Feature Selection Using Generalized Multi-Granulation Dominance Neighborhood Rough Set Based on Weight Partition

Weihua Xu and Qinyuan Bu

***Abstract*—Rough set theory, as an academic hotspot in the field of artificial intelligence, has provided a solid theoretical foundation for feature selection. However, with the continuous updating of large datasets, classical rough set theory is no longer applicable. Multi-granulation rough set theory is an extension of rough set theory that can better handle complex datasets. Therefore, this paper proposes a generalized multi-granulation dominance neighborhood rough set model based on weight distribution and discusses some relevant properties of this model. Furthermore, a new information entropy is constructed based on this model to handle uncertainty in data. This approach enhances the ability to describe uncertainty and enables more effective feature selection. As a result, a forward heuristic feature selection algorithm is developed to find the optimal feature subset. Finally, the effectiveness of the proposed algorithm is validated through instance analysis on twelve publicly available datasets.**

***Index Terms*—Generalized multigranultion rough set, feature selection, information entropy, ordered information system, heuristic feature selection algorithm.**

I. Introduction

RECENTLY, feature selection, as a crucial approach of data preprocessing, has been extensively applied in natural language processing, image processing and data mining. Its primary objective is to identify the most informative and discriminative features from the original set of features to enhance the model's performance and generalization ability. Utilizing feature selection can effectively lower model complexity, mitigate overfitting, and enhance model interpretability by eliminating irrelevant and redundant features [1], [2], [3], [4].

Rough set theory (RST), introduced by Pawlak, is a significant mathematical tool for handling uncertain and fuzzy data, which are pervasive in many real-world applications [5]. Rough set theory is extensively applied in feature selection due to its ability to extract features without relying on any prior knowledge. By relying on equivalence relations, rough set theory is capable of partitioning data objects into distinct equivalence classes, even in scenarios where the relationships between attributes are unknown a priori. This enables the effective analysis and processing of data. Nevertheless, real-world data is typically diverse, and classical rough set theory requires the discretization of continuous data, resulting in information loss and increased computational complexity. Thus, with the deepening of research on rough set theory, scholars have introduced neighborhood relations, dominance relations, and other concepts to overcome the limitations of classical rough sets. As a result, many improved rough set models have been proposed, such as neighborhood rough sets (NRS), dominance rough sets (DRS), variable precision rough sets (VPRS), and multi-granulation rough sets(MRS) [6], [7], [8], [9]. These models aim to overcome the deficiencies of classical rough sets and enhance their practicality. This paper initiates our research by delving into the exploration of the dominance rough set model, neighborhood rough set model, and multi-granulation rough set model.

It is widely acknowledged that RST unavoidably leads to the loss of certain data information and distortion of data when discretizing data. To address this issue, NRS theory introduces the concept of graining neighborhood to replace the equivalence class partition used in RST, thereby avoiding the loss of information caused by the processing of continuous data. Hu et al. put forward neighborhood rough set model and devised a feature subset selection algorithm using neighborhood dependency, demonstrating its flexibility in handling heterogeneous data [10]. Xu et al. introduced a novel composite entropy based on NRS and designed a local heuristic algorithm for selecting the optimal feature subset [11]. Xu et al. integrated the advantages of δ-neighborhood and K-nearest neighbor to build a model that considers heterogeneous data and proposed a highly effective feature selection algorithm [12]. Guo et al. devised a change-based three-way decision model that operates on the basis of confidence level and explored its application in rough set model [13]. In practical problems, object attribute values are not only limited to categorical data, but also include many data sets with partial order relations. The classic rough set model based on equivalence relation is inappropriate for such scenarios. Han et al. proposed an improvement to 3WD models by introducing a linear neighborhood membership-based rough set approach [14]. Sun et al. proposed an adaptive fuzzy multi-neighborhood feature selection method that combines

Manuscript received 7 November 2023; revised 25 March 2024; accepted 23 April 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62376229 and in part by the Natural Science Foundation of Chongqing under Grant CSTB2023NSCQLZX0027. *(Corresponding author: Weihua Xu.)*

The authors are with the College of Artificial Intelligence, Southwest University, Chongqing 400715, China (e-mail: chxuwh@gmail.com; bqy990515@163.com).

Recommended for acceptance by N. Al Moubayed.

Digital Object Identifier 10.1109/TETCI.2024.3398027

distance-based hybrid sampling to address the classification problem in imbalanced datasets [15].

Granular Computing (GrC) [16], as a problem-solving approach that incorporates multiple perspectives and hierarchical levels, has emerged as a rapidly developing discipline. It integrates theories and methods from different fields, such as rough sets and concept lattices. Guo et al. proposed a concept cognitive learning method based on a memory mechanism. By integrating memory, forgetting mechanisms, and concept cognitive learning, this method enables the representation and processing of embedded knowledge in continuous data [17]. Xu et al. proposed a Two-Way Concept-Cognitive Learning (TCCL) approach to enhance the flexibility and evolutionary capability of Two-Way Learning in concept learning [18]. Guo et al. proposed the Fuzzy Granulation Three-Branch Decision Concept Cognitive Learning method, which is employed for concept modeling and dynamic knowledge learning [19]. Both classical rough set models and their extensions can be considered as single granularity models. However, in real-life scenarios, single granularity models have notable limitations when addressing certain problems. For instance, it is clearly unrealistic to perform intersection operations on two decisions that are independent of each other. Qian et al. proposed an extension of Pawlak's rough set model, known as the multi-granulation rough set model (MGRS), and applied this model to practical problems. In this model, the target concept can be characterized using multiple relationships [9]. In order to enhance the problem-solving capability of multi-granulation rough set model, a significant body of scholars has undertaken extensive research and proposed various extended models of multi-granulation rough set. Sun et al. proposed a sparse feature selection method based on local features and high-order label correlations. This method improves multi-label classification performance by considering the relationships between features and labels [20]. Zhang et al. propoesd a methodology for calculating a multi-granulation fusion operator based on matrices and devised an associated dynamic updating algorithm [21]. Sun et al. proposed a multi-objective sparrow search feature selection method that combines sparrow ranking and preference information. This method is applied to high-dimensional data to reduce dimensionality and enhance classification accuracy [22]. Pan et al. conducted research on integrating the additive consistent fuzzy preference relation into the MRS [23]. Kang et al. combined grey system theory with MGRS and designed a variable precision grey multi-granulation rough set [24]. In this paper, we employ the generalized multi-granulation rough set model.

Uncertainty measures [25], as an evaluation of the importance of features and quantification of the numerical characteristics of data inconsistency, plays a pivotal role in feature selection. However, different uncertainty measures have varying capabilities in characterizing attribute dependency and similarity, which can significantly impact feature selection. Zheng et al. investigated the measurement of uncertainty in neighborhood rough set theory and put forth the notion of fuzzy intuitionistic entropy [26]. Liang et al. conducted a study on the measures from RST from the perspective of distance [27]. An et al. integrated relative measure with the lower approximation of fuzzy rough set, presenting a novel relative uncertainty measure [28]. Due to its ability to measure uncertainty or information content, information entropy and its variations are widely applied in the field of feature selection. Xu et al. defined a fuzzy dominated conditional entropy for the fusion of multi-source interval-valued data [29]. Deng et al. introduced dual-similarity and investigated a fresh neighborhood fuzzy entropy to address the label distribution learning problem in feature selection [30].

Inspired by the aforementioned studies, we aim to create a novel information entropy that enhances its capability in measuring uncertainty in feature selection. This paper makes the following key contributions:

1) The concept of the generalized multi-granulation dominance neighborhood rough set is proposed in dominance neighborhood information system. This model allows for the description of objects in the target concept from multiple perspectives. Based on this, weights are assigned to each knowledge granule, and the knowledge granules within a threshold range set are combined into a granularity. The generalized multi-granulation dominance neighborhood rough set model is then constructed based on weight distribution.
2) To improve the capability of describing uncertainty in information entropy, the generalized multi-granulation dominance neighborhood rough set information entropy is designed to describe the features of a dataset. Compared to existing methods, it demonstrates better classification performance.
3) The designed generalized multi-granulation dominance neighborhood rough set information entropy, based on uncertainty measurement and decision distribution, enhances the capability of describing uncertainty and enables better feature selection. Based on this, a new forward heuristic feature selection algorithm is designed, which significantly improves classification accuracy and avoids unnecessary computations, thereby enhancing efficiency.

The subsequent sections of this paper are organized as follows: Section II introduces some relevant concepts and related works. Building upon this foundation, in Section III, we present a method for calculating the weights of knowledge granules and construct the generalized multi-granulation dominance neighborhood rough set model based on weight distribution. We also discuss the relevance properties of this model. Section IV delves into the generalized multi-granulation dominance neighborhood rough set based on weight distribution information entropy and its accompanying properties. Section V presents a heuristic algorithm applied to feature selection and analyzes its complexity. To assess the effectiveness of the proposed algorithm, Section VI presents experimental results on 12 publicly available datasets, highlighting its efficacy and robustness in feature selection. Finally, Section VII summarizes the work presented in this paper and provides an outlook for future research. The framework of this paper can be clearly seen from Fig. 1.

II. Related Work

This section reviews some basic concepts about RST, dominance neighborhood rough set, multi-granulation rough set,

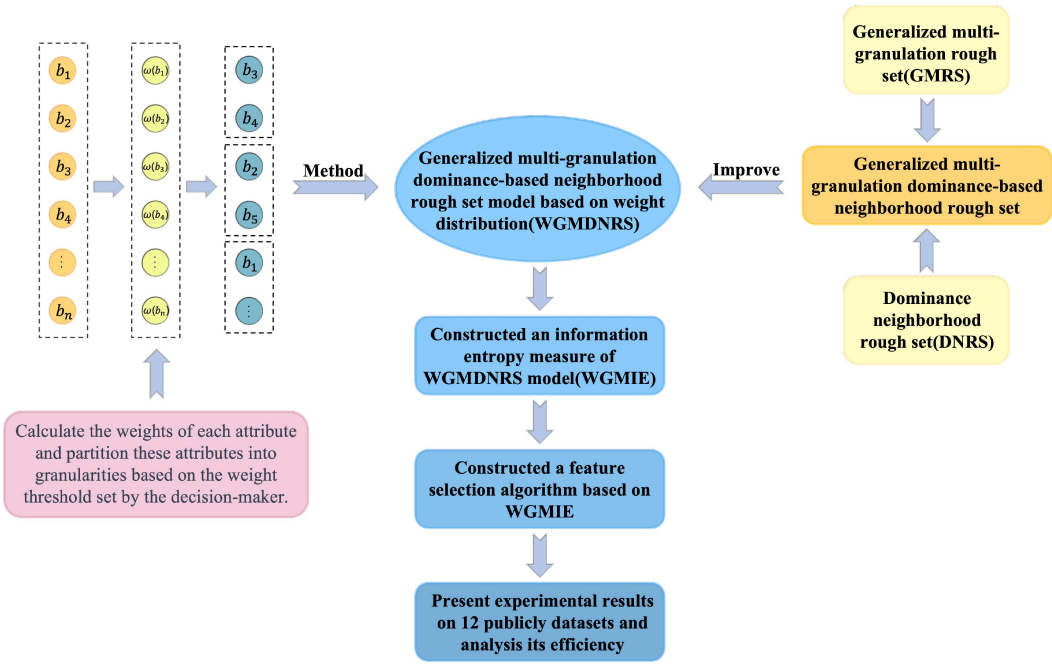


Fig. 1. The framework of this paper.

generalized multi-granulation rough set and some measures of rough set model.

A. Rough Set

RST is a mathematical tool that effectively handles data uncertainty [31]. It uses upper and lower approximations to describe the uncertainty of data projects. The definition of RST [5] is as follows:

Given an information system $IS = (U, N)$, where $U = \{x_1, x_2, \ldots, x_n\}$ denotes a nonempty and finite objects set, the attribute set $N = AT \cup DT$, in which $AT = \{b_1, b_2, \ldots, b_s\}$ is the conditional attribute set and $DT = \{d\}$ is the decision attribute set, and additionally, $AT \cap DT = \emptyset$. If R is an equivalence relation on U, then for any target concept $X \subseteq U$, the upper approximation $\overline{R}(X)$ and the lower approximation $\underline{R}(X)$ of X on R are defined by

$$\begin{aligned} \underline{R}(X) &= \left\{x \in U \,\middle|\, [x]_R \subseteq X\right\}, \\ \overline{R}(X) &= \left\{x \in U \,\middle|\, [x]_R \cap X \neq \emptyset\right\}, \end{aligned} \tag{1}$$

where $[x]_R$ is the equivalence classes of object x with respect to R. Among them, X is a definable set when $\underline{R}(X) = \overline{R}(X)$. On the contrary, X is called a rough set.

