ldots, D_s\}$, $V = \cup_{a \in C \cup D} V_a$, and $f : U \times C \cup D \to V$. Given a parameter δ, the neighborhood binary relation with $B \subseteq C$ is defined as follows:

$$\mathcal{R}_B(x,y) = \begin{cases} 0, & d(x,y) > \delta \\ 1, & d(x,y) \leq \delta \end{cases} \tag{1}$$

where $d(x,y)$ is the Euclidean distance between x and y.

It is easily obtained from (1) that this neighborhood binary relation $\mathcal{R}_B$ has the following properties:

1) reflexivity: for $\forall x \in U$, $\mathcal{R}_B(x,x) = 1$;
2) symmetry: for $\forall x, y \in U$, $\mathcal{R}_B(x,y) = \mathcal{R}_B(y,x)$.

Thus, $\mathcal{R}_B$ is a neighborhood similarity relation, and the neighborhood similarity class of object x is denoted as $\delta_B(x) = \{y \in U | x\mathcal{R}y\}$.

Definition 1 (see [25]): Let $DIS = (U, C \cup D, V, f)$ be a decision information system. Given the parameter δ, for any $X \subseteq U$, the neighborhood lower approximation and upper approximation of X on $B \subseteq C$ are represented as follows:

$$\underline{\mathcal{R}}_B(X) = \{x \in U | \delta_B(x) \subseteq X\}$$
$$\overline{\mathcal{R}}_B(X) = \{x \in U | \delta_B(x) \cap X \neq \emptyset\}. \tag{2}$$

The pair $< \underline{\mathcal{R}}_B(X), \overline{\mathcal{R}}_B(X) >$ is called an NRS. The boundary region of X on B is $\text{Bon}_B(X) = \overline{\mathcal{R}}_B(X) - \underline{\mathcal{R}}_B(X)$. Moreover, the neighborhood lower and upper approximations of D on B are represented as follows:

$$\underline{\mathcal{R}}_B(D) = \{\underline{\mathcal{R}}_B(D_1), \underline{\mathcal{R}}_B(D_2), \ldots, \underline{\mathcal{R}}_B(D_s)\}$$
$$\overline{\mathcal{R}}_B(D) = \{\overline{\mathcal{R}}_B(D_1), \overline{\mathcal{R}}_B(D_2), \ldots, \overline{\mathcal{R}}_B(D_s)\}. \tag{3}$$

The boundary regions of D on B is defined as $\text{Bon}_B(D) = \sum_{q=1}^{s} \overline{\mathcal{R}}_B(D_q) - \sum_{q=1}^{s} \underline{\mathcal{R}}_B(D_q)$.

B. Uncertainty in NRS

Similar to the Pawlak RS, the uncertainty of the NRS is caused by the boundary region that can be described by lower approximation and upper approximation. For convenience, the lower approximation and upper approximation is called approximation level in the approximation process. To depict the uncertainty in NRS, the dependency degree based on lower approximation is proposed.

Definition 2 (see [27]): Let $DIS = (U, C \cup D, V, f)$ be a decision information system. For $B \subseteq C$, the dependency degree of D on B is defined as follows:

$$\gamma_B(D) = \frac{|\sum_{i=1}^{s} \underline{R}_B(D_i)|}{|U|}. \quad (4)$$

The dependency degree defined by the ratio of the cardinalities of the positive region to all objects in U is usually used to characterize the relevancy between conditional attribute subset B and decision attribute set D [27]. Note that this measure mainly focuses on the information presented at the lower approximation while ignoring the elements influencing the approximation process.

C. Uncertainty Measures Based on Entropy

Entropy theory proposed by Shannon [23] gives a measure of uncertainty in the probabilistic domain. Given a random variable, the Shannon entropy can be represented as follows.

Definition 3 (see [23]): Let $W = \{w_1, w_2, \ldots, w_r\}$ be a random variable, where $w_i (i = 1, 2, \ldots, r)$ is its possible value. The probability $p(w_i)$ is the probability of w_i, then the Shannon entropy is defined as follows:

$$E(W) = -\sum_{i=1}^{r} p(w_i) \log p(w_i) \quad (5)$$

where $-logp(w_i)$ is the information content associated with the value w_i having a probability $p(w_i)$.

Currently, many scholars researched Shannon's concept and its variants to measure uncertainty. Some representative entropy-based measures are usually adopted for model evaluation in NRS theory, such as neighborhood entropy and neighborhood conditional entropy (NCE).

Definition 4 (see [10]): Let $DIS = (U, C \cup D, V, f)$ be a decision information system. For $B \subseteq C$ and δ, $\delta_B(x)$ is the neighborhood class of object x under attribute subset B. Then, the neighborhood entropy of B can be defined as follows:

$$NE_B(D) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log \frac{|\delta_B(x_i)|}{|U|}. \quad (6)$$

This neighborhood entropy mainly focuses on the size of neighborhood similarity classes while ignoring the relationship with decisions. Thus, it is usually not adopted in decision tasks. To characterize the uncertainty in the decision information system, the NCE [10] is further proposed.

Definition 5 (see [10]): Let $DIS = (U, C \cup D, V, f)$ be a decision information system. For $B \subseteq C$ and δ, $\delta_B(x)$ is the neighborhood class of object x under attribute subset B. Then, the neighborhood conditional entropy NCE$(D|B)$ of D related to B is defined as follows:

$$NCE(D|B) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log \frac{|\delta_B(x_i) \cap D(x_i)|}{|\delta_B(x_i)|} \quad (7)$$

where $D(x_i)$ is the decision class of x_i.

The abovementioned entropy-based measures mainly concentrate on the information presented at the neighborhood similarity class to construct uncertainty measures in an information system. It is a finer method to analyze and describe uncertainty but ignores the data distribution and the interaction with other information levels. This incomplete characterization could lead to the poor stability and accuracy of related methods.

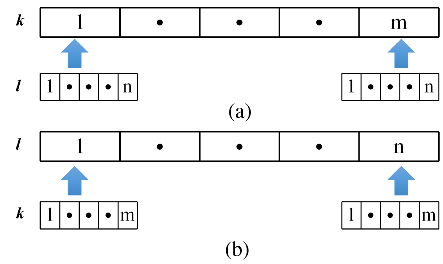


Fig. 1. Two scenarios of configurations of a system. (a) Configuration $k \in \{1, 2, \ldots, m\}$ includes n subconfigurations. (b) Subconfiguration $l \in \{1, 2, \ldots, n\}$ can be explained to the average of m configurations.

TABLE I
EXAMPLE OF DECISION INFORMATION SYSTEM

Objects	a	b	c	d
x_1	0.47	0.48	0.58	0
x_2	0.43	0.39	0.68	0
x_3	0.31	0.26	0.44	0
x_4	0.26	0.17	0.38	0
x_5	0.30	0.25	0.29	0
x_6	0.52	0.49	0.74	1
x_7	0.45	0.57	0.67	1
x_8	0.67	0.71	0.73	1

D. Motivation

As is well known, the external expression of anything results from the collective influence of its parts. Each configuration can be considered a subsystem with subconfigurations as proposed in [15]. When evaluating the disorder of a system, the configurations and all their subconfigurations should be considered. Thus, the zentropy theory considering multilevel information is proposed and applied to actual application. In this theory, two scenarios of system configurations can be depicted in Fig. 1, where (a) depicts a configuration consisting of n subconfigurations, while (b) suggests that all subconfigurations can be explained by the average of m configurations, indicating an interaction between different levels of configurations. This figure shows that the configurations interact, and the disorder of a system should be the overall performance of all configurations at different levels.

In NRS theory, the uncertainty of a vague concept is usually approximated by lower approximation and upper approximation. This approximation process is related to the target concept, approximation level, neighborhood similarity classes, and specific objects, which are multiple granular levels and influence each other in the approximation process. Nevertheless, the existing uncertainty measures, such as dependency degree and NCE, mostly focus on the information presented at a single level, which is incomplete for characterizing uncertainty knowledge, especially when the information system changes. Example 1 is given to illustrate this limitation in decision information.

Example 1: A decision information system $DIS = (U, C \cup D, V, f)$ is shown in Table I , where $U = \{x_1, x_2, \ldots, x_8\}$ is the universe of 8 patients, $C = \{a, b, c\}$ includes three tumor genes, $D = \{d\}$ is the diagnosis results, "1" and "0" are "Sickness" and "Health," and $D_1 = \{x_1, x_2, x_3, x_4, x_5\}, D_2 =$

$\{x_6, x_7, x_8\}$. At present, gene a is known to be necessary, and the other gene only needs to be selected due to the cost. In this example, the dependency degree and NCE are adopted to evaluate genes. For a fair comparison, the neighborhood radius is uniformly set to 0.1. Let $A = \{a, b\}$ and $B = \{a, c\}$, the Euclidean distance matrix D_1 under A can be computed as follows:

$$D = \begin{bmatrix} 0.000 & 0.099 & 0.272 & 0.374 & 0.286 & 0.051 & 0.092 & 0.305 \\ 0.099 & 0.000 & 0.177 & 0.278 & 0.191 & 0.135 & 0.181 & 0.400 \\ 0.272 & 0.177 & 0.000 & 0.103 & 0.014 & 0.311 & 0.340 & 0.576 \\ 0.374 & 0.278 & 0.103 & 0.000 & 0.089 & 0.412 & 0.443 & 0.678 \\ 0.286 & 0.191 & 0.014 & 0.089 & 0.000 & 0.326 & 0.353 & 0.590 \\ 0.051 & 0.135 & 0.311 & 0.412 & 0.326 & 0.000 & 0.106 & 0.266 \\ 0.092 & 0.181 & 0.340 & 0.443 & 0.353 & 0.106 & 0.000 & 0.261 \\ 0.305 & 0.400 & 0.576 & 0.678 & 0.590 & 0.266 & 0.261 & 0.000 \end{bmatrix}.$$

The neighborhood similarity classes of each object under A are obtained as follows:

$\delta_A(x_1) = \{x_1, x_2, x_6, x_7\}, \delta_A(x_2) = \{x_1, x_2\}, \delta_A(x_3) = \{x_3, x_5\}, \delta_A(x_4) = \{x_4, x_5\}, \delta_A(x_5) = \{x_3, x_4, x_5\}, \delta_A(x_6) = \{x_1, x_6\}, \delta_A(x_7) = \{x_1, x_7\}, \delta_A(x_8) = \{x_8\}$.

Then, the lower approximations of D_1 and D_2 under A are

$\underline{R}_A(D_1) = \{x_2, x_3, x_4, x_5\}, \quad \underline{R}_A(D_2) = \{x_8\}$.

Similarly, the neighborhood similarity classes of each object under B are obtained as follows:

$\delta_B(x_1) = \{x_1, x_7\}, \delta_B(x_2) = \{x_2, x_7\}, \delta_B(x_3) = \{x_3, x_4\}$,

$\delta_B(x_4) = \{x_3, x_4\}, \delta_B(x_5) = \{x_3, x_4, x_5\}, \delta_B(x_6) = \{x_6, x_7\}, \delta_B(x_7) = \{x_1, x_2, x_6, x_7\}, \delta_B(x_8) = \{x_8\}$.

Then, the lower approximations of D_1 and D_2 under B are

$\underline{R}_B(D_1) = \{x_3,\ x_4,\ x_5\}, \quad \underline{R}_B(D_2) = \{x_6,\ x_8\}$.

According to the definitions 2 and 5, the measure values could be computed as follows:

$\gamma_A(D) = \gamma_B(D) = \frac{5}{8}, NCE(D|A) = NCE(D|B) = \frac{3}{8}log2$.

It can be known that the genes subset A and B cannot be distinguished based on the dependency degree or NCE. That is because they only consider the information presented at a single level while not considering the connection between different levels, thus leading to an incomplete and imprecision description of uncertainty. Combining the multiple granular levels in the approximation process to investigate a systematic method to characterize uncertainty is necessary for decision information systems.

III. Zentropy-Based Uncertainty Measure

Considering the incomplete description and lower accuracy of uncertainty measures on a single granular level, a novel zentropy-based uncertainty measures combining granular level analysis and its properties are investigated in this section.

A. Granular Level Analysis of Concept Approximation

As a new computing theoretical paradigm, granular computing can simulate the human brain to realize multiperspective and multilevel cognition [34]. Especially the RS model provides a mathematical expression for the cognition of uncertainty and describes uncertainty quantitatively. In RS theory, a target concept is approximated by the lower approximation and upper approximation, and the model's precision is characterized by boundary region. Many methods based on approximation level have been proposed to describe uncertainty.

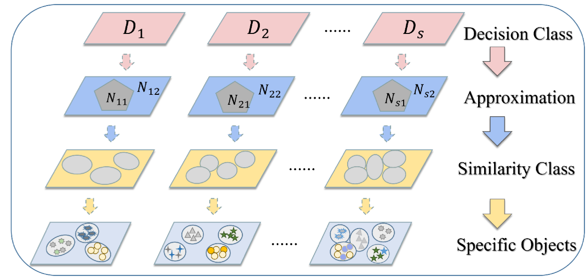


Fig. 2. Granular levels in concept approximation. There are s decision classes and each target decision D_i can be described by lower approximation N_{i1} and nonlower approximation N_{i2}. To further investigate the elements influencing approximation level, the objects' similarity classes are investigated. Moreover, the specific object in each similarity class can also reflect the properties of the corresponding similarity class at a finer level.

