Applied Soft Computing Journal 185 (2025) 113851

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

A multi-granularity decision tree algorithm based on variable precision rough sets and Zentropy

Hui Dong [a], Caihui Liu [a,*], Xiying Chen [a], Duoqian Miao [b]

[a] Department of Mathematics and Computer Science, Gannan Normal University, Ganzhou, 341000 Jiangxi, China
[b] Department of Computer Science and Technology, Tongji University, Shanghai, 201804, China

ARTICLE INFO

Keywords:
Variable precision neighborhood rough sets
Decision tree
Granular computing
Zentropy
Uncertainty measure

ABSTRACT

The existing decision tree algorithms often use a single-layer measure to process data, which cannot fully consider the complex interactions and dependencies between different granularity levels. In addition, decision tree algorithms inevitably face the issue of multi-value preference, which may lead to the selection of unreasonable attributes in the process of partition, thereby affecting the performance of the algorithms. Therefore, this paper proposes an improved decision tree algorithm, called Ze-VNDT, which combines variable precision rough sets with Zentropy. First, to avoid the information loss caused by data discretization, this paper introduces variable precision neighborhood rough sets for data processing. Second, by analyzing the granularity level structure within the variable precision neighborhood rough set model, knowledge uncertainty is analyzed from three granularity levels: decision classes, approximate relations, and similarity classes. We describe the uncertain knowledge from the overall to the internal using the idea of going from coarse to fine, and design a Zentropy to measure uncertainty. To address the issue of multi-value preference, an adaptive weighted Zentropy uncertainty measure is designed based on the definition of uncertainty measure based on Zentropy. Third, when constructing the improved decision tree algorithm, the optimal attributes are selected based on the designed uncertainty measure. Finally, numerical experiments on 18 UCI datasets validated the effectiveness and rationality of the proposed algorithm. The experimental results showed that, compared to traditional algorithms and the latest improved algorithms, the proposed algorithm achieved an average accuracy of 94.79%, an average precision of 85.77%, an average recall rate of 84.68%, and an F1-score of 84.97% across the 18 datasets. It ranked first in all five evaluation metrics, demonstrating higher stability and accuracy.

1. Introduction

Classification problems are crucial in daily life and scientific research. Decision tree algorithms have become a key tool in the field of artificial intelligence for solving various classification problems due to their precise classification effects and transparent processes [1]. Decision tree algorithms are based on tree structures, and their core idea is to form different subsets through different divisions of the features of a dataset, and then obtain the final classification result through the classification of these subsets [2]. The intuitive and easy-to-understand classification process of decision tree algorithms [3,4], as well as their ability to efficiently process large amounts of data, make them highly adaptable in practical applications [5,6]. The classification process of decision trees is recursive, starting from the root node, dividing each sample in the dataset, and assigning samples to different child nodes based on test results, until the leaf nodes, which are the final classification results of the dataset. Decision trees play an indispensable role in data analysis with their intuitive structure and efficient processing capabilities [7]. At the same time, their excellent explanatory characteristics also provide an important guarantee for their application. In the process of constructing decision trees, the choice of decision rules is crucial, as it not only determines the way data is divided in the trees but also affects the decisions made by each node based on input feature values. An excellent decision rule can improve the classification accuracy of decision trees, while incorrect decision rules may lead to complex and redundant decision tree structures, thus affecting their predictive performance. Therefore, in-depth research and optimization of decision rules are of extremely significance for the learning and application of decision trees.

* Corresponding author.
E-mail addresses: 1799762217@qq.com (H. Dong), liucaihui@gnnu.edu.cn (C. Liu), xiying_chen@163.com (X. Chen), dqmiao@tongji.edu.cn (D. Miao).

https://doi.org/10.1016/j.asoc.2025.113851
Received 5 November 2024; Received in revised form 16 June 2025; Accepted 27 August 2025
Available online 11 September 2025

Classic decision tree algorithms include the ID3 algorithm based on information gain [8], the C4.5 algorithm based on information gain ratio [9], and the CART algorithm based on the Gini index [10]. Based on the study of classical algorithms, the academic community is developing new types of decision tree algorithms that integrate interdisciplinary theories to achieve efficient and precise decision support, reflecting the pursuit of algorithm optimization and contributions to the development of the decision science field. For example, Laber et al. [11] proposed a shallow decision tree algorithm that uses a loss function with a penalty term. Klaus et al. [12] proposed a gradient-based model tree algorithm, which significantly improves classification accuracy while maintaining a high degree of interpretability. Deng et al. [13] incorporated singular value decomposition into the calculation of entropy and proposed an adaptive entropy decision tree algorithm. Alast and Keyvanpour [14] proposed a new fuzzy decision tree improvement of the ID3 algorithm using a new entropy formula. When processing continuous data, the aforementioned decision tree algorithms often rely on discretization strategies. However, this discretization process may lead to information loss, which in turn can adversely affect classification results. In light of this, researchers have turned to seeking other models that can directly process continuous data to ensure the efficiency and accuracy of the data processing process.

The classical rough set theory was proposed by Pawlak [15] in 1982 as a mathematical tool for dealing with incompleteness and uncertainty. Its core idea is to use the knowledge in the known knowledge base to approximately describe the imprecise or uncertain knowledge. The classical rough set theory does not require any prior information beyond the data set to be processed, thus the description or treatment of uncertainty is more objective [16,17]. The variable precision neighborhood rough set theory is an extension of the classical rough set theory. By setting a threshold parameter β, it relaxes the strict definition of the approximation boundary in the classical rough set theory and allows for probabilistic classification. This method improves the concept of approximation space, facilitates the discovery of relevant information from data that is considered irrelevant, and to some extent resists noise interference and enhances tolerance for differences between objects in continuous data processing [18,19]. Many scholars have combined variable precision neighborhood rough sets with decision trees to improve the performance of decision tree algorithms. For example, Xie et al. [20] constructed a neighborhood Gini index based on variable precision neighborhood equivalence granules, and used it as a measure for decision tree attribute partitioning. Liu et al. [21] proposed two extended measures, called variable precision neighborhood information gain and variable precision neighborhood Gini index, based on variable precision neighborhood rough sets. They improved the decision tree algorithm by combining these two measures with the variable precision neighborhood dependency. Liu et al. [22] adopted a variable precision neighborhood rough set model to define a boundary domain coefficient, and used it as a measure for decision tree attribute partitioning to improve the decision tree algorithm. However, current methods mostly focus on a single granularity level such as similarity relations or similarity classes, failing to fully consider the interaction between granules. This limitation significantly affects the stability and accuracy of the related methods.

In addition, the current algorithms combining variable precision neighborhood rough sets with decision trees lack research on the multi-value preference problem in decision trees. Multi-value preference is a common problem in decision trees. It refers to a phenomenon in the decision-making process where there is a preference for attributes with a larger number of values. For example, in decision tree algorithms, when facing multiple attributes, if there is an attribute with a particularly large number of values, the algorithm may incorrectly assume that this attribute is more important and thus prioritize splitting based on this attribute, even though it is actually not very important for the classification task, such as ID number, student number, employee number, unique identification number, etc. This preference may lead to the induction of incorrect knowledge or models from the data set. Therefore, developing a new decision tree measure for characterizing uncertain knowledge and solving the multi-value preference problem is very important.

Zentropy is a systematic concept [23] that originates from the concept of entropy in physics [24]. Zentropy can describe the chaos of a system at multiple scales, where a larger granularity level reflects the whole of a lower granularity level [25], and a lower granularity level is a refinement of a larger granularity level [26]. Therefore, it can accurately characterize the entropy change of system variations at different granularity levels [27,28]. This concept is similar to the granular structure of rough approximations [29] and has been successfully applied to feature selection [23,30]. Inspired by the above observations, this paper delves into the expression of uncertainty at different granularity levels in variable precision neighborhood rough sets, and applies the concept of Zentropy to the attribute partitioning measurement method in decision trees. To address the problem of multi-value preference, an adaptive weighted Zentropy uncertainty measure is designed based on the definition of uncertainty measure using Zentropy.

In summary, this paper proposes an improved decision tree algorithm (called Ze-VNDT) that combines variable precision rough sets with Zentropy. It performs uncertainty analysis at three granularity levels: decision classes, approximate relations, and similarity classes. It designs a comprehensive measure of uncertainty by integrating the uncertainty at these three granularity levels, aiming to accurately describe the granular structure characteristics from micro granules to macro wholes. Moreover, an adaptive weight is designed for the measure to address the problem of multi-value preferences in decision trees, thereby constructing the decision tree algorithm. The core contributions of this paper can be summarized as follows.

(1) An uncertainty measure based on Zentropy theory is proposed, which analyzes uncertainty from three granularity levels: decision classes, approximate relations, and similarity classes, to design a measurement function that addresses the limitations of a single granularity level.

(2) An adaptive weight is designed based on the Zentropy uncertainty measure, which increases the importance of attributes and solves the problem of multi-value preferences in decision trees.

(3) A decision tree improvement algorithm is constructed, which uses the proposed adaptive weighted Zentropy uncertainty measure as the attribute division measurement for decision trees, i.e., uses it to select the optimal attributes for attribute division.

(4) Experiments were conducted on 18 public datasets, and the proposed algorithm was evaluated using accuracy, precision, recall, F1 score, and the number of leaves. Compared with the other seven algorithms, the proposed algorithm demonstrated superior performance advantages.

The rest of this paper is organized as follows. Section 2 briefly reviews the relevant theoretical knowledge. Section 3 analyzes uncertainty at three granularity levels and constructs an improved decision tree algorithm by using the adaptive weighted Zentropy uncertainty measure. Section 4 analyzes the experimental results on public datasets. Finally, Section 5 summarizes this paper.

2. Preliminaries

2.1. Neighborhood rough set model

The neighborhood rough set theory is an extension of the classical rough set theory, which integrates the concepts of neighborhood granulation and spatial metrics, combining the information granules in continuous data space with the covering relations in rough set theory, achieving a unified treatment of both discrete and continuous data [31].

A Decision Information System $(DIS) =< U, A, V, I >$ is an organic whole, consisting of the following main components:

(1)Object Set (U): $U = \{x_1, x_2, x_3, \ldots, x_n\}$. The set U contains all the entities or objects involved in the decision problem.

(2)Attribute Set (A): $A = C \cup D$, where C represents the set of condition attributes, which includes all the relevant attributes that affect the decision-making outcomes. D represents the set of decision attributes, which contains the results or objectives of the decision.

(3)Attribute Value Domain (V): $V = \cup_{a \in A} V_a$, where V defines the range or set of all possible values that attributes can take. For each attribute $a \in C \cup D$, there is a corresponding value domain $V_a \subseteq V$, which represents the set of all possible values that attribute a can take.

(4)Information Function (I): $I : U \times (C \cup D) \to V$ is a mapping function such that for each $x \in U$ and each $a \in C \cup D$, $I(x, a)$ returns the value of object x on attribute a, which is taken from the value domain V_a.

For any $x_i, x_j, x_z \in U$ and any $B \subseteq C$, we use Δ^B to measure the difference between these objects, which satisfies the following conditions:

(1)Non-negativity: $\Delta^B(x_i, x_j) \geqslant 0$, and $\Delta^B(x_i, x_i)=0$;

(2)Symmetry: $\Delta^B(x_i, x_j) = \Delta^B(x_j, x_i)$;

(3)Triangle inequality: $\Delta^B(x_i, x_z) \leqslant \Delta^B(x_i, x_j) + \Delta^B(x_j, x_z)$.

Definition 1 ([32]). Let $DIS =< U, A, V, I >$ be a decision information system, where $U = \{x_1, x_2, x_3, \ldots, x_n\}$, $A = C \cup D$, $C = \{a_1, a_2, a_3, \ldots, a_m\}$, $U/D = \{D_1, D_2, D_3, \ldots, D_s\}$ represents the set of decision classes, $V = \cup_{a \in A} V_a$ denotes the set of attribute values, and $I : U \times (C \cup D) \to V$ represents an information function. For any $x_i, x_j \in U$ and any $B \subseteq C$, the Minkowski distance between samples x_i and x_j under B is defined as follows:

$$\Delta^B(x_i, x_j) = \left(\sum_{g \in B} \left| x_i^g - x_j^g \right|^p \right)^{\frac{1}{p}}. \tag{1}$$

where x_i^g and x_j^g represent the attribute values of samples x_i and x_j on attribute g, and Δ^B represents the distance measure. To simplify the calculation of the distance between samples, in this paper, we set $p = 1$. In this case, Δ^B is also known as the Manhattan distance.

For any $x_i \in U$ and any $B \subseteq C$, the neighborhood granule of sample x_i under B, denoted as $\delta_B(x_i)$, is defined as:

$$\delta_B(x_i) = \{x_j \in U \mid \Delta^B(x_i, x_j) \leq \delta\}. \tag{2}$$

It can be seen that when $\delta = 0$, the neighborhood rough set model will degenerate into the classical rough set model.

Definition 2 ([32]). Let $DIS =< U, A, V, I >$ be a decision information system, where $U = \{x_1, x_2, x_3, \ldots, x_n\}$, $A = C \cup D$, $C = \{a_1, a_2, a_3, \ldots, a_m\}$, $U/D = \{D_1, D_2, D_3, \ldots, D_s\}$ represents the set of decision classes, $V = \cup_{a \in A} V_a$ denotes the set of attribute values, and $I : U \times (C \cup D) \to V$ represents an information function. For any $X \subseteq U$ and any $B \subseteq C$, the upper and lower approximations of the neighborhood of X under B are defined as follows:

$$\begin{aligned} \overline{NR_B}(X) &= \{x_i \in U | \delta_B(x_i) \cap X \neq \emptyset\}, \\ \underline{NR_B}(X) &= \{x_i \in U | \delta_B(x_i) \subseteq X\}. \end{aligned} \tag{3}$$

where $\delta_B(x_i)$ represents the neighborhood granule of sample x_i under B.

Let $U/D = \{D_1, D_2, D_3, \ldots, D_s\}$ be the set of decision classes. For any $B \subseteq C$, the upper and lower approximations of B with respect to D are defined as follows:

$$\begin{aligned} \overline{NR_B}(D) &= \{\overline{NR_B}(D_1), \overline{NR_B}(D_2), \ldots, \overline{NR_B}(D_s)\}, \\ \underline{NR_B}(D) &= \{\underline{NR_B}(D_1), \underline{NR_B}(D_2), \ldots, \underline{NR_B}(D_s)\}. \end{aligned} \tag{4}$$

For any $B \subseteq C$, the boundary region of B with respect to D is defined as follows:

$$NB_B(D) = \bigcup_{q=1}^{s} \overline{NR_B}(D_q) - \bigcup_{q=1}^{s} \underline{NR_B}(D_q). \tag{5}$$

For any $B \subseteq C$, the dependency degree of B with respect to D is defined as follows:

$$\gamma_B(D) = \frac{\left| \bigcup_{q=1}^{s} \underline{NR_B}(D_q) \right|}{|U|}. \tag{6}$$

For any $B \subseteq C$ and any $D_q \in U/D$, the rough accuracy of B with respect to D_q is defined as follows:

$$\alpha_B(D_q) = \frac{\left| \underline{NR_B}(D_q) \right|}{\left| \overline{NR_B}(D_q) \right|}. \tag{7}$$

2.2. Neighborhood similarity

Similarity measures are constructed by integrating neighborhood algebraic similarity and neighborhood geometric similarity. This method not only considers the algebraic relationships among objects within the neighborhood system but also their geometric relationships, thereby demonstrating higher flexibility and accuracy in dealing with potential contradictions in the transitivity of equivalence relations.