B. Dominance Neighborhood Rough Set

Let $IS = (U, AT \cup DT)$ be an information system, for $\forall x \in U$ and $\forall b \in AT$, the $f(x, b)$ represents the attribute value of object x for attribute b. Then IS is a neighborhood information system [10]. In IS, we choose to represent the distance between x_i and x_j in the attribute set $B \subseteq AT$ using the following distance function:

$$d_B(x_i, x_j) = \left(\sum_{b \in B} |f(x_i, b) - f(x_j, b)|^s\right)^{\frac{1}{s}}. \tag{2}$$

In this paper, Euclidean distance is employed to compute the distance. Given a neighborhood radius, δ, the neighborhood classes of $x \in U$ on R_B is defined as following:

$$[x]_\delta^{R_B} = \left\{x_k \,\middle|\, d_B(x, x_k) \leq \delta, x_k \in U\right\}. \tag{3}$$

Furthermore, when $\delta = 0$, the neighborhood rough set degenerates into classical rough set.

Let $OIS^{\succeq} = (U, AT \cup DT, g)$ be an ordered information system(OIS), where the conditional attribute set AT and the decision attribute set DT have a dominance relationship. In this case, the $OIS^{\succeq}$ is referred to as an ordered information system [7]. For $\forall b \in AT$, establishing a partial order relation $\succeq_b$ on the value range of b. $x \succeq_b y \Leftrightarrow g(x, b) \geq g(y, b)$ represents that object x is superior to object y according to criterion b, and it is also an incrementing partial order, where $\forall x, y \in U$. Similarity, $x \succeq_b y \Leftrightarrow g(x, b) \leq g(y, b)$ represents a decreasing partial order. In this paper, we only consider the incrementing partial order. For $B \subseteq AT$, Let $[x_i]_B^{\succeq}$ denote the dominance set of object x_i with respect to B, which is defined as

$$[x_i]^{\succeq_B} = \left\{x_j \in U \,\middle|\, g(x_i, b) \leq g(x_j, b), \forall b \in B\right\}. \tag{4}$$

In feature selection, NRS can directly select attributes from numerical decision tables, eliminating the need for the discretization process of numerical data and ensuring classification performance. DRS can effectively handle the inconsistency in decision analysis caused by preference attributes. Therefore, we

choose to combine them and utilize the dominance neighborhood rough set model(DNRS) to process data.

Let $OIS^{\succeq}=(U,AT\cup DT,g)$ be an OIS, for $\forall B\subseteq AT$, $\forall b\in B$, $x\in U$, the dominance neighborhood class [32] of object x_i under the dominance relation $\succeq_B$ is defined as

$$[x_i]_{\delta}^{\succeq_B}=\{x_j\in U\,|\,d_B(x_i,x_j)\leq\delta\wedge g(x_i,b)\leq g(x_j,b),\forall b\in B\}. \tag{5}$$

Moreover, the dominance neighborhood relation on B is defined as

$$\mathbb{R}_{\delta}^{\succeq_B}=\{(x_i,x_j)\in U\times U\,|\,d_B(x_i,x_j)\leq\delta\wedge g(x_i,b)\leq g(x_j,b),\forall b\in B\}. \tag{6}$$

Let $OIS^{\succeq}=(U,AT\cup DT,g)$ be an OIS, for $\forall B\subseteq AT$, $X\subseteq U$, $\overline{\mathbb{R}_{\delta}^{\succeq_B}}(X)$ and $\underline{\mathbb{R}_{\delta}^{\succeq_B}}(X)$ respectively represent the lower and upper approximations of X under $\mathbb{R}_{\delta}^{\succeq_B}$. Their definitions are as follows:

$$\begin{aligned}\underline{\mathbb{R}_{\delta}^{\succeq_B}}(X)&=\left\{x\in U\,|\,[x_i]_{\delta}^{\succeq_B}\subseteq X\right\},\\ \overline{\mathbb{R}_{\delta}^{\succeq_B}}(X)&=\left\{x\in U\,|\,[x_i]_{\delta}^{\succeq_B}\cap X\neq\emptyset\right\}.\end{aligned} \tag{7}$$

III. Generalized Multi-Granulation Dominance Neighborhood Rough Set Based on Weight Distribution

When grouping granules based on their weights, the primary task is to assign weight values to each knowledge granule. In the following, we will introduce the method for calculating the weights of knowledge granules.

A. Multi-Granulation Dominance Neighborhood Rough Set(MDNRS)

In practical problems, a domain is often partitioned by multiple relations rather than a single one. The classical rough set theory is inadequate in addressing this situation. Xu et al. proposed the concepts of pessimistic and optimistic upper and lower approximations under the framework of MGRS. We have extended and applied these concepts to DNRS.

Definition 1: Let $OIS^{\succeq}=(U,AT\cup DT)$ be an OIS, for $\forall N_i\subseteq AT(i=1,2,\ldots,m\leq 2^{|AT|})$, $X\subseteq U$. The $[x]_{\delta}^{\succeq_{N_i}}$ is the dominance neighborhood class. Then the upper and lower approximations of X under $\mathbb{R}_{\delta}^{\succeq_B}$ in optimistic MDNRS are respectively defined as

$$\begin{aligned}\overline{\mathbb{OM}_{\delta}^{\succeq\sum_{i=1}^{m}N_i}}(X)&=\left\{x\in U\,\middle|\,\bigwedge_{i=1}^{m}\left([x]_{\delta}^{\succeq_{N_i}}\cap X\neq\emptyset\right)\right\},\\ \underline{\mathbb{OM}_{\delta}^{\succeq\sum_{i=1}^{m}N_i}}(X)&=\left\{x\in U\,\middle|\,\bigvee_{i=1}^{m}\left([x]_{\delta}^{\succeq_{N_i}}\subseteq X\right)\right\},\end{aligned} \tag{8}$$

and the upper and lower approximations of X in pessimistic MDNRS are respectively defined as

$$\begin{aligned}\underline{\mathbb{PM}_{\delta}^{\succeq\sum_{i=1}^{m}N_i}}(X)&=\left\{x\in U\,\middle|\,\bigwedge_{i=1}^{m}\left([x]_{\delta}^{\succeq_{N_i}}\subseteq X\right)\right\},\\ \overline{\mathbb{PM}_{\delta}^{\succeq\sum_{i=1}^{m}N_i}}(X)&=\left\{x\in U\,\middle|\,\bigvee_{i=1}^{m}\left([x]_{\delta}^{\succeq_{N_i}}\cap X\neq\emptyset\right)\right\}.\end{aligned} \tag{9}$$

In this definition, the symbols "$\bigwedge$" and "$\bigvee$" are uesd to represent "and" and "or" respectively. Moreover, X is considered optimistic and precise when $\underline{\mathbb{OM}_{\delta}^{\succeq\sum_{i=1}^{m}N_i}}(X)=\overline{\mathbb{OM}_{\delta}^{\succeq\sum_{i=1}^{m}N_i}}(X)$. Conversely, X is optimistic and rough. The same applies to pessimistic MDNRS.

B. Generalized Multi-Granulation Dominance Neighborhood Rough Set(GMDNRS)

The MDNRS proposed in Definition 1 represents a specific type where the optimistic MDNRS is overly permissive and the pessimistic MDNRS is overly strict in depict approximations. Furthermore, both of these rough set overlook the principle of majority rule, which is frequently encountered in real-life situations. To overcome these limitations, this paper investigates the generalized multi-granulation dominance neighborhood rough set(GMDNRS), which building upon GMRS. To introduce this model, we first define a feature function.

Let $OIS^{\succeq}=(U,AT\cup DT)$ be an OIS, for $\forall N_i\subseteq AT(i=1,2,\ldots,m\leq 2^{|AT|})$, $\forall x\in U$, $X\subseteq U$, the support feature function $S_X^{\delta,\succeq_{N_i}}(x)$ is defined to describe the inclusion relationship between the dominance neighborhood class $[x]_{\delta}^{\succeq_{N_i}}$ and X as follows:

$$S_X^{\delta,\succeq_{N_i}}(x)=\begin{cases}1, & [x]_{\delta}^{\succeq_{N_i}}\subseteq X,\\ 0, & others.\end{cases}\quad(i\leq 2^{|AT|}) \tag{10}$$

According to the given support feature function, for optimistic MDNRS, approximations are obtained only when $S_X^{\delta,\succeq_{N_i}}(x)=1$, which is too strict and may introduce unnecessary information. Similarly, for pessimistic MDNRS, the condition is too loose to accurately characterize concepts. Therefore, we propose GMDNRS, which utilizes a parameter β to control the selectivity of objects. A smaller value of β implies a looser requirement.

Definition 2: Let $OIS^{\succeq}=(U,AT\cup DT)$ be an OIS, for $\forall N_i\subseteq AT(i=1,2,\ldots,m\leq 2^{|AT|})$, $X\subseteq U$, $\beta\in(0.5,1]$, and the support feature function $S_X^{\delta,\succeq_{N_i}}(x)$, the lower and upper approximations of X with respect to N_i are defined as follows:

$$\begin{aligned}\underline{\mathbb{GM}_{\delta,\beta}^{\succeq\sum_{i=1}^{m}N_i}}(X)&=\left\{x\in U\,\middle|\,\frac{\sum_{i=1}^{m}S_X^{\delta,\succeq_{N_i}}(x)}{m}\geq\beta\right\},\\ \overline{\mathbb{GM}_{\delta,\beta}^{\succeq\sum_{i=1}^{m}N_i}}(X)&=\left\{x\in U\,\middle|\,\frac{\sum_{i=1}^{m}\left(1-S_{\sim X}^{\delta,\succeq_{N_i}}(x)\right)}{m}>1-\beta\right\}.\end{aligned} \tag{11}$$

When $\underline{\mathbb{GM}_{\delta,\beta}^{\succeq\sum_{i=1}^{m}N_i}}(X)\neq\overline{\mathbb{GM}_{\delta,\beta}^{\succeq\sum_{i=1}^{m}N_i}}(X)$, X is a rough set with respect to $\sum_{i=1}^{m}N_i$. $<\underline{\mathbb{GM}_{\delta,\beta}^{\succeq\sum_{i=1}^{m}N_i}}(X)$,

TABLE I
AN ORDERED INFORMATION SYSTEM

U	b_1	b_2	b_3	b_4	b_5	b_6	d
x_1	0.36	0.61	0.34	0.54	0.80	0.63	1
x_2	0.46	0.46	0.42	0.57	0.76	0.87	1
x_3	0.45	0.57	0.43	0.55	0.83	0.21	1
x_4	0.35	0.57	0.33	0.63	0.71	0.65	1
x_5	0.36	0.76	0.34	0.52	0.72	0.43	1
x_6	0.44	0.34	0.41	0.61	0.46	0.76	2
x_7	0.40	0.45	0.36	0.73	0.38	0.54	2
x_8	0.42	0.84	0.38	0.57	0.40	0.98	2
x_9	0.36	0.78	0.34	0.72	0.62	0.32	2
x_{10}	0.40	0.69	0.36	0.71	0.72	0.12	2

$\overline{\mathbb{GM}^{\succeq\sum_{i=1}^{m}N_i}_{\delta,\beta}}(X)>$ is called GMDNRS, β is referred to as the information level regarding $\sum_{i=1}^{m}N_i$.

It is worth noting that there is no explicitly fixed inclusion relationship between $\overline{\mathbb{GM}^{\succeq\sum_{i=1}^{m}N_i}_{\delta,\beta}}(X)$ and $\underline{\mathbb{R}^{\succeq B}_{\delta}}$, and the same applies to $\overline{\mathbb{GM}^{\succeq\sum_{i=1}^{m}N_i}_{\delta,\beta}}(X)$ and $\overline{\mathbb{R}^{\succeq B}_{\delta}}$ as well.

Accuracy is an important measure for describing the precision of rough set concepts. It does not require any numerical assumptions beforehand and only considers the degree of approximation from above and below. The higher the accuracy value, the stronger the ability to handle uncertainty [33]. The measure in this paper is based on the GMDNRS, which is positively correlated with the approximation ability.

Example 1: Table I is an ordered information system, where the object set is $U=\{x_1,x_2,\ldots,x_{10}\}$, the conditional attribute set is $AT=\{b_1,b_2,b_3,b_4,b_5,b_6\}$, the decision attribute set is $DT=\{d\}$, and the decision partition is denoted as $U/DT=\{D_1,D_2\}$. Assuming that we divide the set of conditional attribute set into three granules $N_1=\{b_1,b_2\}, N_2=\{b_3,b_4\}$, and $N_3=\{b_5,b_6\}$, a neighborhood radius of δ, we first obtain the distance matrices M_1, M_2, and M_3 corresponding to the three granules.

