A Novel Unsupervised Feature Selection for High-Dimensional Data Based on FCM and k-Nearest Neighbor Rough Sets

Weihua Xu, Yang Zhang, and Yuhua Qian, *Member, IEEE*

***Abstract*— Large amounts of high-dimensional unlabeled data typically contain only a small portion of truly effective information. Consequently, the issue of unsupervised feature selection methods has gained significant attention in research. However, current unsupervised feature selection approaches face limitations when dealing with datasets that exhibit uneven density, and they also require substantial computational time. To address this problem, this research article proposes a feature extraction technique that combines the Fuzzy C-Means (FCM) and k-nearest neighbor rough sets. FCM is a clustering algorithm grounded in fuzzy theory, which takes into account the inherent data structure and the correlations between different features. Consequently, FCM is particularly well-suited for datasets with uneven density. Our proposed method consists of three steps. First, the FCM algorithm is used to cluster the unlabeled data. Second, a measure that evaluates the importance of features is defined and sorted based on the clustering results. Finally, redundant features are filtered using k-nearest neighbor rough sets while retaining important features, significantly reducing the running time. In addition, we designed the feature selection algorithm (KND-UFS) and conducted experiments on 12 public datasets. We compared KND-UFS with eight existing algorithms in terms of running time, classification accuracy, and the number of selected features. The experimental results provided strong evidence supporting the superior performance of the KND-UFS algorithm.**

***Index Terms*— Clustering, feature selection, granular computing (GrC), k-nearest neighbor rough sets.**

I. Introduction

With the rapid development of the information age, the explosive growth of data has posed higher demands on data processing [1], leading to the emergence of data mining [2]. Granular computing (GrC) [3] has demonstrated great advantages in extracting value from massive data. The idea of GrC originates from the process of humans handling complex problems. Humans usually tend to simplify complicated problems by reducing them to simpler ones. The core idea of GrC is to represent [4], abstract [5], and model [6] data, dividing the raw data in the data space into "granules" according to certain relationships, with each granule representing some common features. In this way, complex problems can be transformed into discrete descriptions. Through the processing of granules, hidden information and knowledge [7], [8] can be further extracted from data, and this information can be used for applications such as classification [9], prediction [10], and decision-making [11].

Pawlak [13] from Poland proposed the notion of rough sets, from the standpoint of information granularity [12], building upon the existing relation theory and set theory. Rough sets are used to handle imprecise calculations and incomplete data. They classify elements in a set through equivalence relations [14] and generate corresponding partitions. Equivalent classes within the same partition can be used for information simplification. By means of undistinguished equivalent classes, an approximate space can be established [15]. Within this approximate space, two precise sets (upper and lower approximation sets [16]) are used to approximate a boundary-fuzzy set.

Information systems [17], as a medium for representing information or knowledge, have been a subject of investigation in the field of rough sets. Feature selection [18] is a core aspect in the analysis and processing of information systems. It is well-known that not all the attributes within an information system are equally important, and some attributes may even be irrelevant or redundant [19]. These irrelevant or redundant attributes can often have adverse effects on the analysis and processing of the system. Hence, there is a need to carefully choose a subset of attributes that effectively capture the essential characteristics of the original information system. This selection process involves eliminating irrelevant or redundant attributes to ensure that the chosen attribute subset retains the same classification capability as the original attribute set [20].

Classical rough set theory mainly focuses on complete information systems, where all the attribute values of the objects under consideration are known. However, in real-world scenarios, the identifiers of decision categories are frequently unidentified or incomplete, thereby necessitating the examination of unsupervised attribute reduction [21]. In recent years, researchers have conducted extensive research on unsupervised attributes. Yuan et al. [22] introduced a comprehensive unsupervised mixed attribute reduction approach

Received 5 March 2024; revised 25 June 2024; accepted 29 August 2024. Date of publication 30 September 2024; date of current version 4 June 2025. 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 CSTB2023NSCQ-LZX0027. *(Corresponding author: Weihua Xu.)*

Weihua Xu and Yang Zhang are with the College of Artificial Intelligence, Southwest University, Chongqing 400715, China (e-mail: chxuwh@gmail.com; 2530078847@qq.com).

Yuhua Qian is with the Institute of Big Data Science and Industry, Shanxi University, Taiyuan 030006, China (e-mail: jinchengqyh@126.com).

Digital Object Identifier 10.1109/TNNLS.2024.3460796

leveraging fuzzy rough sets for unsupervised mixed attribute reduction methodologies. Huang et al. [23] proposed a robust unsupervised feature selection algorithm based on data relation learning to address the problem of poor robustness for outliers. To solve the problems of large computational complexity and poor performance stability, Lin et al. [24] proposed an effective method for unsupervised feature selection by combining orthogonal basis clustering and adaptive graph regularization, which exhibits excellent capabilities in cluster separation and preservation of local information. Zhang and Li [25] introduced adaptive loss regularization into least-squares regression and proposed a novel unsupervised feature selection method using sparse fuzzy membership degrees. Yuan et al. [26] used fuzzy rough set theory and designed an unsupervised feature selection algorithm based on fuzzy mutual information, demonstrating its performance through experiments. Tang et al. [27] proposed an unsupervised feature selection approach using multigraph integration and learning of feature weight. Furthermore, they developed a block coordinate descent algorithm with assured convergence to tackle the ultimate optimization challenge.

Cluster analysis [28] is an unsupervised learning algorithm [29] in data mining that automatically groups datasets based on their distribution or concentration patterns when prior knowledge is unavailable. The aim of grouping is to maximize the similarity within the same group and maximize the dissimilarity between different groups. Fuzzy C-Means (FCM) [30] has gained significant attention in cluster analysis due to its robustness and lack of sample data dimensionality restrictions. In this study, we propose an unsupervised feature selection algorithm based on the FCM clustering algorithm [31]. We define the discriminability of each attribute based on the clustering results and sort the attributes to retain those with higher importance [32]. Then, we eliminate redundant features one by one using the dependency relationship [33] of k-nearest neighbor rough sets. Finally, we design an unsupervised feature selection algorithm based on k-nearest neighbor rough sets (KND-UFS).

The main contributions of this article are as follows.

1) This article proposes a measurement method for assessing the importance of features, which takes into account both the degree of overlap in data space and the trends present in the dataset. The effectiveness of this measurement is verified through experimental evaluation, demonstrating improved data separability.
2) This article proposes a feature extraction algorithm that combines feature retention rates with k-nearest neighbor neighborhood dependency. Through this approach, both the classification accuracy and the computational time required by the algorithm are significantly improved, particularly in the context of high-dimensional data feature extraction, where the algorithm exhibits notable advantages in terms of runtime.
3) This article presents an algorithm for unsupervised feature extraction, which is compared with eight other similar algorithms using 12cpublic datasets. The experimental outcomes reveal notable benefits in terms of both time efficiency and accuracy.