Definition 3 ([33]). For any $x_i, x_j \in U$ and any $B \subseteq C$, let $\Delta^B(x_i, x_j)$ be the Minkowski distance between samples x_i and x_j under B, and let δ be the neighborhood radius. The neighborhood Jaccard geometric similarity between x_i and x_j under B is defined as:

$$NRJGS_B^{\delta}(x_i, x_j) = \left(1 - \frac{\Delta^B(x_i, x_j)}{2\delta} \right) \times \frac{|\delta_B(x_i) \cap \delta_B(x_j)|}{|\delta_B(x_i) \cup \delta_B(x_j)|}. \tag{8}$$

where $\delta_B(x_i)$ and $\delta_B(x_j)$ respectively represent the neighborhood granules of samples x_i and x_j under B.

Considering the geometric structure of the neighborhood system, in order to avoid contradictions in the transitivity of neighborhood equivalence relations, this paper adopts the $NRJGS_B^{\delta}$ neighborhood similarity.

2.3. Variable precision neighborhood rough sets

Based on the neighborhood similarity proposed in the previous section, we can further construct a variable precision neighborhood rough set model. This model upgrades the traditional neighborhood equivalence relation to a variable precision neighborhood equivalence relation induced by neighborhood similarity. This transformation aims to address the limitations that strict neighborhood equivalence relations may bring in certain situations. By introducing a variable precision threshold, we can flexibly define and adjust the precision of neighborhood equivalence relations, thereby constructing a variable precision neighborhood equivalence relation. This makes the model more flexible and adaptable when dealing with data. To make a distinction, this paper adopts the variable precision neighborhood decision table $DIS =< U, A, V, I, \delta, \beta >$. It adds the neighborhood radius δ and the variable precision threshold β to the general information system.

Definition 4 ([33]). Let $DIS =< U, A, V, I, \delta, \beta >$ be a variable neighborhood decision table, where $U = \{x_1, x_2, x_3, \ldots, x_n\}$, $A = C \cup D$, $C = \{a_1, a_2, a_3, \ldots, a_m\}$, $U/D = \{D_1, D_2, D_3, \ldots, D_s\}$ represents the set of decision classes, $V = \cup_{a \in A} V_a$ denotes the set of attribute values, $I : U \times (C \cup D) \to V$ represents an information function, $\delta \in [0, 1]$ is the neighborhood radius, and β is the variable precision threshold. For any $B \subseteq C$, the variable precision neighborhood equivalence relation under B is defined as follows:

$$VNER_B^{(\delta,\beta)} = \{(x_i, x_j) \in U \times U | NRJGS_B^{\delta}(x_i, x_j) \geq \beta\}. \quad (9)$$

where $NRJGS_B^{\delta}(x_i, x_j)$ represents the neighborhood similarity between samples x_i and x_j under B.

Definition 5 ([33]). Let $DIS =< U, A, V, I, \delta, \beta >$ be a variable neighborhood decision table, where $U = \{x_1, x_2, x_3, \dots, x_n\}$, $A = C \cup D$, $C = \{a_1, a_2, a_3, \dots, a_m\}$, $U/D = \{D_1, D_2, D_3, \dots, D_s\}$ represents the set of decision classes, $V = \cup_{a \in A} V_a$ denotes the set of attribute values, $I : U \times (C \cup D) \to V$ represents an information function, $\delta \in [0, 1]$ is the neighborhood radius, and β is the variable precision threshold. Let $VNER_B^{(\delta,\beta)}$ be the variable precision neighborhood equivalence relation under B, the variable precision neighborhood equivalence granular structure induced by $VNER_B^{(\delta,\beta)}$ is defined as follows:

$$U/VNER_B^{(\delta,\beta)} = \{[x]_{VNER_B^{(\delta,\beta)}} | x \in U\} = \{V\delta_B(x_1), V\delta_B(x_2), \dots, V\delta_B(x_r)\}. \quad (10)$$

where for any $1 \leq i \leq r$, $V\delta_B(x_i)$ represents the variable precision neighborhood equivalence class.

Definition 6 ([33]). Let $DIS =< U, A, V, I, \delta, \beta >$ be a variable neighborhood decision table, where $U = \{x_1, x_2, x_3, \dots, x_n\}$, $A = C \cup D$, $C = \{a_1, a_2, a_3, \dots, a_m\}$, $U/D = \{D_1, D_2, D_3, \dots, D_s\}$ represents the set of decision classes, $V = \cup_{a \in A} V_a$ denotes the set of attribute values, $I : U \times (C \cup D) \to V$ represents an information function, $\delta \in [0, 1]$ is the neighborhood radius, and β is the variable precision threshold. For any $B \subseteq C$, the variable precision neighborhood lower approximation of B on D is defined as follows:

$$\underline{VNER_B^{(\delta,\beta)}}(D) = \bigcup_{q=1}^{s} \underline{VNER_B^{(\delta,\beta)}}(D_q). \quad (11)$$

where $VNER_B^{(\delta,\beta)}(D_q) = \bigcup\{V\delta_B(x_i) \in U/VNER_B^{(\delta,\beta)} | V\delta_B(x_i) \subseteq D_q\}$, $1 \leq q \leq s$, and $1 \leq i \leq r$. For any $1 \leq i \leq r$, $V\delta_B(x_i)$ represents the variable precision neighborhood equivalence class.

For any $B \subseteq C$, the variable precision neighborhood upper approximation of B on D is defined as follows:

$$\overline{VNER_B^{(\delta,\beta)}}(D) = \bigcup_{q=1}^{s} \overline{VNER_B^{(\delta,\beta)}}(D_q). \quad (12)$$

where $\overline{VNER_B^{(\delta,\beta)}}(D_q) = \bigcup\{V\delta_B(x_i) \in U/VNER_B^{(\delta,\beta)} | V\delta_B(x_i) \cap D_q \neq \emptyset\}$, $1 \leq q \leq s$, and $1 \leq i \leq r$. For any $1 \leq i \leq r$, $V\delta_B(x_i)$ represents the variable precision neighborhood equivalence class.

For any $B \subseteq C$, the variable precision neighborhood boundary region of B with respect to D is defined as follows:

$$BND_B^{(\delta,\beta)}(D) = \bigcup_{q=1}^{s} \overline{VNER_B^{(\delta,\beta)}}(D_q) - \bigcup_{q=1}^{s} \underline{VNER_B^{(\delta,\beta)}}(D_q). \quad (13)$$

For any $B \subseteq C$, the variable precision neighborhood dependency degree of B with respect to D is defined as follows:

$$\gamma_B(D) = \frac{\left|\bigcup_{q=1}^{s} \underline{VNER_B^{(\delta,\beta)}}(D_q)\right|}{|U|}. \quad (14)$$

For any $B \subseteq C$ and any $D_q \in U/D$, the variable precision neighborhood rough accuracy of B with respect to D_q is defined as follows:

$$\alpha_B^{(\delta,\beta)}(D_q) = \frac{\left|\underline{VNER_B^{(\delta,\beta)}}(D_q)\right|}{\left|\overline{VNER_B^{(\delta,\beta)}}(D_q)\right|}. \quad (15)$$

2.4. Entropy-based uncertainty measure

The Shannon entropy, proposed by Shannon [34], provides a measure of uncertainty in the field of probability. For a given random variable, Shannon entropy (also known as information entropy) can be represented as a measure of the uncertainty of the values that the random variable can take. Shannon entropy quantifies the amount of information content or uncertainty of all possible values of a random variable, which is calculated based on the probability distribution of the values that the random variable can take. The higher the value of Shannon entropy, the greater the uncertainty of the random variable; conversely, if the value of Shannon entropy is smaller, it means that the uncertainty of the random variable is lower.

Definition 7 ([34]). Let $M = \{m_1, m_2, \dots, m_r\}$ be a random variable, where $m_1, m_2, \dots, m_r$ are its possible values. For any $1 \leq i \leq r$, let $p(m_i)$ be the probability of m_i, the Shannon entropy is defined as follows:

$$E(M) = -\sum_{i=1}^{r} p(m_i) \log(p(m_i)). \quad (16)$$

The neighborhood entropy [35] mainly focuses on the size of neighborhood similarity classes. In order to measure the uncertainty of continuous data, this model defines neighborhood entropy by introducing neighborhood relations for information granulation.

Definition 8 ([35]). Let $DIS =< U, A, V, I, \delta, \beta >$ be a variable neighborhood decision table, where $U = \{x_1, x_2, x_3, \dots, x_n\}$, $A = C \cup D$, $C = \{a_1, a_2, a_3, \dots, a_m\}$, $U/D = \{D_1, D_2, D_3, \dots, D_s\}$ represents the set of decision classes, $V = \cup_{a \in A} V_a$ denotes the set of attribute values, $I : U \times (C \cup D) \to V$ represents an information function, $\delta \in [0, 1]$ is the neighborhood radius, and β is the variable precision threshold.

For any $B \subseteq C$, the neighborhood entropy of B is defined as follows:

$$NE_B(D) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log \frac{|V\delta_B(x_i)|}{|U|}. \quad (17)$$

where for any $x_i \in U$, $V\delta_B(x_i)$ represents the variable precision neighborhood equivalence class under B that contains sample x_i.

Neighborhood conditional entropy is a concept defined within the framework of neighborhood rough sets, used to measure the uncertainty or complexity of the decision attribute set given a set of condition attributes.

Definition 9 ([35]). Let $DIS =< U, A, V, I, \delta, \beta >$ be a variable neighborhood decision table, where $U = \{x_1, x_2, x_3, \dots, x_n\}$, $A = C \cup D$, $C = \{a_1, a_2, a_3, \dots, a_m\}$, $U/D = \{D_1, D_2, D_3, \dots, D_s\}$ represents the set of decision classes, $V = \cup_{a \in A} V_a$ denotes the set of attribute values, $I : U \times (C \cup D) \to V$ represents an information function, $\delta \in [0, 1]$ is the neighborhood radius, and β is the variable precision threshold. For any $B \subseteq C$, the neighborhood conditional entropy of D with respect to B is defined as follows:

$$NCE_B(D|B) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log \frac{|V\delta_B(x_i) \cap [x_i]_D|}{|V\delta_B(x_i)|}. \quad (18)$$

where for any $x_i \in U$, $V\delta_B(x_i)$ represents the variable precision neighborhood equivalence class under B that contains sample x_i and $[x_i]_D \in U/D$ represents the decision class that contains sample x_i.

The aforementioned entropy-based measures are commonly used to construct uncertainty measures in information systems, but they primarily focus on similarity within neighborhoods while neglecting the overall hierarchical information of the data distribution. This incomplete representation method may limit the stability and accuracy of related analytical tools and algorithms.

3. The Ze-VNDT algorithm

3.1. Shortcomings of the existing uncertainty measurement methods

In the theory of NRS (neighborhood rough set), the uncertainty of knowledge concepts is often approximated through their lower and upper approximations. This approximation process is not isolated, but

Table 1
Example of a decision information system.

Sample	a	b	d
x_1	0.47	0.48	0
x_2	0.43	0.39	0
x_3	0.31	0.26	0
x_4	0.26	0.17	0
x_5	0.30	0.25	0
x_6	0.52	0.49	1
x_7	0.45	0.57	1
x_8	0.67	0.71	1

Table 2
The Manhattan distance between any two samples under $\{a\}$.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	0	0.04	0.16	0.21	0.17	0.05	0.02	0.2
x_2	0.04	0	0.12	0.17	0.13	0.09	0.02	0.24
x_3	0.16	0.12	0	0.05	0.01	0.21	0.14	0.36
x_4	0.21	0.17	0.05	0	0.04	0.26	0.19	0.41
x_5	0.17	0.13	0.01	0.04	0	0.22	0.15	0.37
x_6	0.05	0.09	0.21	0.26	0.22	0	0.07	0.15
x_7	0.02	0.02	0.14	0.19	0.15	0.07	0	0.22
x_8	0.2	0.24	0.36	0.41	0.37	0.15	0.22	0

Table 3
The Manhattan distance between any two samples under $\{b\}$.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	0	0.09	0.22	0.31	0.23	0.01	0.09	0.23
x_2	0.09	0	0.13	0.22	0.14	0.10	0.18	0.32
x_3	0.22	0.13	0	0.09	0.01	0.23	0.31	0.45
x_4	0.31	0.22	0.09	0	0.08	0.32	0.40	0.54
x_5	0.23	0.14	0.01	0.08	0	0.24	0.32	0.46
x_6	0.01	0.10	0.23	0.32	0.24	0	0.08	0.22
x_7	0.09	0.18	0.31	0.40	0.32	0.08	0	0.14
x_8	0.23	0.32	0.45	0.54	0.46	0.22	0.14	0

Table 4
The neighborhood similarity classes of each sample under $\{a\}$ and $\{b\}$.

Sample	$\{a\}$	$\{b\}$
x_1	$\{x_1, x_2, x_6, x_7\}$	$\{x_1, x_2, x_6, x_7\}$
x_2	$\{x_1, x_2, x_6, x_7\}$	$\{x_1, x_2, x_6\}$
x_3	$\{x_3, x_4, x_5\}$	$\{x_3, x_4, x_5\}$
x_4	$\{x_3, x_4, x_5\}$	$\{x_3, x_4, x_5\}$
x_5	$\{x_3, x_4, x_5\}$	$\{x_3, x_4, x_5\}$
x_6	$\{x_1, x_2, x_6, x_7\}$	$\{x_1, x_2, x_6, x_7\}$
x_7	$\{x_1, x_2, x_6, x_7\}$	$\{x_1, x_6, x_7\}$
x_8	$\{x_8\}$	$\{x_8\}$

is deeply influenced by multiple factors, such as the target concept itself, the granularity level of approximation, the category of neighborhood similarity, and specific objects, which are intertwined at different granularity levels and interact with each other. However, the existing uncertainty measurement methods, such as dependency and rough accuracy, often focus only on the information representation at a specific granularity level, which is particularly inadequate in characterizing complex and variable uncertain knowledge. Especially when the information system faces dynamic changes, this single-granularity-level measurement method is difficult to fully capture the essence of uncertainty. The limitations of this approach in decision-making information are specifically illustrated through Example 1 below.

Example 1 (*[28]*). Let $DIS = (U, C \cup D, V, f)$ be a decision information system, as shown in Table 1, where $U = \{x_1, x_2, \ldots, x_8\}$ represents the set of eight patients, $C = \{a, b, c\}$ includes three tumor genes, $D = \{d\}$ represents the diagnosis result, "1" and "0" respectively represent "disease" and "healthy condition", $D_1 = \{x_1, x_2, x_3, x_4, x_5\}$, and $D_2 = \{x_6, x_7, x_8\}$. In this case, the neighborhood rough set theory is adopted for data processing, and the dependency is used for evaluation. For a fair comparison, the neighborhood radius is uniformly set to 0.1.