For $N_1=\{b_1,b_2\}$, the distance matrix M_1 is:

$$M_1=\begin{bmatrix}
0 & 0.18 & 0.10 & 0.04 & 0.15 & 0.28 & 0.17 & 0.24 & 0.17 & 0.09\\
0.18 & 0 & 0.11 & 0.16 & 0.32 & 0.12 & 0.06 & 0.38 & 0.34 & 0.24\\
0.10 & 0.11 & 0 & 0.10 & 0.21 & 0.23 & 0.13 & 0.27 & 0.23 & 0.13\\
0.04 & 0.16 & 0.10 & 0 & 0.19 & 0.25 & 0.13 & 0.28 & 0.21 & 0.13\\
0.15 & 0.32 & 0.21 & 0.19 & 0 & 0.43 & 0.31 & 0.10 & 0.02 & 0.09\\
0.28 & 0.12 & 0.23 & 0.25 & 0.43 & 0 & 0.12 & 0.50 & 0.45 & 0.35\\
0.17 & 0.06 & 0.13 & 0.13 & 0.31 & 0.12 & 0 & 0.39 & 0.33 & 0.24\\
0.23 & 0.38 & 0.27 & 0.28 & 0.10 & 0.50 & 0.39 & 0 & 0.09 & 0.15\\
0.17 & 0.34 & 0.23 & 0.21 & 0.02 & 0.45 & 0.33 & 0.09 & 0 & 0.10\\
0.10 & 0.24 & 0.13 & 0.13 & 0.08 & 0.35 & 0.24 & 0.15 & 0.10 & 0
\end{bmatrix}.$$

Let $\delta=0.16$, based on the M_1, we can obtain that the dominance neighborhood classes on N_1 are:

$$\begin{aligned}
&[x_1]_{\delta}^{\succeq N_1}=\{x_1,x_5,x_{10}\}, [x_2]_{\delta}^{\succeq N_1}=\{x_2\},\\
&[x_3]_{\delta}^{\succeq N_1}=\{x_3\}, [x_4]_{\delta}^{\succeq N_1}=\{x_1,x_3,x_4,x_{10}\},\\
&[x_5]_{\delta}^{\succeq N_1}=\{x_5,x_8,x_9\}, [x_6]_{\delta}^{\succeq N_1}=\{x_2,x_6\},\\
&[x_7]_{\delta}^{\succeq N_1}=\{x_2,x_3,x_7\}, [x_8]_{\delta}^{\succeq N_1}=\{x_8\},\\
&[x_9]_{\delta}^{\succeq N_1}=\{x_8,x_9\}, [x_{10}]_{\delta}^{\succeq N_1}=\{x_8,x_{10}\}.
\end{aligned}$$

For $N_2=\{b_3,b_4\}$, the distance matrix M_2 is:

$$M_2=\begin{bmatrix}
0 & 0.09 & 0.09 & 0.09 & 0.02 & 0.10 & 0.19 & 0.05 & 0.18 & 0.17\\
0.09 & 0 & 0.02 & 0.11 & 0.09 & 0.04 & 0.17 & 0.04 & 0.17 & 0.15\\
0.09 & 0.02 & 0 & 0.13 & 0.10 & 0.06 & 0.19 & 0.05 & 0.19 & 0.18\\
0.09 & 0.11 & 0.13 & 0 & 0.11 & 0.08 & 0.10 & 0.08 & 0.09 & 0.09\\
0.02 & 0.09 & 0.10 & 0.11 & 0 & 0.11 & 0.21 & 0.06 & 0.20 & 0.19\\
0.10 & 0.04 & 0.06 & 0.08 & 0.11 & 0 & 0.13 & 0.05 & 0.13 & 0.11\\
0.19 & 0.17 & 0.19 & 0.10 & 0.21 & 0.13 & 0 & 0.16 & 0.02 & 0.02\\
0.05 & 0.04 & 0.05 & 0.08 & 0.06 & 0.05 & 0.16 & 0 & 0.16 & 0.14\\
0.18 & 0.17 & 0.19 & 0.09 & 0.20 & 0.13 & 0.02 & 0.16 & 0 & 0.02\\
0.17 & 0.15 & 0.18 & 0.09 & 0.19 & 0.11 & 0.02 & 0.14 & 0.02 & 0
\end{bmatrix}.$$

In the same way, we can obtain that the dominance neighborhood classes on N_2 based on the M_2 are:

$$\begin{aligned}
&[x_1]_{\delta}^{\succeq N_2}=\{x_1,x_2,x_3,x_6,x_8\}, [x_2]_{\delta}^{\succeq N_2}=\{x_2\},\\
&[x_3]_{\delta}^{\succeq N_2}=\{x_3\}, [x_4]_{\delta}^{\succeq N_2}=\{x_4,x_7,x_9,x_{10}\},\\
&[x_5]_{\delta}^{\succeq N_2}=\{x_1,x_2,x_3,x_5,x_6,x_8\}, [x_6]_{\delta}^{\succeq N_2}=\{x_6\},\\
&[x_7]_{\delta}^{\succeq N_2}=\{x_7\}, [x_8]_{\delta}^{\succeq N_2}=\{x_2,x_6,x_8\},\\
&[x_9]_{\delta}^{\succeq N_2}=\{x_7,x_9\}, [x_{10}]_{\delta}^{\succeq N_2}=\{x_7,x_{10}\}.
\end{aligned}$$

For $N_3=\{b_5,b_6\}$, the distance matrix M_3 is:

$$M_3=\begin{bmatrix}
0 & 0.24 & 0.42 & 0.09 & 0.22 & 0.36 & 0.43 & 0.53 & 0.36 & 0.52\\
0.24 & 0 & 0.66 & 0.22 & 0.44 & 0.32 & 0.50 & 0.38 & 0.57 & 0.75\\
0.42 & 0.66 & 0 & 0.46 & 0.25 & 0.66 & 0.56 & 0.88 & 0.24 & 0.14\\
0.09 & 0.23 & 0.46 & 0 & 0.22 & 0.27 & 0.35 & 0.45 & 0.34 & 0.53\\
0.22 & 0.44 & 0.25 & 0.22 & 0 & 0.42 & 0.36 & 0.64 & 0.15 & 0.31\\
0.36 & 0.32 & 0.66 & 0.27 & 0.42 & 0 & 0.23 & 0.23 & 0.47 & 0.69\\
0.43 & 0.50 & 0.56 & 0.35 & 0.36 & 0.23 & 0 & 0.44 & 0.33 & 0.54\\
0.53 & 0.38 & 0.88 & 0.45 & 0.64 & 0.23 & 0.44 & 0 & 0.70 & 0.92\\
0.36 & 0.57 & 0.24 & 0.34 & 0.15 & 0.47 & 0.33 & 0.70 & 0 & 0.22\\
0.51 & 0.75 & 0.14 & 0.53 & 0.31 & 0.69 & 0.54 & 0.92 & 0.22 & 0
\end{bmatrix}.$$

The dominance neighborhood classes on N_3 based on the M_3 are:

$$\begin{aligned}
&[x_1]_{\delta}^{\succeq N_3}=\{x_1\}, [x_2]_{\delta}^{\succeq N_3}=\{x_2\}, [x_3]_{\delta}^{\succeq N_3}=\{x_3\},\\
&[x_4]_{\delta}^{\succeq N_3}=\{x_4\}, [x_5]_{\delta}^{\succeq N_3}=\{x_5\}, [x_6]_{\delta}^{\succeq N_3}=\{x_6,x_8\},\\
&[x_7]_{\delta}^{\succeq N_3}=\{x_7\}, [x_8]_{\delta}^{\succeq N_3}=\{x_8,x_9\}, [x_9]_{\delta}^{\succeq N_3}=\{x_2,x_9\},\\
&[x_{10}]_{\delta}^{\succeq N_3}=\{x_3,x_{10}\}.
\end{aligned}$$

Let $\beta=0.6$, according to (10) and (11), we can calculate the upper and lower approximations of $D_i(i=1,2)$ as follows:

$$\begin{aligned}
&\overline{\mathbb{GM}^{\succeq\sum_{i=1}^{m}N_i}_{\delta,\beta}}(D_1)=\{x_1,x_2,x_3,x_4,x_5\};\\
&\overline{\mathbb{GM}^{\succeq\sum_{i=1}^{m}N_i}_{\delta,\beta}}(D_2)=\{x_1,x_4,x_5,x_6,x_7,x_8,x_9,x_{10}\}.\\
&\underline{\mathbb{GM}^{\succeq\sum_{i=1}^{m}N_i}_{\delta,\beta}}(D_1)=\{x_2,x_3\},\\
&\underline{\mathbb{GM}^{\succeq\sum_{i=1}^{m}N_i}_{\delta,\beta}}(D_2)=\{x_6,x_7,x_8,x_9,x_{10}\},
\end{aligned}$$

However, the GMDNRS model simply divides the knowledge granules based on their quantity without considering the quality of the granules. In real-life situations, the importance of each knowledge granule differs. For highly important knowledge granules, their corresponding weights are larger, while less

important knowledge granules have smaller weights. The impact of multiple low-weight knowledge granules on decision cannot be considered equivalent to the impact of a single high-weight knowledge granule. Therefore, we propose a screening and combination approach for knowledge granules based on their weights. Based on the weights associated with each knowledge granule, different threshold sets are clustered, and knowledge granules within the threshold set range are combined to form a granularity, ensuring the practicality of the decision-making process.

C. The Generation of Weights

Let $OIS^{\succeq} = (U, AT \cup DT, g)$ be an OIS, where the decision attribute set $DT = \{d\}$, the conditional attribute set $AT = \{b_1, b_2, \ldots, b_s\}$. For $\forall b \in AT$, $\forall x \in U$, $g(x, b)$ represents the value of an object x under b. Let the coefficient matrix of AT be

$$M_{AT} = \begin{bmatrix} g(x_1, b_1) & g(x_1, b_2) & \cdots & g(x_1, b_s) \\ g(x_2, b_1) & g(x_2, b_2) & \cdots & g(x_2, b_s) \\ \vdots & \vdots & \vdots & \vdots \\ g(x_n, b_1) & g(x_n, b_2) & \cdots & g(x_n, b_s) \end{bmatrix},$$

the partition coefficients of AT be

$$\gamma = \big(\gamma(b_1), \gamma(b_2), \ldots, \gamma(b_s)\big)^T,$$

the vector of DT be

$$M_{DT} = \big(g(x_1, d), g(x_2, d), \ldots, g(x_n, d)\big)^T.$$

To find the optimal partition coefficients, we will solve the following optimization problem:

$$\gamma^* = argmin \|M_{AT}\gamma - M_{DT}\|^2, \tag{12}$$

where the 2-norm of a vector is represented as $\| \bullet \|^2$. To solve 12, we first assume $M_{AT}\gamma = M_{DT}$ and then multiply both sides by $(M_{AT})^T$, resulting in $(M_{AT})^T M_{AT}\gamma = (M_{AT})^T M_{DT}$. Finally, by solving the equation, we can obtain $\gamma = [(M_{AT})^T M_{AT}]^{-1}(M_{AT})^T M_{DT}$.

Specifically, when the matrix $(M_{AT})^T M_{AT}$ is non invertible matrix or a penalty term is needed in 12, we convert the 12 to: $p(\gamma) = \|M_{AT}\gamma - M_{DT}\|^2 + \|\gamma\|^2$. Since $p(\gamma)$ is a convex function, when $p'(\gamma) = 0$, we can obtain its minimum value, which implies $p'(\gamma) = (M_{AT})^T(M_{AT}\gamma - M_{DT}) + \gamma = 0$. Then we can get $[(M_{AT})^T M_{AT} + E]\gamma = (M_{AT})^T M_{DT}$, where E is an identify matrix. In the end, $\gamma = [(M_{AT})^T M_{AT} + E]^{-1}(M_{AT})^T M_{DT}$. If $(M_{AT})^T M_{AT}$ or $[(M_{AT})^T M_{AT} + E]$ is high-dimensional or nearly ill-conditioned, we no longer calculate the inverse matrix, but instead use the "np.linalg.solve" function from Numpy to solve it.

Definition 3: Let $OIS^{\succeq} = (U, AT \cup DT)$ be an OIS, for $\forall b \in AT$, the weight definition of b is as follows:

$$\omega(b) = \frac{|\gamma(b)|}{\sum_{b_s \in AT} |\gamma(b_s)|}, \tag{13}$$

where $|\gamma(b)|$ is the absolute value of $\gamma(b)$, reflecting the relationship between conditional attribute b and decision attribute d. A larger value signifies a more intense innate relationship between the two.

Property 1: Let $OIS^{\succeq} = (U, AT \cup DT)$ be an OIS, for $\forall b \in AT$, the weight-vector for conditional attributes $\omega = \big(\omega(b_1), \omega(b_2), \ldots, \omega(b_s)\big)^T$, there are:

1) $\omega(b) \geq 0$;
2) $\sum_{b_i \in AT} = |AT|$.

Proof 1: According to Definition 3, 1) and 2) can be directly proven.

From (13), we can conclude the weight of conditional attribute b. The higher the weight value, the higher its correlation with the decision attribute.

D. Generalized Multi-Granulation Dominance Neighborhood Rough Set Based on Weight Distribution

Definition 4: Let $OIS^{\succeq} = (U, AT \cup DT)$ be an OIS, for $\forall b \in AT$, Let the weight-vector for knowledge granules $\omega = \big(\omega(b_1), \omega(b_2), \ldots, \omega(b_s)\big)^T$, where $\sum_{i=1}^{s} \omega(b_i) = 1$. The mapping function h form the weight of knowledge granule to its corresponding granule, is defined as follows:

$$h(\omega(b_s)) = b_s. \tag{14}$$

Let $OIS^{\succeq} = (U, AT \cup DT)$ be an OIS, for $\forall b \in AT$, Let the mapping function h, the weight-vector for knowledge granules $\omega = \big(\omega(b_1), \omega(b_2), \ldots, \omega(b_s)\big)^T$, the threshold set $\alpha' = \{\alpha_1, \alpha_2, \ldots, \alpha_p\}$, where $\sum_{i=1}^{s} \omega(b_i) = 1$, $\forall \alpha_q \subseteq (0, 1]$. The distribution of the set of knowledge granules is defined as:

$$G_i = \left\{ h(w(b_m)) \in AT \,\middle|\, w(b_m) \in \alpha_q \right\}, \tag{15}$$

where $G \subseteq AT, 0 \leq |G| \leq |AT|$, and the "$|\bullet|$" represents the number of knowledge granules. The threshold set is determined based on Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [34]. It utilizes grid search to find the optimal neighborhood radius and density threshold, organizing data points into high-density regions and considering low-density regions as noise points for clustering, thus determining the partitioning of the threshold set.