This article is organized as follows. In Section II, we presented the fundamental knowledge of the FCM and the k-nearest neighbor rough set model. In Section III, we defined homogeneity measures for assessing attribute importance from two perspectives: overlap coefficient (OC) and distance (DIS). In Section IV, we designed pertinent algorithms to extract important features based on attribute homogeneity measures, followed by combining them with k-nearest neighborhood dependency measures for feature selection. In Section V, we validated the effectiveness of the algorithms using 12 datasets, by considering both the runtime and classification accuracy. Finally, we have summarized this article in Section VI.

II. Related Work and Foundations

In this section, we will begin by providing a review of some fundamental concepts related to FCM and k-nearest neighborhood rough set.

A. Fuzzy C-Means

Given a dataset $X = \{x_1, x_2, \ldots, x_n\}$ as n dimensional vector with each object having s attributes that describe it. Let the dataset be partitioned into $c(2 \leqslant c \leqslant n)$ categories and the clustering centers of each category are $V = \{v_1, v_2, \ldots, v_c\}$. The objective function of the FCM (see Algorithm 1) clustering algorithm is defined as follows:

$$J_m = \sum_{i=1}^{n} \sum_{j=1}^{c} u_{ij}^m \left\| x_i - v_j \right\|^2 \tag{1}$$

where m is the fuzziness parameter is a parameter that governs the degree of fuzziness in the clustering outcomes. A larger value of m leads to more fuzzy clustering results, while a smaller value of m leads to more distinct clustering results. u_{ij} represents the membership degree of the ith object x_i in relation to the jth cluster center v_j. $\|x_i - v_j\| = (\sum_{k=1}^{s} |x_{ik} - v_{jk}|^{1/p})^p$ represents the DIS between x_i and v_j, where the most commonly used DIS metric is the Euclidean DIS with $p = 2$ and s denotes the number of features, and x_{ik} represents the feature value of the ith object for the kth feature.

The FCM iteratively optimizes the objective function by updating the membership degrees and cluster centroids, until convergence criteria are satisfied. The updated expressions for u_{ij} and v_j are as follows:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\left\| x_i - v_j \right\|}{\left\| x_i - v_k \right\|} \right)^{\frac{2}{m-1}}} \tag{2}$$

$$v_j = \frac{\sum_{i=1}^{n} u_{ij}^m \cdot x_i}{\sum_{i=1}^{n} u_{ij}^m}. \tag{3}$$

In the case of a single object x_i, the sum of its degrees of membership to all the clusters is equal to 1. This implies that an object must have a complete allocation of membership across all the clusters. Consequently, each object is assigned to the cluster with the highest degree of membership, indicating the cluster to which it predominantly belongs.

A threshold ε is predetermined for the variation in the objective function, and after t iterations, the algorithm terminates

Algorithm 1 Fuzzy C-Means

Input: A dataset $X = \{x_1, \ldots, x_n\}$; the number of clusters c; the fuzziness parameter m; and the threshold ε.
Output: Cluster assignments D.

```
Set D ← ∅, J_m^(0) ← 0, t ← 0.
Initialize the membership degrees U randomly.
Compute the cluster centroids V using the updated membership degrees U by formula (3).
Compute J_m^(1) by formula (1).
while J_m^(t+1) − J_m^(t) ⩾ ε do
    t ← t + 1.
    Update U by formula (2) and V by formula (3).
    Compute J_m^(t+1) by formula (1).
end while
for all x_i ∈ X do
    Select j = argmax(u_ij).
    D ← D ∪ {j}.
end for
return D.
```

once the threshold condition is met. The expression for the termination criterion is given below

$$J_m^{(t+1)} - J_m^{(t)} < \varepsilon. \tag{4}$$

B. k-Nearest Neighborhood Rough Set

In $S = (U, C \cup D, F)$, let x be an object in U. For an attribute a belonging to C, the k-nearest neighbor neighborhood of x on a can be defined as follows:

$$\kappa_a(x) = \big\{x_j \in U \;\big|\; |f(x_j, a) - f(x, a)| \leqslant |f(x_i, a) - f(x, a)|, x_i \in U, i \neq j, |\kappa_a(x)| = k\big\} \tag{5}$$

where $\kappa_a(x)$ denotes the top k objects that are closest to x in relation to a.

In $S = (U, C \cup D, F)$, let x be an object in U, and $B \subseteq C$. The k-nearest neighbor neighborhood of x in relation to B can be defined as follows:

$$\kappa_B(x) = \bigcap_{a \in B} \kappa_a(x) \tag{6}$$

where $\kappa_B(x)$ denotes the top k objects that are closest to x in relation to attribute subset B.

Definition 1: Let $S = (U, C \cup D, F)$ be a decision information system, $x \in U$, $B \subseteq C$, $U/D = \{D_1, D_2, \ldots, D_r\}$, given a neighborhood relation R, the upper and lower approximations of set D on subset B in relation to R can be defined as follows:

$$\overline{R}_B^{\kappa}(D) = \bigcup_{i=1}^{r} \overline{R}_B^{\kappa}(D_i) \tag{7}$$

$$\underline{R}_B^{\kappa}(D) = \bigcup_{i=1}^{r} \underline{R}_B^{\kappa}(D_i) \tag{8}$$

where

$$\overline{R}_B^{\kappa}(X) = \{x | \kappa_B(x) \cap X \neq \emptyset, x \in U\} \tag{9}$$

$$\underline{R}_B^{\kappa}(X) = \{x | \kappa_B(x) \subseteq X, x \in U\}. \tag{10}$$

The lower approximation of set D on subset B, commonly known as the positive region, is denoted as POS_B^{κ}, i.e., $\mathrm{POS}_B^{\kappa}(D) = \underline{R}_B^{\kappa}(D)$.

The degree of neighborhood dependency of set D in relation to subset B and relation R can be defined as follows:

$$\gamma_B^{\kappa}(D) = \frac{|\mathrm{POS}_B^{\kappa}(D)|}{|U|} \tag{11}$$

where $\gamma_B^{\kappa}(D)$ reflects the ability of the conditional attribute set B to approximate the decision attribute set D, $0 \leqslant \gamma_B^{\kappa}(D) \leqslant 1$.