The Manhattan distance between any two samples under $\{a\}$ is shown in Table 2, and the Manhattan distance between any two samples under $\{b\}$ is shown in Table 3. The neighborhood similarity classes of each sample under $\{a\}$ and $\{b\}$ are shown in Table 4.

Then, the upper and lower approximations of D_1 and D_2 under $\{a\}$ are as follows:

$$\overline{NR_{\{a\}}}(D_1) = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}, \quad \underline{NR_{\{a\}}}(D_1) = \{x_3, x_4, x_5\},$$

$$\overline{NR_{\{a\}}}(D_2) = \{x_1, x_2, x_6, x_7, x_8\}, \quad \underline{NR_{\{a\}}}(D_2) = \{x_8\}.$$

Then, the upper and lower approximations of D_1 and D_2 under $\{b\}$ are as follows:

$$\overline{NR_{\{b\}}}(D_1) = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}, \quad \underline{NR_{\{b\}}}(D_1) = \{x_3, x_4, x_5\},$$

$$\overline{NR_{\{b\}}}(D_2) = \{x_1, x_2, x_6, x_7, x_8\}, \quad \underline{NR_{\{b\}}}(D_2) = \{x_8\}.$$

According to the definition of neighborhood dependency, the calculation method for the measurement values is as follows:

$$\gamma_{\{a\}}(D) = \gamma_{\{b\}}(D) = 1/2.$$

In neighborhood rough sets, the discrimination of attributes a and b cannot be achieved through mere dependency degrees, since dependency only focuses on information at a single level and overlooks the complex connections that may exist between different levels. This limitation leads to one-sidedness and inaccuracy when describing uncertainty. Therefore, in order to more comprehensively understand and accurately characterize the uncertainty in decision-making information systems, we need to develop a systematic method for uncertainty representation that combines multiple granulation levels.

3.2. Adaptive weighted Zentropy uncertainty measure

Most existing entropy-based methods for measuring uncertainty characterize the uncertainty of a system from the granularity level of objects. Although these methods are relatively precise, they overlook the distribution of target decisions. These measurement methods are often confined to a single granularity level and cannot comprehensively reflect the complexity of uncertain knowledge. In fact, from the granularity of objects to granules, to approximate spaces and target decisions, the approximation process itself is a multi-level and multi-granularity process [30]. Therefore, in order to achieve a comprehensive representation of uncertain knowledge, this paper proposes a novel method for measuring uncertainty. This method employs variable precision neighborhood rough sets for data processing, integrates information from different granularity levels, and evaluates uncertainty based on the principle of Zentropy. As shown in Fig. 1, this method comprehensively measures uncertainty in a decision information system, covering the information presented at various granularity levels, including target concepts, approximation levels, and similarity classes. This comprehensive measurement method helps to understand and deal with uncertainty in the system more accurately.

Multi-value preference is a common issue in decision tree algorithms. Multi-value preference refers to the tendency of decision tree algorithms to preferentially select attributes with a larger number of values when choosing split attributes. The problem with this preference is that it associates the importance of an attribute in classification with the number of its values, which may ultimately lead to the induction of incorrect knowledge from the dataset. To address this problem, this paper proposes a new adaptive weighted Zentropy uncertainty measure.

Definition 10. Let $DIS =< U, A, V, I, \delta, \beta >$ be a variable neighborhood decision table, where $U = \{x_1, x_2, x_3, \ldots, x_n\}$, $A = C \cup D$, $C = \{a_1, a_2, a_3, \ldots, a_m\}$, $U/D = \{D_1, D_2, D_3, \ldots, D_s\}$ represents the set of decision classes, $V = \cup_{a \in A} V_a$ denotes the set of attribute values, $I : U \times (C \cup D) \to V$ represents an information function, $\delta \in [0, 1]$ is

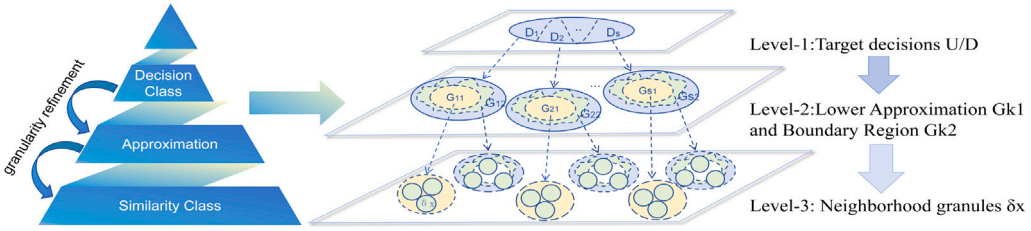


Fig. 1. Granularity levels in conceptual approximation.

the neighborhood radius, and β is the variable precision threshold. For any $B \subseteq C$, the adaptive weighted Zentropy uncertainty measure of B on D is defined as follows:

$$E_B(D) = \varphi \times (-\sum_{k=1}^{s} p_k \log p_k + \sum_{k=1}^{s} p_k F_k). \tag{19}$$

where φ is an adaptive weight (Note: the definition of φ is given below), and for any $1 \le k \le s$, $p_k = \frac{|D_k|}{|U|}$ denotes the probability of the kth decision class D_k at the decision level, and F_k denotes the entropy of D_k.

To address the issue of multi-value preferences in measurement, we introduce the adaptive weight φ to reduce the impact of multi-value partitioning, which is defined as follows:

$$\begin{aligned}\varphi &= \gamma_B(D) \times \frac{1}{\sum_{i=1}^{|U|} \frac{|V\delta_B(x_i)|}{|\bigcup_{i=1}^{|U|}(V\delta_B(x_i))|}} \\ &= \frac{\left|\bigcup_{k=1}^{s} \underline{VNER_B^{(\delta,\beta)}}(D_k)\right|}{|U|} \times \frac{1}{\sum_{i=1}^{|U|} \frac{|V\delta_B(x_i)|}{|\bigcup_{i=1}^{|U|}(V\delta_B(x_i))|}} \\ &= \frac{\sum_{k=1}^{s} \frac{\left|\underline{VNER_B^{(\delta,\beta)}}(D_k)\right|}{\sum_{i=1}^{|U|}|V\delta_B(x_i)|}}{\sum_{i=1}^{|U|}|V\delta_B(x_i)|}.\end{aligned} \tag{20}$$

where for any $B \subseteq C$, $\gamma_B(D)$ denotes the dependency degree of B with respect to D, $\underline{VNER_B^{(\delta,\beta)}}(D_k)$ denotes the lower approximation of the kth decision class D_k under B, and for any $x_i \in U$, $V\delta_B(x_i)$ denotes the variable precision neighborhood equivalence class under B that contains sample x_i.

Proposition 1. *Let $DIS =< U, A, V, I, \delta, \beta >$ be a variable neighborhood decision table. For any $B \subseteq C$ and any $x_i \in U$, let $V\delta_B(x_i)$ represents the variable precision neighborhood equivalence class under B that contains sample x_i. The adaptive weight φ satisfies the following properties:*

(1) $\varphi \in (0, 1]$;

(2) If $|V\delta_B(x_i)| = 1$, then $\varphi = 1$.

Proof. (1) It is obvious that $\underline{VNER_B^{(\delta,\beta)}}(D_k) \neq \emptyset$ and $V\delta_B(x_i) \neq \emptyset$, hence we have that $\left|\bigcup_{k=1}^{s} \underline{VNER_B^{(\delta,\beta)}}(D_k)\right| \in (0, U]$ and $|U| \le \sum_{i=1}^{|U|}|V\delta_B(x_i)|$. According to the definition of φ, we can obtain that $\varphi \in (0, 1]$.

(2) If $|V\delta_B(x_i)| = 1$, then we have that $\bigcup_{k=1}^{s} \underline{VNER_B^{(\delta,\beta)}}(D_k) = U$ and $\sum_{i=1}^{|U|}|V\delta_B(x_i)| = |U|$. According to the definition of φ, we can obtain that $\varphi = 1$.

F_k, which reflects the uncertainty at an approximate level, is defined as follows:

$$F_k = -\sum_{f=1}^{2} p_{kf} \log p_{kf} + \sum_{f=1}^{2} p_{kf} G_{kf} - \sum_{f=1}^{2} q_{kf} \log q_{kf} + \sum_{f=1}^{2} q_{kf} H_{kf}. \tag{21}$$

where $p_{k1} = \frac{\left|\underline{VNER_B^{(\delta,\beta)}}(D_k)\right|}{\left|\overline{VNER_B^{(\delta,\beta)}}(D_k)\right|}, p_{k2} = \frac{\left|\overline{BND_B^{(\delta,\beta)}}(D_k)\right|}{\left|\overline{VNER_B^{(\delta,\beta)}}(D_k)\right|}, q_{k1} = \frac{\left|(BND_B^{(\delta,\beta)}(D_k)) \cap (D_k)\right|}{\left|BND_B^{(\delta,\beta)}(D_k)\right|}$,

Table 5
The neighborhood similarity between any two samples under $\{a\}$.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	1	0.8	0	0	0	0.75	0.9	0
x_2	0.8	1	0	0	0	0.55	0.9	0
x_3	0	0	1	0.75	0.95	0	0	0
x_4	0	0	0.75	1	0.8	0	0	0
x_5	0	0	0.95	0.8	1	0	0	0
x_6	0.75	0.55	0	0	0	1	0.65	0
x_7	0.9	0.9	0	0	0	0.65	1	0
x_8	0	0	0	0	0	0	0	1

$q_{k2} = \frac{\left|(BND_B^{(\delta,\beta)}(D_k)) - \left((BND_B^{(\delta,\beta)}(D_k)) \cap (D_k)\right)\right|}{\left|BND_B^{(\delta,\beta)}(D_k)\right|}$, $\underline{VNER_B^{(\delta,\beta)}}(D_k)$ represents the lower approximation of the kth decision class D_k under B, $\overline{VNER_B^{(\delta,\beta)}}(D_k)$ represents the upper approximation of D_k under B, and $BND_B^{(\delta,\beta)}(D_k)$ represents the boundary region of D_k under B.

In this layer, the uncertainty information of the positive domain and boundary domain has been fully considered.

G_{kf}, which reflects the uncertainty on the neighborhood granularity of the positive domain, is defined as follows:

$$G_{kf} = -\sum_{w=1}^{|N_{kf}|} p_{kfw} \log p_{kfw}. \tag{22}$$

where $N_{k1} = \underline{VNER_B^{(\delta,\beta)}}(D_k)$, $N_{k2} = BND_B^{(\delta,\beta)}(D_k)$, and $p_{kfw} = \frac{|V\delta_B(w)|}{\sum_{w=1}^{|N_{kf}|}|V\delta_B(w)|}$.

H_{kf}, which reflects the uncertainty on the neighborhood granularity of the boundary domain, is defined as follows:

$$H_{kf} = -\sum_{t=1}^{|Q_{kf}|} p_{kft} \log p_{kft}. \tag{23}$$

where $Q_{k1} = (BND_B^{(\delta,\beta)}(D_k)) \cap (D_k)$, $Q_{k2} = (BND_B^{(\delta,\beta)}(D_k)) - \left((BND_B^{(\delta,\beta)}(D_k)) \cap (D_k)\right)$, and $p_{kft} = \frac{|V\delta_B(t)|}{\sum_{t=1}^{|Q_{kf}|}|V\delta_B(t)|}$.

3.3. Properties of the adaptive weighted Zentropy uncertainty measure

This section illustrates the calculation process of the proposed measure through an example, and then presents and analyzes some properties to study its rationality. Continuous data processing generally uses neighborhood rough set calculations. In order to improve the flexibility of the model, this paper adopts variable precision neighborhood rough sets for calculations. The calculation of the adaptive weighted Zentropy uncertainty measure is shown in Example 2.

Example 2. To obtain variable precision neighborhood granules, we continue to process the neighborhood granules in Example 1. This paper adopts the $NRJGS_B^{\delta}$ neighborhood similarity to construct variable precision neighborhood granules. The neighborhood similarity between any two samples under $\{a\}$ is shown in Table 5, and the neighborhood similarity between any two samples under $\{b\}$ is shown in Table 6.

Table 6
The neighborhood similarity between any two samples under $\{b\}$.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	1	0.41	0	0	0	0.95	0.41	0
x_2	0.41	1	0	0	0	0.38	0.05	0
x_3	0	0	1	0.55	0.95	0	0	0
x_4	0	0	0.55	1	0.6	0	0	0
x_5	0	0	0.95	0.6	1	0	0	0
x_6	0.95	0.38	0	0	0	1	0.3	0
x_7	0.41	0.05	0	0	0	0.3	1	0
x_8	0	0	0	0	0	0	0	1

Table 7
The variable precision neighborhood classes of each sample under $\{a\}$ and $\{b\}$.

Sample	$\{a\}$	$\{b\}$
x_1	$\{x_1, x_2, x_7\}$	$\{x_1, x_6\}$
x_2	$\{x_1, x_2, x_7\}$	$\{x_2\}$
x_3	$\{x_3, x_5\}$	$\{x_3, x_5\}$
x_4	$\{x_4, x_5\}$	$\{x_4\}$
x_5	$\{x_3, x_4, x_5\}$	$\{x_3, x_5\}$
x_6	$\{x_6\}$	$\{x_1, x_6\}$
x_7	$\{x_1, x_2, x_7\}$	$\{x_7\}$
x_8	$\{x_8\}$	$\{x_8\}$

For a fair comparison, the variable precision neighborhood similarity threshold is uniformly set to 0.8. According to the results listed in Tables 5 and 6, we can calculate the variable precision neighborhood classes of each sample under $\{a\}$ and $\{b\}$, as shown in Table 7.