After obtaining the distribution of knowledge granules based on weights, we apply it to GMDNRS to construct the generalized multi-granulation dominance neighborhood rough set model based on weight distribution(WGMDNRS), and provide its definition and some related properties.

Let $OIS^{\succeq} = (U, AT \cup DT)$ be an OIS, for $\forall b \in AT, X \subseteq U$, let the mapping function h, the weight-vector for knowledge granules $\omega = \big(\omega(b_1), \omega(b_2), \ldots, \omega(b_s)\big)^T$, sets of knowledge granules $G_i \subseteq AT (i = 1, 2, \ldots, m \leq 2^{|AT|})$, $\beta \in (0.5, 1]$, the support feature function $S_X^{\delta, \succeq_{G_i}}(x)$ is defined as follows:

$$S_X^{\delta, \succeq_{G_i}}(x) = \begin{cases} 1, & [x]_\delta^{\succeq_{G_i}} \subseteq X, \\ 0, & others, \end{cases} \quad (i \leq 2^{|AT|}) \tag{16}$$

the lower and upper approximations of X with respect to G_i are defined as follows:

$$\underline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta,\omega}(X)=\left\{x\in U\,\middle|\,\frac{\sum\limits_{i=1}^{m}S_X^{\delta,\succeq_{G_i}}(x)}{m}\geq\beta\right\},$$

$$\overline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta,\omega}(X)=\left\{x\in U\,\middle|\,\frac{\sum\limits_{i=1}^{m}\left(1-S_{\sim X}^{\delta,\succeq_{G_i}}(x)\right)}{m}>1-\beta\right\}. \tag{17}$$

When $\underline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta,\omega}(X)\neq\overline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta,\omega}(X)$, X is a rough set. $<\underline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta,\omega}(X),\overline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta,\omega}(X)>$ is called generalized multi-granulation dominance neighborhood rough set model based on weight distribution, β is referred to as the information level regarding $\sum_{i=1}^{m}G_i$.

In the actual decision-making process, the selection and distribution of weights are usually related to the practical significance of the decision. Therefore, it is necessary to choose attributes with higher weights, indicating their greater importance. The distribution of weights should ideally align with the importance of attributes in the practical problem to ensure that the weight selection and distribution are effective and scientifically grounded. This model not only allows for controlling the number of granules that must be satisfied in approximating the target decision but also enables control over the quality of the granules that must be satisfied.

Property 2: Let $OIS^{\succeq}=(U,AT\cup DT)$ be an OIS, for $\forall b\in AT, X\subseteq U$, let the weight-vector for knowledge granules $\omega=\big(\omega(b_1),\omega(b_2),\ldots,\omega(b_s)\big)^T$, sets of knowledge granules $G_i\subseteq AT(i=1,2,\ldots,m\leq 2^{|AT|}),\beta\in(0.5,1]$, the following properties hold true:

1) $\underline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta}(\sim X)=\sim\overline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta,\omega}(X)$;
2) $X\subseteq Y\Rightarrow\underline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta,\omega}(X)\subseteq\underline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta}(Y)$;
3) $\underline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta}(X\cap Y)\subseteq\underline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta,\omega}(X)\cap\underline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta}(Y)$.

Proof 2:

1) From $x\in\overline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta,\omega}(X)\Leftrightarrow\frac{\sum_{i=1}^{m}(1-S_{\sim X}^{\delta,\succeq_{G_i}}(x))}{m}>1-\beta$, it can be obtained that: $x\in\sim\overline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta,\omega}(X)\Leftrightarrow\frac{\sum_{i=1}^{m}(1-S_X^{\delta,\succeq_{G_i}}(x))}{m}\leq 1-\beta\Leftrightarrow\frac{\sum_{i=1}^{m}S_X^{\delta,\succeq_{G_i}}(x)}{m}\geq\beta\Leftrightarrow x\in\overline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta,\omega}(X)$. Thus, 1) is proved.
2) For $\forall x\in\underline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta,\omega}(X)$, there exists $\frac{\sum_{i=1}^{m}S_X^{\delta,\succeq_{G_i}}(x)}{m}\geq\beta$. Also, from $X\subseteq Y$, it follows that $S_X^{\delta,\succeq_{G_i}}\leq S_Y^{\delta,\succeq_{G_i}}$. Thus it can be concluded that $\frac{\sum_{i=1}^{m}S_Y^{\delta,\succeq_{G_i}}(x)}{m}\geq\frac{\sum_{i=1}^{m}S_X^{\delta,\succeq_{G_i}}(x)}{m}\geq\beta$, resulting in $x\in\underline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta,\omega}(Y)$. So 2) is proved.
3) Based on the property of $S_X^{\delta,\succeq_{G_i}}(x)$, for $\forall x\in\underline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta,\omega}(X\cap Y)$, it can prove that $x\in\underline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta,\omega}(X\cap Y)\Leftrightarrow\frac{\sum_{i=1}^{m}S_{X\cap Y}^{\delta,\succeq_{G_i}}(x)}{m}=\frac{\sum_{i=1}^{m}S_X^{\delta,\succeq_{G_i}}(x)\wedge\sum_{i=1}^{m}S_Y^{\delta,\succeq_{G_i}}(x)}{m}\geq\beta\Leftrightarrow\frac{\sum_{i=1}^{m}S_X^{\delta,\succeq_{G_i}}(x)}{m}\geq\beta$ and $\frac{\sum_{i=1}^{m}S_Y^{\delta,\succeq_{G_i}}(x)}{m}\geq\beta\Leftrightarrow x\in\underline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta,\omega}(X)$ and $x\in\underline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta,\omega}(Y)\Leftrightarrow x\in\underline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta,\omega}(X)\cap\underline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta,\omega}(Y)$. Thus, 3) is proved.

Property 3: Let $OIS^{\succeq}=(U,AT\cup DT)$ be an OIS, for $\forall b\in AT, X\subseteq U$, let the weight-vector for knowledge granules $\omega=\big(\omega(b_1),\omega(b_2),\ldots,\omega(b_s)\big)^T$, sets of knowledge granules $G_i\subseteq AT(i=1,2,\ldots,m\leq 2^{|AT|}),\beta\in(0.5,1]$, the following properties hold true:

1) $\underline{\mathbb{PM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\omega}(X)\subseteq\underline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta,\omega}(X)\subseteq\underline{\mathbb{OM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\omega}(X)$;
2) $\overline{\mathbb{OM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\omega}(X)\subseteq\overline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta,\omega}(X)\subseteq\overline{\mathbb{PM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\omega}(X)$.

Proof 3: According to Definitions 4, 1) and 2) can be directly proven.

Property 4: Let $OIS^{\succeq}=(U,AT\cup DT)$ be an OIS, for $\forall b\in AT, X\subseteq U$, let the weight-vector for knowledge granules $\omega=\big(\omega(b_1),\omega(b_2),\ldots,\omega(b_s)\big)^T$, sets of knowledge granules $G_i\subseteq AT(i=1,2,\ldots,m\leq 2^{|AT|}),\ \beta\in(0.5,1],\ t\leq m$, the following properties hold true:

1) $\underline{\mathbb{GM}}^{\succeq\sum_{i=1}^{t}G_i}_{\delta,\beta,\omega}(X)\subseteq\underline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta,\omega}(X)$;
2) $\overline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta,\omega}(X)\subseteq\overline{\mathbb{GM}}^{\succeq\sum_{i=1}^{t}G_i}_{\delta,\beta,\omega}(X)$.

Proof 4: According to Definitions 4 and property 3, 1) and 2) can be directly proven.

Example 2: Let's continue to Example 1. To demonstrate the effectiveness of WGMDNRS, let's reviewing Example 1. The ω can be calculated by using (13), where $\omega=(0.3635,0.0648,0.1258,0.2192,0.1926,0.0341)^T$. According to Definition 4, we applied the DBSCAN algorithm to cluster all the knowledge granules and obtained three granules: $G_1=\{b_2,b_6\},G_2=\{b_3,b_4,b_5\},G_3=\{b_1\}$. Then we can calculate the upper and lower approximations of $D_i(i=1,2)$ as follows:

$$\overline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta}(D_1)=\{x_1,x_2,x_3,x_4,x_5\};$$
$$\overline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta}(D_2)=\{x_4,x_6,x_7,x_8,x_9,x_{10}\}.$$
$$\underline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta}(D_1)=\{x_1,x_2,x_3,x_5\},$$
$$\underline{\mathbb{GM}}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta}(D_2)=\{x_6,x_7,x_8,x_9,x_{10}\},$$

For the granules distraction in Example 1 and Example 2, we use KNN(K-Nearest Neighbors) classifier to classify the unknown region, and the results are shown in the figure. In Fig. 2, $(a)-(c)$ correspond to the classification results of $N_1=\{b_1,b_2\},N_2=\{b_3,b_4\},N_3=\{b_5,b_6\}$ in Example 1, and $(d)-(f)$ correspond to the classification results of $G_1=\{b_2,b_6\},G_2=\{b_3,b_4,b_5\},G_3=\{b_1\}$ in

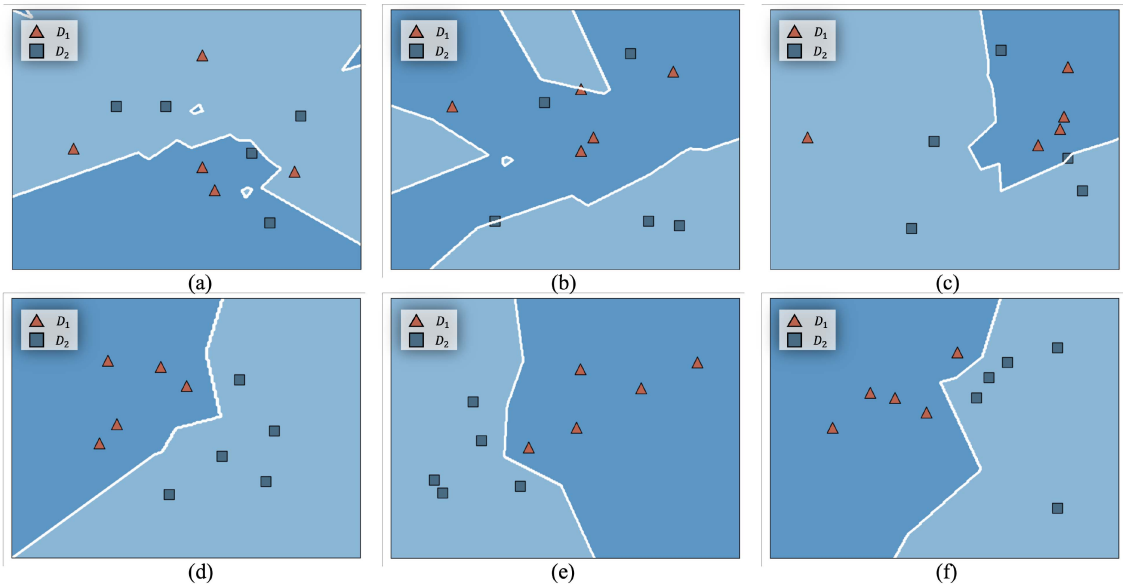


Fig. 2. Classification results under KNN on N_1, N_2, N_3 and G_1, G_2, G_3.

Example 2. From the figure, for Example 1, it is evident that the decision boundary is not particularly clear, and the objects are not correctly classified. However, for Example 2, each object is correctly classified, and the decision boundary is also very clear. It is evident that the granularity partitioning based on the weight of attributes yields excellent results. Therefore, we will choose to construct a new entropy based on WGMDNRS.

IV. An Information Entropy Measure of WGMDNRS

Since Shannon initially introduced the concept of utilizing information entropy to assess the uncertainty within a discrete sample space, entropy and its extensions have found extensive application across diverse fields. In the context of feature selection, entropy plays a crucial role [29], [35], [36]. However, in real-life scenarios, our data often exhibit imbalances or missing values, and the evaluation of features can be influenced by the distribution of such data. Therefore, in this paper, we propose an improved information entropy based on the WGMDNRS model(WGMIE) to measure rough set models, taking into account the distribution characteristics of the data.