Property 1: Let $B_1, B_2 \subseteq C$, $B_1 \subseteq B_2$, then

1) $\forall X \subseteq U, \underline{R}_{B_1}^{\kappa}(X) \subseteq \underline{R}_{B_2}^{\kappa}(X)$;
2) $\mathrm{POS}_{B_1}^{\kappa}(D) \subseteq \mathrm{POS}_{B_2}^{\kappa}(D), \gamma_{B_1}^{\kappa}(D) \leqslant \gamma_{B_2}^{\kappa}(D)$.

III. Feature Ranking Based on Homogeneity

To select effective features in the data, we adopt a ranking-filtering feature selection method. The ranking-filtering feature selection method calculates the score of each feature based on a predefined scoring function, sorts the features according to the scores, selects the top-ranked features to form a feature subset, and the final number of retained features can be manually determined. Due to the constant-level traversal of features, the time complexity is extremely low. We define the homogeneity (DI) as the scoring function to measure the importance of individual features, and then start the search from the most important feature in the feature ranking, thereby improving the efficiency of feature selection. DI reflects the OC and DIS between different categories under a certain feature. We define that the feature with a smaller OC and a larger DIS is more important [34]. By sorting the importance of individual features based on DI, we select and retain a percentage ($\alpha\%$) of top-ranked objects, effectively filtering out relatively unimportant features.

In $S = (U, C \cup D, F)$, the OC of objects D_i and D_j in relation to a can be defined as follows:

$$\mathrm{OC}_a(D_i, D_j) = \frac{\big|[f_{\min}^a(D_i), f_{\max}^a(D_i)] \cap [f_{\min}^a(D_j), f_{\max}^a(D_j)]\big|}{\big|[f_{\min}^a(D_i), f_{\max}^a(D_i)] \cup [f_{\min}^a(D_j), f_{\max}^a(D_j)]\big|} \tag{12}$$

where $f_{\min}^a(D_i) = \min(f(x, a) : x \in D_i)$ and $f_{\max}^a(D_i) = \max(f(x, a) : x \in D_i)$.

The OC_a reflects the degree of spatial overlap of objects between any two different decision classes under the conditional attribute a. When the value of OC_a is larger, it indicates a higher degree of overlap between objects. If the denominator of OC_a is 0, there are two possible scenarios. One is when $[f_{\min}^a(D_i), f_{\max}^a(D_i)] = [f_{\min}^a(D_j), f_{\max}^a(D_j)]$, in which case the OC is 1. The other is when $[f_{\min}^a(D_i), f_{\max}^a(D_i)] \neq [f_{\min}^a(D_j), f_{\max}^a(D_j)]$, in which case the OC is 0.

Property 2: Let $a \in C$, objects D_i and D_j belonging to U/D, then $0 \leqslant \mathrm{OC}_a(D_i, D_j) \leqslant 1$.

Proof: $\forall D_i, D_j \in U/D$, it follows that $[f_{\min}^a(D_i), f_{\max}^a(D_i)] \cap [f_{\min}^a(D_j), f_{\max}^a(D_j)] \subseteq [f_{\min}^a(D_i), f_{\max}^a(D_i)] \cup [f_{\min}^a(D_j), f_{\max}^a(D_j)]$, such that $|[f_{\min}^a(D_i), f_{\max}^a(D_i)] \cap [f_{\min}^a(D_j), f_{\max}^a(D_j)]| \leqslant |[f_{\min}^a(D_i), f_{\max}^a(D_i)] \cup [f_{\min}^a(D_j), f_{\max}^a(D_j)]|$, therefore $0 \leqslant \mathrm{OC}_a(D_i, D_j) \leqslant 1$.

In $S = (U, C \cup D, F)$, the OC of S for a is defined as follows:

$$\mathrm{OC}_a(S) = \sum_{D_i \neq D_j} \mathrm{OC}_a(D_i, D_j). \tag{13}$$

The $\mathrm{OC}_a(S)$ reflects the degree of spatial overlap of all the objects under the conditional attribute a by calculating the sum of OC between any two different decision classes. The higher the value of $\mathrm{OC}_a(S)$, the greater the degree of overlap among all the objects under attribute a, indicating lower separability of attribute a.

In $S = (U, C \cup D, F)$, the DIS between D_i and D_j for attribute a is defined as follows:

$$\mathrm{DIS}_a(D_i, D_j) = \frac{\left|\bar{f}_a(D_i) - \bar{f}_a(D_j)\right|}{f^a_{\max}(D_i, D_j) - f^a_{\min}(D_i, D_j)} \tag{14}$$

where $f^a_{\max}(D_i, D_j) = \max(f(x, a) : x \in D_i \cup D_j)$, $f^a_{\min}(D_i, D_j) = \min(f(x, a) : x \in D_i \cup D_j)$, $\bar{f}_a(D_i) = (1/|D_i|) \sum_{x \in D_i} f(x, a)$, and $\bar{f}_a(D_i)$ represents the central tendency of objects belonging to the ith decision class under attribute a.

The DIS_a reflects the degree of deviation of objects between any two different decision classes under the conditional attribute a. The larger the value of DIS_a, the greater the difference between the two categories of objects. If $f^a_{\max}(D_i, D_j) = f^a_{\min}(D_i, D_j)$, then $\mathrm{DIS}_a = 0$.

Property 3: Let $a \in C$, objects D_i and D_j belonging to U/D, then $0 \leqslant \mathrm{DIS}_a(D_i, D_j) \leqslant 1$.

Proof: $\forall D_i, D_j \in U/D$, it follows that $f^a_{\min}(D_i) \leqslant \bar{f}_a(D_i) \leqslant f^a_{\max}(D_i)$ and $f^a_{\min}(D_j) \leqslant \bar{f}_a(D_j) \leqslant f^a_{\max}(D_j)$, such that $|\bar{f}_a(D_i) - \bar{f}_a(D_j)| \leqslant f^a_{\max}(D_i, D_j) - f^a_{\min}(D_i, D_j)$, therefore $0 \leqslant \mathrm{DIS}_a(D_i, D_j) \leqslant 1$.

In $S = (U, C \cup D, F)$, the DIS of S for a is defined as follows:

$$\mathrm{DIS}_a(S) = \sum_{D_i \neq D_j} \mathrm{DIS}_a(D_i, D_j). \tag{15}$$

The $\mathrm{DIS}_a(S)$ reflects the separability of the decision system in relation to the conditional attribute a by calculating the sum of DIS between any two different decision classes. The larger the value of $\mathrm{DIS}_a(S)$, the higher the separability of objects under attribute a; the smaller the value of $\mathrm{DIS}_a(S)$, the lower the separability of objects under attribute a.