Then, the upper and lower approximations and boundary regions of D_1 and D_2 under $\{a\}$ are as follows:

$$\overline{VNER}^{(\delta,\beta)}_{\{a\}}(D_1) = \{x_1, x_2, x_3, x_4, x_5, x_7\}, \quad \underline{VNER}^{(\delta,\beta)}_{\{a\}}(D_1) = \{x_3, x_4, x_5\},$$
$$\overline{VNER}^{(\delta,\beta)}_{\{a\}}(D_2) = \{x_1, x_6, x_7, x_8\}, \quad \underline{VNER}^{(\delta,\beta)}_{\{a\}}(D_2) = \{x_8\},$$
$$BND^{(\delta,\beta)}_{\{a\}}(D_1) = \{x_1, x_2, x_7\}, \quad BND^{(\delta,\beta)}_{\{a\}}(D_2) = \{x_1, x_2, x_6, x_7\}.$$

The upper and lower approximations and boundary regions of D_1 and D_2 under $\{b\}$ are as follows:

$$\overline{VNER}^{(\delta,\beta)}_{\{b\}}(D_1) = \{x_1, x_2, x_4, x_5, x_6\}, \quad \underline{VNER}^{(\delta,\beta)}_{\{b\}}(D_1) = \{x_2, x_3, x_4, x_5\},$$
$$\overline{VNER}^{(\delta,\beta)}_{\{b\}}(D_2) = \{x_1, x_6, x_7, x_8\}, \quad \underline{VNER}^{(\delta,\beta)}_{\{b\}}(D_2) = \{x_7, x_8\},$$
$$BND^{(\delta,\beta)}_{\{b\}}(D_1) = \{x_1, x_6\}, \quad BND^{(\delta,\beta)}_{\{b\}}(D_2) = \{x_1, x_6\}.$$

According to Definition 10, the uncertainty measure of $\{a\}$ is calculated as follows:

Firstly, based on the partition above, we can obtain that:

$$D_1 = \{x_1, x_2, x_3, x_4, x_5\}, \quad D_2 = \{x_6, x_7, x_8\},$$
$$N_{11} = \underline{VNER}^{(\delta,\beta)}_{\{a\}}(D_1) = \{x_3, x_4, x_5\}, \quad N_{12} = BND^{(\delta,\beta)}_{\{a\}}(D_1) = \{x_1, x_2, x_7\},$$
$$N_{21} = \underline{VNER}^{(\delta,\beta)}_{\{a\}}(D_2) = \{x_8\}, \quad N_{22} = BND^{(\delta,\beta)}_{\{a\}}(D_2) = \{x_1, x_2, x_6, x_7\},$$
$$Q_{11} = (BND^{(\delta,\beta)}_{\{a\}}(D_1)) \cap (D_1) = \{x_1, x_2\},$$
$$Q_{12} = (BND^{(\delta,\beta)}_{\{a\}}(D_1)) - \left(\left(BND^{(\delta,\beta)}_{\{a\}}(D_1)\right) \cap (D_1)\right) = \{x_7\},$$
$$Q_{21} = (BND^{(\delta,\beta)}_{\{a\}}(D_2)) \cap (D_2) = \{x_6, x_7\},$$
$$Q_{22} = (BND^{(\delta,\beta)}_{\{a\}}(D_2)) - \left((BND^{(\delta,\beta)}_{\{a\}}(D_2)) \cap (D_2)\right) = \{x_1, x_2\}.$$

Then, we can obtain the probability distribution of the target decision level as follows:

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

The approximate horizontal probability distribution is as follows:

$$p_{11} = \frac{\left|\underline{VNER}^{(\delta,\beta)}_{\{a\}}(D_1)\right|}{\left|\overline{VNER}^{(\delta,\beta)}_{\{a\}}(D_1)\right|} = \frac{1}{2}, \quad p_{12} = \frac{\left|BND^{(\delta,\beta)}_{\{a\}}(D_1)\right|}{\left|\overline{VNER}^{(\delta,\beta)}_{\{a\}}(D_1)\right|} = \frac{1}{2},$$
$$p_{21} = \frac{\left|\underline{VNER}^{(\delta,\beta)}_{\{a\}}(D_2)\right|}{\left|\overline{VNER}^{(\delta,\beta)}_{\{a\}}(D_2)\right|} = \frac{1}{5}, \quad p_{22} = \frac{\left|BND^{(\delta,\beta)}_{\{a\}}(D_2)\right|}{\left|\overline{VNER}^{(\delta,\beta)}_{\{a\}}(D_2)\right|} = \frac{4}{5},$$
$$q_{11} = \frac{\left|(BND^{(\delta,\beta)}_{\{a\}}(D_1)) \cap (D_1)\right|}{\left|BND^{(\delta,\beta)}_{\{a\}}(D_1)\right|} = \frac{2}{3},$$
$$q_{12} = \frac{\left|(BND^{(\delta,\beta)}_{\{a\}}(D_1)) - \left((BND^{(\delta,\beta)}_{\{a\}}(D_1)) \cap (D_1)\right)\right|}{\left|BND^{(\delta,\beta)}_{\{a\}}(D_1)\right|} = \frac{1}{3},$$
$$q_{21} = \frac{\left|(BND^{(\delta,\beta)}_{\{a\}}(D_2)) \cap (D_2)\right|}{\left|BND^{(\delta,\beta)}_{\{a\}}(D_2)\right|} = \frac{1}{2},$$
$$q_{22} = \frac{\left|(BND^{(\delta,\beta)}_{\{a\}}(D_2)) - \left((BND^{(\delta,\beta)}_{\{a\}}(D_2)) \cap (D_2)\right)\right|}{\left|BND^{(\delta,\beta)}_{\{a\}}(D_2)\right|} = \frac{1}{2}.$$

The probability distribution of similar class levels is as follows:

$$p_{111} = \frac{|V\delta_B(x_3)|}{\sum_{w\in N_{11}}|V\delta_B(x_3)|} = \frac{2}{7}, \quad p_{112} = \frac{|V\delta_B(x_4)|}{\sum_{w\in N_{11}}|V\delta_B(x_4)|} = \frac{2}{7},$$
$$p_{113} = \frac{|V\delta_B(x_5)|}{\sum_{w\in N_{11}}|V\delta_B(x_5)|} = \frac{3}{7},$$
$$p_{121} = \frac{|V\delta_B(x_1)|}{\sum_{w\in N_{12}}|V\delta_B(x_1)|} = \frac{1}{3}, \quad p_{122} = \frac{|V\delta_B(x_2)|}{\sum_{w\in N_{12}}|V\delta_B(x_2)|} = \frac{1}{3},$$
$$p_{123} = \frac{|V\delta_B(x_7)|}{\sum_{w\in N_{12}}|V\delta_B(x_7)|} = \frac{1}{3},$$
$$p_{211} = \frac{|V\delta_B(x_8)|}{\sum_{w\in N_{21}}|V\delta_B(x_8)|} = 1, \quad p_{221} = \frac{|V\delta_B(x_1)|}{\sum_{w\in N_{22}}|V\delta_B(x_1)|} = \frac{3}{10},$$
$$p_{222} = \frac{|V\delta_B(x_2)|}{\sum_{w\in N_{22}}|V\delta_B(x_2)|} = \frac{3}{10},$$
$$p_{223} = \frac{|V\delta_B(x_6)|}{\sum_{w\in N_{22}}|V\delta_B(x_6)|} = \frac{1}{10}, p_{224} = \frac{|V\delta_B(x_7)|}{\sum_{w\in N_{22}}|V\delta_B(x_7)|} = \frac{3}{10},$$
$$q_{111} = \frac{|V\delta_B(x_1)|}{\sum_{w\in Q_{11}}|V\delta_B(x_1)|} = \frac{1}{2},$$
$$q_{112} = \frac{|V\delta_B(x_2)|}{\sum_{w\in Q_{11}}|V\delta_B(x_2)|} = \frac{1}{2}, \quad q_{121} = \frac{|V\delta_B(x_7)|}{\sum_{w\in Q_{12}}|V\delta_B(x_7)|} = 1,$$
$$q_{211} = \frac{|V\delta_B(x_6)|}{\sum_{w\in Q_{21}}|V\delta_B(x_6)|} = \frac{1}{4},$$
$$q_{212} = \frac{|V\delta_B(x_7)|}{\sum_{w\in Q_{21}}|V\delta_B(x_7)|} = \frac{3}{4}, \quad q_{221} = \frac{|V\delta_B(x_1)|}{\sum_{w\in Q_{22}}|V\delta_B(x_1)|} = \frac{1}{2},$$
$$q_{222} = \frac{|V\delta_B(x_2)|}{\sum_{w\in Q_{22}}|V\delta_B(x_2)|} = \frac{1}{2}.$$

Therefore, the adaptive weighted Zentropy uncertainty measure of $\{a\}$ on D can be calculated as follows:

$$E_{\{a\}}(D) = \varphi \times (-\sum_{k=1}^{s} p_k \log p_k + \sum_{k=1}^{s} p_k F_k) = 1.1346.$$

Similarly, the adaptive weighted Zentropy uncertainty measure of $\{b\}$ on D can be calculated as follows:

$$E_{\{b\}}(D) = \varphi \times (-\sum_{k=1}^{s} p_k \log p_k + \sum_{k=1}^{s} p_k F_k) = 2.3031.$$

Compared to the dependency in Example 1, the sets $\{a\}$ and $\{b\}$ cannot be distinguished. However, according to the adaptive weighted Zentropy uncertainty measure, $\{a\}$ and $\{b\}$ can be distinguished. Assuming that the numbered item x_i is treated as an attribute for partitioning, according to Definition 12, the measurement calculation for

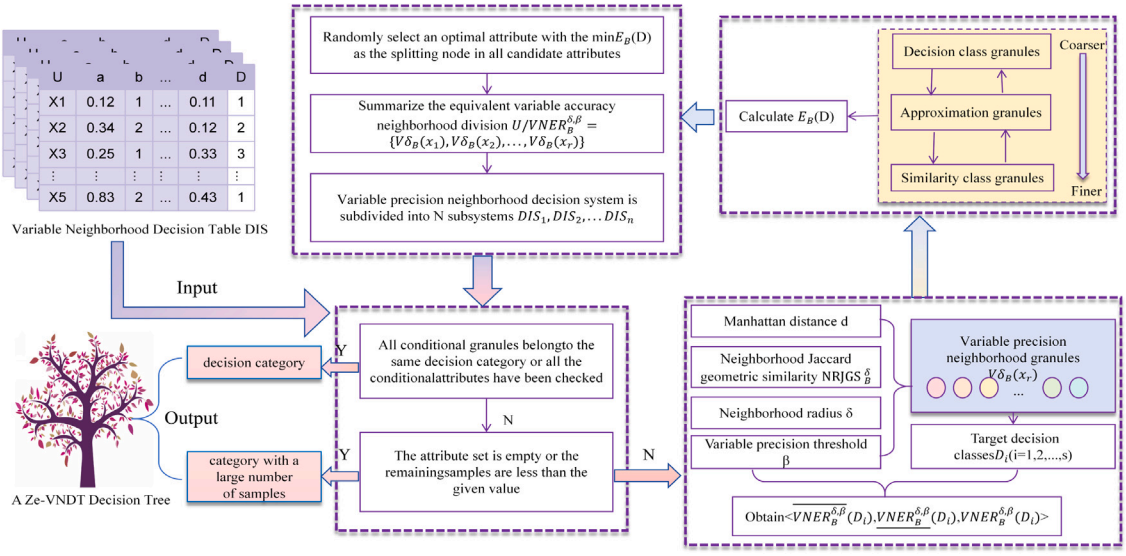


Fig. 2. The framework diagram of the Ze-VNDT algorithm.

Table 8
The description of the 18 UCI datasets.

No.	Dataset	Abbreviation	Sample	Feature	Class	Missing value
1	CrayoDataset	Crayo	90	6	2	no
2	Iris	Iris	150	4	3	no
3	wine	wine	178	13	3	no
4	plrx	plrx	182	18	2	no
5	wpbc	wpbc	194	33	2	no
6	Sheesegmentationt	Shees	210	19	7	no
7	seeds	seeds	210	7	3	no
8	glass	glass	214	10	6	no
9	heart	heart	270	13	2	no
10	ecoli	ecoli	336	7	7	no
11	balance-scale	balance	625	4	3	no
12	glass-identification	glass-id	214	9	7	no
13	lung-cancer	lung	32	56	3	yes
14	Breast-cancer	Breast	286	9	2	yes
15	ENB2012	ENB	768	9	6	no
16	CMC	CMC	1473	9	3	no
17	magic	magic	2501	10	2	no
18	segment	segment	2310	19	7	no

the attribute of the numbered item x_i is as follows:

$$E_{\{b\}}(D) = \varphi \times (-\sum_{k=1}^{s} p_k \log p_k + \sum_{k=1}^{s} p_k F_k) = 2.4056.$$

It can be seen that if there are multiple values, the measure proposed in this paper can effectively reduce the impact of multiple values and increase the importance of attributes.

According to the above definition, the adaptive weighted Zentropy uncertainty measure has the following properties.

Proposition 2. *Let $DIS =< U, A, V, I, \delta, \beta >$ be a variable neighborhood decision table. For any $B \subseteq C$, let $E_B(D)$ be the adaptive weighted Zentropy uncertainty measure of B on D. The following properties hold.*

(1) $E_B(D) \geq 0$.

(2) If $\underline{VNER_B^{(\delta,\beta)}}(D_k) = D_k$, then $F_k = G_{k1}$.

(3) If $\underline{VNER_B^{(\delta,\beta)}}(D_k) = \emptyset$, then $F_k = G_{k2} - \sum_{f=1}^{2} q_{kf} \log q_{kf} + \sum_{f=1}^{2} q_{kf} H_{kf}$.

(4) For any $x_r \in U$, if $V\delta_B(x_r) = \{x_r\}$, then $F_k = \log |D_k|$.

Proof. (1) According to Definition 10, it is known that $0 < p_k, p_{kf}, p_{kfw}, p_{kft} < 1; k = 1, 2, \ldots, s;\ f = 1, 2;\ w = 1, 2, \ldots, |N_{kf}|;\ t = 1, 2, \ldots, |Q_{kf}|$. Therefore, we have that $E_B(D) \geq 0$.

(2) If $\underline{VNER_B^{(\delta,\beta)}}(D_k) = D_k$, then we can obtain that $\overline{VNER_B^{(\delta,\beta)}}(D_k) = D_k$,

$$p_{k1} = \frac{\left|\underline{VNER_B^{(\delta,\beta)}}(D_k)\right|}{\left|\overline{VNER_B^{(\delta,\beta)}}(D_k)\right|} = 1, \quad p_{k2} = \frac{\left|BND_B^{(\delta,\beta)}(D_k)\right|}{\left|\overline{VNER_B^{(\delta,\beta)}}(D_k)\right|} = 0,$$

$$qk1 = \frac{\left|(BND_B^{(\delta,\beta)}(D_k)) \cap (D_k)\right|}{\left|BND_B^{(\delta,\beta)}(D_k)\right|} = 0,$$

and $q_{k2} = \frac{\left|(BND_B^{(\delta,\beta)}(D_k)) - \left((BND_B^{(\delta,\beta)}(D_k)) \cap (D_k)\right)\right|}{\left|BND_B^{(\delta,\beta)}(D_k)\right|} = 0$. Therefore, it can be concluded that $F_k = G_{k1}$.

(3) If $\underline{VNER_B^{(\delta,\beta)}}(D_k) = \emptyset$, then we can obtain that $\overline{VNER_B^{(\delta,\beta)}}(D_k) = BND_B^{(\delta,\beta)}(D_k)$,

$$p_{k1} = \frac{\left|\underline{VNER_B^{(\delta,\beta)}}(D_k)\right|}{\left|\overline{VNER_B^{(\delta,\beta)}}(D_k)\right|} = 0, p_{k2} = \frac{\left|BND_B^{(\delta,\beta)}(D_k)\right|}{\left|\overline{VNER_B^{(\delta,\beta)}}(D_k)\right|} = 1,$$

$$q_{k1} = \frac{\left|(BND_B^{(\delta,\beta)}(D_k)) \cap (D_k)\right|}{\left|BND_B^{(\delta,\beta)}(D_k)\right|} \neq \emptyset,$$

and $q_{k2} = \frac{\left|(BND_B^{(\delta,\beta)}(D_k)) - \left((BND_B^{(\delta,\beta)}(D_k)) \cap (D_k)\right)\right|}{\left|BND_B^{(\delta,\beta)}(D_k)\right|} \neq \emptyset$. Therefore, it can be concluded that $F_k = G_{k2} - \sum_{f=1}^{2} q_{kf} \log q_{kf} + \sum_{f=1}^{2} q_{kf} H_{kf}$.