Definition 5: Let $OIS^{\succeq}=(U, AT\cup DT)$ be an OIS, for $B\subseteq AT$, $G_i\subseteq B(i=1,2,\ldots,m\leq 2^{|B|})$, $\forall X\in U$, $\beta\in(0.5,1]$, $\forall D_i\in U/DT$. Let neighborhood radius δ, $\omega=\big(\omega(b_1),\omega(b_2),\ldots,\omega(b_s)\big)^T$, the upper approximation $\overline{\mathbb{GM}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta,\omega}}(X)$ and lower approximation $\underline{\mathbb{GM}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta,\omega}}(X)$. Then the information entropy based on the WGMDNRS model $\mathcal{W}(B,DT)$ is defined as:

$$\mathcal{W}(B,DT)=-\sum_{i=1}^{s}\frac{|D_i|}{|U|}log\frac{\left|\underline{\mathbb{GM}^{\succeq\sum_{i=1}^{m}DG_i}_{\delta,\beta}}_{\omega}(D_i)\right|}{\left|\overline{\mathbb{GM}^{\succeq\sum_{i=1}^{m}DG_i}_{\delta,\beta}}^{\omega}(D_i)\right|}. \tag{18}$$

Property 5: Let $OIS^{\succeq}=(U, AT\cup DT)$ be an OIS, for $B\subseteq AT$, $DG_i\subseteq B(i=1,2,\ldots,m\leq 2^{|B|})$, $\beta\in(0.5,1]$, the following properties hold true:

1) If $\frac{|\underline{\mathbb{GM}^{\succeq\sum_{i=1}^{m}DG_i}_{\delta,\beta,\omega}}(D_i)|}{|\overline{\mathbb{GM}^{\succeq\sum_{i=1}^{m}DG_i}_{\delta,\beta,\omega}}(D_i)|}=1(i=1,2,\ldots,s)$, then $\mathcal{W}(B, DT)=0$;
2) If $\frac{|\underline{\mathbb{GM}^{\succeq\sum_{i=1}^{m}DG_i}_{\delta,\beta,\omega}}(D_i)|}{|\overline{\mathbb{GM}^{\succeq\sum_{i=1}^{m}DG_i}_{\delta,\beta,\omega}}(D_i)|}=0(i=1,2,\ldots,s)$, then $\mathcal{W}(B, DT)\to\infty$.

Proof 5: According to Definition 5, 1) and 2) can be directly proven.

Property 6: Let $OIS^{\succeq}=(U, AT\cup DT)$ be an OIS. For $B\subseteq AT$, $DG_i\subseteq B(i=1,2,\ldots,m\leq 2^{|B|})$, $\beta\in(0.5,1]$. Let $B\subseteq B'$, $DG_i{}'\subseteq B'(i=1,2,\ldots,m\leq 2^{|B'|})$, then $\mathcal{W}(B,DT)\geq\mathcal{W}(B',DT)$.

Proof 6: From $B\subseteq B'$, then $\underline{\mathbb{GM}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta,\omega}}(D_i)\subseteq\underline{\mathbb{GM}^{\succeq\sum_{i=1}^{m}G_i{}'}_{\delta,\beta,\omega}}(D_i)$ and $\overline{\mathbb{GM}^{\succeq\sum_{i=1}^{m}G_i{}'}_{\delta,\beta,\omega}}(D_i)\subseteq\overline{\mathbb{GM}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta,\omega}}(D_i)$ can be obtained, we can further get $\frac{|\underline{\mathbb{GM}^{\succeq\sum_{i=1}^{m}G_i{}'}_{\delta,\beta,\omega}}(D_i)|}{|\overline{\mathbb{GM}^{\succeq\sum_{i=1}^{m}G_i{}'}_{\delta,\beta,\omega}}(D_i)|}\geq\frac{|\underline{\mathbb{GM}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta,\omega}}(D_i)|}{|\overline{\mathbb{GM}^{\succeq\sum_{i=1}^{m}G_i}_{\delta,\beta,\omega}}(D_i)|}$. Finally, according to Definition 5, $\mathcal{W}(B,DT)\geq\mathcal{W}(B',DT)$ is proved.

Example 3: Let's continue to Example 2. Since we have already calculated the accuracy values in Example 1 and Example 2, the values of $\mathcal{W}(B,DT)$ before and after applying the weight distribution are as follows:

$$\mathcal{W}(B,DT)=1.00,\qquad \mathcal{W}(B^{\omega},DT)=0.61.$$

The smaller the value of information entropy, the lower the uncertainty in the information. Therefore, WGMIE will select the attributes after using weight distribution to partition the granules. This also demonstrates the feasibility of using WGMDNRS model.

V. Feature Selection Algorithm Based on WGMIE

The WGMIE combines the approximation ability of the target and provides a comprehensive reflection of the rough set model's approximation capability on a particular feature subset. With the increasing number of features, more information is provided, and the accuracy of the RST improves. Consequently, the value of $\mathcal{W}(B, DT)$ will decrease.

We can evaluate the importance of an attribute b with respect to feature subset $B(B \subseteq AT)$ by calculating the value of $\mathcal{W}(B \cup \{b\}, DT)$. A smaller value of $\mathcal{W}(B \cup \{b\}, DT)$ indicates a greater reduction in uncertainty, which implies that the attribute is more important in the process of approximation. It is known that there may be multiple reduction sets, but in some cases, a single attribute reduction is sufficient. To tackle this issue, we employ a heuristic forward greedy feature selection algorithm to identify features that possess similar approximation abilities to the original data. To achieve this, we firstly introduce interior and exterior significance measures that characterize each feature as follows.

Definition 6: Let $OIS^{\succeq} = (U, AT \cup DT)$ be an OIS. For $\forall b \in AT$, the interior significance measure of b in relation to AT is defined as following:

$$\mathcal{S}_{in}(b, AT, DT) = \mathcal{W}(AT - \{b\}, DT) - \mathcal{W}(AT, DT). \tag{19}$$

When the value of $\mathcal{S}_{in}(b, AT, DT)$ is higher, it indicates a larger increase in information entropy compared to $\mathcal{W}(AT, DT)$. This suggests that feature b is relatively most important to AT. Consequently, b is selected firstly which satisfying $argmax_{b \in AT}\, \mathcal{S}_{in}(b, AT, DT)$.

Let $OIS^{\succeq} = (U, AT \cup DT)$ be an OIS, $B \subseteq AT$. For $\forall b \in AT - B$, the exterior significance measure of b in relation to B is defined as following:

$$\mathcal{S}_{out}(b, B, DT) = \mathcal{W}(B, DT) - \mathcal{W}(B \cup \{b\}, DT). \tag{20}$$

When the value of $\mathcal{S}_{out}(b, B, DT)$ is higher, the decrease in $\mathcal{W}(B, DT)$ is greater, indicating that feature b is important relative to B.

To enhance the efficiency of feature selection in datasets, a heuristic forward algorithm is implemented for selecting the important feature subset. The algorithm selects the first feature, b_o, that satisfies $argmax_{b \in AT}\, \mathcal{S}_{in}(b, AT, DT)$ and subsequently identifies other excellent features based on the maximum principle of indicator $\mathcal{S}_{out}(b, B, DT)$. It is noteworthy that the importance of feature b relative to B increases with the larger value of $\mathcal{S}_{out}(b, B, DT)$. Feature b cannot be selected if $\mathcal{S}_{out}(b, B, DT) \to 0$. In this paper, we let $\nu = 0.06$, when the judgment condition is not met, the selection process will terminate. The pseudocode for the selection process is shown in Algorithm 1.

Algorithm 1: Feature Selection Algorithm Based on WGMIE(W-GMDA).

Input: $OIS^{\succeq} = (U, AT \cup DT)$, δ, β, threshold $\alpha' = \{\alpha_1, \alpha_2, \ldots, \alpha_p\}$; where $\forall \alpha_q \subseteq (0, 1]$, $\beta \in (0.5, 1]$, $U/DT = \{D_1, D_2, \ldots, D_n\}$.
Output: Feature subset B.
1: Initialize $B \leftarrow \emptyset$.
2: Compute the weights of each attribute in the condition set based on Definitoin 3.
3: Partition the conditional attributes into granulations based on their weights.
4: Calculate the $\mathcal{W}(AT, DT)$ for feature set AT.
5: Calculate the $\mathcal{W}(AT - \{b\}, DT)$ for $b \in AT$.
6: Select $b_o = argmax_{b \in AT}\, \mathcal{S}_{in}(b, AT, DT)$.
7: Let $B \leftarrow b_o$.
8: **while** flag $= 1$ **do**
9: **for** each $b \in AT - B$ **do**
10: **for** $i = 1 : n$ **do**
11: Calculate $\frac{\mathbb{GM}^{\succeq}_{\delta,\beta,\omega}{}^{\sum_{i=1}^{m} G_i}(D_i)}{\overline{\mathbb{GM}^{\succeq}_{\delta,\beta,\omega}{}^{\sum_{i=1}^{m} G_i}(D_i)}}$;
12: **end for**
13: Calculate $\mathcal{W}(B \cup \{b\}, DT)$ and $\mathcal{S}_{out}(b, B, DT)$;
14: **end for**
15: Select $b_t = argmax_{b \in AT - B}\, \mathcal{S}_{out}(b, B, DT)$;
16: **if** $\mathcal{S}_{out}(b_t, B, DT) \geq \nu$ **then**
17: $B = B \cup \{b_t\}$;
18: $flag = 1$;
19: **else**
20: $flag = 0$;
21: **end if**
22: **end while**
23: **return:** B.

Moving forward, we will examining the time complexity of the W-GMDA. Here, m denotes the quantity of conditional attributes and n denotes the amount of objects. The weights of each attribute are calculated and granularity is partitioned based on these weights in steps 2-3. The time complexity is $O(n * m + m^3 + m * \log m)$. Steps 4-7 first calculate the entropy value of AT, then calculate the entropy values corresponding to each attribute after it is removed, and finally select the attribute b_o based on the interior significance measure $\mathcal{S}_{in}$, and use it as the feature subset B. The time complexity is $O(n^2 * m + n^2 + n * m)$. The time complexity for computing the corresponding upper and lower approximations to obtain the exterior significance measure $\mathcal{S}_{out}$ in steps 9-15 is $O(m * n + n^2 * m)$. The time complexity for steps 16-21 is $O(n)$.

VI. Experimental Decision and Analysis

In this section, we selected twelve datasets from the UCI Machine Learning Repository and conducted a series of experiments as illustrated in Table II, to showcase the efficacy of W-GMDA. The algorithm was implemented using Python 3.9. All experiments were executed on a single PC with an Apple

TABLE II
THE DETAILED INTRODUCTION OF THE DATASET

No.	Dataset	Abbreviation	Objects	Features	Classes	Type
set1	Appendicitis	App	106	8	2	Real
set2	Wine	Wine	178	13	3	Real
set3	Parkinsons	Par	197	23	2	Real
set4	Wisconsin Prognostic Breast Cancer	WPBC	198	33	2	Real
set5	Whole Scale Customers	WSC	440	8	3	Real
set6	Climate Model Simulation Crashes	CMSC	540	18	2	Real
set7	Wisconsin Diagnostic Breast Cancer	WDBC	569	31	2	Real
set8	HCV	HCV	615	14	4	Real
set9	Maternal Health Risk	MHR	1014	7	3	Real
set10	Wireless Indoor Localization	WIL	2000	8	4	Real
set11	Cardiotocography	Car	2126	23	3	Real
set12	Abalone	Aba	4178	9	3	Real

M1 CPU, 16GB of RAM, and macOS Ventura 13.6 operating system. Three classifiers, namely K-Nearest Neighbors(KNN), Random Forest(RF), and Decision Tree(DT), were employed to evaluate the classification accuracy of the feature selection algorithm.

A. Experimental Design

Firstly, we investigated the impact of different neighborhood radii on the clustering effectiveness of the algorithm and selected the optimal neighborhood radius based on the reduction results from twelve datasets. Secondly, we examined the influence of different threshold values, B, on the clustering performance of the algorithm. Additionally, we compared the classification accuracy of the W-GMDA with six different algorithms to validate its effectiveness. Lastly, we designed a p-value test method to verify the significance of the W-GMDA. We normalized the values of the conditional attributes to the range [0,1]. The selection of the neighborhood radius significantly affects the results of feature selection. Hence, we varied the parameter δ from 0.10 to 0.30 with a step size of 0.02, parameter β from 0.55 to 0.85 with a step size of 0.15 Based on experimental results, we identified the optimal neighborhood radius. Furthermore, we employed three classical classifiers, KNN, RF, DT, to evaluate the classification accuracy and further validate the superiority of the W-GMDA. In the classification experiment, we implemented a ten-fold cross-validation methodology, whereby the dataset was randomly partitioned into ten mutually exclusive subsets. In each iteration, one of these subsets was designated as the validation set, while the remaining nine subsets were used for model training and evaluation. This process was repeated ten times, with each iteration utilizing a different subset as the validation set. Ultimately, the mean of the ten validation results was computed and adopted as the final evaluation metric. Subsequently, we conducted a comparative analysis between the W-GMDA and six feature selection algorithms based on RST. The specific details are outlined below:

1) Infinite Feature Selection (INF-FS) [37]: The feature subset is regarded as paths in a graph, where features are represented as nodes and pairwise relationships between features are represented as edges. By employing different interpretations, the values of paths of arbitrary lengths are evaluated, ultimately leading to infinity. Subsequently, an unsupervised strategy is proposed for feature selection.
2) h-Rough Degree Method(θ-M) [38]: Through the given θ-similarity rate, similarity relationships are constructed, defining h-similarity classes. The concept of θ-roughness is introduced to quantify uncertainty, feature selection based on this measure.
3) Feature Selection Based On α-Approximate Equal Relation(AER) [39]: This algorithm is an unsupervised attribute reduction approach for IVIS, which is constructed by combining the α-based fuzzy similarity class approximation equivalence with mutual information.
4) k-nearest Neighborhood Conditional Mutual Information Method(KNI) [12]: By amalgamating the strengths of the δ-neighborhood and K-nearest neighbors methodologies, an iterative strategy is employed to define the importance of features, and information entropy is utilized for feature selection.
5) The Weighted Rough Set Method(WDNCE) [40]: By initially assigning weights to features, a matrix-based conditional entropy is subsequently proposed to evaluate the significance of attributes and perform feature selection.
6) The Hybrid Fernel Fuzzy Complementary Entropy Approach(HKI). [41]: A hybrid kernel function fuzzy complementary entropy is proposed, which serves as the basis for constructing uncertainty measures. Three unsupervised feature selection criteria are defined, and a feature selection algorithm is designed.