Note that the approximation process of target concept is related to the target concept, classes, and objects, which is a gradually refined granular level structure and each granular level influenced each other. Thus, the depiction of uncertainty should be multilevel and gradually refined from the target concept to samples. However, the existing uncertainty measures mostly focus on the information presented at a single level, leading to a limited and incomplete characterization of uncertainty, which motivates this study.

Specifically, the recognition process in NRS can be shown in Fig. 2. Given the dataset with s decision classes, it can roughly depict the target decision from its lower approximation and upper approximation. To further investigate the factors affecting the model's precision, it is necessary to analyze the similarity class and their specific objects. The approximation process is multiple granular levels from coarser to finer. Thus, the uncertainty of a decision information system should be the overall performance of different granular levels. This section investigates a novel uncertainty measure combined with the granular level structure and systematically discusses its representation and properties.

B. Uncertainty Measure With Granular Level

Based on the analysis in Section III-A, in this section, a zentropy-based uncertainty measure combining information in different granular levels is proposed.

As the approximation process shown in Fig. 2, the uncertainty measure in a decision information system should represent the information presented at the target concept, approximation level, similarity class, and specific objects, which can be shown as follows.

Definition 6: Let $DIS = (U, C \cup D, V, f)$ be a decision information system. For $B \subseteq C$, the zentropy-based uncertainty measure($\mathcal{Z}$) of D on B is defined as follows:

$$\mathcal{Z}_B(D) = -\sum_{k=1}^{s} p_k \text{log} p_k + \sum_{k=1}^{s} p_k \mathcal{Z}_k \tag{8}$$

where $p_k = \frac{|D_k|}{|U|}$ is the probability of the kth class at the decision level, $\mathcal{Z}_k$ represents the entropy of kth class, which can be decomposed into granular in the lower levels with the same (8).

Next, the $\mathcal{Z}_k$ reflected the uncertainty at approximation level can be represented as follows:

$$\mathcal{Z}_k = -\sum_{l=1}^{2} p_{kl}\log p_{kl} + \sum_{l=1}^{2} p_{kl}\mathcal{Z}_{kl} \tag{9}$$

where $p_{k1} = \frac{|\underline{\mathcal{R}}_B(D_k)|}{|D_k|}$, $p_{k2} = \frac{|D_k - \underline{\mathcal{R}}_B(D_k)|}{|D_k|}$, depicting the distribution of certainty set and noncertainty set in class D_k.

The certainty of an object depends on the relationship between its similar class and target concept. Thus, the $\mathcal{Z}_{k1}$ and $\mathcal{Z}_{k2}$ can be further defined from the similarity classes, which are defined as follows:

$$\mathcal{Z}_{kl} = -\sum_{w=1}^{|N_{kl}|} p_{klw}\log p_{klw} + \sum_{w=1}^{|N_{kl}|} p_{klw}\mathcal{Z}_{klw} \tag{10}$$

where $N_{k1} = \underline{\mathcal{R}}_B(D_k), N_{k2} = D_k - \underline{\mathcal{R}}_B(D_k)$, and $p_{klw} = \frac{|\delta_B(w)|}{\sum_{w=1}^{|N_{kl}|}|\delta_B(w)|}$ represents the probability of wth similarity class $\delta_B(w)$ among all similarity classes of objects in N_{kl}.

Moreover, the uncertainty of finer specific object level needs to be considered, which can be described as follows:

$$\mathcal{Z}_{klw} = -\sum_{o=1}^{2} p_{klwo}\log p_{klwo} \tag{11}$$

where $p_{klw1} = \frac{|\delta_B(w)\cap D_k|}{|\delta_B(w)|}$, $p_{klw2} = \frac{|\delta_B(w)\cap D_k^c|}{|\delta_B(w)|}$ reflecting the distribution of objects with decision labels in similarity classes.

From the abovementioned definition, it can be obtained that the zentropy at each granular level consists of two parts, where the first one describes the uncertainty at this level, and the others reflect the information presented at the interior of this level. This representation reflects the interaction of granular level in uncertainty measure.

C. Properties About Zentropy-Based Uncertainty Measure

This section shows an example to illustrate the computational process of the proposed measure, and then some properties are given and analyzed to investigate its rationality. Specifically, the computational details of the zentropy-based uncertainty measure are shown in Example 2.

Example 2: Continue to the Example 1, we know $D_1 = \{x_1, x_2, x_3, x_4, x_5\}$, $D_2 = \{x_6, x_7, x_8\}, N_{11} = \underline{\mathcal{R}}_A(D_1) = \{x_2, x_3, x_4, x_5\}$, $N_{12} = D_1 - N_{11} = \{x_1\}$, $N_{21} = \underline{\mathcal{R}}_A(D_2) = \{x_8\}, N_{22} = D_1 - N_{21} = \{x_6, x_7\}$. Then, we could obtain the probability distribution at the target decision level is

$p_1 = \frac{|D_1|}{|U|} = \frac{5}{8}, p_2 = \frac{|D_2|}{|U|} = \frac{3}{8}$.

The probability distribution at the approximation level is

$p_{11} = \frac{|\underline{\mathcal{R}}_A(D_1)|}{|D_1|} = \frac{4}{5}, p_{12} = \frac{|D_1 - \underline{\mathcal{R}}_A(D_1)|}{|D_1|} = \frac{1}{5}$

$p_{21} = \frac{|\underline{\mathcal{R}}_A(D_2)|}{|D_2|} = \frac{1}{3}, p_{22} = \frac{|D_2 - \underline{\mathcal{R}}_A(D_2)|}{|D_2|} = \frac{2}{3}$.

The probability distribution at the similarity class level is

$p_{111} = \frac{|\delta(x_2)|}{\sum_{w\in N_{1l}}|\delta(x_w)|} = \frac{2}{9}$, $p_{112} = \frac{|\delta(x_3)|}{\sum_{w\in N_{1l}}|\delta(x_w)|} = \frac{2}{9}$

$p_{113} = \frac{|\delta(x_4)|}{\sum_{w\in N_{1l}}|\delta(x_w)|} = \frac{2}{9}$, $p_{114} = \frac{|\delta(x_5)|}{\sum_{w\in N_{1l}}|\delta(x_w)|} = \frac{3}{9}$

$p_{121} = \frac{|\delta(x_1)|}{\sum_{w\in N_{12}}|\delta(x_w)|} = 1$, $p_{211} = \frac{|\delta(x_8)|}{\sum_{w\in N_{2l}}|\delta(x_w)|} = 1$

$p_{221} = \frac{|\delta(x_6)|}{\sum_{w\in N_{22}}|\delta(x_w)|} = \frac{1}{2}$, $p_{222} = \frac{|\delta(x_7)|}{\sum_{w\in N_{22}}|\delta(x_w)|} = \frac{1}{2}$.

The probability distribution at the specific object level is

$p_{1111} = \frac{|\delta_A(x_2)\cap D_1|}{|\delta_A(x_2)|} = 1$, $p_{1112} = \frac{|\delta_A(x_2)\cap D_1^c|}{|\delta_A(x_2)|} = 0$

$p_{1121} = \frac{|\delta_A(x_3)\cap D_1|}{|\delta_A(x_3)|} = 1$, $p_{1122} = \frac{|\delta_A(x_3)\cap D_1^c|}{|\delta_A(x_3)|} = 0$

$p_{1131} = \frac{|\delta_A(x_4)\cap D_1|}{|\delta_A(x_4)|} = 1$, $p_{1132} = \frac{|\delta_A(x_4)\cap D_1^c|}{|\delta_A(x_4)|} = 0$

$p_{1141} = \frac{|\delta_A(x_5)\cap D_1|}{|\delta_A(x_5)|} = 1$, $p_{1142} = \frac{|\delta_A(x_5)\cap D_1^c|}{|\delta_A(x_5)|} = 0$

$p_{1211} = \frac{|\delta_A(x_1)\cap D_1|}{|\delta_A(x_1)|} = \frac{1}{2}$, $p_{1212} = \frac{|\delta_A(x_1)\cap D_1^c|}{|\delta_A(x_1)|} = \frac{1}{2}$

$p_{2111} = \frac{|\delta_A(x_8)\cap D_2|}{|\delta_A(x_8)|} = 1$, $p_{2112} = \frac{|\delta_A(x_8)\cap D_2^c|}{|\delta_A(x_8)|} = 0$

$p_{2211} = \frac{|\delta_A(x_6)\cap D_2|}{|\delta_A(x_6)|} = \frac{1}{2}$, $p_{2212} = \frac{|\delta_A(x_6)\cap D_2^c|}{|\delta_A(x_6)|} = \frac{1}{2}$

$p_{2221} = \frac{|\delta_A(x_7)\cap D_2|}{|\delta_A(x_7)|} = \frac{1}{2}$, $p_{2222} = \frac{|\delta_A(x_7)\cap D_2^c|}{|\delta_A(x_7)|} = \frac{1}{2}$.

Based on the probability at different granular levels, the zentropy values at specific object level are computed as

$\mathcal{Z}_{111} = 0$, $\mathcal{Z}_{112} = 0, \mathcal{Z}_{113} = 0, \mathcal{Z}_{114} = 0, \mathcal{Z}_{121} = \log 2$

$\mathcal{Z}_{211} = 0, \mathcal{Z}_{221} = \log 2, \mathcal{Z}_{222} = \log 2$.

Then, the entropy values at similarity class can be computed according to (10)

$$\begin{aligned}\mathcal{Z}_{11} &= -\sum_{w=1}^{|N_{11}|} p_{11w}\log p_{11w} + \sum_{w=1}^{|N_{11}|} p_{11w}\mathcal{Z}_{11w}\\ &= -\left(\frac{2}{9}\log\frac{2}{9} + \frac{2}{9}\log\frac{2}{9} + \frac{2}{9}\log\frac{2}{9} + \frac{3}{9}\log\frac{3}{9}\right) + 0\\ &= -\frac{1}{3}\log\frac{2^2}{3^5}.\end{aligned}$$

Similarly, we obtain $\mathcal{Z}_{12} = \log 2, \mathcal{Z}_{21} = 0, \mathcal{Z}_{22} = 2\log 2$. Moreover, the zentropy values at approximation level are obtained according to (9)

$$\mathcal{Z}_1 = -\sum_{l=1}^{2} p_{1l}\log p_{1l} + \sum_{l=1}^{2} p_{1l}\mathcal{Z}_{1l} = -\frac{29}{15}\log 2 + \frac{4}{3}\log 3 + \log 5$$

$$\mathcal{Z}_2 = -\sum_{l=1}^{2} p_{2l}\log p_{2l} + \sum_{l=1}^{2} p_{2l}\mathcal{Z}_{2l} = \frac{2}{3}\log 2 + \log 3.$$

Therefore, the zentropy-based uncertainty measure of D on A can be computed as follows:

$$\begin{aligned}\mathcal{Z}_A(D) &= -\sum_{k=1}^{s} p_k\log p_k + \sum_{k=1}^{s} p_k\mathcal{Z}_k\\ &= -\left(\frac{5}{8}\log\frac{5}{8} + \frac{3}{8}\log\frac{3}{8}\right) + \left(\frac{5}{8}\mathcal{Z}_1 + \frac{3}{8}\mathcal{Z}_2\right)\\ &= \frac{49}{24}\log 2 + \frac{5}{6}\log 3.\end{aligned}$$

Similarly, the zentropy-based uncertainty measure of D on B can be computed as $\mathcal{Z}_B(D) = \frac{461}{168}\log 2 - \frac{2}{7}\log 3 + \frac{3}{8}\log 7$. Compared with the dependency degree and NCE in Example 1, the gene B could be selected according to the minimum values of the zentropy-based uncertainty measure.

Property 1: Let $DIS = (U, C\cup D, V, f)$ be a decision information system. For $B, B_1, B_2 \subseteq C$, $\mathcal{Z}_B(D)$ is the zentropy-based uncertain measure. Then, the following properties hold.

1) Nonnegative: $\mathcal{Z}_B(D) \geq 0$.
2) For $B_1 \subseteq B_2$, the $\mathcal{Z}_{B_1}(D)$ and $\mathcal{Z}_{B_2}(D)$ is nonmonotonic.

3) For $B, P \subseteq C$ and $B \prec P$, the $\mathcal{Z}_B(D)$ and $\mathcal{Z}_P(D)$ is incomparable.

Proof: 1) From definition 6, it is easily obtain that $0 \le p_k, p_{kl}, p_{klw}, p_{klwo} \le 1$ for $k = 1, 2, \ldots, s$, $l = 1, 2, w = 1, 2, \ldots, |N_{kl}|, o = 1, 2$, thus the $\mathcal{Z}_B(D) \ge 0$ holds.

2) According to (1), for $D_k \in U/D$ and $w \in D_k$, when $B_1 \subseteq B_2$, the $\frac{|\delta_{B_1}(w) \cap D_k|}{|\delta_{B_1}(w)|}$ and $\frac{|\delta_{B_2}(w) \cap D_k|}{|\delta_{B_2}(w)|}$ is incomparable. Then, based on definition 6, the $\mathcal{Z}_{B_2}(D)$ and $\mathcal{Z}_{B_1}(D)$ is non-monotonic.