(4) For any $x_r \in U$, if $V\delta_B(x_r) = \{x_r\}$, then we can obtain that $F_k = G_{k1}$ and $G_{k1} = log|D_k|$. Therefore, it can be concluded that $F_k = log|D_k|$.

3.4. Algorithm design

In this section, we propose a new decision tree algorithm. First, we process the data using a variable precision rough set model to avoid information loss. Second, we analyze the granularity level of variable precision neighborhood rough sets and design a measure of uncertainty based on Zentropy. At the same time, we design an adaptive weight for this measure. Finally, we construct a decision tree algorithm using the proposed measure as the basis for partitioning. When building this algorithm, we need to pay special attention to the key role of the neighborhood radius and the variable precision neighborhood threshold in the neighborhood granulation process. However, in the application of existing neighborhood rough set models, the setting of the neighborhood radius and the variable precision neighborhood threshold often

Table 9
Sensitivity analysis of the variable precision threshold β in Ze-VNDT under acc.

Dataset	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1	aver
Crayo	1	1	1	0.88	1	1	1	1	1	1	1	**0.99**
Iris	0.93	1	0.93	0.86	0.93	0.93	0.93	0.86	0.8	0.93	1	**0.92**
Seeds	0.9	0.95	0.85	0.9	1	0.8	0.8	0.95	0.8	0.95	0.9	**0.89**
Ecoli	0.94	0.97	0.97	0.94	0.94	0.91	0.88	0.97	0.94	0.97	0.97	**0.95**

Table 10
Sensitivity analysis of the variable precision threshold β in Ze-VNDT under pre.

Dataset	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1	aver
Crayo	1	1	1	0.93	1	1	1	1	1	1	1	**0.99**
Iris	0.93	1	0.93	0.88	0.75	0.96	0.95	0.7	0.64	0.75	1	**0.86**
Seeds	0.75	0.75	0.75	0.75	1	0.8	0.8	0.75	0.75	0.75	0.75	**0.78**
Ecoli	0.83	0.98	0.98	0.8	0.81	0.94	0.83	0.98	0.95	0.85	0.85	**0.89**

relies on the subjective judgment and experience of the researchers. To enhance the objectivity and applicability of the algorithm, this paper uses the improved neighborhood radius from [33].

Definition 11 (*[33]*). Let $DIS =< U, A, V, I, \delta, \beta >$ be a variable neighborhood decision table. For any $B \subseteq C$, let $std(B)$ and $\bar{b}$ be the standard deviation and mean value of B, respectively, and let NRC be the neighborhood radius coefficient. The improved neighborhood radius is defined as follows:

$$\delta = \frac{std(B)}{NRC}. \tag{24}$$

where $std(B) = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(b_n - \bar{b})}$, and b_n represents the attribute value of sample b on attribute n.

In Definition 11, the neighborhood radius coefficient NRC is a constant used to control the size of the neighborhood radius. In this paper, we set NRC=2. The framework diagram of the Ze-VNDT algorithm is shown in Fig. 2.

Fig. 2 illustrates the framework of the Ze-VNDT algorithm. The algorithm begins by inputting a variable precision neighborhood decision table (DIS) into the judgment conditions. It then divides the variable precision neighborhoods and calculates $E_B(D)$. A random optimal attribute is selected as the split node for the decision tree. The equivalent variable precision neighborhood division is then summarized and divided into multiple subsystems. Finally, it checks whether all condition granularities belong to the same decision category or if all condition attributes have been detected to determine if further division is necessary. If the attribute set is empty or the remaining samples are fewer than the given threshold, the category with the most samples is output. Otherwise, the process continues until the conditions are met, ultimately constructing a decision tree.

The time complexity of generating neighborhood granules is $O(M \times N^2)$, where M represents the number of features and N represents the number of samples within the domain. The time complexity for calculating the adaptive weighted Zentropy uncertainty measure is also $O(M \times N^2)$. In order to find the equivalent partition of a decision tree, it is necessary to identify the equivalent classes involving all samples under each attribute, which has a time complexity of $O(M \times N^2)$. Therefore, the time complexity of Ze-VNDT is $O(M \times N^2)$. The specific process is shown in Table 8.

Algorithm 1: Ze-VNDT Variable Precision Neighborhood Decision Tree Algorithm

Input: Variable precision neighborhood decision table $DIS = \langle U, A, V, I, \delta, \beta \rangle$
Output: Variable precision neighborhood decision tree T

Initialize decision tree T as empty, processing queue $Q = [DIS]$;
while *Q is not empty* **do**
  $DIS_{curr} \leftarrow Q$.dequeue();
  if *All conditional granules in DIS_{curr} belong to same decision class* ***or*** *all conditional attributes are checked* **then**
    Mark DIS_{curr} as leaf node, record decision class;
  else
    $C \leftarrow$ Conditional attribute set of DIS_{curr};
    Initialize metric dictionary metric_dict = {};
    foreach *Attribute $b \in C$* **do**
      $B \leftarrow \{b\}$;
      $E_B(D) = \varphi \times (-\sum_{k=1}^{s} p_k \log p_k + \sum_{k=1}^{s} p_k F_k) \leftarrow$ Compute Variable Precision Metric($DIS_{curr}, B, \delta, \beta$);
      metric_dict[b] $\leftarrow E_B(D)$;
    end
    min_E $\leftarrow$ min(metric_dict.values());
    Optimal Attribute Set $\leftarrow [b \mid b \in C \wedge$ metric_dict[b] == min_E];
    Randomly select $b_{opt} \in$ Optimal Attribute Set;
    Create split node for DIS_{curr} in T, marked with attribute b_{opt};
    $U/VNER_{\delta,\beta}^{b_{opt}} \leftarrow$ Induce Neighborhood Partition($DIS_{curr}.U, b_{opt}, \delta, \beta$);
    foreach *Equivalent branch $V \in U/VNER_{\delta,\beta}^{b_{opt}}$* **do**
      Construct sub-decision system DIS_{sub};
      Q.enqueue(DIS_{sub});
    end
  end
end
return T;

4. Experimental analysis

4.1. Experimental design

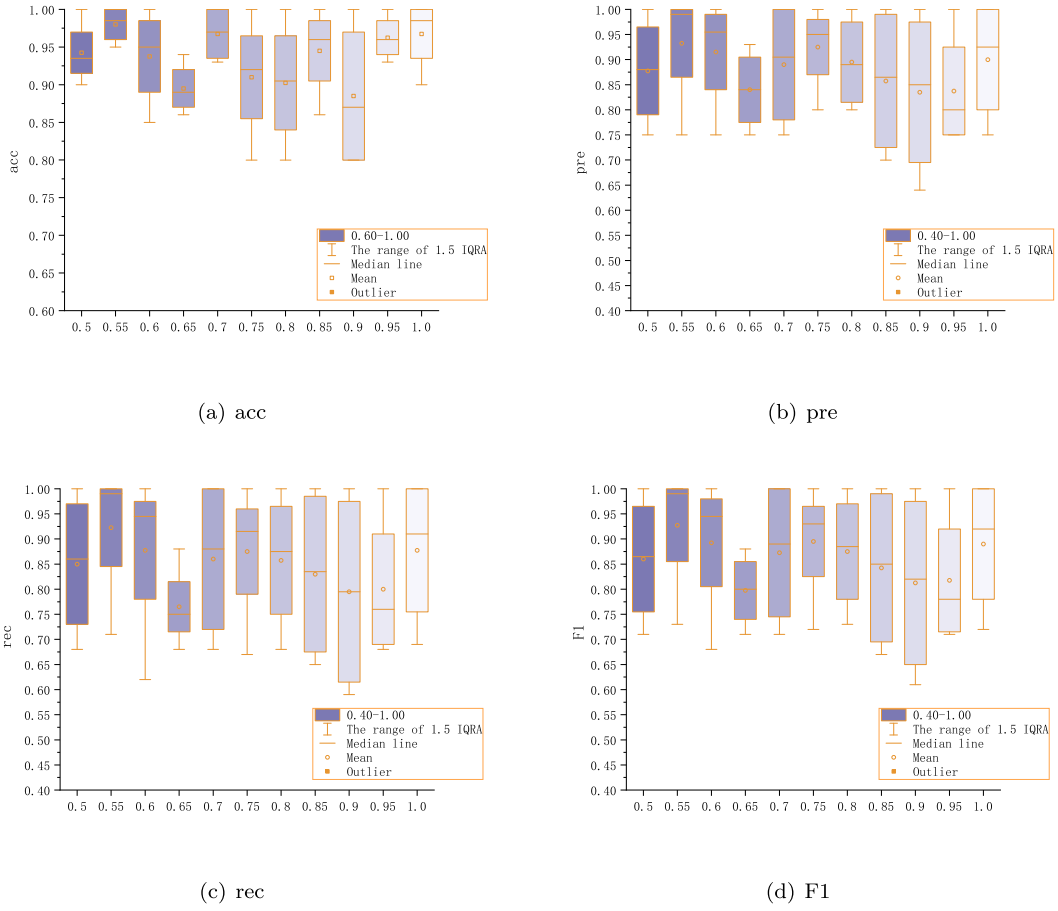
In the experiments, we utilized 18 public UCI datasets to verify the effectiveness of the Ze-VNDT algorithm. Table 8 provides the detailed information of the 18 UCI datasets. We selected three traditional decision tree algorithms and four recently developed decision tree algorithms with higher accuracy for comparative experiments. The three traditional decision tree algorithms are ID3 [8], C4.5 [9], and CRAT [10], while the four improved decision tree algorithms are NID3 [36], VPNDT [33], NDT [20], and BMAD [22]. The evaluation metrics used are accuracy, precision, recall, F1 score, and the number of leaves, with a ten-fold cross-validation method for verification. All average results and standard deviations are abbreviated as aver and st, respectively. Moreover, accuracy, precision, recall, F1 score, and number of leaves are represented by acc, pre, rec, F1 and leaf, respectively. The experimental environment is as follows: an AMD Ryzen 7 4700U with Radeon Graphics at 2.00 GHz, with 16.0 GB RAM.

Table 11
Sensitivity analysis of the variable precision threshold β in Ze-VNDT under rec.

Dataset	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1	**aver**
Crayo	1	1	1	0.75	1	1	1	1	1	1	1	**0.98**
Iris	0.94	1	0.94	0.88	0.68	0.91	0.93	0.65	0.59	0.68	1	**0.84**
Seeds	0.68	0.71	0.62	0.68	1	0.67	0.68	0.7	0.64	0.7	0.69	**0.71**
Ecoli	0.78	0.98	0.95	0.75	0.76	0.92	0.82	0.97	0.95	0.82	0.82	**0.87**

Table 12
Sensitivity analysis of the variable precision threshold β in Ze-VNDT under F1.

Dataset	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1	**aver**
Crayo	1	1	1	0.83	1	1	1	1	1	1	1	**0.98**
Iris	0.93	1	0.93	0.88	0.71	0.93	0.94	0.67	0.61	0.71	1	**0.85**
Seeds	0.71	0.73	0.68	0.71	1	0.72	0.73	0.72	0.69	0.72	0.72	**0.74**
Ecoli	0.8	0.98	0.96	0.77	0.78	0.93	0.83	0.98	0.95	0.84	0.84	**0.88**



(a) acc (b) pre

(c) rec (d) F1

Fig. 3. Sensitivity analysis of the variable precision threshold β in Ze-VNDT for acc, pre, rec, and F1.

4.2. Sensitivity analysis of the variable precision threshold β in the Ze-VNDT algorithm

Since the choice of neighborhood radius and variable precision threshold directly affects the effect of neighborhood granulation, which in turn affects the performance of the entire model, this paper adopted an adaptive radius based on standard deviation and adaptive coefficients instead of the traditional neighborhood radius. The parameter β was set between 0.5 and 1, with a step size of 0.05, and four datasets were randomly selected to analyze the sensitivity of the accuracy, precision, recall, and F1 score of parameter β in Ze-VNDT. In this experiment, the results for each dataset are the average of 10 runs. The experimental results are shown in Tables 9–12 and Fig. 3, with the last column displaying the average values.

According to Fig. 3, we can conclude that Ze-VNDT is not highly sensitive to different values of the variable precision threshold β within the same dataset in terms of acc, but shows fluctuations in terms of the other three metrics: pre, rec, and F1. Overall, the choice of the variable precision threshold β has a relatively minor impact on the performance of Ze-VNDT.

4.3. Experimental results

Tables 13–17 and Fig. 4 present the comparative experimental results of accuracy, precision, recall, F1 score, and number of leaves for various algorithms across 18 datasets, with the highest values highlighted in bold.

Table 13
Accuracy of the Ze-VNDT algorithm and seven other comparative algorithms.