B. Experimental Analysis

In this section, we aim to demonstrate the effectiveness of W-GMDA by selecting the two optimal parameters, δ and β, and comparing its classification results with those of six other algorithms.

Before commencing the experiments, we conducted data preprocessing by employing the technique of max-min scaling normalization to rescale the data within the range of 0 to 1. The procedure is as follows:

$$b(x_j) = \frac{b(x_j) - \min b(x)}{\max b(x) - \min b(x)},$$

where $\min b(x)$ and $\max b(x)$ represent the maximum and minimum values of attribute b, respectively.

For handling missing values in the dataset, we utilized the K-Nearest Neighbors algorithm to estimate the missing values

TABLE III
CLASSIFICATION ACCURACY(%) UNDER DIFFERENT NEIGHBORHOOD RADIUS δ ON KNN CLASSIFIER

Dataset	0.10	0.12	0.14	0.16	0.18	0.20	0.22	0.24	0.26	0.28	0.30
App	86.00±0.51	86.00±0.51	86.00±0.34	**87.91±0.99**	87.91±0.99	87.91±0.99	87.91±0.99	87.91±0.99	87.91±0.99	87.91±0.99	87.91±0.99
Wine	85.46±0.68	85.46±0.68	85.46±0.68	**94.38±0.25**	80.98±0.73	80.98±0.73	72.45±0.96	72.45±0.96	71.31±0.98	87.71±0.29	71.31±0.98
Par	**85.08±0.67**	85.08±0.67	85.08±0.67	81.95±0.73	81.95±0.73	81.95±0.73	81.95±0.73	81.95±0.73	81.95±0.73	81.95±0.73	81.95±0.73
WPBC	76.79±0.90	78.34±1.34	77.29±0.70	75.34±0.85	75.79±0.82	75.34±0.85	**78.34±1.34**	75.26±0.68	77.82±0.65	77.82±0.65	77.82±0.65
WSC	89.55±0.14	89.55±0.14	89.55±0.14	**89.55±0.14**	89.09±0.11	89.09±0.11	89.09±0.11	89.55±0.11	89.55±0.11	89.55±0.14	89.55±0.14
CMSC	92.41±0.18	92.41±0.18	92.41±0.18	**93.15±0.18**	92.04±0.20	92.04±0.20	92.04±0.20	92.04±0.20	91.67±0.25	92.59±0.16	92.22±0.14
WDBC	92.45±0.18	92.45±0.18	92.45±0.18	91.74±0.19	**92.49±0.15**	89.45±0.16	92.27±0.16	92.44±0.15	92.44±0.15	92.44±0.15	92.44±0.15
HCV	91.06±0.14	91.06±0.14	92.04±0.10	**92.04±0.10**	92.04±0.10	92.04±0.10	92.04±0.10	91.06±0.14	91.06±0.14	91.06±0.14	91.06±0.14
MHR	71.61±0.25	61.74±0.13	67.66±0.17	67.66±0.17	**73.38±0.19**	71.61±0.25	74.26±0.16	71.61±0.25	71.61±0.25	71.61±0.25	71.61±0.25
WIL	94.20±0.03	95.05±0.04	98.05±0.01	**98.05±0.01**	89.15±0.09	89.15±0.09	95.10±0.02	89.15±0.09	89.15±0.09	95.90±0.02	95.90±0.02
Car	91.16±0.03	91.16±0.03	96.38±0.01	96.38±0.01	96.38±0.01	91.16±0.03	91.16±0.03	91.21±0.04	**96.99±0.01**	96.85±0.01	96.69±0.01
Aba	46.23±0.32	49.01±0.35	49.65±0.27	**57.41±0.31**	53.52±0.35	49.57±0.36	50.42±0.41	50.37±0.27	49.01±0.37	48.69±0.30	47.28±0.46
Avg	83.50±0.32	83.25±0.35	84.34±0.27	**85.42±0.31**	83.51±0.35	82.63±0.36	83.13±0.41	82.13±0.37	82.70±0.37	84.69±0.30	82.28±0.36

The best results are highlighted in bold entities.

TABLE IV
CLASSIFICATION ACCURACY(%) UNDER DIFFERENT NEIGHBORHOOD RADIUS δ ON RF CLASSIFIER

Dataset	0.10	0.12	0.14	0.16	0.18	0.20	0.22	0.24	0.26	0.28	0.30
App	87.82±0.69	87.82±0.69	88.73±0.82	**88.73±0.82**	86.82±0.73	88.73±0.82	88.73±0.82	86.82±0.73	86.82±0.73	86.82±0.73	88.73±0.82
Wine	97.22±0.14	97.75±0.13	97.78±0.14	**98.89±0.13**	97.78±0.26	97.22±0.26	98.33±0.13	97.78±0.14	97.22±0.20	97.22±0.14	97.22±0.20
Par	91.74±0.44	91.71±0.29	**92.74±0.40**	91.24±0.16	90.71±0.26	90.71±0.15	91.24±0.16	90.71±0.26	90.71±0.15	90.18±0.24	91.24±0.22
WPBC	80.74±0.37	80.24±0.55	79.76±0.42	80.26±0.69	80.76±0.41	79.26±0.64	**81.76±0.48**	80.74±0.52	80.24±0.45	80.24±0.45	80.21±0.62
WSC	90.68±0.16	90.91±0.17	91.14±0.18	**91.36±0.19**	91.36±0.41	91.14±0.24	91.14±0.24	90.91±0.17	90.68±0.17	90.45±0.18	90.23±0.17
CMSC	92.41±0.13	**92.59±0.14**	92.22±0.13	92.04±0.13	92.41±0.13	92.22±0.14	92.22±0.14	92.22±0.14	92.59±0.13	92.41±0.13	91.67±0.13
WDBC	95.08±0.10	95.43±0.09	95.26±0.10	95.26±0.07	**95.96±0.11**	94.91±0.08	95.79±0.08	95.08±0.12	95.26±0.07	95.08±0.09	94.91±0.08
HCV	93.17±0.09	93.49±0.09	94.32±0.10	94.64±0.10	94.15±0.04	**94.96±0.08**	93.99±0.07	93.33±0.11	93.01±0.12	92.68±0.11	92.52±0.12
MHR	76.25±0.28	63.12±0.20	69.04±0.21	68.45±0.20	**76.94±0.16**	75.96±0.39	75.75±0.19	76.55±0.34	76.25±0.28	76.15±0.26	76.05±0.16
WIL	93.55±0.03	94.15±0.02	97.10±0.02	**97.25±0.02**	88.70±0.07	88.95±0.05	93.25±0.01	88.80±0.08	88.90±0.06	94.35±0.02	94.30±0.02
Car	92.95±0.04	93.46±0.03	94.40±0.02	97.46±0.02	**97.60±0.01**	93.09±0.04	92.99±0.04	90.45±0.03	97.41±0.02	94.69±0.01	93.13±0.01
Aba	50.95±0.35	51.33±0.52	51.92±0.37	**55.61±0.31**	55.54±0.20	54.66±0.32	53.58±0.26	51.33±0.11	52.41±0.28	51.32±0.25	51.05±0.62
Avg	86.88±0.24	86.00±0.24	87.03±0.24	**87.59±0.24**	87.30±0.23	86.82±0.27	87.30±0.22	86.22±0.23	86.79±0.22	86.86±0.22	87.85±0.26

The best results are highlighted in bold entities.

TABLE V
CLASSIFICATION ACCURACY(%) UNDER DIFFERENT NEIGHBORHOOD RADIUS δ ON DT CLASSIFIER

Dataset	0.10	0.12	0.14	0.16	0.18	0.20	0.22	0.24	0.26	0.28	0.30
App	86.91±0.84	86.91±0.84	86.91±0.84	**87.00±0.84**	87.00±0.67	86.91±0.84	86.91±0.84	86.91±0.84	86.91±0.84	86.91±0.84	86.91±0.84
Wine	89.35±0.55	89.90±0.43	89.35±0.35	91.50±0.61	92.12±0.26	91.57±0.39	**93.27±0.36**	92.68±0.44	89.90±0.85	87.68±0.43	91.60±0.88
Par	**92.76±0.23**	92.26±0.18	92.26±0.23	83.39±0.64	83.39±0.64	82.89±0.55	83.39±0.64	82.89±0.55	82.89±0.55	83.39±0.64	83.39±0.64
WPBC	**76.79±0.50**	76.79±0.50	74.79±0.58	76.29±0.54	75.29±0.41	76.79±0.50	76.29±0.54	78.29±0.45	75.26±0.47	74.76±0.69	74.76±0.59
WSC	89.77±0.11	90.00±0.11	90.23±0.10	90.23±0.10	**91.82±0.10**	91.59±0.10	91.59±0.10	90.45±0.18	90.00±0.17	89.77±0.11	89.77±0.11
CMSC	91.11±0.10	91.11±0.10	90.93±0.13	**91.30±0.14**	91.30±0.09	91.11±0.12	91.11±0.12	91.11±0.12	90.93±0.10	90.93±0.13	90.93±0.13
WDBC	92.09±0.11	92.27±0.14	92.27±0.11	93.32±0.11	**93.67±0.13**	93.32±0.12	93.50±0.12	93.32±0.12	92.79±0.18	93.32±0.13	93.32±0.12
HCV	91.55±0.14	91.55±0.14	93.33±0.09	**93.33±0.09**	93.33±0.09	93.33±0.09	93.17±0.08	91.55±0.14	91.39±0.14	91.23±0.15	91.07±0.13
MHR	60.76±0.36	51.78±0.17	61.54±0.23	61.54±0.23	**65.49±0.09**	60.76±0.36	65.49±0.09	60.76±0.36	60.76±0.36	60.76±0.36	60.76±0.36
WIL	90.45±0.01	90.60±0.02	96.65±0.01	**96.65±0.01**	83.40±0.03	83.40±0.03	90.45±0.01	83.40±0.03	83.40±0.03	90.60±0.02	90.60±0.02
Car	90.59±0.02	90.59±0.02	98.21±0.01	**98.40±0.01**	98.40±0.01	90.59±0.02	90.59±0.02	91.16±0.03	98.40±0.00	98.40±0.01	98.21±0.03
Aba	47.19±0.22	49.78±0.21	49.32±0.34	**54.68±0.37**	54.27±0.14	53.77±0.45	53.05±0.08	51.33±0.21	50.73±0.41	50.04±0.35	49.32±0.20
Avg	83.28±0.27	82.80±0.24	84.64±0.25	**84.80±0.30**	84.12±0.22	83.01±0.30	84.07±0.25	82.82±0.28	82.78±0.34	83.15±0.32	83.39±0.34

The best results are highlighted in bold entities.

based on the available neighboring samples. The similarity between a missing value sample and other samples was computed using the Euclidean distance. By calculating the similarity measure, we identified the K most similar samples as the nearest neighbors to the missing value sample. Finally, the missing values were imputed using the mean of the nearest neighbor samples.

Classification accuracy is commonly used to measure the performance of a classification task and is considered one of the most effective metrics. To avoid the impact of data sparsity and computational randomness on experimental results, The classification accuracy of W-GMDA on various datasets is averaged and presented in rows labeled as "Avg". in designed tables. Also, the experimental results of raw data are represented as "Raw". The left side of the "±" represents the mean, and the right side represents the variance. The optimal classification performance is denoted in bold.

Tables III, IV, V records the classification accuracy of the W-GMDA based on different neighborhood radius δ in KNN, RF, and DT classifiers. The corresponding classification accuracy bar chart is shown in Fig. 3. In this figure, the blue bar represents the KNN classifier, the red bar represents the RF classifier, and the green bar represents the DT classifier. By comparing the classification accuracy shown in Tables III, IV, and V, the following conclusions can be drawn. The classification performance is relatively better when δ falls within the range of 0.14 to 0.18. Specifically, when $\delta = 0.16$, it achieves the highest classification accuracy on seven datasets under the KNN classifier, five datasets under the RF classifier, and six datasets under the DT classifier. Considering the average classification accuracy across twelve datasets, $\delta = 0.16$ yields the highest accuracy for all three classifiers. Therefore, we determine the optimal neighborhood radius to be 0.16.