Definition 2: Let $S = (U, C \cup D, F)$ be a decision information system, $a \in C$, objects D_i and D_j belonging to U/D, the degree of homogeneity of S for a is defined as follows:

$$\mathrm{DI}_a(S) = \frac{\mathrm{DIS}_a(S)}{\mathrm{OC}_a(S)}. \tag{16}$$

It is obvious that $\mathrm{DIS}_a(S)$ is inversely proportional to $\mathrm{OC}_a(S)$. We measure the homogeneity of attribute a from these two aspects. The larger the value of $\mathrm{DI}_a(S)$, the more important the attribute a; the smaller the value of $\mathrm{DI}_a(S)$, the less important the attribute a.

Example 1: Table I illustrates the decision information system derived from the utilization of the FCM algorithm. The number of clusters is predefined as $c = 2$, the fuzziness parameter is assigned the value of $m = 2$, and the convergence threshold is set at $\varepsilon = 0.0001$. This decision information system encompasses the universe $U = \{x_1, x_2, \ldots, x_{12}\}$ and attributes $A = \{a_1, a_2, , a_3, a_4\}$. Subsequently, $\mathrm{DIS}_a(S)$, $\mathrm{OC}_a(S)$, $\mathrm{DI}_a(S)$ values for each attribute can be obtained, as illustrated in Table II. Based on the descending order of $\mathrm{DI}_a(S)$ values, the resulting $order(C)$ is $\{a_4, a_3, a_1, a_2\}$. Consequently, we have obtained the attribute ranking in descending order of importance. In this case, we set the retention rate α to 75%. Consequently, the retained features $retention(C) = \{a_4, a_3, a_1\}$, and attribute a_2 is directly filtered out.

TABLE I
DECISION INFORMATION SYSTEM

U	a_1	a_2	a_3	a_4	d
x_1	0.03	0.06	0.24	0.28	0
x_2	0.04	0.14	0.21	0.64	1
x_3	0.04	0.12	0.48	0.48	0
x_4	0.05	0.13	0.41	0.38	0
x_5	0.04	0.12	0.19	0.86	1
x_6	0.07	0.11	0.73	0.26	0
x_7	0.09	0.09	0.28	0.33	0
x_8	0.11	0.04	0.59	0.29	0
x_9	0.11	0.05	0.67	0.68	0
x_{10}	0.1	0.07	0.38	0.96	1
x_{11}	0.12	0.09	0.72	0.52	0
x_{12}	0.15	0.08	0.89	0.36	0

TABLE II
$\mathrm{DIS}_a(S)$, $\mathrm{OC}_a(S)$, $\mathrm{DI}_a(S)$ OF EACH ATTRIBUTE

	a_1	a_2	a_3	a_4
$DIS_a(S)$	0.2130	0.2444	0.4238	0.6032
$OC_a(S)$	0.5000	0.6000	0.2000	0.0571
$DI_a(S)$	0.4259	0.4074	2.1190	10.5556

IV. FEATURE SELECTION ALGORITHM

Although the ranking-filtering feature selection method is computationally efficient, it only considers the importance of individual features and does not take into account the relationships between features, resulting in feature subsets with high redundancy. To address this redundancy issue, we use the dependency relationship of k-nearest neighbor rough sets to establish the significance indicator for attribute a.

In $S = (U, C \cup D, F)$, $a \in C$, $B \subseteq C$, the significance of attribute a in relation to D is defined as follows:

$$\mathrm{SIG}(a, B, D) = \gamma^{\kappa}_{B \cup \{a\}}(D) - \gamma^{\kappa}_{B}(D). \tag{17}$$

$\mathrm{SIG}(a, B, D)$ is an indicator that describes the significance of attribute a relative to the attribute subset B under decision D. According to Property 1, the significance indicator $\mathrm{SIG}(a, B, D)$ is always greater than or equal to 0. When $\mathrm{SIG}(a, B, D)$ equals 0, attribute a is considered redundant and removable. The magnitude of $\mathrm{SIG}(a, B, D)$ reflects the importance of attribute a within the attribute subset B. Leveraging this concept, we have devised an unsupervised feature selection algorithm using the dependency of k-nearest neighbor rough sets, as depicted in Algorithm 2.

In Algorithm 2, steps 3–5 involve calculating the homogeneity for each feature in C, resulting in a time complexity

TABLE III
ENTIRE CALCULATION PROCESS

step	C	B	$\gamma_B^\kappa(D)$	selected features
1	a_4	$\{a_4\}$	0.5833	$\{a_4\}$
2	a_3	$\{a_4, a_3\}$	1.0	$\{a_4, a_3\}$
3	a_1	$\{a_4, a_3, a_2\}$	1.0	$\{a_4, a_3\}$
4	a_2	$\{a_4, a_3, a_1\}$	1.0	$\{a_4, a_3\}$

Algorithm 2 Unsupervised Feature Selection Algorithm Based on k-Nearest Neighborhood Dependency

Input: A dataset $S = (U, C, F)$; the number of clusters r; the feature retention rate α and k-nearest parameters k.
Output: Feature selection subset A.
1: Compute cluster assignments D by Algorithm 1 and $U/D = \{D_1, D_2, \ldots, D_r\}$.
2: Set $A \leftarrow \emptyset$.
3: **for all** $a \in C$ **do**
4: Compute $DI_a(S)$ by formula (16).
5: **end for**
6: Order attributes in descending order based on $\mathrm{DI}_a(S)$ and denote the result as $order(C)$.
7: Put the top $\alpha\%$ objects from $order(C)$ into $retention(C)$.
8: **for all** $a \in retention(C)$ **do**
9: Compute $\gamma_{A\cup\{a\}}^\kappa(D)$ and $\gamma_A^\kappa(D)$ by formula (11).
10: $SIG(a, A, D) = \gamma_{A\cup\{a\}}^\kappa(D) - \gamma_A^\kappa(D)$.
11: $A \leftarrow A \cup \{a\}$ where $SIG(a, A, D) \neq 0$.
12: **end for**
13: **return** A.

of O(|U|×|C|). Step 6, which sorts the features in descending order based on their homogeneity values to obtain order(C), has a time complexity of O(1). Step 7, which retains the corresponding proportion of features to obtain retention(C), also has a time complexity of O(1). Steps 8–12 involve removing redundant attributes based on the dependency in a sequential manner, resulting in a time complexity of O(|U|×log(k)×|U/D|×|C|).