Dataset	ID3	C45	CART	NID3	NDT	VPNDT	BMAD	Ze-VNDT
Crayo	55.55 ± 0.05	86.88 ± 0.09	84.88 ± 0.09	**100 ± 0.10**	**100 ± 0.10**	**100 ± 0.10**	**100 ± 0.10**	**100 ± 0.10**
Iris	80.00 ± 0.06	93.33 ± 0.07	86.66 ± 0.07	86.66 ± 0.07	93.33 ± 0.07	86.66 ± 0.07	86.36 ± 0.07	**100 ± 0.08**
wine	88.98 ± 0.06	86.87 ± 0.06	**94.44 ± 0.07**	88.38 ± 0.06	**94.44 ± 0.07**	83.33 ± 0.06	93.44 ± 0.07	**94.44 ± 0.07**
plrx	84.21 ± 0.06	84.21 ± 0.06	**89.47 ± 0.06**	84.21 ± 0.06	84.86 ± 0.06	**89.47 ± 0.06**	88.95 ± 0.06	**89.47 ± 0.06**
wpbc	80.00 ± 0.05	90.00 ± 0.06	80.00 ± 0.05	85.00 ± 0.06	89.47 ± 0.06	85.00 ± 0.06	85.00 ± 0.06	**95.00 ± 0.06**
Shees	81.43 ± 0.05	85.71 ± 0.05	85.71 ± 0.05	90.47 ± 0.06	95.00 ± 0.06	**95.23 ± 0.06**	90.48 ± 0.06	**95.23 ± 0.06**
seeds	90.47 ± 0.06	85.71 ± 0.05	90.47 ± 0.06	90.47 ± 0.06	76.19 ± 0.05	95.23 ± 0.06	90.47 ± 0.06	**98.00 ± 0.06**
glass	85.18 ± 0.05	63.63 ± 0.04	72.72 ± 0.04	90.90 ± 0.06	80.95 ± 0.05	86.36 ± 0.05	**93.46 ± 0.06**	90.90 ± 0.06
heart	92.96 ± 0.05	92.59 ± 0.05	88.48 ± 0.05	78.88 ± 0.05	92.59 ± 0.05	92.59 ± 0.05	91.85 ± 0.05	**96.29 ± 0.05**
ecoli	80.59 ± 0.04	88.23 ± 0.04	82.53 ± 0.04	94.11 ± 0.04	**97.05 ± 0.05**	91.17 ± 0.04	88.23 ± 0.04	**97.05 ± 0.05**
balance	74.44 ± 0.03	74.60 ± 0.02	88.89 ± 0.03	96.82 ± 0.03	**98.41 ± 0.03**	96.82 ± 0.03	96.83 ± 0.03	93.65 ± 0.03
glass-id	74.55 ± 0.05	90.90 ± 0.06	90.90 ± 0.06	81.81 ± 0.04	81.81 ± 0.05	80.30 ± 0.05	81.81 ± 0.05	**95.45 ± 0.06**
lung	95.00 ± 0.17	75.40 ± 0.13	75.80 ± 0.13	75.10 ± 0.10	77.70 ± 0.13	78.30 ± 0.13	97.50 ± 0.17	**100 ± 0.10**
Breast	74.30 ± 0.13	73.00 ± 0.13	74.67 ± 0.13	75.00 ± 0.13	50.00 ± 0.13	75.50 ± 0.13	73.32 ± 0.04	**93.45 ± 0.03**
ENB	24.15 ± 0.01	25.97 ± 0.01	35.06 ± 0.01	94.80 ± 0.03	93.50 ± 0.03	**98.70 ± 0.03**	85.71 ± 0.03	89.61 ± 0.03
CMC	49.32 ± 0.01	53.33 ± 0.03	55.33 ± 0.01	66.66 ± 0.07	75.20 ± 0.13	73.00 ± 0.13	77.42 ± 0.06	**87.70 ± 0.02**
magic	84.11 ± 0.06	85.24 ± 0.04	86.76 ± 0.03	86.36 ± 0.07	85.98 ± 0.06	90.10 ± 0.01	90.07 ± 0.04	**95.44 ± 0.07**
segment	73.44 ± 0.03	72.10 ± 0.13	72.92 ± 0.04	81.31 ± 0.04	70.95 ± 0.05	85.20 ± 0.06	91.54 ± 0.04	**94.54 ± 0.07**
aver	76.04 ± 0.05	78.20 ± 0.06	79.76 ± 0.05	85.94 ± 0.06	85.41 ± 0.06	87.94 ± 0.07	89.02 ± 0.06	**94.79 ± 0.58**

Table 14
Precision of the Ze-VNDT algorithm and seven other comparative algorithms.

Dataset	ID3	C45	CART	NID3	NDT	VPNDT	BMAD	Ze-VNDT
Crayo	71.42 ± 0.07	83.33 ± 0.08	92.85 ± 0.09	**100 ± 0.10**	75.00 ± 0.06	75.00 ± 0.07	**100 ± 0.10**	**100 ± 0.10**
Iris	76.66 ± 0.06	95.83 ± 0.07	90.47 ± 0.07	75.00 ± 0.06	75.00 ± 0.05	71.42 ± 0.05	62.50 ± 0.05	**100 ± 0.08**
wine	90.47 ± 0.06	88.57 ± 0.06	**91.66 ± 0.06**	75.00 ± 0.05	75.00 ± 0.05	70.83 ± 0.05	75.10 ± 0.04	75.00 ± 0.05
plrx	82.85 ± 0.06	**82.95 ± 0.06**	66.66 ± 0.04	66.66 ± 0.04	66.66 ± 0.04	66.66 ± 0.04	66.46 ± 0.04	66.66 ± 0.04
wpbc	42.10 ± 0.03	**93.75 ± 0.06**	66.66 ± 0.04	66.66 ± 0.04	66.66 ± 0.04	66.67 ± 0.04	66.67 ± 0.04	66.67 ± 0.04
Shees	66.67 ± 0.04	89.28 ± 0.06	80.20 ± 0.05	87.50 ± 0.06	75.00 ± 0.05	**96.42 ± 0.06**	87.45 ± 0.06	75.00 ± 0.05
seeds	90.47 ± 0.06	85.18 ± 0.05	91.90 ± 0.06	60.00 ± 0.04	75.00 ± 0.05	75.00 ± 0.05	75.00 ± 0.04	**97.67 ± 0.06**
glass	85.18 ± 0.05	59.39 ± 0.04	60.04 ± 0.04	**85.71 ± 0.05**	83.67 ± 0.05	83.33 ± 0.05	83.33 ± 0.05	**85.71 ± 0.05**
heart	85.16 ± 0.05	92.58 ± 0.05	61.53 ± 0.03	89.16 ± 0.05	66.67 ± 0.04	64.10 ± 0.03	66.67 ± 0.04	**97.05 ± 0.05**
ecoli	68.01 ± 0.03	79.90 ± 0.04	76.00 ± 0.04	83.67 ± 0.04	**97.77 ± 0.06**	83.33 ± 0.04	77.14 ± 0.04	83.33 ± 0.04
balance	87.50 ± 0.03	66.67 ± 0.02	76.72 ± 0.03	97.77 ± 0.03	93.33 ± 0.03	90.62 ± 0.03	**98.95 ± 0.03**	85.55 ± 0.03
glass-id	68.88 ± 0.04	81.66 ± 0.05	**93.46 ± 0.06**	71.42 ± 0.04	80.00 ± 0.05	80.30 ± 0.05	83.33 ± 0.05	80.00 ± 0.05
lung	95.00 ± 0.17	37.50 ± 0.06	66.67 ± 0.11	37.50 ± 0.06	37.50 ± 0.06	33.33 ± 0.05	93.22 ± 0.03	**97.34 ± 0.10**
Breast	**83.33 ± 0.14**	37.50 ± 0.06	50.00 ± 0.08	66.66 ± 0.11	22.22 ± 0.06	50.00 ± 0.08	57.63 ± 0.03	68.32 ± 0.03
END	65.77 ± 0.02	61.87 ± 0.02	45.67 ± 0.01	92.59 ± 0.03	89.58 ± 0.03	**96.66 ± 0.03**	87.20 ± 0.03	87.14 ± 0.03
CMC	40.38 ± 0.01	54.39 ± 0.04	51.00 ± 0.08	72.42 ± 0.04	78.39 ± 0.03	89.03 ± 0.05	87.15 ± 0.06	**89.52 ± 0.04**
magic	53.21 ± 0.03	58.36 ± 0.06	62.11 ± 0.04	66.35 ± 0.09	82.67 ± 0.05	57.88 ± 0.01	54.62 ± 0.03	**93.33 ± 0.05**
segment	72.83 ± 0.02	75.55 ± 0.05	78.31 ± 0.03	79.00 ± 0.05	64.67 ± 0.04	81.03 ± 0.04	80.31 ± 0.04	**95.68 ± 0.02**
aver	73.66 ± 0.05	73.57 ± 0.05	72.32 ± 0.05	76.28 ± 0.05	72.48 ± 0.05	73.97 ± 0.05	77.93 ± 0.04	**85.77 ± 0.05**

Table 15
Recall of the Ze-VNDT algorithm and seven other comparative algorithms.

Dataset	ID3	C45	CART	NID3	NDT	VPNDT	BMAD	Ze-VNDT
Crayo	66.66 ± 0.07	92.85 ± 0.09	83.33 ± 0.08	**100 ± 0.1**	70.83 ± 0.05	93.75 ± 0.09	**100 ± 0.10**	**100 ± 0.1**
Iris	88.88 ± 0.7	93.33 ± 0.07	88.88 ± 0.07	65.00 ± 0.05	68.75 ± 0.05	66.67 ± 0.05	58.33 ± 0.04	**100 ± 0.01**
wine	90.47 ± 0.06	88.57 ± 0.06	**95.23 ± 0.07**	67.72 ± 0.05	71.15 ± 0.05	58.14 ± 0.04	68.75 ± 0.05	72.72 ± 0.05
plrx	79.48 ± 0.06	**84.52 ± 0.06**	60.60 ± 0.04	60.00 ± 0.04	56.11 ± 0.04	59.63 ± 0.04	61.11 ± 0.04	60.60 ± 0.04
wpbc	47.05 ± 0.03	**83.33 ± 0.06**	44.44 ± 0.03	54.16 ± 0.03	64.70 ± 0.04	60.00 ± 0.04	60.10 ± 0.04	64.91 ± 0.04
Shees	78.57 ± 0.05	89.28 ± 0.06	77.08 ± 0.05	82.50 ± 0.05	63.75 ± 0.04	**95.23 ± 0.06**	81.25 ± 0.05	83.33 ± 0.05
seeds	92.59 ± 0.06	84.92 ± 0.05	91.90 ± 0.06	56.00 ± 0.03	62.85 ± 0.04	71.87 ± 0.04	67.26 ± 0.04	**99.36 ± 0.06**
glass	**85.18 ± 0.05**	65.50 ± 0.04	59.44 ± 0.04	76.53 ± 0.05	72.14 ± 0.04	69.84 ± 0.04	77.77 ± 0.05	83.51 ± 0.05
heart	85.16 ± 0.05	92.58 ± 0.05	59.44 ± 0.03	88.73 ± 0.05	62.50 ± 0.03	61.72 ± 0.03	62.22 ± 0.03	**95.45 ± 0.05**
ecoli	65.00 ± 0.03	76.66 ± 0.04	72.44 ± 0.03	81.29 ± 0.04	**98.66 ± 0.05**	75.94 ± 0.04	71.90 ± 0.03	75.00 ± 0.04
balance	61.75 ± 0.02	70.90 ± 0.02	84.75 ± 0.03	94.59 ± 0.03	98.71 ± 0.03	97.76 ± 0.03	**98.81 ± 0.03**	85.58 ± 0.03
glass-id	80.00 ± 0.05	73.14 ± 0.05	**93.46 ± 0.06**	62.22 ± 0.04	68.33 ± 0.04	77.08 ± 0.05	76.01 ± 0.05	77.50 ± 0.05
lung	95.00 ± 0.17	50.00 ± 0.08	55.55 ± 0.09	50.00 ± 0.08	50.00 ± 0.08	33.33 ± 0.05	93.12 ± 0.03	**99.07 ± 0.10**
Breast	**75.00 ± 0.13**	50.00 ± 0.08	50.00 ± 0.08	50.00 ± 0.08	22.22 ± 0.08	37.50 ± 0.06	55.48 ± 0.04	70.03 ± 0.03
ENB	50.36 ± 0.01	39.09 ± 0.01	37.80 ± 0.01	94.81 ± 0.03	94.43 ± 0.03	**98.71 ± 0.03**	87.19 ± 0.03	90.89 ± 0.03
CMC	35.31 ± 0.01	58.37 ± 0.06	46.44 ± 0.03	73.66 ± 0.09	82.67 ± 0.04	89.03 ± 0.05	**89.12 ± 0.03**	73.75 ± 0.01
magic	57.45 ± 0.04	57.92 ± 0.03	62.11 ± 0.04	68.39 ± 0.09	84.67 ± 0.04	59.53 ± 0.04	56.91 ± 0.03	**95.38 ± 0.03**
segment	74.94 ± 0.04	75.55 ± 0.05	80.57 ± 0.03	78.33 ± 0.03	68.66 ± 0.05	85.03 ± 0.03	82.47 ± 0.04	**97.22 ± 0.04**
aver	72.71 ± 0.09	73.69 ± 0.05	69.08 ± 0.05	72.44 ± 0.05	70.06 ± 0.05	71.7 ± 0.05	74.88 ± 0.04	**84.68 ± 0.05**

From Table 13 and Fig. 4(a), it can be seen that the Ze-VNDT algorithm proposed in this paper has improved accuracy on most datasets, with an average accuracy of 94.79%. This indicates that Ze-VNDT exhibits good generalization capabilities when dealing with different types of datasets. The improvement in accuracy is evident both in datasets with high complexity and those with lower complexity, fully illustrating the effectiveness of Ze-VNDT. The accuracy of other comparative algorithms shows significant fluctuations, while the Ze-VNDT algorithm proposed in this paper has a smaller range of accuracy fluctuations, indicating higher stability.

Table 16
F1 scores of the Ze-VNDT algorithm and seven other comparative algorithms.

Dataset	ID3	C45	CART	NID3	NDT	VPNDT	BMAD	Ze-VNDT
Crayo	68.96 ± 0.07	87.83 ± 0.09	87.83 ± 0.09	**100 ± 0.10**	72.85 ± 0.05	83.33 ± 0.08	**100 ± 0.10**	**100 ± 0.10**
Iris	82.32 ± 0.6	93.33 ± 0.07	89.67 ± 0.07	69.64 ± 0.05	71.73 ± 0.05	68.96 ± 0.05	60.34 ± 0.04	**100 ± 0.08**
Wine	90.47 ± 0.06	88.57 ± 0.06	**93.41 ± 0.07**	71.17 ± 0.05	73.02 ± 0.05	63.86 ± 0.04	71.73 ± 0.05	73.84 ± 0.05
Plrx	81.13 ± 0.06	**83.73 ± 0.06**	63.49 ± 0.04	63.15 ± 0.04	60.93 ± 0.04	62.95 ± 0.04	63.76 ± 0.04	63.49 ± 0.04
WPBC	44.44 ± 0.03	**88.23 ± 0.06**	53.33 ± 0.03	59.77 ± 0.04	65.67 ± 0.04	63.15 ± 0.04	63.17 ± 0.04	65.77 ± 0.04
Shees	72.13 ± 0.04	89.28 ± 0.06	78.61 ± 0.05	84.92 ± 0.05	68.91 ± 0.04	**95.83 ± 0.06**	84.25 ± 0.05	78.94 ± 0.05
seeds	91.52 ± 0.06	85.05 ± 0.05	91.90 ± 0.06	57.93 ± 0.04	68.39 ± 0.04	73.40 ± 0.05	70.92 ± 0.04	**98.75 ± 0.06**
glass	**85.16 ± 0.05**	62.29 ± 0.04	59.74 ± 0.04	80.86 ± 0.05	77.48 ± 0.05	75.99 ± 0.05	80.46 ± 0.05	84.60 ± 0.05
heart	85.16 ± 0.05	92.58 ± 0.05	60.47 ± 0.03	88.95 ± 0.05	64.51 ± 0.03	62.89 ± 0.03	64.36 ± 0.03	**96.25 ± 0.05**
ecoli	66.47 ± 0.03	78.25 ± 0.04	74.18 ± 0.04	82.46 ± 0.04	**98.22 ± 0.05**	79.46 ± 0.04	74.43 ± 0.04	78.94 ± 0.05
balance	72.41 ± 0.02	68.72 ± 0.02	80.53 ± 0.03	96.16 ± 0.03	95.95 ± 0.03	94.06 ± 0.03	**98.88 ± 0.03**	85.57 ± 0.03
glass-id	74.03 ± 0.05	77.17 ± 0.05	**93.46 ± 0.06**	66.50 ± 0.04	73.70 ± 0.05	78.66 ± 0.05	79.50 ± 0.05	78.73 ± 0.05
Lung	95 ± 0.17	42.85 ± 0.07	60.67 ± 0.09	42.85 ± 0.07	42.85 ± 0.07	33.33 ± 0.05	93.15 ± 0.04	**98.72 ± 0.10**
Breast	**78.94 ± 0.13**	42.85 ± 0.07	50.00 ± 0.08	57.14 ± 0.10	22.22 ± 0.08	42.85 ± 0.07	56.52 ± 0.04	69.35 ± 0.03
ENB	57.04 ± 0.02	47.91 ± 0.01	41.37 ± 0.01	93.68 ± 0.03	91.94 ± 0.03	**97.68 ± 0.03**	87.19 ± 0.03	88.98 ± 0.03
CMC	37.68 ± 0.01	56.48 ± 0.06	48.56 ± 0.07	72.89 ± 0.09	80.61 ± 0.04	**89.03 ± 0.05**	88.14 ± 0.04	76.61 ± 0.01
Magic	55.43 ± 0.04	56.78 ± 0.06	62.11 ± 0.04	67.82 ± 0.09	83.61 ± 0.04	58.92 ± 0.04	55.27 ± 0.03	**94.59 ± 0.04**
Segment	73.59 ± 0.04	75.55 ± 0.05	79.03 ± 0.04	78.75 ± 0.04	66.57 ± 0.04	83.03 ± 0.03	81.83 ± 0.04	**96.45 ± 0.05**
aver	72.88 ± 0.08	73.19 ± 0.05	70.46 ± 0.05	74.14 ± 0.05	71.06 ± 0.04	72.63 ± 0.04	76.33 ± 0.04	**84.97 ± 0.05**

Table 17
Number of leaves of the Ze-VNDT algorithm and seven other comparative algorithms.