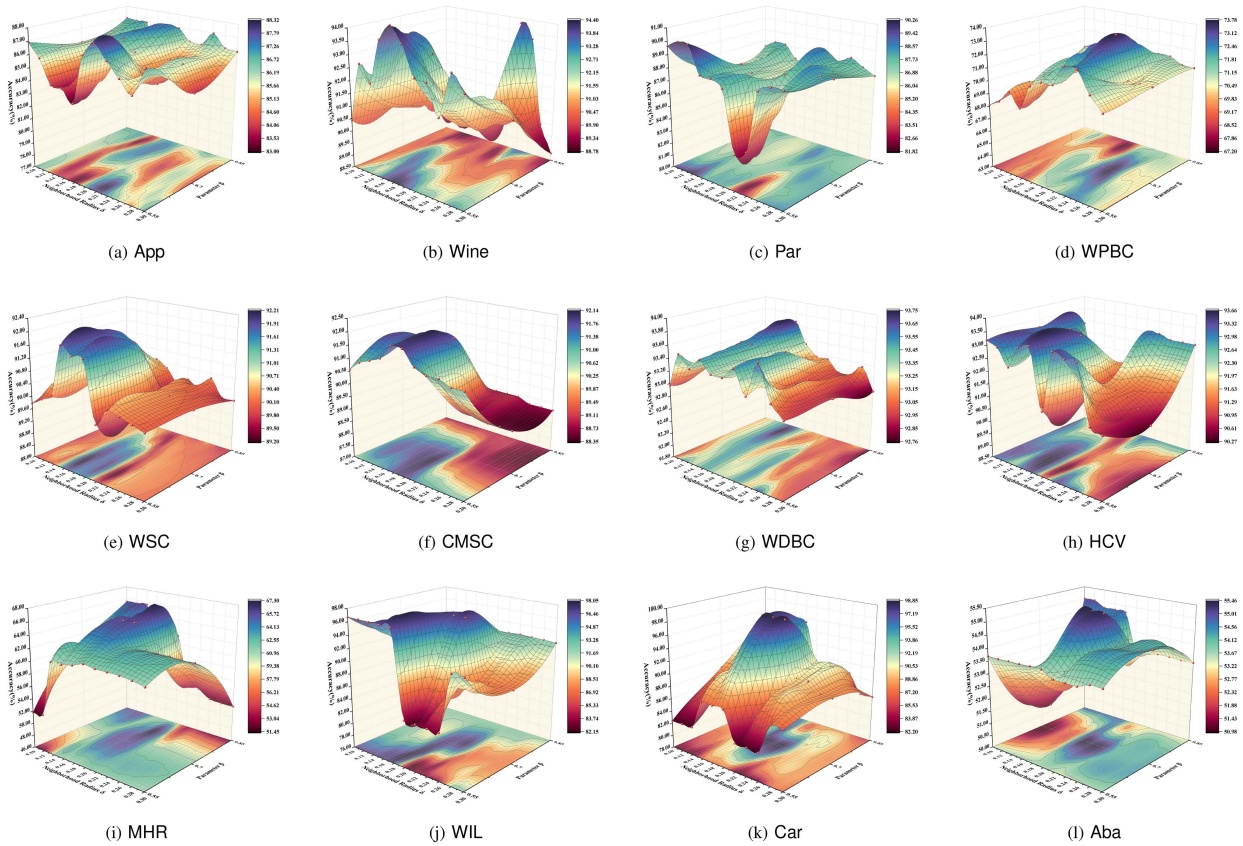
Fig. 3. Classification accuracies of twelve datasets with different neighborhood radius δ on KNN, RF and DT classifiers.

Fig. 4 illustrates the classification results using the DT classifier on twelve datasets under different values of δ and β. Subplots (a)–(l) represent the results for each dataset. The three axes correspond to parameter δ, parameter β, and classification accuracy, respectively. Through the analysis of the twelve subplots, we can conclude that it is necessary to set two parameters, δ and β, in this model. From the graph, we can visually perceive the significant impact of different parameter values on the classification results. The highest classification accuracy and the best clustering performance are achieved when δ ranges from 0.14 to 0.18 and β is around 0.7. For example, for Par, the utmost accuracy and superior performance are obtained when $\delta = 0.14$ and $\beta = 0.55$. For MHR, the utmost accuracy and superior performance are obtained when $\delta = 0.16$ and $\beta = 0.85$. For Car, the utmost accuracy and superior performance are achieved when $\delta = 0.20$ and $\beta = 0.70$. Hence, it is vital to determine the most appropriate parameter values for each dataset to attain optimal performance. Based on the comprehensive analysis of the twelve datasets, the W-GMDA exhibits relatively superior performance when $\delta = 0.20$ and $\beta = 0.70$. These findings highlight the significance of parameter selection in enhancing the performance of the algorithm.

Table VI showcases the quantity of features chosen by the seven algorithms across the twelve datasets. It is worth noting that all datasets used in this study are of numerical type. From Table VI, it is observed that the average number of selected features by all algorithms consistently exhibits a lower count compared to that of the original dataset, indicating the efficacy of attribute reduction. When considering the number of selected features, the average number of selected features by W-GMDA is 9.42, ranking third after KNI with 8.42 and

TABLE VI
SELECTION OF FEATURE QUANTITIES BY SEVEN ALGORITHMS ON DIFFERENT DATASET

Dataset	WDNCE	INF-UFS	HKI	θ-M	AER	KNI	W-GMDA
App	5	5	6	7	5	7	5
Wine	10	8	10	12	11	8	8
Par	19	14	22	16	18	12	12
WPBC	14	19	21	24	18	19	18
WSC	2	5	6	7	7	5	5
CMSC	5	10	16	14	9	7	5
WDBC	14	18	29	21	27	14	15
HCV	8	9	12	10	12	6	12
MHR	6	5	3	6	6	4	4
WIL	6	5	5	7	7	5	5
Car	12	14	21	17	22	8	17
Aba	8	6	7	8	6	6	7
Avg	9.08	9.83	13.17	12.42	12.33	8.42	9.42



(a) App (b) Wine (c) Par (d) WPBC

(e) WSC (f) CMSC (g) WDBC (h) HCV

(i) MHR (j) WIL (k) Car (l) Aba

Fig. 4. Classification accuracies with different neighborhood radius δ and parameter β on DT classifier.

TABLE VII
CLASSIFICATION ACCURACY(%) OF DIFFERENT DLGORITHMS ON KNN CLASSIFIER

Dataset	RAW	WDNCE	INF-UFS	HKI	θ-M	AER	KNI	W-GMDA
App	85.18±1.00	86.00±0.74	75.63±0.10	74.63±0.10	82.94±0.07	83.94±0.09	85.18±0.99	**87.91±0.99**
Wine	72.45±0.96	83.20±0.60	60.80±0.06	50.35±0.12	66.22±0.07	88.33±0.01	94.08±0.13	**94.38±0.25**
Par	85.08±0.67	85.08±0.67	68.21±0.06	67.70±0.07	83.08±0.03	**88.72±0.04**	77.82±0.49	78.34±1.34
WPBC	71.55±0.37	70.03±0.75	82.06±0.06	82.11±0.04	71.17±0.05	76.76±0.07	84.85±0.06	**85.08±0.67**
WSC	61.14±0.39	61.14±0.39	70.46±0.06	70.46±0.06	32.05±0.14	64.32±0.04	60.00±0.50	**89.55±0.14**
CMSC	91.67±0.21	92.78±0.17	90.33±0.02	89.31±0.03	91.85±0.01	92.78±0.03	93.06±0.06	**93.15±0.18**
WDBC	92.27±0.16	92.12±0.11	54.66±0.02	54.84±0.02	84.18±0.02	91.39±0.01	87.17±0.31	**92.49±0.15**
HCV	92.20±0.12	89.43±0.14	**96.20±0.01**	84.55±0.04	82.28±0.04	88.62±0.02	92.03±0.08	92.04±0.10
MHR	73.77±0.13	73.27±0.13	40.51±0.02	38.31±0.04	54.64±0.06	69.72±0.04	72.89±0.16	**73.38±0.19**
WIL	98.25±0.01	**98.45±0.00**	40.17±0.04	41.92±0.01	71.30±0.02	98.00±0.01	96.25±0.01	98.05±0.01
Car	82.08±0.05	93.70±0.03	72.34±0.01	69.31±0.02	76.39±0.01	82.08±0.05	92.10±0.02	**96.99±0.01**
Aba	52.86±0.04	52.86±0.04	33.92±0.01	34.44±0.01	57.35±0.01	52.81±0.06	50.25±0.03	**57.41±0.31**
Avg	79.87±0.34	81.51±0.31	65.44±0.04	63.16±0.05	71.12±0.04	81.46±0.04	82.14±0.24	**86.56±0.36**

The best results are highlighted in bold entities.

WDNCE with 9.08. It can be observed that W-GMDA demonstrates a certain advantage in terms of the number of selected features.

By analyzing Tables VI, VII, VIII, and IX, it is evident that when using the KNN classifier, Par and AER achieved the highest classification accuracy of 88.72%; $WDBC$ and INF-UFS attained the highest accuracy of 96.20%; and for MHR, WDNCE-IAR exhibited the highest accuracy of 98.45%. Furthermore, W-GMDA demonstrated the highest classification accuracy among the remaining nine datasets, accounting for 75% of all datasets. When employing the RF classifier, W-GMDA achieved the highest classification accuracy on eight datasets, representing 66.67% of all datasets. Similarly, when utilizing the DT classifier, W-GMDA obtained the highest classification accuracy on eight datasets, accounting for 66.67% of all datasets. Considering the average classification accuracy obtained by each algorithm using the KNN, RF, and DT classifiers across twelve datasets, W-GMDA consistently outperformed other algorithms, with the highest accuracies being 86.56%, 88.70%, and 86.26%, respectively.

In terms of the number of selected features, while W-GMDA ranks third in the number of selected features, the difference is

TABLE VIII
CLASSIFICATION ACCURACY(%) OF DIFFERENT DLGORITHMS ON RF CLASSIFIER

Dataset	RAW	WDNCE	INF-UFS	HKI	θ-M	AER	KNI	W-GMDA
App	86.91±0.70	88.73±0.65	69.74±0.12	63.25±0.07	**92.38±0.06**	87.82±0.69	87.82±0.69	88.73±0.82
Wine	97.22±0.20	97.22±0.20	52.80±0.07	51.20±0.06	94.95±0.03	97.78±0.14	96.08±0.19	**98.89±0.13**
Par	85.58±0.59	86.11±0.27	65.37±0.05	65.90±0.06	**98.97±0.01**	87.16±0.49	84.55±0.26	92.74±0.40
WPBC	76.68±0.86	78.71±1.26	75.97±0.05	91.79±0.05	**97.45±0.03**	76.13±1.08	76.18±0.85	81.76±0.48
WSC	71.82±0.46	71.82±0.46	67.80±0.06	69.30±0.04	90.28±0.01	71.82±0.46	71.82±0.46	**91.36±0.19**
CMSC	91.48±0.10	91.48±0.10	89.35±0.01	87.65±0.03	92.35±0.01	91.48±0.10	91.48±0.10	**92.59±0.14**
WDBC	95.08±0.08	95.08±0.08	53.61±0.03	49.28±0.06	90.49±0.00	95.44±0.07	92.63±0.14	**95.96±0.11**
HCV	91.55±0.12	91.38±0.11	94.77±0.01	94.77±0.01	92.20±0.02	92.04±0.13	91.38±0.10	**94.96±0.08**
MHR	69.82±0.13	69.73±0.08	39.81±0.02	36.29±0.03	69.62±0.00	70.21±0.07	65.48±0.14	**76.94±0.16**
WIL	96.10±0.03	**97.35±0.03**	37.25±0.03	94.21±0.02	95.56±0.01	70.21±0.07	95.15±0.03	97.25±0.02
Car	93.60±0.01	90.45±0.03	70.74±0.01	90.76±0.03	97.47±0.00	91.05±0.57	94.69±0.02	**97.60±0.01**
Aba	53.65±0.03	54.25±0.03	33.66±0.01	52.88±0.07	51.27±0.11	52.87±0.18	53.51±0.05	**55.61±0.31**
Avg	84.12±0.28	84.36±0.28	62.57±0.04	70.61±0.04	88.58±0.02	83.83±0.34	83.40±0.25	**88.70±0.24**

The best results are highlighted in bold entities.