3) Since $B \prec P$, then $\delta_B(w) \subseteq \delta_P(w)$ for any $w \in D_k (k = 1, 2, \ldots, s)$. Since $\frac{|\delta_B(w) \cap D_k|}{|\delta_B(w)|}$ and $\frac{|\delta_P(w) \cap D_k|}{|\delta_P(w)|}$ are incomparable, and the relationship between $\underline{\mathcal{R}}_B(D_k)$ and $\underline{\mathcal{R}}_P(D_k)$, $\overline{\mathcal{R}}_B(D_k)$ and $\overline{\mathcal{R}}_P(D_k)$ are unclear, the $\mathcal{Z}_B(D)$ and $\mathcal{Z}_P(D)$ are incomparable according to definition 6. ■

Property 2: Let $DIS = (U, C \cup D, V, f)$ be a decision information system. For $B \subseteq C$, $D_k \in U/D$, the following properties hold.

1) $\mathcal{Z}_{k1} = -\sum_{w=1}^{|N_{k1}|} p_{klw} \log p_{klw}$.
2) If $\underline{\mathcal{R}}_B(D_k) = D_k$, $\mathcal{Z}_k = \mathcal{Z}_{k1}$.
3) If $\underline{\mathcal{R}}_B(D_k) = \emptyset$, $\mathcal{Z}_k = \mathcal{Z}_{k2}$.
4) For $Q \subseteq C$, $\mathcal{R}_Q(w) = w$ for $w \in U$, then $\mathcal{Z}_Q = \log n$.

Proof: 1) According to definition 1, for the wth object in $\underline{\mathcal{R}}_B(D_k)$, $\delta_B(w) \subseteq D_k$ holds, thus, $p_{k1w1} = \frac{|\delta_B(w) \cap D_k|}{|\delta_B(w)|} = 1, p_{k1w2} = \frac{|\delta_B(w) \cap D_k^c|}{|\delta_B(w)|} = 0$. Then, $\mathcal{Z}_{k1w} = 0$ and $\mathcal{Z}_{k1} = -\sum_{w=1}^{|N_{k1}|} p_{klw} \log p_{klw}$ holds.

2) According to (9), $p_{k1} = \frac{|\underline{\mathcal{R}}_B(D_k)|}{|D_k|} = 1$ and $p_{k2} = \frac{|D_k - \underline{\mathcal{R}}_B(D_k)|}{|D_k|} = 0$ when $\underline{\mathcal{R}}_B(D_k) = D_k$, thus, $\mathcal{Z}_k = \mathcal{Z}_{k1}$.

3) Similarly, according to definition 1, if $\underline{\mathcal{R}}_B(D_k) = \emptyset$, then $p_{k1} = 0$ and $p_{k2} = 1$. Therefore, $\mathcal{Z}_k = \mathcal{Z}_{k2}$ holds.

4) From definition 1, it can be obtained that $\underline{\mathcal{R}}_Q(D_k) = D_k$ when $\mathcal{R}_Q(w) = w$ for $w \in U$, then $\mathcal{Z}_k = \mathcal{Z}_{k1}$ from the 2) in this property. Furthermore, from (10), $\mathcal{Z}_{k1} = \log|D_k|$. Therefore, $\mathcal{Z}_Q = \log n$ according to (8). ■

According to the abovementioned analysis, the proposed zentropy-based uncertainty measure considers the information presented at multiple granular levels and is systematically defined by the probability distribution at different levels. Thus, when the information system changes, the changes of the proposed measure will depend on the whole changes of different levels. Different levels exhibit diverse responses to system changes, and the increase or decrease of entropy at different levels might offset each other. Therefore, compared with other entropy measures focusing on a single level, the proposed measure from the whole system is relatively stable for changes.

IV. Feature Selection Via Uncertainty Measure $\mathcal{Z}$

In this section, a feature selection algorithm based on the proposed zentropy uncertainty measure $\mathcal{Z}$ is designed for selecting the optimal feature subset.

A. Features Evaluation in DIS

As the zentropy-based uncertainty measure increases, it indicates a higher level of confusion within the information system. Therefore, the ultimate goal of feature selection is to choose a subset of features that reduces the zentropy measure compared to the original information system.

Definition 7: Let $DIS = (U, C \cup D, V, f)$ be a decision information system. For $R \subseteq C$, R is called a feature reduct of $(U, C \cup D, V, f)$ if it has the following properties.

1) $\mathcal{Z}_R(D) \le \mathcal{Z}_C(D)$.
2) $\mathcal{Z}_{R-\{r\}}(D) > \mathcal{Z}_R(D)$ for any $r \in R$.

The first item guarantees that the uncertainty degree of the reduced system $(U, R \cup D, V, f)$ is not higher than the original $(U, C \cup D, V, f)$. The second item confirms that there are no redundant features in reduct R.

In the feature selection process, the inner and outer importance measures based on the proposed measure are defined to evaluate features.

Definition 8: Let $DIS = (U, C \cup D, V, f)$ be a decision information system. For any $c \in C$, and parameter δ, the inner importance measure of c relative to C is defined as

$$IM(c, C) = \mathcal{Z}_C(D) - \mathcal{Z}_{C-c}(D). \quad (12)$$

This inner measure is used to compute the increment of the system's certainty after reducing an attribute.

Definition 9: For any $b \in C - B$, the outer importance measure of b relative to B is defined as

$$OM(b, B) = \mathcal{Z}_B(D) - \mathcal{Z}_{B \cup b}(D). \quad (13)$$

Similarly, the outer measure can also be used to calculate the reduction of system uncertainty introduced by attributes. In this selection process, the relative most important feature could be selected according to the maximum principle.

B. Feature Selection Algorithm

This section designs a heuristic algorithm for feature selection in Algorithm 1. The relatively important features satisfying $\text{IM}(c, C) > 0$ for $c \in C$ are first selected. Then, the following excellent features are selected when $\mathcal{Z}_R(D) > \mathcal{Z}_C(D)$. Finally, steps 18–22 are also adopted to remove relative redundant features. Fig. 3 shows more details about this algorithm.

In Algorithm 1, step 2 first calculates the $\mathcal{Z}_C(D)$, which needs to compute all similar classes in U. Thus, the time complexity of step 2 is $O(n^2 m)$. Steps 3–8 select some important features relative to the original information system, whose time complexity is $O(|C|n^2(m-1))$. Suppose there are l_1 and l_2 features selected in steps 3–8 and 10–17, respectively. To obtain the ith feature in steps 10–17, similar class of all objects under $i + 1$ features needs to be computed. Thus, its complexity is $O(n^2(i+1))$ and the whole complexity of steps 10–17 is $O(\sum_{i=l_1}^{l_2}(m - i + 1)n^2(i+1))$. Similarly, steps 18–22 for removing some redundant features is $O(|R|n^2(|R|-1))$. Therefore, the whole time complexity of Algorithm 1 is $O(n^2 m^2)$.

C. Wrapper Technique of Searching a Best Subset

It can be obtained from Algorithm 1 that the feature reduction preserves the complete information of multiple information scales. Algorithm 1 offered a heuristic feature selection method based on the proposed measure. This way focuses on the maximum information gain in a forward and greedy strategy and

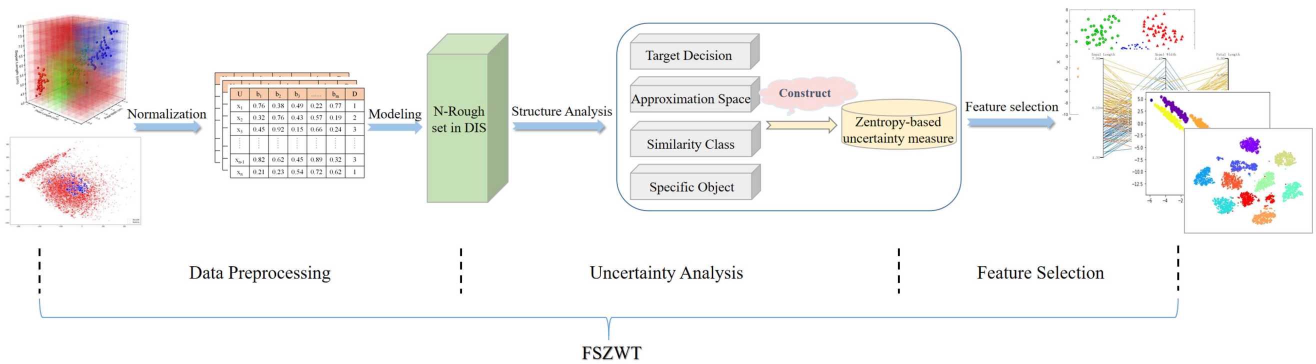


Fig. 3. Schematic flow chart of FSZWT. This chart describes three stages: firstly, the high dimension is processed as a DIS, and then the granular structure of the N-RS model and zentropy-based measure is analyzed and proposed. Finally, a feature selection method is designed for reduction.

Algorithm 1: Feature Selection Via Zentropy-Based Measure.

Input: Decision information system(DIS), radius δ.
Output: The feature reduction R.

Initialize $R \leftarrow \emptyset$, $start = 1$;
Compute $\mathcal{Z}_C(D)$ according to definition 6;
for $c \in C$ **do**
  Calculate $IM(c, C)$ by definition 8;
  if $IM(c, C) > 0$ **then**
    $R \leftarrow c$;
  end if
end for
Compute $\mathcal{Z}_R(D)$ according to definition 6;
while $\mathcal{Z}_R(D) > \mathcal{Z}_C(D)$ **do**
  for $b \in C - R$ **do**
    Calculate $OM(b, R)$ according to definition 9;
  end for
  Obtain $b_0 = argmax_{b \in C-R} OM(b, R)$;
  $R \leftarrow b_0$;
  Compute $\mathcal{Z}_R(D)$;
end while
for $r \in R$ **do**
  Calculate $\mathcal{Z}_{R-r}(D)$;
  if $\mathcal{Z}_{R-r}(D) < \mathcal{Z}_R(D)$ **then**
    $R \leftarrow R - \{r\}$;
  end if
end for
return R.

easily causes redundant features in the feature subset, disturbing the classification performance. Therefore, To avoid this situation as much as possible, the wrapper technique [38] is also added to the proposed method for selecting the best excellent feature subset, the detail as shown in Algorithm 2.

Algorithm 2 displays the process of selecting the best feature subset with the highest classification accuracy. This algorithm needs to compare $|R|$ feature subsets, including $R_1, R_2, \ldots, R_{|R|}$, thus, the whole time complexity is $O(|R|)$. In this article, Algorithm 1 is first adopted to select a feature reduction, and a better feature subset with the optimal classification performance is determined according to

Algorithm 2: An Optimal Subset Based on Wrapper Technique.

Input: The selected feature subset $R = \{r_1, r_2, \ldots, r_{|R|}\}$ obtained by Algorithm 1.
Output: The best feature subset R^* for DIS.

Initialize $R_1 = \{r_1\}, R_2 = \{r_1, r_2\}, \ldots, R_{|R|} = \{r_1, r_2, \ldots, r_{|R|}\}$;
for $i = 1$ to $|R|$ **do**
  Select the feature subset R_i and denote $DIS_i = (U, R_i \cup D, V, f)$, for $i = 1, 2, \ldots, |R|$;
  Compute the average classification accuracy $Accuracy_i$ of DIS^i by two employed classifiers;
end for
Select the best subset $R^* = argmax_{1 \le i \le |R|} Accuracy_i$;
return R.

Algorithm 2. The whole feature selection process is called FSZWT.

V. Experimental Design and Analysis

In this section, various comparative experiments are conducted to validate the effectiveness and superiority of the proposed uncertainty measure. Specifically, the $\mathcal{Z}$ is compared with other existing IE measures in robustness for noisy data and then compare the performance of FSZWT with the other eight representative feature selection methods in classification performance. All the experiments are carried out on a public computer with OS: Microsoft WIN10; Processor: Intel(R) Core(TM) i7-6800 K CPU @ 3.4 GHz×12; Memory: 62.7 GB; Programming language: MATLAB.

A. Experimental Design

A total of 12 public datasets from UCI Repository (https://www.uci.edu/) are employed to analyze the performance of different methods, whose information is shown in Table II. For each dataset, the conditional attribute value of each object $x_i \in U$ is normalized as follows:

$$\widehat{f}(x_i, a_j) = \frac{f(x_i, a_j) - \min(V_{a_j})}{\max(V_{a_j}) - \min(V_{a_j})} \quad \forall a_j \in C \tag{14}$$

TABLE II
DATASET DESCRIPTION

No.s	Dataset	Abbreviation	Sample	Feature	Classe
1	Cardiotocography dataset	Card	2126	21	3
2	Hill	Hill	606	101	2
3	Ionosphere	Ionosphere	351	34	2
4	Libras movement	Movement	360	91	15
5	Mice protein expression dataset	MPED	1077	69	8
6	Nursery	Nursery	12960	9	5
7	Sonar	Sonar	208	61	2
8	Spambase	Spambase	4601	58	2
9	Urban Land Cover Data Set	ULCD	168	148	9
10	Colonok	Colonok	63	2001	3
11	DryBean dataset	DryBean	13611	17	6
12	Dermatology	Derma	366	35	6

where $f(x_i, a_j)$ is the value of object x_i under conditional attribute $a_j \in C$, the $\max(V_{a_j})$ and $\min(V_{a_j})$ denote the maximum and minimum value of all objects in a_j.