Dataset	ID3	C45	CART	NID3	NDT	VPNDT	BMAD	Ze-VNDT
Crayo	34	31	97	55	**116**	56	70	54
Iris	16	19	49	102	**160**	117	99	130
Wine	66	45	**294**	171	170	162	156	148
plrx	181	160	**384**	159	159	159	160	160
wpbc	**223**	123	160	172	172	172	173	163
Shees	185	151	**343**	200	204	203	194	220
Seeds	39	48	124	**184**	183	182	180	183
Glass	181	104	154	205	**209**	204	204	188
Heart	181	146	240	258	259	**281**	257	261
Ecoli	79	93	181	344	336	**346**	331	337
Balance	30	96	254	492	501	**502**	495	498
Glass-Id	97	64	180	203	**208**	203	202	193
Lung	11	12	24	30	**37**	34	34	26
Breast	26	28	63	41	38	40	35	**44**
ENB	223	593	504	978	974	984	**993**	962
CMC	374	476	432	987	1011	**1134**	1128	1112
Magic	967	1146	1386	**2133**	2056	2123	2116	2028
Segment	632	845	1132	**1998**	1789	1895	1884	1834
aver	196.94	232.22	333.39	484.00	476.78	**488.72**	483.94	474.50

Table 18
Rank of accuracy ordinal values and average ordinal values.

Dataset	ID3	C45	CART	NID3	NDT	VPNDT	BMAD	Ze-VNDT
Crayo	8	6	7	3	3	3	3	3
Iris	8	2.5	5	5	2.5	5	7	1
wine	5	7	2	6	2	8	4	2
plrx	7	7	2	7	5	2	4	2
wpbc	7.5	2	7.5	5	3	5	5	1
Shees	8	6.5	6.5	5	3	1.5	4	1.5
seeds	4.5	7	4.5	4.5	8	2	4.5	1
glass	5	8	7	2.5	6	4	1	2.5
heart	2	4	7	8	4	4	6	1
ecoli	8	5.5	7	3	1.5	4	5.5	1.5
balance	8	7	6	3.5	1	3.5	2	5
glass-id	8	2.5	2.5	5	5	7	5	1
lung	3	7	6	8	5	4	2	1
Breast	5	7	4	3	8	2	6	1
ENB	8	7	6	2	3	1	5	4
CMC	8	7	6	5	3	4	2	1
magic	8	7	4	5	6	2	3	1
segment	5	7	6	4	8	3	2	1
rank	6.44	5.94	5.33	4.69	4.28	3.61	3.94	1.75

In Table 14 and Fig. 4(b), we conducted an in-depth analysis of the precision of eight algorithms across 18 different datasets. Precision, which measures the proportion of true positive instances among those predicted as positive by the classifier, shows that on average, the Ze-VNDT algorithm achieves the highest precision across all tested datasets. This result indicates that Ze-VNDT has a higher accuracy in distinguishing true positive samples.

In Table 15 and Fig. 4(c), we evaluated the recall of various algorithms on 18 datasets. Recall is a measure of the classifier's ability to correctly identify positive samples, and across all datasets, the average recall of the Ze-VNDT algorithm is the highest, demonstrating its strong capability in identifying positive samples.

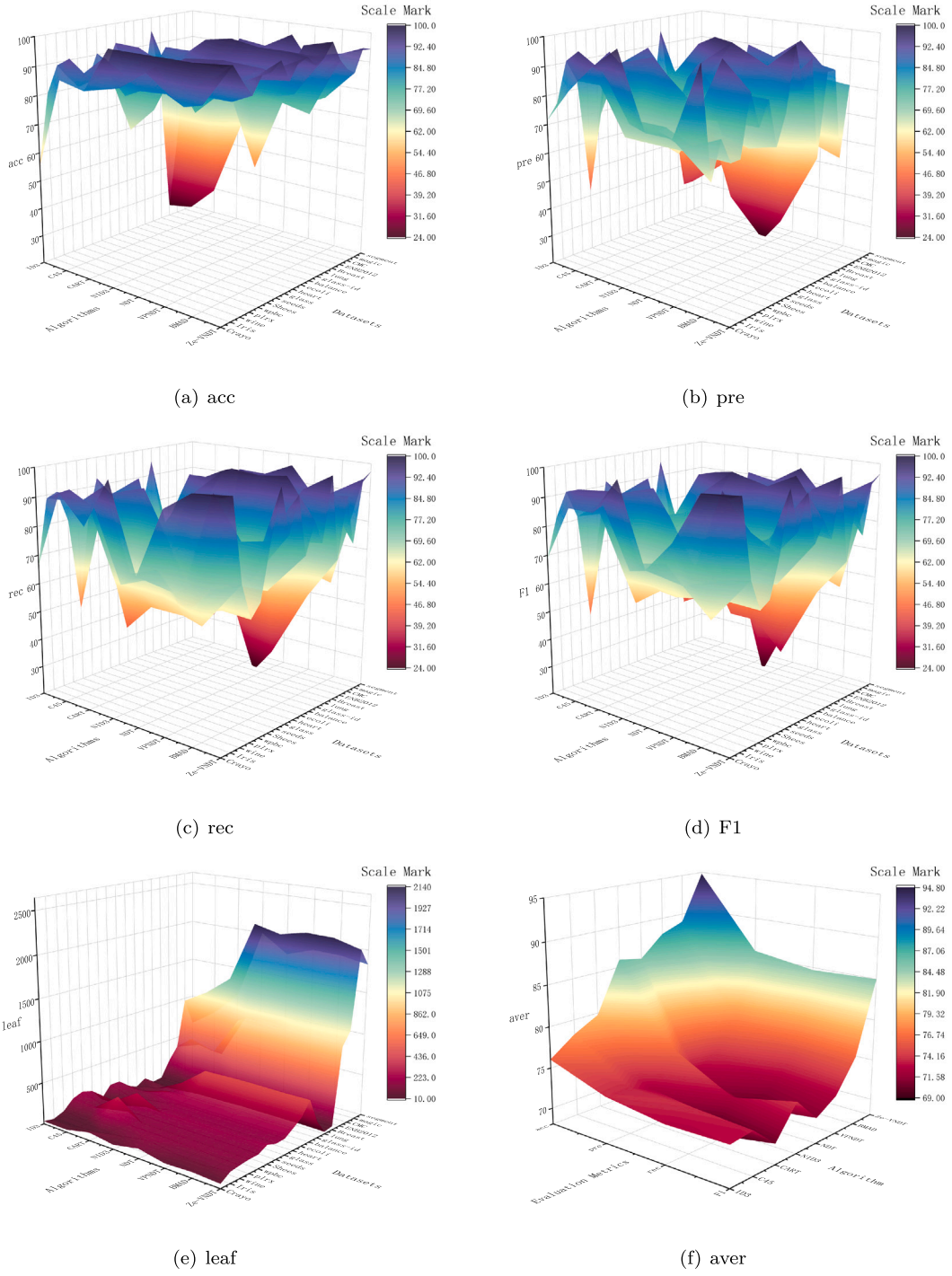
In Table 16 and Fig. 4(d), we evaluated the F1 scores of various algorithms on 18 datasets. The F1 score is a metric that combines precision and recall to measure the performance of classifiers. Across all datasets, the Ze-VNDT algorithm achieves the highest average F1 score, indicating that this algorithm can generally achieve more accurate classification in most cases.

In Table 17 and Fig. 4(e), through a comparative analysis of the number of leaves produced by different algorithms on 18 datasets, we observed that the Ze-VNDT algorithm exhibits fewer leaves on most datasets, significantly outperforming existing algorithms. Although the number of leaves for Ze-VNDT is relatively high on some large datasets, even in these cases, the number of leaves remains within a reasonable and acceptable range.

Fig. 4(f) shows the comparison of the average accuracy, precision, recall, and F1 score of eight algorithms. It can be seen that compared with the other seven algorithms, the Ze-VNDT algorithm has higher average values in all four evaluation metrics (i.e., accuracy, precision, recall, and F1 score), which also validates the effectiveness of the algorithm proposed in this paper.

4.4. Statistical testing

To test whether there are significant differences in performance between the Ze-VNDT algorithm and other comparison methods, the Friedman test [37] and Nemenyi post-hoc test [38] were conducted on 18 datasets to determine the performance differences among the eight algorithms, using accuracy, precision, recall, and F1 score as performance metrics. The Friedman test is a non-parametric method used to evaluate the overall performance of S algorithms on E datasets. If the hypothesis that "all algorithms have the same performance" is rejected, it indicates that there are significant differences in algorithm performance. In such cases, further testing is required to differentiate the algorithms, and in this paper, the Nemenyi post-hoc test [39] was adopted.



(a) acc (b) pre

(c) rec (d) F1

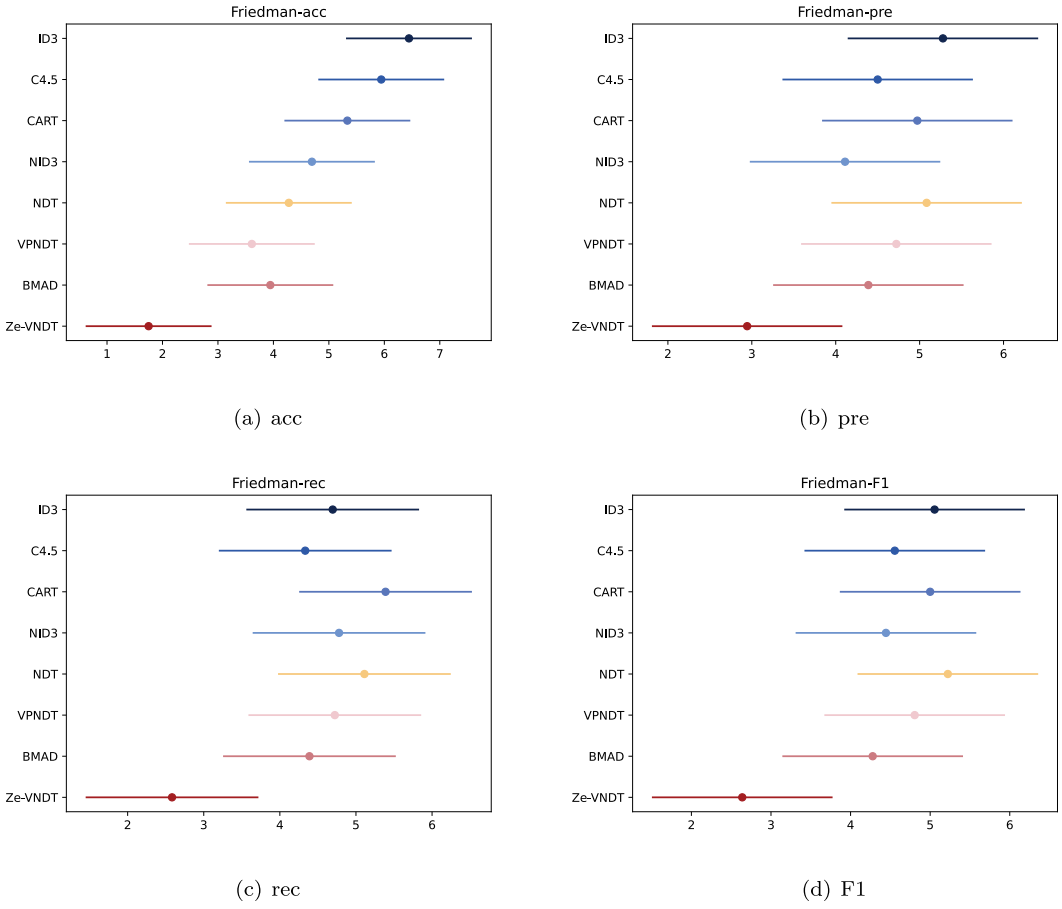
(e) leaf (f) aver

Fig. 4. Comparison of experimental results in terms of accuracy, precision, recall, F1 score, and number of leaves for different algorithms across 18 datasets, along with the average values of accuracy, precision, recall, and F1 score.

Firstly, the four evaluation metrics on each dataset were ranked from best to worst and assigned rank values (Note: algorithms that have the same performance share the rank values). The significance level α was set to 0.1, and the ranking and average ranking of the rank values for the eight algorithms across the four evaluation metrics on the 18 datasets are shown in Tables 18–21 with the average ranking represented by rank.