Example 2: Based on Example 1, we have obtained the retained features $retention(C) = \{a_4, a_3, a_1\}$. Now, we need to establish significance indicators for each feature using the k-nearest neighbor rough set dependency relationship. In this case, we set the value of k–5. Under attribute a_4, the nearest neighbor neighborhood of x_1 is [x_1, x_4, x_5, x_6, x_9]. Similarly, we can determine the nearest neighbor neighborhoods for the other objects. Using formulas 8 and 11, we can calculate $\gamma_{a4}^\kappa(D) = 0.5833$. Consequently, we retain feature a4. Similarly, for $B = \{a_4, a_3\}$, $\gamma_B^\kappa(D) = 1$, indicating that we retain feature a_3. The remaining results are presented in Table III. It can be observed from the table that we can eliminate the redundant attributes a_1 and a_2. Since we have already filtered out the relatively less important feature a_2 in the previous section, it can be disregarded when calculating the significance indicators, thereby significantly enhancing computational efficiency while maintaining classification capability.

TABLE IV
DESCRIPTION OF DATASETS

N0.	Abbreviation	Objects	Attributes	Classes
1	seeds	210	8	3
2	australian	690	15	2
3	concrete	1030	9	14
4	cardiotocography	2126	21	3
5	rice	3810	8	2
6	waveform	5000	22	3
7	texture	5500	41	11
8	thyroid	7200	22	3
9	magic	19020	11	2
10	nervousSystem	60	7130	2
11	leukemia	72	12583	3
12	lungCancer	181	12534	2

V. EXPERIMENTAL DECISION AND ANALYSIS

In this section, our initial focus was to conduct experiments aimed at validating the efficacy of the homogeneity measure DI. Subsequently, a series of comparative algorithms were used to validate the superiority of the unsupervised feature selection method proposed in this article. Finally, a significance test was performed to ascertain the reliability and statistical significance of the algorithm.

A. Experimental Design

The experimental setup for this study involved using a computer equipped with a 64-bit Windows 10 operating system and an Intel Xeon W-2123 CPU. The processor had a base frequency of 3.6 GHz, and the system had 64 GB of memory capacity. All the algorithms, including the proposed one and the comparative ones, were implemented in Python and executed in Python 3.8. To validate the effectiveness of unsupervised feature selection, some labeled datasets were chosen for experimentation. Throughout the experimental process, the original labels were deliberately not used to prevent any potential interference caused by the original labels in the feature selection procedure. Instead, they were solely used as a point of reference to evaluate the effectiveness of the feature selection. We selected 12 datasets from two renowned repositories, namely, UCI and the KEEL-dataset repository as the experimental subjects, as shown in Table IV. All the 12 datasets were of numeric type. Prior to the experiments, we normalized the conditioning attributes of all the datasets using the Min–Max normalization method. For datasets with a runtime exceeding 48 h, we will extract a subset of objects or attributes for reference, such as 10% of objects or 1% of attributes. The results of the execution will be annotated with an asterisk (*) to indicate the distinction.

First, to validate the effectiveness of DI, we first sorted the features in descending order based on their homogeneity measure. Subsequently, we retained different proportions of features and compared the classification accuracy. To investigate the impact of DI on classification accuracy, we set the range of the feature retention rate α from 0.5 to 1, with a step size of 0.05. When $\alpha = 1$, the retained features are identical to the original dataset features.

Second, to demonstrate the superiority of our proposed algorithm, we selected the following experiments for comparison.

1) *Mixed-Core Fuzzy Reciprocal Information (HKFCMI) [35]:* Generalized attribute reduction algorithm based on mixed kernel function.
2) *Infinite Unsupervised Feature Selection (INF-UFS) [36]:* Utilization of filter feature selection framework for handling correlation and redundancy principles, with introduction of a simple unsupervised pruning strategy.
3) *Attribute Reduction Based on θ-Roughness (UM) [37]:* Construction of similarity relations using similarity rate θ, and introduction of the concepts of θ-precision and θ-roughness to measure uncertainty problem.
4) *Graph-Based Unsupervised Feature Selection (GBUFS) [32]:* A novel graph-theory-based feature selection method, leveraging the properties of matrix power series effectively.
5) *Attribute Reduction Based on Neighborhood Conditional Mutual Information (KNCMI) [38]:* Using novel NRS for constructing feature-target evaluation function and corresponding interactive feature selection algorithm.
6) *Feature Selection Based on Contradictory State Sequence (OSFCS) [39]:* Introducing contradictory objects to describe contradictory states and fuzzy contradictory states, and proposing a feature selection method based on contradictory state sequences and fuzzy contradictory state sequences.
7) *Feature Selection Based on Genetic Algorithm (CHCQX) [40]:* Constructing a lightweight qualitative metamodel and using it for feature selection.
8) *Feature Selection Based on Particle Swarm Optimization (PSO-MI) [41]:* Integration of filter and wrapper techniques with the introduction of a novel merger for merging effective features.

The proposed algorithm is compared and analyzed with the eight mentioned algorithms in terms of runtime, remaining number of features, and classification accuracy.

Finally, we discussed the influence of two parameters, the feature retention rate α and the number of neighbors k, on the final classification accuracy under different parameter combinations to determine the most effective parameter combination. Since a large value of k may not demonstrate the superiority of the k-nearest neighbor rough set, we set the range of the parameter k to be between 0.05 and 0.5, with a step size of 0.05.

Moreover, to eliminate data randomness and enhance the persuasiveness of the experimental results, we use four classifiers, namely, KNN, SVM, DT, and BYS, along with tenfold cross-validation to determine the final classification accuracy. Tenfold crossvalidation involves partitioning the dataset into ten equal segments. In this process, each segment is sequentially used as the test set, while the remaining nine segments serve as the training set. The results are reported as the average value (μ) with the corresponding standard deviation (σ) presented in the form of $\mu \pm \sigma$. In addition, to validate the reliability of the results, we use the Wilcoxon statistical test to verify the significance of algorithm comparisons.

B. Experimental Analysis

1) Effectiveness Analysis of DI: The impact of different feature retention rates on classification accuracy after attribute sorting using DI is shown in Fig. 1. From the figure, it can be observed that maintaining a certain proportion of features can either improve or maintain classification accuracy within a certain range. This demonstrates the effectiveness of DI measure for attribute importance assessment proposed in this article. Furthermore, removing attributes with low importance can effectively enhance classification accuracy and reduce the time required for feature selection. When the number of attributes in the dataset is large, different values of α can help maintain stable classification accuracy. Therefore, it is possible to lower α appropriately. For instance, in the case of the leukemia dataset with 12 583 attributes, a value of α equal to 0.5 yields better classification accuracy than α equal to 1 for KNN, while the accuracy remains unchanged for SVM and BYS, and only decreases from 90.36% to 87.68% for DT. Similarly, for datasets that contain a substantial amount of instances, different values of α can also maintain stable classification accuracy. For example, in the magic dataset, classification accuracy remains relatively unchanged when α is within the range of 0.5–1. However, for datasets with relatively fewer attributes, classification accuracy may suddenly change within a certain range, necessitating careful selection. For instance, in the seeds dataset, when α is set to 0.6, classification accuracy drops abruptly from 92.38% to 81.43% for KNN. In comparison to other algorithms, the distinguishing advantage of KND-UFS lies in its flexibility to adjust the selected number or proportion of features according to specific requirements. This attribute provides users with the convenience of customizing the feature selection process to best suit their specific needs and preferences. If the goal is to maximize reduction efficiency, the selected proportion of features can be appropriately reduced. In contrast, many other feature selection methods, such as UM, are limited to providing a single unique subset of features and lack the flexibility to adjust according to specific requirements. Unlike these methods, KND-UFS offers the advantage of adaptability.