In this section, to evaluate the robustness of the proposed zentropy-based measure, four representative entropy-based measures under the same parameter $\delta = 0.9$, including IE based on approximation (IEA) [25], composite information entropy (CIE) [31], IE [10] based on neighborhood class, and conditional entropy (CE) [1] are adopted to make comparisons in noise environment.

Meanwhile, five representative feature methods, including feature selection based on neighborhood self-information (FSNSI) [25], feature selection with fuzzy rough set (FS-FRS) [26], local neighborhood rough set (LNRS) [27], feature selection based on fuzzy neighborhood rough set (FSFNRS) [28], fuzzy feature selection using composite entropy-based measure (FFSCE) [31], and other three entropy-based feature selection methods, including uncertainty measure for incomplete decision table (UMIDT) [1], conditional entropy for incomplete decision system (CEIDS) [2], and rough entropy knowledge granulation (REKG) [13] are selected to illustrate the superiority of the proposed FSZWT method in classification performance. For these above-compared methods, they need to compute the similarity classes of n objects under m features to select the first important feature. Thus, the time complexity of this step is $O(n^2 m)$. Subsequently, in a heuristic algorithm, they also need to obtain the similarity classes under $i+1 (i = 1, 3, \ldots, m-1)$ features at most and repeated $m-i+1$ times in each cycle; thus, the complexity is $O(\frac{1}{2}n^2 m^2)$. Therefore, the whole time complexity of compared methods is $O(n^2 m^2)$. The running time would differ due to the actual selection mechanism, and the related results are recorded and analyzed in the experimental section.

All the compared methods rerun with the same optimization by K-nearest neighbor (KNN) and naive Bayes (NB) classifiers in Algorithm 2. Meanwhile, ten-fold cross-validation is adopted to evaluate the classification performance of each method on 12 datasets for a fair comparison.

B. Robustness Analysis Between Different Measures

This section evaluates the robustness of the proposed zentropy-based measure compared with IEA [25], CIE [31], IE [10], and CE [1]. In each dataset, we add random noise to different sample proportions on the conditional attribute set, and the proportion changes from 2.5% to 25% with a step of 2.5%. For each conditional attribute $a_j \in C$, the noise data are obtained as follows:

$$\widehat{v}(x_i, a_j) = \begin{cases} 0 & , \ \widehat{f}(x_i, a_j) + r_{ij} < 0 \\ \widehat{f}(x_i, a_j) + r_{ij}, & 0 \leq \widehat{f}(x_i, a_j) + r_{ij} \leq 1 \\ 1 & , \ \widehat{f}(x_i, a_j) + r_{ij} > 1 \end{cases} \quad (15)$$

where $\widehat{f}(x_i, a_j)$ is the normalized data and all r_{ij} obeys a Gaussian distribution with mean value 0 and variance 1.

The values of compared entropy-based measures at different noise levels and their standard deviation (std) are shown in Fig. 4, where the bar figure in each subfigure describes the std. Note that the CE's values shown in this figure are scaled to describe clearly the trends of different measures, where the reduction multiple is shown in Table III. It could be obtained that the value of $\mathcal{Z}$ is relatively stable with the noise data increasing compared with other entropy measures. The smallest column height of our measure variance in most datasets also illustrates this issue. Moreover, the detailed values of compared entropy-based measures are shown in Table III, where the smallest std of compared measures is in bold on each dataset. From this table, the IE measure increases with the noise data increasing except on the Ionosphere, while the other compared measures show a decreasing trend in most datasets. That is because the Gaussian noise significantly enhances the dataset's discreteness, and the neighborhood class is refined with the same neighborhood parameter. Therefore, the IE measure defined on neighborhood class shows an upward trend with noise data increasing. By contrast, the refined neighborhood class will reduce the uncertainty in neighborhood approximation and improve the approximation accuracy. Thus, other entropy measures only focusing on the approximation sets mostly decrease with noise data due to the uncertainty reduction in the approximation process. Under the interaction of different levels, the proposed measure increases with noise data on the Ionosphere, Sonar, urban land cover data set (ULCD), and Derma and shows a decreasing trend on other datasets. Moreover, these compared measures are sensitive to noise data because these single granular levels are easily influenced. The systematic thought in zentropy theory could avoid this issue by considering the interaction of changes between different levels. These experimental results illustrate the robustness of the proposed measure in a noise environment.

C. Classification Performance Evaluation of FSZWT

To further verify the effectiveness of the FSZWT method based on zentropy, eight other representative feature selection methods are selected to make comparisons. In the FSZWT method, δ is an important parameter that directly affects the computational process of FSZWT. Its range is set from 0.05 to 0.5 with a step of 0.05 to learn the optimal classification performance of FSZWT on two classifiers. In this section, three aspects are analyzed to analyze the performance of FSZWT, including the number of selected features, classification performance, and statistical test.

1) Consuming Time and Number of Selected Features: Table IV records the time consumption and number of the selected

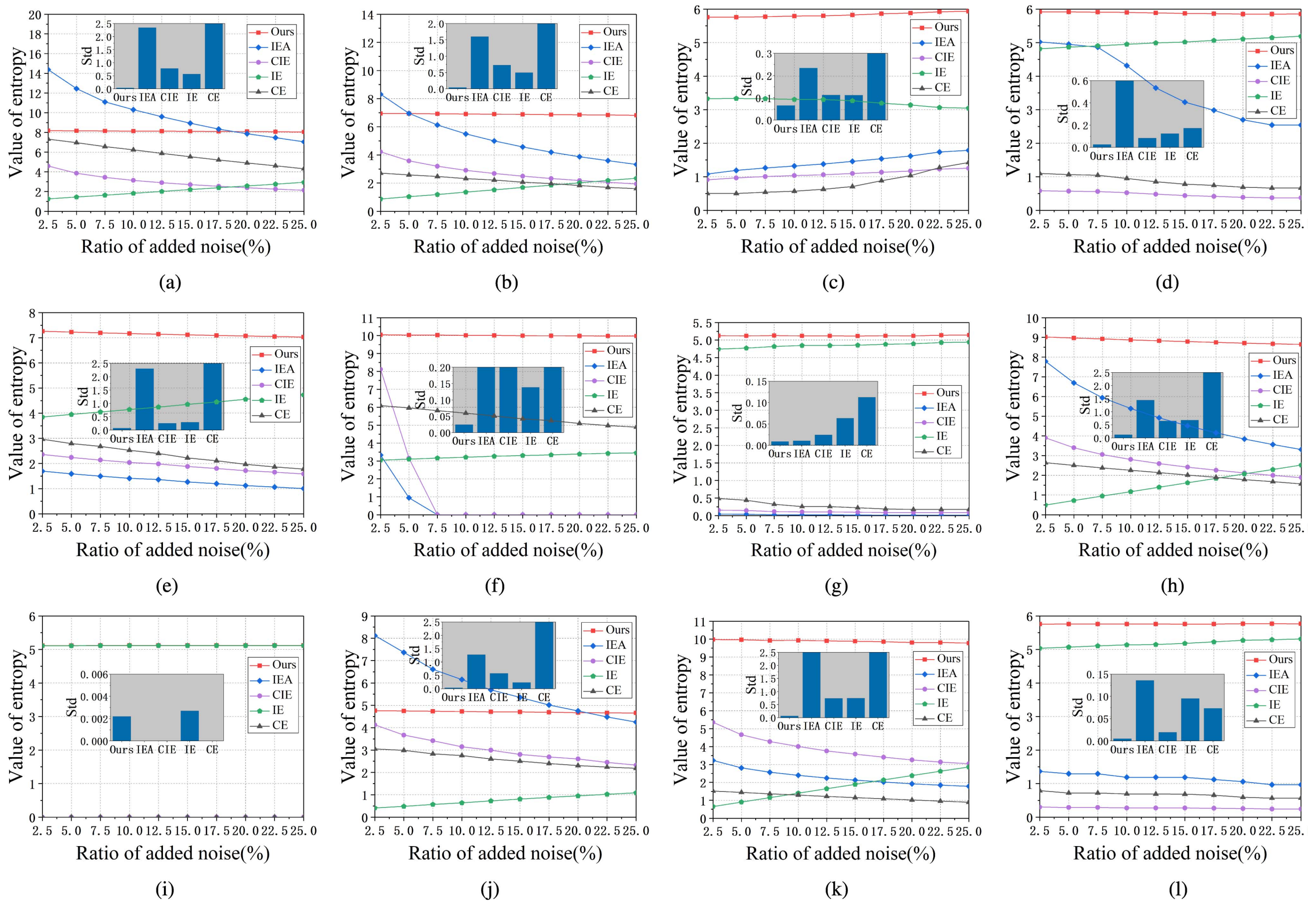


Fig. 4. Comparison of robustness of different entropy under different noise levels. (a) Card. (b) Hill. (c) Ionosphere. (d) Movement. (e) MPED. (f) Nursery. (g) Sonar. (h) Spambase. (i) ULCD. (j) Colonok. (k) DryBean. (l) Derma.

features for different methods, where the minimum time and feature number are in bold. This table shows that our method is superior to REKG, CEIDS, and UMIDT, while it is inferior to other compared methods from the average cardinality of selected features. Meanwhile, our method is more effective than FSFNI, CEIDS, and FSFRS from the average consuming time. From this table, the FSFNRS and FSFRS can only select 1 or 2 features in most cases, which significantly reduces the number of features, but their effect is limited to the classification performance. Since LNRS, FFSCE, and FFSCE only focus on the information presented at the approximation level while ignoring the other details in the approximation process, they effectively select fewer features to approximate the original information. However, they are not good at classification performance due to the incomplete characterization of uncertainty knowledge. Therefore, FSZWT has suitable reduction abilities in feature selection.

2) Classification Performance: This part mainly verifies the classification performance of different feature selection methods on 12 datasets. The results are expressed as $\mu \pm \sigma$, where μ and σ represent different methods' average values and std in ten-fold cross-validation experiments.

Tables V and VI record the classification accuracy of different methods on KNN and NB classifiers. Among these tables, rank is the average order of twelve datasets under different methods, and the excellent results are in bold. It can be seen from these tables that our method achieves the highest accuracy eight times on KNN and NB classifiers, which illustrates the excellent classification performance of our method. Moreover, the average accuracy of FSFNRS and FSFRS is much lower than that of other methods because fewer essential features are selected. The REKG, CEIDS, and UMIDT methods also have poor classification performance because the redundant features of interference classification still need to be removed. Fig. 5 additionally represents the excellence of FSZWT on average accuracy on KNN and NB classifiers. A common observation from all the subgraphs in Fig. 5(a) and (b) is that our designed method has the best classification accuracy in almost dataset. The FSZWT is relatively effective in feature selection.

Tables VII–XII also record the precision, recall, and $F1$-score of nine methods under two classifiers, where the excellent classification results are also in bold. On the KNN classifier, FSZWT achieves the highest values at 8, 7, and 7 times precision, recall, and $F1$-score in 12 datasets. On the NB classifier, the FSZWT performs very well at 7 times of 12 datasets in these three indicators and has the best behavior in average performance and rank. From the average classification performance, the FSFNRS and FSFRS still perform poorly on these 12 datasets due to the failure to filter out sufficiently essential features. Although many features are selected to feature subsets for REKG, CEIDS, and UMIDT methods, the classification performance could not be excellent. Thus,

TABLE III
DIFFERENT VALUES OF COMPARED ENTROPY-BASED MEASURES AT DIFFERENT NOISE LEVELS(%)

Noise(%)	Card					Hill					Ionosphere					Movement				
	ZUM	IEA	CIE	IE	CE(/100)	ZUM	IEA	CIE	IE	CE(/100)	ZUM	IEA	CIE	IE	CE(/10)	ZUM	IEA	CIE	IE	CE
2.5	8.179	14.372	4.590	1.252	7.319	6.949	8.314	4.222	0.873	2.711	5.760	1.081	0.910	3.330	0.501	5.918	5.021	0.582	4.814	1.101
5.0	8.166	12.444	3.862	1.442	6.953	6.935	6.953	3.577	1.036	2.579	5.760	1.190	0.966	3.335	0.502	5.917	4.951	0.571	4.866	1.067
7.5	8.152	11.099	3.440	1.630	6.565	6.924	6.130	3.197	1.190	2.467	5.775	1.263	1.006	3.327	0.538	5.908	4.863	0.560	4.903	1.044
10.0	8.139	10.307	3.135	1.819	6.239	6.909	5.495	2.907	1.360	2.323	5.793	1.321	1.035	3.306	0.571	5.901	4.313	0.524	4.951	0.950
12.5	8.125	9.597	2.899	2.011	5.869	6.897	4.997	2.683	1.517	2.214	5.801	1.380	1.060	3.303	0.629	5.887	3.649	0.480	4.990	0.851
15.0	8.111	8.938	2.697	2.199	5.529	6.880	4.568	2.491	1.692	2.067	5.824	1.458	1.099	3.265	0.710	5.873	3.226	0.439	5.017	0.779
17.5	8.102	8.343	2.533	2.380	5.209	6.867	4.196	2.326	1.850	1.961	5.861	1.536	1.134	3.193	0.881	5.865	2.985	0.416	5.065	0.743
20.0	8.087	7.867	2.382	2.567	4.893	6.852	3.873	2.183	2.019	1.831	5.883	1.618	1.172	3.141	1.035	5.852	2.697	0.387	5.107	0.690
22.5	8.073	7.465	2.249	2.751	4.610	6.835	3.592	2.059	2.192	1.696	5.920	1.737	1.230	3.065	1.271	5.852	2.542	0.374	5.147	0.664
25.0	8.060	7.045	2.127	2.935	4.305	6.823	3.324	1.941	2.343	1.609	5.938	1.790	1.255	3.046	1.425	5.855	2.542	0.374	5.190	0.663
Std	**0.040**	2.345	0.782	0.566	1.016	**0.043**	1.599	0.729	0.498	0.377	**0.065**	0.234	0.113	0.112	0.335	**0.027**	1.025	0.084	0.124	0.173