TABLE IX
CLASSIFICATION ACCURACY(%) OF DIFFERENT DLGORITHMS ON DT CLASSIFIER

Dataset	RAW	WDNCE	INF-UFS	HKI	θ-M	AER	KNI	W-GMDA
App	84.91±0.87	86.91±0.67	76.45±0.07	75.00±0.12	91.51±0.06	85.91±0.70	83.00±0.64	**87.00±0.84**
Wine	89.31±0.71	91.01±0.32	54.40±0.05	60.00±0.07	71.89±0.06	90.46±0.81	90.42±0.88	**93.27±0.36**
Par	87.00±0.42	91.21±0.17	76.41±0.07	80.51±0.04	**96.41±0.00**	87.53±0.41	78.37±0.78	92.76±0.23
WPBC	69.45±0.52	71.00±0.60	75.97±0.02	93.69±0.03	97.47±0.02	69.97±0.53	72.61±0.64	**76.79±0.50**
WSC	71.59±0.34	70.00±0.50	69.55±0.05	69.68±0.04	95.23±0.01	71.59±0.34	70.91±0.54	**91.82±0.10**
CMSC	90.93±0.13	91.11±0.10	90.50±0.03	88.89±0.02	91.89±0.01	90.93±0.15	91.67±0.10	**91.30±0.14**
WDBC	92.79±0.12	92.97±0.05	60.28±0.04	54.27±0.05	91.24±0.00	93.50±0.10	91.91±0.12	**93.67±0.13**
HCV	92.85±0.09	89.44±0.16	**95.49±0.01**	96.03±0.01	88.13±0.04	93.01±0.10	89.93±0.11	93.33±0.09
MHR	65.49±0.09	65.38±0.09	43.53±0.05	44.02±0.03	62.92±0.03	**65.49±0.09**	64.79±0.14	**65.49±0.09**
WIL	91.45±0.73	**96.65±0.01**	43.85±0.02	89.24±0.16	85.45±0.02	92.85±0.09	92.20±0.06	96.65±0.01
Car	95.30±0.01	90.55±0.02	77.24±0.01	89.02±0.01	90.82±0.02	98.33±0.06	98.40±0.00	**98.40±0.01**
Aba	53.41±0.04	53.89±0.04	36.73±0.02	51.12±0.04	53.11±0.02	53.28±0.06	52.02±0.03	**54.68±0.37**
Avg	82.04±0.33	82.52±0.23	66.70±0.04	74.29±0.05	84.67±0.03	82.74±0.31	81.35±0.34	**86.26±0.24**

The best results are highlighted in bold entities.

TABLE X
THE P-VALUES OF W-GMDA AND OTHER SIX ALGORITHMS BASED ON WILCOXON TEST

Classifier	WDNCE	INF-UFS	HKI	θ-M	AER	KNI
KNN	0.043	0.002	0.001	0.002	0.021	0.004
RF	0.004	0.001	0.007	0.042	0.001	0.001
DT	0.005	0.002	0.021	0.042	0.003	0.001

only one feature compared to KNI. However, when it comes to classification accuracy, W-GMDA significantly outperforms KNI and other algorithms in terms of average classification accuracy across KNN, RF, and DT classifiers. Therefore, we conclude that W-GMDA surpasses the performance of the other six algorithms across the KNN, DT, and RF classifiers, and consistently outperforms the accuracy of the original datasets. These experimental results validate the effectiveness of W-GMDA and suggest it as a robust feature selection algorithm.

We conducted a Wilcoxon test to assess the presence of differences between the W-GMDA and six other algorithms. Initially, we set the significance level at 0.06. Subsequently, we calculated the P-values for the Wilcoxon tests performed on W-GMDA and each of the six algorithms when employing KNN, DT, and RF classifiers. The results are summarized in Table X. Upon analyzing the P-values for each algorithm, we observed that all P-values were below the chosen significance level of 0.06. This suggests that the original null hypothesis can be rejected, indicating a statistically significant difference between the W-GMDA and the other six algorithms.

VII. CONCLUSION AND FUTURE WORK

With the advancement of science and technology, the size of data is increasing, accompanied by a growing number of redundant attributes. Efficiently handling such large-scale data poses a significant challenge in classical rough set theory. dominance neighborhood rough set not only extracts similarities and reduces noisy objects but also handles partial order problems effectively. Multi-granulation rough set overcomes the limitations of single-granulation models by considering multiple perspectives. In this paper, we leverage these advantages and, based on the generalized multi-granulation dominance neighborhood rough set, propose a novel approach that partitions the granularity using weights to better select the quality of knowledge granules. Traditional information entropy has certain limitations in problem-solving, prompting us to create a new entropy, WGMIE, based on the WGMDNRS model. We further introduce a heuristic forward greedy feature selection algorithm, W-GMDA, that identifies features with similar approximation abilities for feature selection. Experimental results demonstrate that W-GMDA exhibits excellent classification performance and robustness. Moreover, based on the research findings in this paper, there is potential for further exploring the calculation process of feature indicators and the selection mechanism for dynamic data.

REFERENCES

[1] W. Xu, M. Huang, Z. Jiang, and Y. Qian, "Graph-based unsupervised feature selection for interval-valued information system," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Apr. 17, 2023, doi: 10.1109/TNNLS.2023.3263684.

[2] Y. Xue, Y. Tang, X. Xu, J. Liang, and F. Neri, "Multi-objective feature selection with missing data in classification," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 6, no. 2, pp. 355–364, Apr. 2022.

[3] H. Chen, T. Li, X. Fan, and C. Luo, "Feature selection for imbalanced data based on neighborhood rough sets," *Inf. Sci.*, vol. 483, pp. 1–20, 2019.

[4] W. Ding, I. Triguero, and C.-T. Lin, "Coevolutionary fuzzy attribute order reduction with complete attribute-value space tree," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 5, no. 1, pp. 130–142, Feb. 2021.

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

[6] J. Zhang, T. Li, D. Ruan, and D. Liu, "Neighborhood rough sets for dynamic data mining," *Int. J. Intell. Syst.*, vol. 27, no. 4, pp. 317–342, 2012.

[7] X. Yang, J. Yang, C. Wu, and D. Yu, "Dominance-based rough set approach and knowledge reductions in incomplete ordered information system," *Inf. Sci.*, vol. 178, no. 4, pp. 1219–1234, 2008.

[8] M. Inuiguchi, Y. Yoshioka, and Y. Kusunoki, "Variable-precision dominance-based rough set approach and attribute reduction," *Int. J. Approx. Reasoning*, vol. 50, no. 8, pp. 1199–1214, 2009.

[9] Y. Qian, J. Liang, Y. Yao, and C. Dang, "MGRS: A multi-granulation rough set," *Inf. Sci.*, vol. 180, no. 6, pp. 949–970, 2010.

[10] Q. Hu, D. Yu, J. Liu, and C. Wu, "Neighborhood rough set based heterogeneous feature subset selection," *Inf. Sci.*, vol. 178, no. 18, pp. 3577–3594, 2008.

[11] W. 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.

[12] W. Xu, Z. Yuan, and Z. Liu, "Feature selection for unbalanced distribution hybrid data based on k-nearest neighborhood rough set," *IEEE Trans. Artif. Intell.*, vol. 5, no. 1, pp. 229–243, Jan. 2024.

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

[14] M. Han, Y. Qu, N. MacParthaláin, C. Shang, Z. Yao, and Q. Shen, "Representation-based decision-theoretic rough sets for three-way classification," *IEEE Trans. Emerg. Topics Comput. Intell.*, early access, Jul. 27, 2023, doi: 10.1109/TETCI.2023.3296467.

[15] L. Sun, M. Li, W. Ding, and J. Xu, "Adaptive fuzzy multi-neighborhood feature selection with hybrid sampling and its application for class-imbalanced data," *Appl. Soft Comput.*, vol. 149, 2023, Art. no. 110968.

[16] Y. Yao et al., "Granular computing: Basic issues and possible solutions," in *Proc. 5th Joint Conf. Inf. Sci.*, 2000, pp. 186–189.

[17] D. Guo, W. Xu, Y. Qian, and W. Ding, "M-FCCL: Memory-based concept-cognitive learning for dynamic fuzzy data classification and knowledge fusion," *Inf. Fusion*, vol. 100, 2023, Art. no. 101962.

[18] W. 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.

[19] D. Guo, W. Xu, Y. Qian, and W. Ding, "Fuzzy-granular concept-cognitive learning via three-way decision: Performance evaluation on dynamic knowledge discovery," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 3, pp. 1409–1423, Mar. 2024.

[20] L. Sun, Y. Ma, W. Ding, and J. Xu, "Sparse feature selection via local feature and high-order label correlation," *Appl. Intell.*, vol. 54, no. 1, pp. 565–591, 2024.

[21] X. Zhang, X. Huang, and W. Xu, "Matrix-based multi-granulation fusion approach for dynamic updating of knowledge in multi-source information," *Knowl.-Based Syst.*, vol. 262, 2023, Art. no. 110257.

[22] L. Sun, S. Si, W. Ding, X. Wang, and J. Xu, "Multiobjective sparrow search feature selection with sparrow ranking and preference information and its applications for high-dimensional data," *Appl. Soft Comput.*, vol. 147, 2023, Art. no. 110837.

[23] W. Pan, K. She, and P. Wei, "Multi-granulation fuzzy preference relation rough set for ordinal decision system," *Fuzzy Sets Syst.*, vol. 312, pp. 87–108, 2017.

[24] Y. Kang, S. Wu, Y. Li, J. Liu, and B. Chen, "A variable precision grey-based multi-granulation rough set model and attribute reduction," *Knowl.-Based Syst.*, vol. 148, pp. 131–145, 2018.

[25] I. Düntsch and G. Gediga, "Uncertainty measures of rough set prediction," *Artif. Intell.*, vol. 106, no. 1, pp. 109–137, 1998.

[26] T. Zheng and L. Zhu, "Uncertainty measures of neighborhood system-based rough sets," *Knowl.-Based Syst.*, vol. 86, pp. 57–65, 2015.

[27] J. Liang, R. Li, and Y. Qian, "Distance: A more comprehensible perspective for measures in rough set theory," *Knowl.-Based Syst.*, vol. 27, pp. 126–136, 2012.

[28] S. An, J. Liu, C. Wang, and S. Zhao, "A relative uncertainty measure for fuzzy rough feature selection," *Int. J. Approx. Reasoning*, vol. 139, pp. 130–142, 2021.

[29] W. Xu, Y. Pan, X. Chen, W. Ding, and Y. Qian, "A novel dynamic fusion approach using information entropy for interval-valued ordered datasets," *IEEE Trans. Big Data*, vol. 9, no. 3, pp. 845–859, Jun. 2023.

[30] Z. Deng et al., "Feature selection for label distribution learning using dual-similarity based neighborhood fuzzy entropy," *Inf. Sci.*, vol. 615, pp. 385–404, 2022.

[31] Y. Qian et al., "Local rough set: A solution to rough data analysis in Big Data," *Int. J. Approx. Reasoning*, vol. 97, pp. 38–63, 2018.

[32] H. Chen, T. Li, Y. Cai, C. Luo, and H. Fujita, "Parallel attribute reduction in dominance-based neighborhood rough set," *Inf. Sci.*, vol. 373, pp. 351–368, 2016.

[33] Y. Yao and X. Zhang, "Class-specific attribute reducts in rough set theory," *Inf. Sci.*, vol. 418, pp. 601–618, 2017.

[34] B. Ma, C. Yang, A. Li, Y. Chi, and L. Chen, "A faster DBSCAN algorithm based on self-adaptive determination of parameters," *Procedia Comput. Sci.*, vol. 221, pp. 113–120, 2023.

[35] 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.

[36] X. Zhang, Y. Fan, and J. Yang, "Feature selection based on fuzzy-neighborhood relative decision entropy," *Pattern Recognit. Lett.*, vol. 146, pp. 100–107, 2021.

[37] G. Roffo, S. Melzi, U. Castellani, A. Vinciarelli, and M. Cristani, "Infinite feature selection: A graph-based feature filtering approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 12, pp. 4396–4410, Dec. 2021.

[38] J. Dai, W. Wang, and J.-S. Mi, "Uncertainty measurement for interval-valued information systems," *Inf. Sci.*, vol. 251, pp. 63–78, 2013.

[39] X. Liu, J. Dai, J. Chen, and C. Zhang, "Unsupervised attribute reduction based on α-approximate equal relation in interval-valued information systems," *Int. J. Mach. Learn. Cybern.*, vol. 11, no. 9, pp. 2021–2038, 2020.

[40] Y. Pan, W. Xu, and Q. Ran, "An incremental approach to feature selection using the weighted dominance-based neighborhood rough sets," *Int. J. Mach. Learn. Cybern.*, vol. 14, no. 4, pp. 1217–1233, 2023.

[41] Z. Yuan, H. Chen, X. Yang, T. Li, and K. Liu, "Fuzzy complementary entropy using hybrid-kernel function and its unsupervised attribute reduction," *Knowl.-Based Syst.*, vol. 231, 2021, Art. no. 107398.

Weihua Xu received the M.Sc. degree in mathematics from the School of Mathematics and Information Sciences, Guangxi University, Nanning, China, in 2004, and the Ph.D. degree in mathematics from the School of Sciences, Xi'an Jiaotong University, Xi'an, China, in 2007. He is currently a Professor with the College of Artificial Intelligence, Southwest University, Chongqing, China. He has authored or coauthored five monographs and more than 200 articles in international journals. His research interests include granular computing, cognitive computing, and information fusion. Dr. Xu is a Senior Member of Chinese Association for Artificial Intelligence. He is an Associate Editor for *International Journal of Machine Learning and Cybernetics and Journal of Intelligent and Fuzzy Systems*.

Qinyuan Bu received the B.E. degree from the School of Information Science and Technology, Gansu Agricultural University, Lanzhou, China, in 2022. He is currently working toward the master's degree with the School of Artificial Intelligence, Southwest University, Chongqing, China. His research interests include feature selection, granular computing, and rough set theory.