2) Analysis of Comparative Experiments: We conducted a comparative analysis between our proposed methodology and other eight unsupervised feature selection methods. The average running times of each algorithm are presented in Table V, the number of selected features is shown in Table VI, and the classification accuracy of the selected features on KNN and SVM classifiers is displayed in Tables VII and VIII, respectively. It can be observed that the KND-UFS algorithm exhibits significant advantages on high-dimensional datasets. For instance, for the lungCancer dataset, the KND-UFS algorithm only takes 63.03 s to select four highly discriminative features from a pool of 12 533 features. In contrast, the IGUFS algorithm requires 34 h, and the INF-USF algorithm requires 30 h, with selected feature proportions of 50%, which are much higher than the 0.03% of the KND-UFS algorithm. Moreover, the KND-UFS algorithm achieves higher classification accuracy when using the selected features compared to the original dataset and other algorithms. The KND-UFS algorithm efficiently removes redundant features in a short

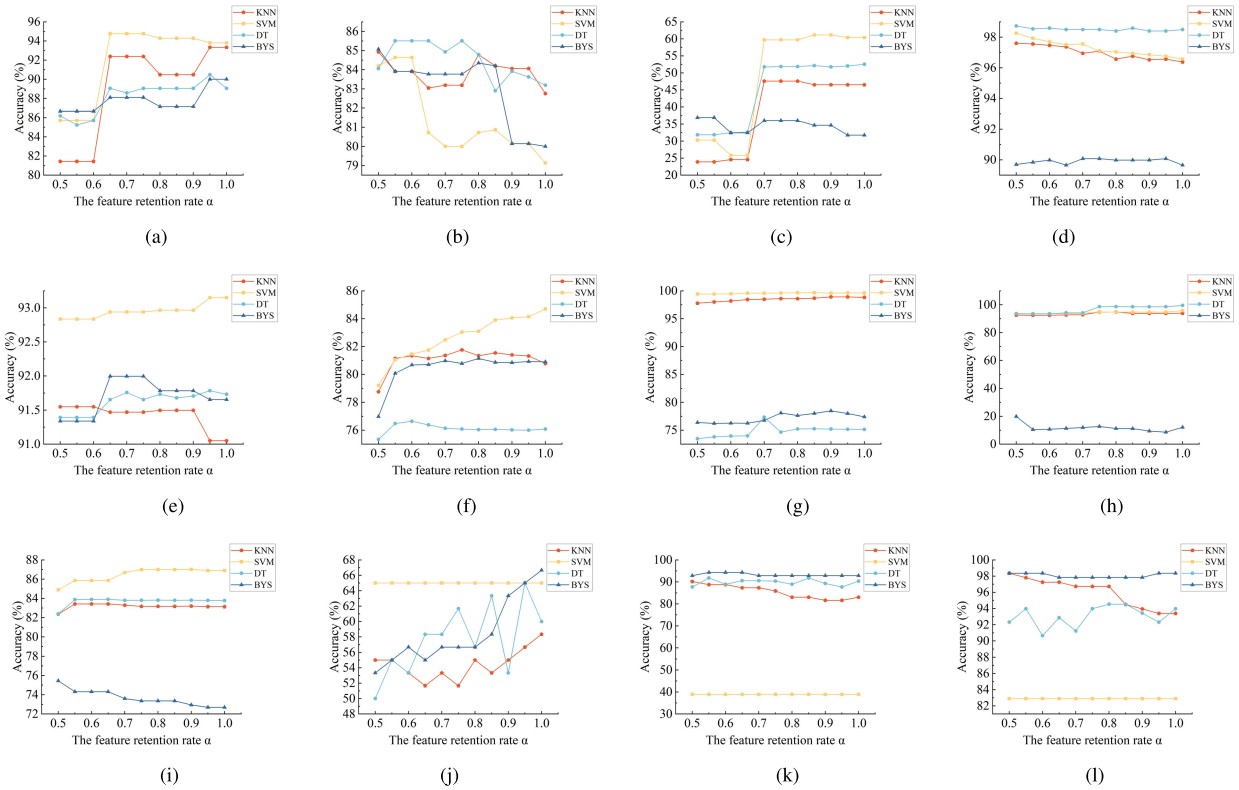


Fig. 1. Classification accuracy of different feature retention rates α under KNN, SVM, DT, and BYS classifiers. (a) Seeds. (b) Australian. (c) Concrete. (d) Cardiotocography. (e) Rice. (f) Waveform. (g) Texture. (h) Thyroid. (i) Magic. (j) nervousSystem. (k) Leukemia. (l) lungCancer.

TABLE V
COMPARISON OF RUNTIME OF DIFFERENT ALGORITHMS (S)

Datasets	HKCMI	INF-UFS	UM	IGUFS	KNCMI	OSFCS	CHCQX	PSO-MI	KND-UFS
seeds	21.70	**0.05**	75.14	0.69	3.97	0.05	1.14	4.20	0.08
australian	1035.39	**0.16**	1660.19	3.77	3.02	0.80	2.13	10.84	1.52
concrete	700.25	**0.05**	1971.48	3.83	90.02	1.23	2.84	15.16	3.55
cardiotocography	20684.25	**2.06**	23641.78	23.88	15.20	5.47	5.63	32.30	13.98
rice	7217.16	**0.38**	27609.13	16.92	1752.48	49.08	11.36	53.09	36.94
waveform	118672.52	**3.69**	143491.08	80.44	61.86	182.56	42.39	337.25	133.02
texture	533471.28	**13.48**	364296.52	209.80	30833.22	251.33	65.81	637.69	324.20
thyroid	2123.75*	**3.39**	341244.47	112.78	30240.55	150.52	5.78	194.75	263.64
magic	2775.70*	**2.14**	4784.25*	319.13	399.33*	1684.80	150.31	282.05	1251.09
nervousSystem	94.22*	32219.47	43.61*	34572.45	153.14*	**0.28**	737.72	115.06	9.84
leukemia	410.73*	103673.72	8.36*	106538.88	278.02*	**0.77**	3026.16	668.44	18.97
lungCancer	2583.05*	108810.42	54.67*	123060.39	350.94*	**1.64**	4557.00	810.80	63.03
Average	57482.50*	20394.08	75740.06*	22078.58	5348.48*	194.04	717.36	263.47	**176.65**