Noise(%)	MPED					Nursery					Sonar					Spambase				
	ZUM	IEA(/10)	CIE	IE	CE(/10)	ZUM	IEA(/10)	CIE	IE	CE(/100)	ZUM	IEA	CIE	IE	CE	ZUM	IEA	CIE	IE	CE(/1000)
2.5	7.255	1.696	2.359	3.847	2.965	10.045	3.329	8.119	3.048	6.095	5.127	0.044	0.154	4.743	0.489	9.017	7.786	3.904	0.493	2.643
5.0	7.225	1.589	2.241	3.951	2.791	10.036	0.946	3.155	3.104	5.965	5.121	0.039	0.144	4.772	0.438	8.967	6.693	3.400	0.721	2.508
7.5	7.195	1.497	2.137	4.047	2.677	10.025	0.000	0.000	3.161	5.811	5.130	0.025	0.115	4.819	0.324	8.921	5.939	3.057	0.950	2.376
10.0	7.165	1.410	2.039	4.145	2.524	10.016	0.000	0.000	3.210	5.669	5.122	0.021	0.106	4.846	0.259	8.875	5.390	2.806	1.171	2.257
12.5	7.138	1.360	1.982	4.243	2.400	10.006	0.000	0.000	3.265	5.512	5.122	0.021	0.106	4.846	0.259	8.829	4.918	2.595	1.397	2.135
15.0	7.115	1.268	1.882	4.350	2.221	9.999	0.000	0.000	3.307	5.354	5.117	0.018	0.096	4.856	0.222	8.788	4.517	2.417	1.622	2.016
17.5	7.090	1.198	1.802	4.445	2.110	9.993	0.000	0.000	3.340	5.253	5.125	0.015	0.087	4.881	0.189	8.747	4.161	2.260	1.846	1.902
20.0	7.069	1.121	1.716	4.553	1.967	9.984	0.000	0.000	3.390	5.097	5.121	0.015	0.087	4.894	0.178	8.710	3.842	2.121	2.075	1.784
22.5	7.049	1.065	1.654	4.641	1.865	9.979	0.000	0.000	3.428	4.979	5.140	0.015	0.087	4.931	0.176	8.675	3.564	1.998	2.295	1.678
25.0	7.026	1.004	1.585	4.728	1.776	9.975	0.000	0.000	3.455	4.891	5.146	0.015	0.087	4.942	0.176	8.645	3.309	1.885	2.523	1.569
Std	**0.077**	0.230	0.257	0.299	0.408	**0.024**	1.062	2.649	0.138	0.418	**0.009**	0.011	0.025	0.063	0.113	**0.126**	1.448	0.652	0.682	0.360

Noise(%)	ULCD					Colonok					DryBean					Derma				
	ZUM	IEA	CIE	IE	CE	ZUM	IEA	CIE	IE	CE(/10)	ZUM	IEA(/10)	CIE	IE	CE(/10000)	ZUM	IEA	CIE	IE	CE
2.5	5.115	0.000	0.000	5.109	0.000	4.759	8.121	4.127	0.411	30.509	9.980	32.289	5.357	0.653	1.516	5.757	1.365	0.301	5.035	0.791
5.0	5.115	0.000	0.000	5.109	0.000	4.752	7.367	3.670	0.481	29.871	9.968	28.110	4.663	0.904	1.441	5.760	1.294	0.289	5.069	0.720
7.5	5.120	0.000	0.000	5.116	0.000	4.735	6.611	3.418	0.567	28.311	9.924	25.693	4.280	1.153	1.364	5.759	1.294	0.289	5.106	0.720
10.0	5.120	0.000	0.000	5.116	0.000	4.727	6.152	3.148	0.640	27.542	9.927	23.982	4.005	1.401	1.293	5.759	1.189	0.274	5.136	0.691
12.5	5.120	0.000	0.000	5.116	0.000	4.709	5.692	2.996	0.726	26.014	9.903	22.452	3.758	1.649	1.220	5.759	1.189	0.274	5.146	0.687
15.0	5.120	0.000	0.000	5.116	0.000	4.703	5.355	2.803	0.811	25.048	9.880	21.282	3.578	1.892	1.153	5.758	1.189	0.274	5.180	0.685
17.5	5.120	0.000	0.000	5.116	0.000	4.688	5.018	2.691	0.883	23.995	9.852	20.200	3.407	2.139	1.086	5.758	1.123	0.265	5.224	0.653
20.0	5.120	0.000	0.000	5.116	0.000	4.672	4.749	2.602	0.951	23.055	9.809	19.241	3.257	2.382	1.021	5.769	1.060	0.256	5.271	0.593
22.5	5.120	0.000	0.000	5.116	0.000	4.665	4.479	2.451	1.023	22.376	9.811	18.465	3.138	2.624	0.958	5.769	0.968	0.242	5.290	0.567
25.0	5.120	0.000	0.000	5.116	0.000	4.659	4.254	2.326	1.091	21.861	9.781	17.858	3.043	2.864	0.897	5.768	0.968	0.242	5.313	0.566
Std	0.002	**0.000**	**0.000**	0.003	**0.000**	**0.036**	1.277	0.573	0.232	3.104	**0.069**	4.629	0.740	0.744	0.208	**0.005**	0.136	0.020	0.095	0.073

The minimum time and feature number are in bold.

TABLE IV
TIME CONSUMPTION(S) AND NUMBER(NUM) OF SELECTED FEATURES BY DIFFERENT METHODS

Dataset	LNRS		FSNSI		REKG		CEIDS		UMIDT		FFSCE		FSFNRS		FSFRS		**Ours**	
	Time	Num	Time	Num	Time	Num	Time	Num	Time	Num	Time	Num	Time	Num	Time	Num	Time	Num
Card	7.10	(4)	139.09	(16)	16.16	(19)	99.84	(14)	23.76	(15)	23.09	(6)	4.12	(1)	4.54	(1)	24.77	(13)
Hill	91.16	(50)	2.31	(1)	53.63	(95)	54.51	(25)	27.69	(96)	97.80	(5)	16.57	(14)	1796.46	(13)	663.20	(49)
Ionosphere	0.72	(4)	0.21	(1)	1.02	(27)	1.14	(25)	1.11	(8)	2.34	(4)	0.15	(1)	42.69	(9)	32.22	(12)
Movement	1.48	(2)	4318.67	(8)	6.79	(79)	593.12	(62)	5.32	(74)	18.27	(5)	2.02	(1)	2.58	(1)	794.45	(59)
MPED	1044.84	(3)	8.27	(1)	4383.49	(48)	5214.49	(63)	71.23	(48)	83.06	(7)	15.32	(1)	10.47	(11)	733.69	(42)
Nursery	142.08	(7)	379.48	(7)	135.34	(7)	139.71	(7)	272.03	(7)	214.43	(2)	219.60	(2)	840.77	(7)	289.95	(7)
Sonar	0.26	(1)	4.53	(1)	0.47	(10)	1.57	(26)	0.55	(3)	4.14	(4)	4.03	(1)	9103.89	(6)	5.35	(5)
Spambase	620.25	(12)	2148.50	(57)	765.55	(56)	809.21	(11)	1691.30	(56)	1205.40	(8)	83.92	(1)	119.35	(1)	2205.50	(17)
ULCD	0.50	(2)	922.61	(16)	0.73	(7)	5.01	(43)	44.79	(7)	18.42	(4)	1.28	(2)	1.12	(1)	31.97	(9)
Colonok	7.13	(4)	19420.00	(40)	104.46	(28)	1182.40	(54)	2998.70	(9)	1290.90	(6)	2.31	(1)	1.89	(1)	2346.90	(4)
DryBean	248.65	(4)	1880.30	(8)	604.72	(9)	624.41	(13)	1281.80	(11)	742.88	(7)	590.92	(2)	1019.30	(2)	1545.20	(8)
Derma	0.82	(7)	77.75	(32)	0.74	(12)	1.24	(29)	24.08	(12)	2.48	(6)	0.62	(2)	0.56	(1)	16.65	(14)
Average	180.41	(8.33)	2441.81	(15.67)	506.09	(33.08)	727.22	(31.00)	536.86	(28.83)	308.60	(5.33)	**78.40**	**2.42**	1078.63	(4.50)	724.16	(19.92)
Rank	3.00	(5.75)	6.75	(4.17)	3.67	(7.67)	5.42	(6.50)	5.75	(2.17)	5.75	(2.33)	**2.50**	**(3.00)**	4.67	(6.33)	7.50	(3.75)

The minimum time and feature number are in bold.

effectively choosing essential features is crucial to improving classification performance.

3) Statistical Test: To test whether there is a statistical difference from compared methods in classification performance, the Friedman test is first adopted at a significant level of $P = 0.1$. Table XIII shows the average rankings of these nine methods and the Friedman test's results, including χ and the corresponding p-value, where the minimum average ranking is in bold. This table shows that all the p-values are much less than 0.1, indicating significant differences among the nine methods on the two classifiers. Therefore, the Nemenyi post hoc test is necessary to determine whether there is a substantial difference between any two methods. In the Nemenyi test, the critical distance obtained as follows:

$$\mathrm{CD} = q_\alpha \sqrt{\frac{k(k+1)}{6N}} \tag{16}$$

where the $q_{\alpha=0.1} = 2.9716$ when $k = 12, N = 9$. A significant difference exists in the classification performance when the distance between two compared methods exceeds the $\mathrm{CD} = 3.192$.

Fig. 6 is the CD diagram that reflects the ranking of nine methods, in which the more minor the rank of the method, the

TABLE V
CLASSIFY ACCURACY OF DIFFERENT ALGORITHMS ON KNN BAYESIAN CLASSIFIER

Dataset	LNRS	FSNSI	REKG	CEIDS	UMIDT	FFSCE	FSFNRS	FSFRS	**Ours**
Card	72.10±0.03	81.19±0.03	81.33±0.03	80.72±0.02	81.52±0.02	74.51±0.04	48.73±0.03	47.79±0.04	**82.78±0.01**
Hill	56.78±0.05	51.30±0.07	57.26±0.04	51.64±0.08	52.30±0.06	53.62±0.06	46.21±0.06	52.48±0.07	**59.22±0.06**
Ionosphere	86.91±0.05	35.05±0.12	87.17±0.04	86.61±0.05	88.87±0.04	**90.02±0.08**	38.19±0.14	88.33±0.04	88.04±0.07
Movement	38.89±0.08	78.06±0.05	79.44±0.09	79.17±0.06	79.44±0.06	70.28±0.06	16.39±0.06	19.44±0.05	**81.67±0.07**
MPED	62.12±0.07	25.08±0.03	99.07±0.01	99.26±0.01	99.54±0.01	95.64±0.02	24.13±0.07	98.05±0.02	**99.81±0.00**
Nursery	90.86±0.01	91.05±0.01	90.68±0.01	90.96±0.01	91.02±0.01	49.65±0.06	28.35±0.03	90.93±0.00	**91.10±0.01**
Sonar	51.48±0.09	51.83±0.11	72.12±0.08	79.36±0.13	75.95±0.08	79.79±0.08	54.33±0.10	76.36±0.10	**81.26±0.08**
Spambase	87.37±0.02	90.33±0.02	90.24±0.01	80.00±0.02	**90.46±0.02**	73.96±0.02	39.90±0.02	39.93±0.03	86.10±0.00
ULCD	32.61±0.09	**77.98±0.15**	61.32±0.16	46.03±0.14	56.51±0.10	49.34±0.12	35.04±0.13	24.96±0.14	75.59±0.05
Colonok	71.19±0.10	65.24±0.15	61.43±0.22	73.10±0.15	73.1±0.22	67.86±0.12	55.00±0.14	51.90±0.17	**74.30±0.20**
DryBean	88.98±0.01	91.80±0.01	91.78±0.01	91.82±0.01	91.71±0.01	91.80±0.01	79.72±0.01	79.64±0.01	**91.90±0.00**
Derma	64.53±0.11	96.16±0.03	76.49±0.07	**96.99±0.02**	74.32±0.05	68.87±0.07	34.40±0.08	29.26±0.07	85.80±0.06
Average	66.53±0.06	61.15±0.06	78.56±0.07	79.69±0.06	79.67±0.06	72.36±0.06	42.22±0.07	58.65±0.06	**83.31±0.05**
Rank	6.67	4.92	4.67	4.08	3.17	4.83	7.92	6.67	**1.83**

The minimum time and feature number are in bold.