Secondly, we calculated the results of Friedman test using the following formula:

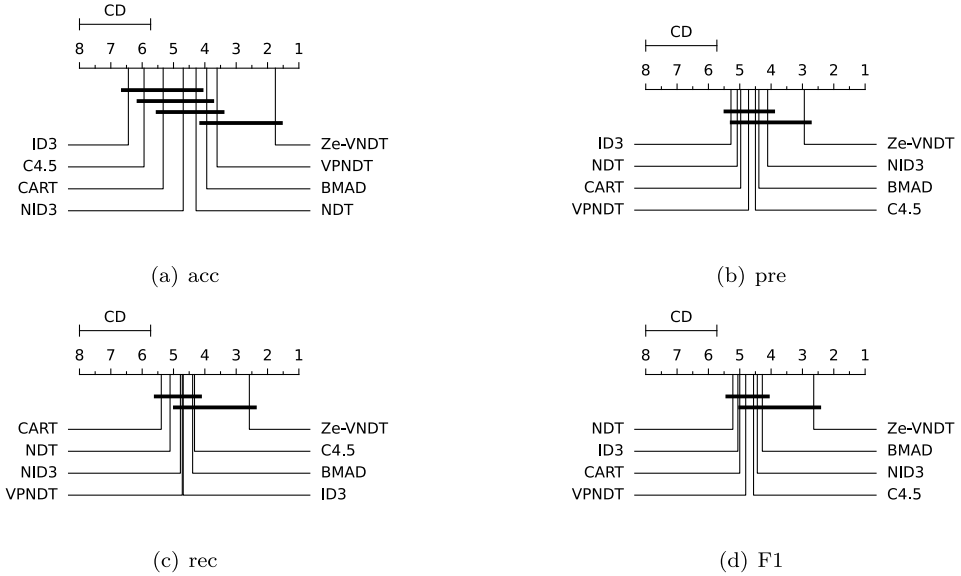
$$\tau_F = \frac{(N-1)\tau_{x^2}}{N(k-1)-\tau_{x^2}}, \quad \tau_{x^2} = \frac{12N}{k(k+1)}\left(\sum_{i=1}^{k} r_i^2 - \frac{k(k+1)^2}{4}\right). \tag{25}$$



(a) acc (b) pre

(c) rec (d) F1

Fig. 5. Friedman test results.



(a) acc (b) pre

(c) rec (d) F1

Fig. 6. Nemenyi test results.

where for any $1 \leq i \leq k$, r_i represents the average rank value of the ith algorithm, and τ_F follows an F-distribution with degrees of freedom $k-1$ and $(k-1)(N-1)$ (Note: in this paper, $N = 18$ and $k = 8$).

The calculated τ_F values in terms of accuracy, precision, recall, and F1-score are 9.7518, 1.6916, 2.3191, and 2.1384, respectively. According to the Friedman test, when $\alpha = 0.1$, the τ_F value is 1.7679. Since all the above results are greater than the critical value, the null hypothesis that "all algorithms have the same performance" is rejected.

Table 19
Rank of precision order values and average order values.

Dataset	ID3	C45	CART	NID3	NDT	VPNDT	BMAD	Ze-VNDT
Crayo	8	5	4	2	6.5	6.5	2	2
Iris	4	2	3	5.5	5.5	7	8	1
wine	2	3	1	6	6	8	4	6
plrx	2	1	5	5	5	5	8	5
wpbc	8	1	6	6	6	3	3	3
Shees	8	2	5	3	6.5	1	4	6.5
seeds	3	4	2	8	6	6	6	1
glass	3	8	7	1.5	4	5.5	5.5	1.5
heart	4	2	8	3	5.5	7	5.5	1
ecoli	8	5	7	2	1	3.5	6	3.5
balance	5	8	7	2	3	4	1	6
glass-id	8	3	1	7	5.5	4	2	5.5
lung	2	6	4	6	6	8	3	1
Breast	1	7	5.5	3	8	5.5	4	2
ENB	6	7	8	2	3	1	4	5
CMC	8	6	7	5	4	2	3	1
magic	8	5	4	3	2	6	7	1
segment	7	6	5	4	8	2	3	1
rank	5.28	4.50	4.97	4.11	5.08	4.72	4.39	2.94

Table 20
Rank order and average rank of recall.

Dataset	ID3	C45	CART	NID3	NDT	VPNDT	BMAD	Ze-VNDT
Crayo	8	5	6	2	7	4	2	2
Iris	3.5	2	3.5	7	5	6	8	1
wine	2	3	1	7	5	8	6	4
plrx	2	1	4.5	6	8	7	3	4.5
wpbc	7	1	8	6	3	5	4	2
Shees	6	2	7	4	8	1	5	3
seeds	2	4	3	8	7	5	6	1
glass	1	7	8	4	5	6	3	2
heart	4	2	8	3	5	7	6	1
ecoli	8	3	6	2	1	4	7	5
balance	8	7	6	4	2	3	1	5
glass-id	2	6	1	8	7	4	5	3
lung	2	5	7	5	5	8	3	1
Breast	1	5	5	5	8	7	3	2
ENB	6	7	8	2	3	1	5	4
CMC	8	6	7	5	3	2	1	4
magic	7	6	4	3	2	5	8	1
segment	7	6	4	5	8	2	3	1
rank	4.69	4.33	5.39	4.78	5.11	4.72	4.39	2.58

Table 21
Rank and average rank of F1.

Dataset	ID3	C45	CART	NID3	NDT	VPNDT	BMAD	Ze-VNDT
Crayo	8	4.5	4.5	2	7	6	2	2
Iris	4	2	3	6	5	7	8	1
wine	2	3	1	7	5	8	6	4
plrx	2	1	4.5	6	8	7	3	4.5
wpbc	8	1	7	6	3	5	4	2
Shees	7	2	6	3	8	1	4	5
seeds	3	4	2	8	7	5	6	1
glass	1	7	8	3	5	6	4	2
heart	4	2	8	3	5	7	6	1
ecoli	8	5	7	2	1	3	6	4
balance	7	8	6	2	3	4	1	5
glass-id	6	5	1	8	7	4	2	3
lung	2	6	4	6	6	8	3	1
Breast	1	6.5	5	3	8	6.5	4	2
ENB	6	7	8	2	3	1	5	4
CMC	8	6	7	5	3	1	2	4
magic	7	6	4	3	2	5	8	1
segment	7	6	4	5	8	2	3	1
rank	5.06	4.56	5.00	4.44	5.22	4.81	4.28	2.64

It can be concluded that there are indeed significant differences in performance among the algorithms. Fig. 5 visually demonstrates these significant differences in performance among the algorithms.

Because the null hypothesis that "all algorithms have the same performance" is rejected, a Nemenyi post-hoc test is needed. The calculation formula is as follows:

$$CD = q_\alpha \sqrt{\frac{k(k-1)}{6N}}. \tag{26}$$

When $k = 8$ and $N = 18$, $q_{\alpha=0.1} = 2.780$. Therefore, according to Eq. (26), we can obtain that CD=2.2699. The results of the Nemenyi post-hoc test are shown in Fig. 6. Although there is no significant difference in some indicators, Ze-VNDT ranks first in accuracy, precision, recall, and F1 score. All results demonstrate that the proposed algorithm performs the best.

5. Conclusions

In decision tree algorithms, a key step in constructing a decision tree is to select appropriate split nodes. Previous attribute partitioning measures were analyzed at a single granularity level, neglecting the connections between different granularity levels. Additionally, the issue of multi-value preferences in decision trees cannot be overlooked. This paper analyzed data uncertainty from different granularity levels, starting from attribute importance, and proposed an improved decision tree algorithm (i.e., Ze-VNDT) that combines variable precision rough sets with Zentropy, to describe the overall and internal uncertainty knowledge. In the experiments, this improved decision tree algorithm was tested on 18 datasets, and the results show that the Ze-VNDT algorithm is effective, with significant improvements in accuracy, precision, recall, and F1 score, indicating its high stability and adaptability. However, considering information from multiple granularity levels increases the computational burden to some extent. Therefore, one of the future work focuses will be on how to reduce the time complexity while maintaining good classification results.

CRediT authorship contribution statement

Hui Dong: Writing – review & editing, Writing – original draft, Software, Methodology, Data curation, Conceptualization. **Caihui Liu:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization. **Xiying Chen:** Software, Data curation. **Duoqian Miao:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by National Natural Science Foundation of China under Grant Nos. : 62566003, 62166001.

Data availability

Data will be made available on request.

References

[1] Z.G. Sun, G.T. Wang, P.F. Li, H. Wang, M. Zhang, X.W. Liang, An improved random forest based on the classification accuracy and correlation measurement of decision trees, Expert. Syst. Appl. 237 (2024) 121549.
[2] V.G. Costa, C.E. Pedreira, Recent advances in decision trees: An updated survey, Artif. Intell. Rev. 56 (5) (2023) 4765–4800.
[3] Y. Feng, H.Z. Li, X.X. Wang, J.X. Tian, Y.Q. Qi, Application of machine learning decision tree algorithm based on big data in intelligent procurement, 2024.
[4] P.H. Putra, A. Azanuddin, B. Purba, Y.A. Dalimunthe, Random forest and decision tree algorithms for car price prediction, J. Mat. Dan Ilmu Pengetah. Alam LLDikti Wil. 1 (JUMPA) 4 (1) (2024) 81–89.

[5] N. Saputra, R. Antoni, A. Widodo, E.M. Solissa, I. Arief, Improving foreign language proficiency in society by decision tree classification, in: AIP Conference Proceedings, vol. 3001, (1) AIP Publishing, 2024.
[6] M. Marudi, I. Ben-Gal, G. Singer, A decision tree-based method for ordinal classification problems, IISE Trans. 56 (9) (2024) 960–974.
[7] O.Z. Maimon, L. Rokach, Data Mining with Decision Trees: Theory and Applications, World scientific, 2014.
[8] J.R. Quinlan, Induction of decision trees, Mach. Learn. 1 (1986) 81–106.
[9] J.R. Quinlan, C4. 5: Programs for Machine Learning, Elsevier, 2014.
[10] L. Breimann, J. Friedman, R. Olshen, C. Stone, Classification and Regression Trees, Pacific Grove, Wadsworth, 1984.
[11] L. Breiman, Classification and Regression Trees, Routledge, 2017.
[12] K. Broelemann, G. Kasneci, A gradient-based split criterion for highly accurate and transparent model trees, 2018, arXiv preprint arXiv:1809.09703.
[13] S.B. Deng, W.L. Yuan, S.J. Guan, X. Lin, Z.M. Liao, M. Li, A decision tree algorithm based on adaptive entropy of feature value importance, Big Data Res. 40 (2025) 100530.
[14] Z.S.H. Alast, M.R. Keyvanpour, Fuzzy-id3 algorithm for breast cancer disease classification, in: 2023 9th International Conference on Web Research, ICWR, IEEE, 2023, pp. 185–188.
[15] Z. Pawlak, Rough sets, Int. J. Comput. Inf. Sci. 11 (1982) 341–356.
[16] C. Wang, Y. Qi, M. Shao, Q.H. Hu, D.G. Chen, Y.H. Qian, A fitting model for feature selection with fuzzy rough sets, IEEE Trans. Fuzzy Syst. 25 (4) (2016) 741–753.
[17] Q. Wang, Y.H. Qian, X.Y. Liang, Q. Guo, J.Y. Liang, Local neighborhood rough set, Knowl.-Based Syst. 153 (2018) 53–64.
[18] C. Luo, S.Z. Wang, T.R. Li, H.M. Chen, J.C. Lv, Z. Yi, Spark rough hypercuboid approach for scalable feature selection, IEEE Trans. Knowl. Data Eng. 35 (3) (2021) 3130–3144.
[19] S. Luo, D. Miao, Z.F. Zhang, Y.J. Zhang, S.D. Hu, A neighborhood rough set model with nominal metric embedding, Inf. Sci. 520 (2020) 373–388.
[20] X. Xie, X.Y. Zhang, W.A.Y. Wang, et al., Neighborhood decision tree construction algorithm based on variable precision neighborhood equivalent granules, Chin. J. Comput. Appl. 42 (2) (2021) 382–388, in Chinese with English Abstract.
[21] C.H. Liu, J.Y. Lai, B.W. Lin, et al., An improved ID3 algorithm based on variable precision neighborhood rough sets, Appl. Intell. 53 (20) (2023) 23641–23654.
[22] B.W. Lin, C.H. Liu, D.Q. Miao, An improved decision tree algorithm based on boundary mixed attribute dependency, Appl. Intell. 54 (2) (2024) 2136–2153.
[23] K.H. Yuan, et al., Feature selection using zentropy-based uncertainty measure, IEEE Trans. Fuzzy Syst. 32 (4) (2023) 2246–2260.
[24] Z.K. Liu, Y. Wang, S.L. Shang, Zentropy theory for positive and negative thermal expansion, J. Phase Equilibria Diffus. 43 (6) (2022) 598–605.
[25] W.H. Xu, Y. Zhang, Y.H. Qian, A novel unsupervised feature selection for high-dimensional data based on FCM and k-nearest neighbor rough sets, IEEE Trans. Neural Netw. Learn. Syst. (2025).
[26] W.H. Xu, M. Huang, Z.Y. Jiang, Y.H. Qian, Graph-based unsupervised feature selection for interval-valued information system, IEEE Trans. Neural Netw. Learn. Syst. 35 (9) (2024) 12576–12589.
[27] W.H. Xu, W.R. Ye, Incremental feature selection: Parallel approach with local neighborhoodrough sets and composite entropy, Pattern Recognit. 159 (2025) 111141.
[28] W.T. Li, H.X. Zhou, W.H. Xu, X.Z. Wang, W. Pedrycz, Interval dominance-based feature selection for interval-valued ordered data, IEEE Trans. Neural Netw. Learn. Syst. 34 (10) (2023) 6898–6912.
[29] B.B. Sang, W.H. Xu, H.M. Chen, T.R. Li, Active antinoise fuzzy dominance rough feature selection using adaptive K-nearest neighbors, IEEE Trans. Fuzzy Syst. 31 (11) (2023) 3944–3958.
[30] K.H. Yuan, D.Q. Miao, W. Pedrycz, et al., Ze-HFS: Zentropy-based uncertainty measure for heterogeneous feature selection and knowledge discovery, IEEE Trans. Knowl. Data Eng. 36 (11) (2024) 7326–7339.
[31] T.Y. Lin, Neighborhood systems: mathematical models of information granulations, in: SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme-System Security and Assurance (Cat. No. 03CH37483), vol. 4, IEEE, 2003, pp. 3188–3193.
[32] Q.H. Hu, et al., Neighborhood rough set based heterogeneous feature subset selection, Inf. Sci. 178 (18) (2008) 3577–3594.
[33] C.H. Liu, B.W. Lin, J.Y. Lai, et al., An improved decision tree algorithm based on variable precision neighborhood similarity, Inf. Sci. 615 (2022) 152–166.
[34] C.E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (3) (1948) 379–423.
[35] Q.H. Hu, et al., Information entropy for ordinal classification, Sci. China Inf. Sci. 53 (2010) 1188–1200.
[36] X. Xie, Information Measure and Decision Tree Based on the Neighborhood Equivalence Relation, Sichuan Normal University, 2023, (in Chinese with English Abstract).
[37] M. Friedman, A comparison of alternative tests of significance for the problem of m rankings, Ann. Math. Stat. 11 (1) (1940) 86–92.
[38] J. Demšar, Statistical comparisons of classifiers over multiple data sets, J. Mach. Learn. Res. 7 (2006) 1–30.
[39] X.Y. Chen, C.H. Liu, B.W. Lin, et al., AHA-3WKM: The optimization of K-means with three-way clustering and artificial hummingbird algorithm, Inf. Sci. 672 (2024) 120661.