TABLE VI
COMPARISON OF THE NUMBER OF SELECTED FEATURES FOR DIFFERENT ALGORITHMS

Datasets	RAW	HKCMI	INF-UFS	UM	IGUFS	KNCMI	OSFCS	CHCQX	PSO-MI	KND-UFS
seeds	7	2	4	**1**	4	7	**1**	5	3	3
australian	14	2	9	**1**	9	**1**	2	5	2	6
concrete	8	2	6	**1**	6	5	**1**	5	3	4
cardiotocography	20	2	10	**1**	10	9	4	14	8	7
rice	7	2	5	**1**	5	6	**1**	4	3	5
waveform	21	2	15	**1**	15	21	3	13	15	14
texture	40	2	36	**1**	36	40	3	22	21	35
thyroid	21	2*	15	**1**	15	18	4	15	6	5
magic	10	2*	7	**1***	7	10*	2	8	4	7
nervousSystem	7129	**1***	3564	**1***	3564	55*	**1**	3598	3377	11
leukemia	12582	**1***	6291	**1***	6291	125*	2	6275	5975	21
lungCancer	12533	**1***	6266	**1***	6266	96*	2	6250	5851	4
Average	2699.33	1.75*	1352.33	**1.00***	1352.33	32.75*	2.17	1351.17	1272.33	10.17

amount of time, improving the classification accuracy. For example, the INF-UFS algorithm exhibits the fastest execution time on low-dimensional datasets, while OSFCS demonstrates superior performance on high-dimensional datasets. However, their final classification accuracy only marginally improves compared with the original dataset. On the other hand, the UM algorithm consistently selects the fewest features on low-dimensional datasets but is computationally expensive and yields significantly lower classification accuracy than the original dataset. Regarding the KNN and SVM classifiers,

TABLE VII
CLASSIFICATION ACCURACY OF DIFFERENT ALGORITHMS UNDER KNN CLASSIFIER (%)

Datasets	RAW	HKCMI	INF-UFS	UM	IGUFS	KNCMI	OSFCS	CHCQX	PSO-MI	KND-UFS
seeds	93.33±4.36	89.05±7.39	92.38±3.81	83.33±6.48	87.62±7.13	93.33±4.36	65.24±10.87	91.43±5.55	91.9±5.24	**94.29±3.56**
australian	84.2±5.12	61.59±5.32	83.77±4.57	51.01±6.79	83.77±4.53	85.51±5.31	66.38±4.62	85.07±5.54	85.36±5.55	**85.65±3.8**
concrete	46.5±3.72	41.26±5.48	49.32±2.94	23.11±2.63	49.32±2.94	50.0±3.23	37.96±4.52	54.08±3.34	53.5±5.22	**55.92±4.1**
cardiotocography	96.99±0.79	72.06±2.39	96.85±1.1	69.61±3.1	97.04±0.63	97.41±0.85	97.6±0.77	96.99±0.7	97.55±0.86	**98.02±0.91**
rice	88.48±1.22	87.53±1.56	88.4±1.31	81.15±1.33	88.4±1.31	88.5±1.07	55.77±2.82	89.08±2.12	89.08±1.98	**89.84±2.36**
waveform	77.48±1.97	56.08±1.43	72.7±2.55	33.36±3.02	76.18±1.74	77.48±1.97	38.9±2.55	77.3±1.51	76.3±1.24	**78.66±1.41**
texture	98.82±0.37	58.62±1.12	98.78±0.38	14.27±0.93	98.8±0.35	98.82±0.37	62.22±2.09	98.71±0.41	98.8±0.36	**98.96±0.45**
thyroid	93.88±0.86	92.92±2.88*	90.99±0.76	91.61±1.21	90.99±0.76	93.92±0.88	92.36±0.94	95.76±0.78	**97.99±0.62**	96.56±0.56
magic	83.14±0.71	68.93±2.65*	78.58±0.91	65.3±3.28*	79.25±1.21	81.23±3.2*	73.01±1.16	82.15±1.05	80.26±0.74	**83.74±0.64**
nervousSystem	60.0±15.28	56.67±15.28*	60.0±20.0	58.33±13.44*	58.33±15.37	63.33±17.95*	56.67±15.28	53.33±14.53	58.33±17.08	**70.0±16.33**
leukemia	83.04±12.51	56.61±18.49*	66.43±23.23	46.61±18.84*	80.18±17.21	80.18±15.98*	51.25±9.78	78.93±14.5	80.18±13.18	**83.57±14.59**
lungCancer	85.12±7.78	79.01±12.12*	85.12±7.78	80.67±7.92*	86.23±7.84	84.53±8.16*	85.67±8.21	84.56±7.68	84.01±8.33	**87.89±9.11**
Average	82.58±4.56	68.36±6.34*	80.28±5.78	58.2±5.75*	81.34±5.09	82.85±5.28*	65.25±5.3	82.28±4.81	82.77±5.03	**85.26±4.82**

TABLE VIII
CLASSIFICATION ACCURACY OF DIFFERENT ALGORITHMS UNDER SVM CLASSIFIER (%)