TABLE VI
CLASSIFY ACCURACY OF DIFFERENT ALGORITHMS ON NAIVE BAYESIAN CLASSIFIER

Dataset	LNRS	FSNSI	REKG	CEIDS	UMIDT	FFSCE	FSFNRS	FSFRS	**Ours**
Card	65.52±0.02	56.83±0.09	51.13±0.06	70.14±0.03	69.76±0.03	65.90±0.03	56.91±0.03	56.82±0.03	**70.65±0.02**
Hill	**48.69±0.03**	47.56±0.07	48.49±0.09	47.01±0.05	48.33±0.06	47.53±0.06	47.84±0.06	47.53±0.08	**48.69±0.07**
Ionosphere	78.36±0.07	44.67±0.23	91.16±0.05	90.90±0.04	88.31±0.05	87.74±0.04	55.81±0.19	90.92±0.07	**92.02±0.03**
Movement	34.72±0.09	54.17±0.07	**70.00±0.06**	64.17±0.08	64.17±0.12	51.11±0.07	16.94±0.05	15.56±0.04	65.00±0.08
MPED	47.62±0.06	32.31±0.05	83.67±0.04	85.15±0.04	83.85±0.04	73.16±0.03	31.85±0.06	77.90±0.07	**88.03±0.04**
Nursery	88.87±0.01	88.90±0.01	88.90±0.01	**88.93±0.01**	88.67±0.01	70.97±0.01	50.05±0.02	88.86±0.01	**88.93±0.01**
Sonar	55.83±0.11	57.64±0.11	64.05±0.12	70.21±0.12	71.12±0.09	71.26±0.14	56.12±0.09	66.79±0.11	**79.33±0.11**
Spambase	53.90±0.01	54.64±0.03	54.51±0.03	46.62±0.05	54.62±0.04	**74.70±0.02**	60.49±0.02	60.27±0.02	65.30±0.00
ULCD	35.51±0.10	81.65±0.13	71.43±0.05	51.18±0.12	68.38±0.09	61.69±0.16	33.31±0.13	25.55±0.11	**84.49±0.09**
Colonok	57.86±0.28	68.33±0.18	69.52±0.24	66.67±0.20	51.90±0.18	68.81±0.15	64.76±0.19	59.29±0.22	**69.80±0.20**
DryBean	88.13±0.01	91.23±0.01	90.27±0.01	89.96±0.01	90.20±0.01	**91.46±0.01**	64.81±0.01	64.84±0.01	91.30±0.00
Derma	51.05±0.07	74.57±0.12	67.24±0.10	**76.26±0.10**	69.65±0.07	66.43±0.06	35.55±0.07	32.53±0.08	72.41±0.06
Average	58.74±0.08	62.73±0.09	70.74±0.07	70.71±0.07	70.68±0.06	69.26±0.07	47.84±0.08	57.32±0.07	**76.33±0.06**
Rank	6.92	4.92	4.58	3.83	4.92	4.33	7.42	6.25	**1.42**

The minimum time and feature number are in bold.

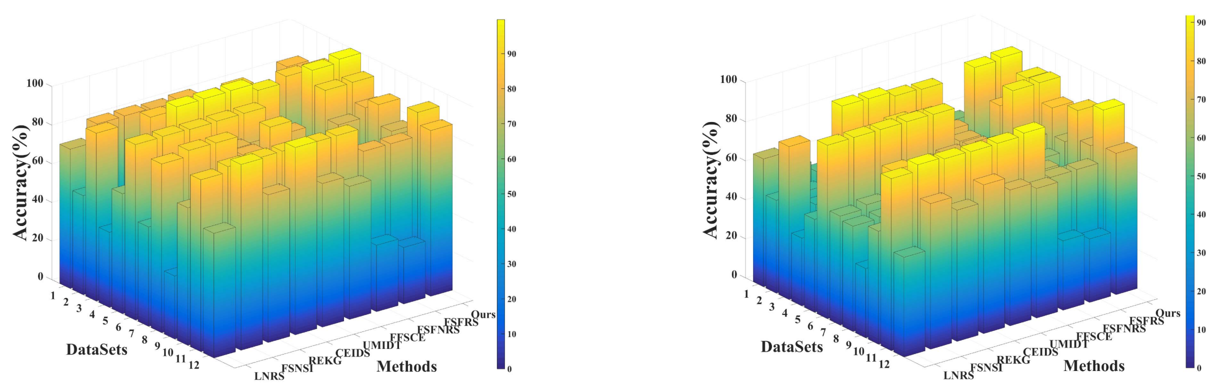


Fig. 5. Accuracy performance of different algorithms on two classifiers. (a) KNN. (b) Naive Bayesian.

better the performance. As shown in Fig. 6, FSZWT ranks first on all metrics and statistically better than the other compared methods in most situations.

D. Parameter Analysis of FSZWT

According to the analysis in Section III, the neighborhood radius δ is an important parameter that influences the construction of the proposed zentropy-based uncertainty measure and further affects the selection process of optimal feature subsets. Therefore, it is necessary to explore the influence of the radius parameter for feature selection on classification performance. The optimal parameter of FSZWT under different datasets is shown in Table XIV, and the classification accuracy of four datasets with different parameters is shown in Fig. 7.

Fig. 7 records the classification accuracy of the proposed method in KNN(C1), NB(C2), and their average accuracy(C3) under different parameters. It can be seen from this figure that in all experimental datasets, the classification accuracy fluctuates significantly with the change in neighborhood parameters.

TABLE VII
CLASSIFY PRECISION OF DIFFERENT ALGORITHMS ON KNN CLASSIFIER

Dataset	LNRS	FSNSI	REKG	CEIDS	UMIDT	FFSCE	FSFNRS	FSFRS	**Ours**
Card	67.08±0.03	79.75±0.05	79.68±0.06	78.35±0.04	80.18±0.04	68.64±0.05	51.59±0.09	43.09±0.08	**81.32±0.03**
Hill	56.73±0.05	51.76±0.07	57.38±0.04	51.51±0.08	52.36±0.06	54.08±0.06	46.18±0.05	52.48±0.07	**59.49±0.06**
Ionosphere	86.43±0.04	22.05±0.18	90.25±0.03	90.00±0.04	89.86±0.05	90.10±0.09	23.93±0.21	89.25±0.05	**90.25±0.06**
Movement	35.29±0.07	75.21±0.11	77.26±0.10	77.82±0.10	77.02±0.08	68.11±0.09	14.57±0.06	16.53±0.07	**84.53±0.05**
MPED	62.80±0.07	25.83±0.03	99.04±0.01	99.04±0.02	99.57±0.01	95.97±0.02	22.52±0.06	98.10±0.02	**99.88±0.00**
Nursery	79.21±0.07	79.24±0.07	78.80±0.07	79.38±0.07	79.43±0.07	28.80±0.01	9.56±0.04	78.66±0.07	**79.99±0.08**
Sonar	52.86±0.09	54.23±0.13	72.03±0.09	80.02±0.13	77.31±0.07	79.94±0.09	55.11±0.10	76.85±0.10	**80.51±0.08**
Spambase	86.72±0.02	90.06±0.02	89.93±0.02	79.09±0.02	**90.18±0.02**	74.44±0.02	53.57±0.10	54.17±0.11	85.80±0.00
ULCD	29.06±0.11	**74.89±0.16**	55.07±0.17	38.54±0.16	53.43±0.12	41.04±0.12	30.94±0.15	20.82±0.13	74.84±0.08
Colonok	64.00±0.21	54.02±0.22	53.92±0.28	66.00±0.22	**67.33±0.30**	66.25±0.17	48.33±0.21	42.75±0.23	62.90±0.30
DryBean	89.16±0.01	92.29±0.01	92.31±0.01	92.32±0.01	92.19±0.01	92.30±0.01	78.64±0.01	78.61±0.01	**92.40±0.00**
Derma	62.25±0.10	95.65±0.04	76.28±0.10	**97.01±0.02**	74.74±0.09	70.85±0.10	26.56±0.10	9.81±0.03	83.80±0.06
Average	63.88±0.07	65.78±0.09	76.34±0.08	77.49±0.07	77.94±0.08	69.43±0.07	38.98±0.10	55.50±0.08	**81.31±0.06**
Rank	6.58	5.17	4.25	3.67	3.33	4.92	7.92	7.17	**2.00**

The excellent classification results are also in bold.

TABLE VIII
CLASSIFY PRECISION OF DIFFERENT ALGORITHMS ON NAIVE BAYESIAN CLASSIFIER

Dataset	LNRS	FSNSI	REKG	CEIDS	UMIDT	FFSCE	FSFNRS	FSFRS	**Ours**
Card	64.05±0.13	57.79±0.08	51.72±0.04	66.13±0.06	65.14±0.06	66.09±0.11	52.90±0.15	64.16±0.10	**68.62±0.06**
Hill	48.23±0.05	47.65±0.08	47.96±0.07	47.39±0.05	47.99±0.06	47.80±0.06	47.83±0.07	48.09±0.08	**48.76±0.07**
Ionosphere	76.32±0.07	22.49±0.11	90.11±0.06	90.65±0.05	88.51±0.05	87.26±0.06	27.92±0.09	89.41±0.07	**91.51±0.04**
Movement	27.18±0.07	50.11±0.06	**69.13±0.06**	65.21±0.10	61.94±0.13	52.57±0.06	10.75±0.05	8.41±0.03	63.61±0.10
MPED	50.09±0.07	33.97±0.07	84.63±0.03	86.02±0.05	85.54±0.03	72.06±0.04	33.91±0.07	79.14±0.06	**88.96±0.03**
Nursery	64.38±0.06	64.35±0.05	**64.46±0.06**	64.40±0.06	64.30±0.06	42.58±0.01	30.09±0.03	**64.46±0.06**	64.43±0.06
Sonar	57.95±0.11	59.55±0.12	65.73±0.12	70.38±0.13	71.58±0.08	71.40±0.13	57.46±0.08	66.92±0.11	**80.13±0.10**
Spambase	71.54±0.01	71.72±0.02	71.92±0.01	60.80±0.05	72.01±0.02	**74.48±0.03**	39.45±0.13	39.37±0.09	65.60±0.00
ULCD	31.29±0.14	76.02±0.17	69.68±0.15	47.97±0.15	63.38±0.07	54.69±0.18	29.04±0.12	15.46±0.09	**83.48±0.14**
Colonok	41.31±0.31	70.17±0.18	67.75±0.26	38.52±0.18	36.08±0.17	**74.17±0.17**	32.38±0.10	30.48±0.09	43.20±0.20
DryBean	88.18±0.01	91.40±0.01	90.57±0.01	90.10±0.01	90.31±0.01	**91.78±0.01**	68.01±0.01	68.04±0.01	91.50±0.00
Derma	49.94±0.13	70.22±0.07	65.51±0.12	68.71±0.08	66.12±0.11	67.95±0.12	17.06±0.07	8.77±0.06	**71.76±0.12**
Average	55.82±0.10	59.63±0.08	69.88±0.08	66.41±0.08	67.73±0.07	66.91±0.08	37.26±0.08	48.55±0.07	**71.80±0.08**
Rank	6.58	5.17	4.33	3.92	4.58	4.00	7.83	6.50	**2.08**

The excellent classification results are also in bold.

TABLE IX
CLASSIFY RECALL OF DIFFERENT ALGORITHMS ON KNN CLASSIFIER

Dataset	LNRS	FSNSI	REKG	CEIDS	UMIDT	FFSCE	FSFNRS	FSFRS	**Ours**
Card	68.19±0.03	78.40±0.03	78.94±0.06	77.91±0.04	78.30±0.04	69.79±0.04	41.08±0.02	41.28±0.04	**79.32±0.05**
Hill	56.59±0.05	51.83±0.07	57.55±0.04	51.64±0.08	52.31±0.06	54.16±0.06	46.21±0.05	52.51±0.07	**59.32±0.06**
Ionosphere	83.45±0.07	51.36±0.04	82.90±0.05	81.96±0.06	86.53±0.04	**88.34±0.09**	51.00±0.03	84.50±0.06	83.94±0.10
Movement	34.99±0.06	75.92±0.08	78.15±0.08	77.60±0.10	78.03±0.09	68.37±0.08	13.97±0.06	17.38±0.04	**82.34±0.07**
MPED	61.13±0.07	24.39±0.03	99.13±0.01	98.92±0.02	99.55±0.01	95.66±0.02	22.96±0.06	98.20±0.02	**99.75±0.01**
Nursery	78.86±0.07	79.57±0.08	78.88±0.07	79.32±0.07	**80.14±0.07**	31.76±0.07	21.36±0.06	79.16±0.07	79.58±0.07
Sonar	52.2±0.08	53.84±0.11	71.69±0.08	79.61±0.12	75.66±0.08	79.93±0.08	55.07±0.09	75.60±0.10	**81.35±0.08**
Spambase	87.01±0.02	89.62±0.02	89.58±0.01	79.92±0.02	**89.77±0.02**	75.55±0.02	50.16±0.01	50.16±0.01	84.90±0.00
ULCD	28.18±0.10	73.50±0.18	57.16±0.15	43.20±0.14	50.92±0.12	45.44±0.09	32.12±0.14	23.60±0.15	**74.69±0.08**
Colonok	66.12±0.16	57.70±0.18	58.33±0.21	63.00±0.22	**69.17±0.22**	62.90±0.16	54.67±0.14	51.17±0.19	65.50±0.20
DryBean	89.14±0.01	92.30±0.01	92.30±0.01	92.38±0.01	92.21±0.01	92.34±0.01	78.86±0.01	78.77±0.01	**92.40±0.00**
Derma	59.23±0.08	95.31±0.04	75.03±0.12	**96.94±0.02**	71.05±0.10	62.99±0.09	25.52±0.06	19.95±0.03	83.74±0.07
Average	63.36±0.07	68.18±0.07	76.15±0.08	76.92±0.07	77.12±0.07	69.14±0.07	41.61±0.06	56.44±0.06	**80.57±0.06**
Rank	6.42	5.08	4.42	4.25	3.00	4.83	8.08	6.67	**2.00**

The excellent classification results are also in bold.

TABLE X
CLASSIFY RECALL OF DIFFERENT ALGORITHMS ON NAIVE BAYESIAN CLASSIFIER

Dataset	LNRS	FSNSI	REKG	CEIDS	UMIDT	FFSCE	FSFNRS	FSFRS	**Ours**
Card	50.22±0.03	59.31±0.05	56.82±0.06	59.15±0.05	59.31±0.05	51.22±0.02	43.34±0.04	43.85±0.04	**59.50±0.04**
Hill	48.52±0.04	48.76±0.06	47.88±0.07	47.35±0.05	47.84±0.05	47.82±0.05	48.00±0.06	48.06±0.07	**48.84±0.07**
Ionosphere	76.63±0.08	45.36±0.08	90.86±0.05	90.42±0.05	86.48±0.04	86.53±0.06	47.50±0.08	90.00±0.07	**91.04±0.04**
Movement	33.07±0.08	52.47±0.06	**67.57±0.05**	62.84±0.09	61.03±0.14	49.83±0.09	17.25±0.06	15.40±0.03	63.77±0.07
MPED	47.10±0.06	32.56±0.04	84.13±0.04	85.41±0.04	83.98±0.04	72.08±0.03	31.28±0.05	78.51±0.06	**88.22±0.05**
Nursery	65.50±0.06	65.48±0.05	**65.58±0.06**	**65.58±0.06**	65.39±0.06	43.50±0.01	31.77±0.03	65.50±0.06	**65.58±0.06**
Sonar	57.68±0.11	59.63±0.12	65.11±0.12	70.77±0.14	70.50±0.09	70.76±0.13	57.14±0.08	66.65±0.12	**77.87±0.09**
Spambase	61.75±0.01	62.35±0.03	62.28±0.02	54.90±0.04	62.40±0.02	**75.30±0.03**	50.16±0.01	50.01±0.01	65.70±0.00
ULCD	30.67±0.10	75.65±0.16	68.84±0.11	48.01±0.13	62.13±0.09	56.28±0.16	31.79±0.13	22.16±0.12	**81.79±0.13**
Colonok	61.08±0.18	67.04±0.19	**68.58±0.24**	51.50±0.14	46.92±0.17	65.06±0.20	50.00±0.00	46.25±0.12	56.30±0.10
DryBean	87.86±0.01	91.56±0.01	90.60±0.01	90.06±0.01	90.41±0.01	**91.82±0.01**	63.41±0.01	63.45±0.01	91.60±0.00
Derma	48.99±0.08	70.09±0.10	63.42±0.12	**71.45±0.06**	62.75±0.11	63.69±0.06	25.19±0.08	20.78±0.06	70.82±0.07
Average	55.78±0.07	60.79±0.08	69.26±0.08	66.54±0.07	66.59±0.07	64.50±0.07	41.41±0.05	50.87±0.06	**71.75±0.06**
Rank	6.17	4.67	3.83	3.75	5.25	4.50	7.83	7.08	**1.92**

The excellent classification results are also in bold.

TABLE XI
CLASSIFY $F1$-SCORE OF DIFFERENT ALGORITHMS ON KNN CLASSIFIER

Dataset	LNRS	FSNSI	REKG	CEIDS	UMIDT	FFSCE	FSFNRS	FSFRS	**Ours**
Card	67.58±0.02	79.03±0.03	79.2±0.05	78.08±0.03	79.17±0.03	69.19±0.05	45.44±0.04	42.03±0.06	**80.25±0.03**
Hill	56.66±0.05	51.79±0.07	57.47±0.04	51.57±0.08	52.34±0.06	54.12±0.06	46.19±0.05	52.49±0.07	**59.40±0.06**
Ionosphere	84.88±0.05	28.69±0.15	86.37±0.04	85.76±0.05	88.12±0.04	**89.20±0.09**	29.89±0.15	86.76±0.06	86.91±0.08
Movement	35.04±0.06	75.48±0.09	77.65±0.08	77.69±0.10	77.45±0.08	68.17±0.08	14.14±0.06	16.53±0.06	**83.36±0.05**
MPED	61.94±0.07	25.06±0.03	99.09±0.01	98.98±0.02	99.56±0.01	95.82±0.02	22.69±0.05	98.15±0.02	**99.82±0.00**
Nursery	79.01±0.07	79.39±0.07	78.83±0.07	79.32±0.07	**79.78±0.07**	29.99±0.03	12.57±0.03	78.90±0.07	79.77±0.07
Sonar	52.51±0.09	54.00±0.12	71.85±0.08	79.79±0.12	76.46±0.07	79.91±0.08	55.08±0.09	76.19±0.10	**80.92±0.08**
Spambase	86.86±0.02	89.84±0.02	89.75±0.01	79.50±0.02	**89.97±0.02**	74.99±0.02	51.46±0.05	51.69±0.05	85.40±0.00
ULCD	27.95±0.10	74.11±0.17	55.99±0.16	40.15±0.15	51.84±0.12	42.80±0.10	31.35±0.14	21.44±0.14	**74.60±0.07**
Colonok	64.45±0.18	54.99±0.19	54.14±0.26	64.39±0.22	**67.32±0.26**	64.35±0.16	49.90±0.18	45.49±0.21	63.20±0.20
DryBean	89.15±0.01	92.29±0.01	92.30±0.01	92.35±0.01	92.20±0.01	92.32±0.01	78.75±0.01	78.69±0.01	**92.40±0.00**
Derma	60.65±0.09	95.47±0.04	75.55±0.10	**96.97±0.02**	72.80±0.09	66.59±0.09	25.74±0.07	13.08±0.03	83.72±0.07
Average	63.48±0.07	66.21±0.08	76.02±0.08	77.11±0.07	77.40±0.07	69.17±0.07	39.12±0.08	55.54±0.07	**80.81±0.06**
Rank	6.42	5.25	4.58	4.00	3.08	4.75	7.92	6.92	**2.08**

The excellent classification results are also in bold.

TABLE XII
CLASSIFY $F1$-SCORE OF DIFFERENT ALGORITHMS ON NAIVE BAYESIAN CLASSIFIER CLASSIFIER

Dataset	LNRS	FSNSI	REKG	CEIDS	UMIDT	FFSCE	FSFNRS	FSFRS	**Ours**
Card	55.97±0.07	58.37±0.06	54.14±0.05	62.76±0.05	62.02±0.05	57.41±0.05	47.17±0.08	51.88±0.05	**63.69±0.05**
Hill	48.37±0.04	48.13±0.07	47.92±0.07	47.37±0.05	47.92±0.05	47.81±0.05	47.91±0.06	48.07±0.07	**48.80±0.07**
Ionosphere	76.43±0.07	29.17±0.12	90.46±0.05	90.50±0.04	87.46±0.04	86.84±0.05	34.69±0.11	89.69±0.07	**91.26±0.04**
Movement	29.67±0.08	51.14±0.05	**68.30±0.06**	63.88±0.09	61.40±0.13	51.03±0.07	12.84±0.06	10.54±0.03	63.61±0.09
MPED	48.49±0.06	33.10±0.05	84.37±0.04	85.70±0.04	84.74±0.03	72.06±0.04	32.40±0.05	78.82±0.06	**88.58±0.04**
Nursery	64.94±0.06	64.91±0.05	**65.02±0.06**	64.99±0.06	64.84±0.06	43.04±0.01	30.90±0.03	64.97±0.06	65.00±0.06
Sonar	57.80±0.11	59.58±0.12	65.40±0.12	70.56±0.13	71.02±0.09	71.07±0.13	57.30±0.08	66.78±0.11	**78.97±0.09**
Spambase	66.28±0.01	66.70±0.02	66.74±0.01	57.65±0.05	66.85±0.02	**74.89±0.03**	43.25±0.08	43.58±0.06	65.70±0.00
ULCD	30.36±0.11	75.73±0.16	69.15±0.12	47.81±0.13	62.57±0.08	55.31±0.17	30.08±0.12	17.61±0.10	**82.60±0.14**
Colonok	46.56±0.28	68.36±0.18	67.39±0.25	43.26±0.16	40.20±0.17	**68.70±0.19**	38.55±0.07	35.99±0.10	47.80±0.20
DryBean	88.02±0.01	91.48±0.01	90.58±0.01	90.08±0.01	90.36±0.01	**91.80±0.01**	65.63±0.01	65.67±0.01	91.50±0.00
Derma	49.03±0.10	69.98±0.08	64.24±0.11	69.96±0.07	64.34±0.11	65.58±0.08	19.82±0.07	11.95±0.07	**70.97±0.09**
Average	55.14±0.09	59.70±0.08	69.43±0.08	66.26±0.07	66.97±0.07	65.47±0.07	38.39±0.07	48.78±0.07	**71.54±0.07**
Rank	6.08	4.83	4.17	4.00	4.67	4.33	8.08	6.83	**2.00**

The excellent classification results are also in bold.

TABLE XIII
STATISTICAL TEST OF NINE ALGORITHMS UNDER KNN AND NB CLASSIFIERS

Classifier	Performance	LNRS	FSNSI	REKG	CEIDS	UMIDT	FFSCE	FSFNRS	FSFRS	**Ours**	Chi-sq	Prob>Chi-sq
KNN	Accuracy	6.67	4.92	4.67	4.08	3.17	4.83	7.92	6.67	**1.83**	44.93	3.80×10^{-7}
	Precision	6.58	5.17	4.25	3.67	3.33	4.92	7.92	7.17	**2.00**	48.24	8.89×10^{-8}
	Recall	6.42	5.08	4.50	4.33	3.08	4.83	8.08	6.67	**2.00**	44.27	5.05×10^{-7}
	$F1$-score	6.42	5.25	4.58	4.00	3.08	4.75	7.92	6.92	**2.08**	44.27	5.07×10^{-7}
NB	Accuracy	6.92	4.92	4.58	3.83	4.92	4.33	7.42	6.25	**1.42**	40.99	2.10×10^{-6}
	Precision	6.58	5.17	4.33	3.92	4.58	4.00	7.83	6.50	**2.08**	38.32	6.57×10^{-6}
	Recall	6.17	4.67	3.75	3.83	5.25	4.50	7.83	7.08	**1.92**	43.41	7.35×10^{-7}
	$F1$-score	6.08	4.83	4.17	4.00	4.67	4.33	8.08	6.83	**2.00**	40.92	2.16×10^{-6}

The minimum average ranking is in bold.

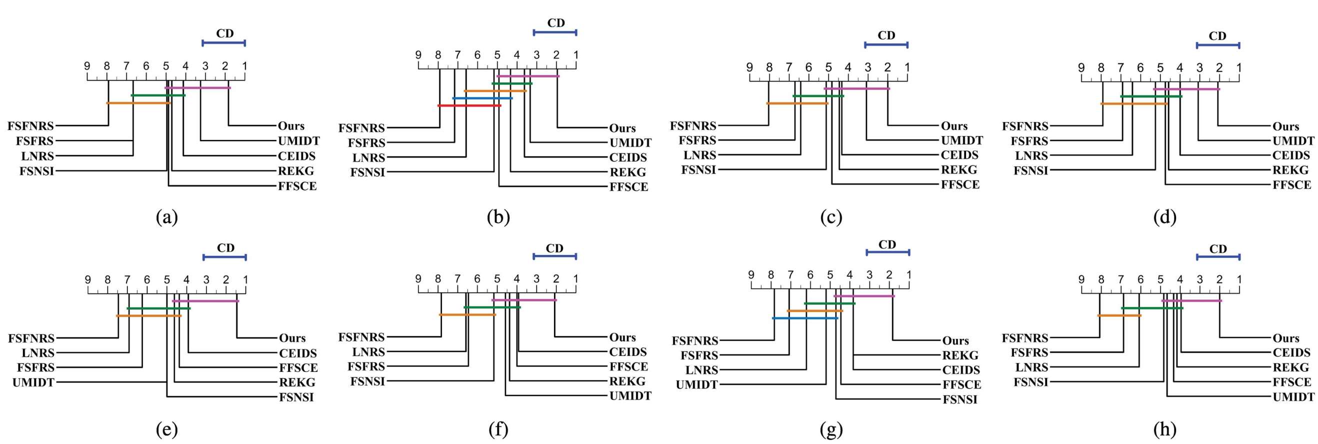


Fig. 6. Accuracy performance with nine algorithms on two classifiers. (a)–(d) are the results on KNN classifier, (e)–(h) are the results on NB classifier. (a) Accuracy_1. (b) Precision_1. (c) Recall_1. (d) $F1\text{-score}_1$. (e) Accuracy_2. (f) Precision_2. (g) Recall_2. (h) $F1\text{-score}_2$.

TABLE XIV
OPTIMAL NEIGHBORHOOD PARAMETERS OF FSZWT METHOD

Dataset	Card	Hill	Ionosphere	Movement	MPED	Nursery
δ	0.05	0.05	0.50	0.45	0.50	0.15
Dataset	Sonar	Spambase	ULCD	Colonok	DryBean	Derma
δ	0.30	0.30	0.20	0.40	0.20	0.40

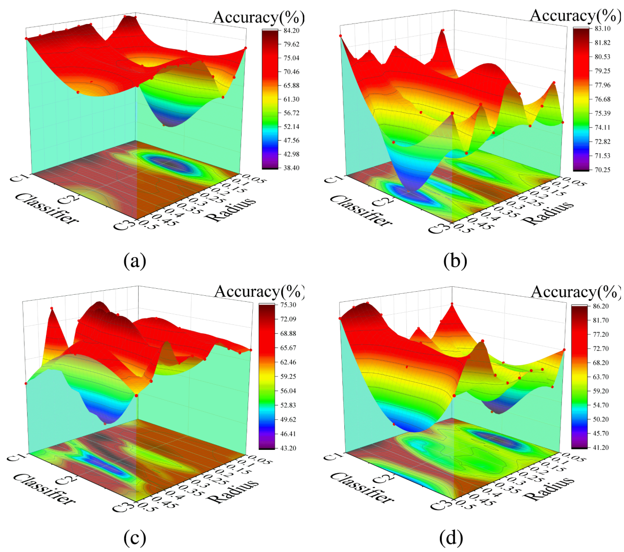


Fig. 7. Parameter analysis of FSZWT on selected four datasets. (a) Card. (b) Sonar. (c) Colonok. (d) Derma.

Therefore, the parameter δ is sensitive to the result of feature selection.

As mentioned above, the radius is an important parameter affecting the model building and feature selection process. Meanwhile, classification performance is susceptible to the change of feature subset. Thus, the value of the radius parameter could be adjusted to further adapt to the distribution structure of features to obtain better classification performance.

According to the abovementioned analysis, we find that the proposed zentropy-based uncertainty measure is stable to noise data due to the interaction of changes between different granular levels, as shown in Example 2 and Section V-B. Moreover, it is efficient for feature selection by considering the information presented at multiple levels in the neighborhood approximation process, which is verified by the excellent classification performance in Section V-C.

VI. CONCLUSION

This article presents a novel zentropy-based uncertainty measure to design a feature selection algorithm by analyzing the granular level structure in knowledge space. By exploiting the granular level in decision information systems, the zentropy-based uncertainty measure is first established to depict uncertain knowledge from whole and internal and further applied to select optimal features. All the designed experimental results show that the proposed zentropy-based uncertainty measure is more stable in noise data and can choose a feasible feature subset with excellent classification performance. Compared with some existing methods, the proposed zentropy-based method could comprehensively depict uncertain knowledge combining granular level structure, providing a novel viewpoint to uncertainty information processing. Nevertheless, the proposed measure considers the information at multiple granular levels, which increases the computational burden to a specific extent, especially in a dynamic environment. Therefore, an efficient strategy for computing the proposed uncertainty measure needs to be developed in further work.

ACKNOWLEDGMENT

The authors would like to thank the Associate Editor and the reviewers for their insightful comments and suggestions.

REFERENCES

[1] J. H. Dai, W. T. Wang, and Q. Xu, "An uncertainty measure for incomplete decision tables and its applications," *IEEE Trans. Cybern.*, vol. 43, no. 4, pp. 1277–1289, Aug. 2013.

[2] J. H. Dai et al., "Conditional entropy for incomplete decision systems and its application in data mining," *Int. J. Gen. Syst.*, vol. 41, no. 7, pp. 713–728, Oct. 2012.

[3] C. Gao et al., "Maximum decision entropy-based attribute reduction in decision-theoretic rough set model," *Knowl.-Based Syst.*, vol. 143, pp. 179–191, Dec. 2018.

[4] I. D. Gediga, "Rough approximation quality revisited," *Artif. Intell.*, vol. 132, no. 2, pp. 219–234, Nov. 2001.

[5] D. D. Guo et al., "A novel outcome evaluation model of three-way decision: A change viewpoint," *Inf. Sci.*, vol. 607, pp. 1089–1110, Aug. 2022.

[6] D. D. Guo and W. H. Xu, "Fuzzy-based concept-cognitive learning: An investigation of novel approach to tumor diagnosis analysis," *Inf. Sci.*, vol. 639, Aug. 2023, Art. no. 118998.

[7] D. D. Guo, W. Xu, Y. Qian, and W. Ding, "Fuzzy-granular concept-cognitive learning via three-way decision: Performance evaluation on dynamic knowledge discovery," *IEEE Trans. Fuzzy Syst.*, 2023, to be published, doi: 10.1109/TFUZZ.2023.3325952.

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

[9] D. D. Guo, C. M. Jiang, and P. Wu, "Three-way decision based on confidence level change in rough set," *Int. J. Approx. Reasoning*, vol. 143, pp. 57–77, Apr. 2022.

[10] Q. H. Hu et al., "Information entropy for ordinal classification," *Sci. China Inf. Sci.*, vol. 53, pp. 1188–1200, 2010.

[11] B. M. Lake, R. Salakhutdinov, and J. B. Tenenbaum, "Human-level concept learning through probabilistic program induction," *Science*, vol. 350, pp. 1332–1338, Dec. 2015.

[12] W. T. Li, H. Zhou, W. Xu, X.-Z. Wang, and W. Pedrycz, "Interval dominance-based feature selection for interval-valued ordered data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 10, pp. 6898–6912, Oct. 2023.

[13] J. Y. Liang et al., "Information entropy, rough entropy and knowledge granulation in incomplete information systems," *Int. J. Gen. Syst.*, vol. 35, no. 6, pp. 641–654, Dec. 2006.

[14] K. P. Liu, Y. Fu, L. Wu, X. Li, C. Aggarwal, and H. Xiong, "Automated feature selection: A reinforcement learning perspective," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 3, pp. 2272–2284, Mar. 2023.

[15] Z. K. Liu, Y. Wang, and S. L. Shang, "Zentropy theory for positive and negative thermal expansion," *J. Phase Equilibria Diffusion*, vol. 43, pp. 598–605, Feb. 2022.

[16] C. Luo, S. Wang, T. Li, H. Chen, J. Lv, and Z. Yi, "Spark rough hypercuboid approach for scalable feature selection," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 3, pp. 3130–3144, Mar. 2023.

[17] S. Luo et al., "A neighborhood rough set model with nominal metric embedding," *Inf. Sci.*, vol. 520, pp. 373–388, May 2020.

[18] S. Pal, S. Ghosh, H. Biswas, and M. Patwari, "Discretization using combination of heuristics for high accuracy with huge noise reduction," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 4, pp. 1710–1722, Apr. 2022.

[19] Z. Pawlak, "Rough sets," *Int. J. Comput. Inf. Sci.*, vol. 11, pp. 341–356, Jun. 1982.

[20] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min redundancy,"

IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[21] B. B. Sang, H. Chen, L. Yang, T. Li, and W. Xu, "Incremental feature selection using a conditional entropy based on fuzzy dominance neighborhood rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 6, pp. 1683–1697, Jun. 2022.

[22] B. B. Sang, H. Chen, L. Yang, J. Wan, T. Li, and W. Xu, "Feature selection considering multiple correlations based on soft fuzzy dominance rough sets for monotonic classification," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 12, pp. 5181–5195, Dec. 2022.

[23] C. Shannon, "The mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3/4, pp. 379–423, Jul. 1948.

[24] L. Sun, T. Yin, W. Ding, Y. Qian, and J. Xu, "Feature selection with missing labels using multilabel fuzzy neighborhood rough sets and maximum relevance minimum redundancy," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 5, pp. 1197–1211, May 2022.

[25] C. Z. Wang, Y. Huang, M. Shao, Q. Hu, and D. Chen, "Feature selection based on neighborhood self-information," *IEEE Trans. Cybern.*, vol. 50, no. 9, pp. 4031–4042, Sep. 2020.

[26] C. Wang et al., "A fitting model for feature selection with fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 4, pp. 741–753, Aug. 2017.

[27] Q. Wang et al., "Local neighborhood rough set," *Knowl.-Based Syst.*, vol. 153, pp. 53–64, Aug. 2018.

[28] C. Z. Wang et al., "Feature subset selection based on fuzzy neighborhood rough sets," *Knowl.-Based Syst.*, vol. 111, pp. 173–179, Nov. 2016.

[29] W. H. Xu, D. Guo, J. Mi, Y. Qian, K. Zheng, and W. Ding, "Two-way concept-cognitive learning via concept movement viewpoint," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 10, pp. 6798–6812, Oct. 2023.

[30] W. H. Xu, D. Guo, Y. Qian, and W. Ding, "Two-way concept-cognitive learning method: A fuzzy-based progressive learning," *IEEE Trans. Fuzzy Syst.*, vol. 31, no. 6, pp. 1885–1899, Jun. 2023.

[31] W. H. Xu, K. Yuan, W. Li, and W. Ding, "An emerging fuzzy feature selection method using composite entropy-based uncertainty measure and data distribution," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 7, no. 1, pp. 76–88, Feb. 2023.

[32] Y. Y. Yao, "The geometry of three-way decision," *Appl. Intell.*, vol. 51, pp. 6298–6325, Sep. 2021.

[33] Y. Y. Yao, "Three-way conflict analysis: Reformulations and extensions of the Pawlak model," *Knowl.-Based Syst.*, vol. 180, pp. 26–37, Sep. 2019.

[34] Y. Y. Yao and L. Q. Zhao, "A measurement theory view on the granularity of partitions," *Inf. Sci.*, vol. 213, pp. 1–13, Dec. 2012.

[35] Z. Yuan, H. Chen, P. Zhang, J. Wan, and T. Li, "A novel unsupervised approach to heterogeneous feature selection based on fuzzy mutual information," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 9, pp. 3395–3409, Sep. 2022.

[36] K. H. Yuan et al., "An incremental learning mechanism for object classification based on progressive fuzzy three-way concept," *Inf. Sci.*, vol. 584, pp. 127–147, Jan. 2022.

[37] P. F. Zhang, T. Li, Z. Yuan, C. Luo, K. Liu, and X. Yang, "Heterogeneous feature selection based on neighborhood combination entropy," *IEEE Trans. Neural Netw. Learn. Syst.*, 2022, to be published, doi: 10.1109/TNNLS.2022.3193929.

[38] X. Zhang, C. Mei, D. Chen, Y. Yang, and J. Li, "Active incremental feature selection using a fuzzy-rough-set-based information entropy," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 5, pp. 901–915, May 2020.

[39] X. H. Zhang, J. Wang, J. Zhan, and J. Dai, "Fuzzy measures and Choquet integrals based on fuzzy covering rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 7, pp. 2360–2374, Jul. 2022.

[40] W. Zheng, S. Chen, Z. Fu, F. Zhu, H. Yan, and J. Yang, "Feature selection boosted by unselected features," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 9, pp. 4562–4574, Sep. 2022.

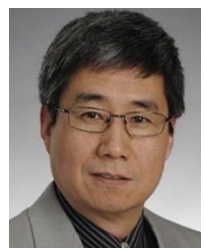
Kehua Yuan received the M.Sc. degree in statistics from the College of Artificial Intelligence, Southwest University, Chongqing, China, in 2022. She is currently working toward the Ph.D. degree with the College of Electronics and Information Engineering, Tongji University, Shanghai, China.

Her current research interests include granular computing, uncertainty analysis, and machine learning.

Duoqian Miao received the Ph.D. degree in pattern recognition and intelligent system from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 1997.

He is currently a Professor with the School of Electronics and Information Engineering, Tongji University, Shanghai, China. He has authored or coauthored more than 180 academic papers, which include prestigious journals and international conferences, and the h-index is 39. His main research interests include rough sets, soft computing, machine learning, and intelligent systems.

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



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

He is currently a Professor of Computer Science with the University of Regina. He proposed a theory of three-way decision, a decision-theoretic rough set model, and a triarchic theory of granular computing. He has authored or coauthored more than 400 papers. He was a highly cited Researcher from 2015 to 2019. His research interests include three-way decision, granular computing, rough sets, formal concept analysis, information retrieval, data mining, and web intelligence.

Dr. Yao is currently an Associate Editor for *Information Sciences* and *International Journal of Approximate Reasoning*, and so on.

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 with 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.

Xuerong Zhao received the Ph.D. degree in computational mathematics from the School of Mathematics and Statistics, Wuhan University, Wuhan, China, in 2015.

She is currently a Postdoctoral Fellow with Tongji University, Shanghai, China. Her major research interests include uncertainty analysis, rough set, three-way decision, and machine learning.