Datasets	RAW	HKCMI	INF-UFS	UM	IGUFS	KNCMI	OSFCS	CHCQX	PSO-MI	KND-UFS
seeds	93.33±3.81	87.62±6.46	93.33±3.81	85.71±4.76	88.57±7.74	93.33±3.81	72.86±5.65	92.38±4.36	92.86±5.73	**95.24±3.69**
australian	84.78±4.77	58.84±5.23	84.49±5.65	55.51±5.94	84.2±5.28	85.51±5.31	59.42±4.44	85.36±4.74	84.93±5.23	**86.09±5.03**
concrete	60.0±6.06	43.59±4.8	59.81±4.56	41.26±3.29	59.81±4.56	56.6±4.53	44.66±3.47	**62.04±4.69**	57.28±4.47	**62.04±4.42**
cardiotocography	97.55±0.59	77.85±1.98	98.02±0.66	77.85±1.98	97.98±0.67	98.49±0.78	97.6±0.71	97.27±0.72	98.02±0.81	**98.78±0.67**
rice	**92.99±1.5**	91.29±1.6	92.78±1.52	88.01±1.22	92.78±1.52	92.76±1.54	63.62±3.64	92.83±1.54	92.86±1.42	**92.99±1.46**
waveform	85.3±1.28	66.08±1.16	82.08±1.66	31.78±1.67	83.7±1.06	85.3±1.28	46.14±1.59	84.64±1.39	84.28±1.24	**85.82±1.16**
texture	98.95±0.29	63.95±1.5	98.93±0.31	15.87±1.75	98.91±0.29	98.95±0.29	61.56±2.02	98.96±0.29	98.93±0.37	**99.07±0.37**
thyroid	95.72±0.55	94.03±2.92*	92.53±0.65	92.58±0.66	92.53±0.65	95.62±0.62	93.65±0.72	**97.04±0.62**	96.4±0.53	95.1±0.82
magic	86.9±0.88	74.5±2.83*	81.41±1.16	71.71±2.1*	82.58±1.04	84.28±2.45*	76.65±0.92	86.23±0.9	82.07±0.52	**86.97±0.88**
nervousSystem	65.0±15.72	65.0±15.72*	65.0±15.72	65.0±15.72*	65.0±15.72	65.0±15.72*	60.0±15.28	65.0±15.72	65.0±15.72	**78.33±13.02**
leukemia	38.93±20.73	54.46±22.03*	38.93±20.73	49.29±21.01*	38.93±20.73	38.93±20.73*	43.04±14.56	38.93±20.73	38.93±20.73	**62.86±14.21**
lungCancer	82.89±8.71	82.89±8.71*	82.89±8.71	81.78±8.58*	82.89±8.71	82.89±8.71*	85.18±9.31	82.89±8.71	82.89±8.71	**92.34±6.42**
Average	81.86±5.41	71.68±6.25*	80.85±5.43	63.03±5.72*	80.66±5.66	81.47±5.48*	67.03±5.19	81.96±5.37	81.2±5.46	**86.3±4.35**

TABLE IX
WILCOXON TEST P-VALUE RESULTS UNDER KNN

Datasets	HKCMI	INF-UFS	UM	IGUFS	KNCMI	OSFCS	CHCQX	PSO-MI
seeds	<0.1	0.13	<0.01	<0.1	0.31	<0.01	<0.1	0.11
australian	<0.01	<0.1	<0.01	<0.1	0.45	<0.01	0.43	0.36
concrete	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.1	<0.1
cardiotocography	<0.01	<0.01	<0.01	<0.1	<0.01	<0.1	<0.1	0.1
rice	<0.1	<0.1	<0.01	<0.1	<0.1	<0.01	<0.1	<0.1
waveform	<0.01	<0.01	<0.01	<0.01	<0.1	<0.01	<0.1	<0.01
texture	<0.01	<0.1	<0.01	<0.1	<0.1	<0.01	<0.1	<0.1
thyroid	<0.01*	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	1
magic	<0.01*	<0.01	<0.01*	<0.01	<0.1*	<0.01	<0.01	<0.01
nervousSystem	<0.1*	0.2	0.15*	0.14	0.17*	<0.1	<0.1	0.13
leukemia	<0.01*	<0.1	<0.01*	0.31	0.31*	<0.01	0.15	0.26
lungCancer	<0.1*	<0.1	<0.1*	0.29	<0.1*	<0.1	<0.1	<0.1

TABLE X
WILCOXON TEST P-VALUE RESULTS UNDER SVM

Datasets	HKCMI	INF-UFS	UM	IGUFS	KNCMI	OSFCS	CHCQX	PSO-MI
seeds	<0.1	<0.1	<0.01	<0.1	<0.1	<0.01	<0.1	0.1
australian	<0.01	<0.1	<0.01	<0.1	0.13	<0.01	0.24	0.11
concrete	<0.01	<0.1	<0.01	<0.1	<0.01	<0.01	0.55	<0.01
cardiotocography	<0.01	<0.1	<0.01	<0.1	<0.1	<0.01	<0.01	<0.1
rice	<0.01	0.25	<0.01	0.25	0.12	<0.01	<0.1	<0.1
waveform	<0.01	<0.01	<0.01	<0.01	<0.1	<0.01	<0.01	<0.01
texture	<0.01	<0.1	<0.01	<0.1	<0.1	<0.01	0.25	0.1
thyroid	0.22*	<0.01	<0.01	<0.01	0.99	<0.01	1	1
magic	<0.01*	<0.01	<0.01*	<0.01	<0.01*	<0.01	<0.01	<0.01
nervousSystem	<0.1*	<0.1	<0.1*	<0.1	<0.1*	<0.1	<0.1	<0.1
leukemia	0.15	<0.01	<0.1*	<0.01	<0.01*	<0.1	<0.01	<0.01
lungCancer	<0.1*	<0.1	<0.1*	<0.1	<0.1*	<0.1	<0.1	<0.1

the KND-UFS algorithm achieves the highest accuracy in 11 datasets, outperforming other algorithms. The KND-UFS algorithm exhibits exceptional performance compared with the other eight algorithms in terms of mean execution time, mean number of chosen features, and mean classification accuracy. These findings underscore the efficacy and efficiency of our suggested algorithm as a superior technique for feature selection.

3) Parameter Analysis: To investigate the impact of two parameters, namely, the feature retention rate α and the neighborhood size k, on the final classification accuracy of the proposed KND-UFS algorithm, we have generated 3-D plots depicting the classification accuracy for 12 datasets, as shown in Fig. 2. It can be observed from the plots that different combinations of α and k have significant effects on the resulting classification accuracy. For example, in the thyroid dataset, the KND-UFS algorithm achieves a maximum classification accuracy of 96.56% when setting α to 0.75 and k to 0.35, indicating the relative optimality of the proposed algorithm. Moreover, the impact of α and k on classification accuracy varies with the size of the dataset. For instance, in the lungCancer dataset, α has minimal effect on classification accuracy, while different values of k result in significant fluctuations in accuracy. Conversely, in the concrete dataset, the setting of α becomes more crucial. Consequently, the parameter combinations that yield optimal classification results tend to differ. For each dataset, it is possible to determine the appropriate values of parameters α and k that, when combined, yield a relatively optimal classification accuracy using the KND-UFS algorithm.

4) Hypothesis Testing: Due to the nonparametric nature and the advantage of not requiring the assumption of normality, we used the Wilcoxon test to validate the effectiveness of algorithm comparisons. To conduct a comprehensive comparison of the experimental results obtained from various algorithms, we conducted hypothesis tests on 12 datasets against the eight comparative algorithms. In each hypothesis test, the null hypothesis was formulated to assert that the

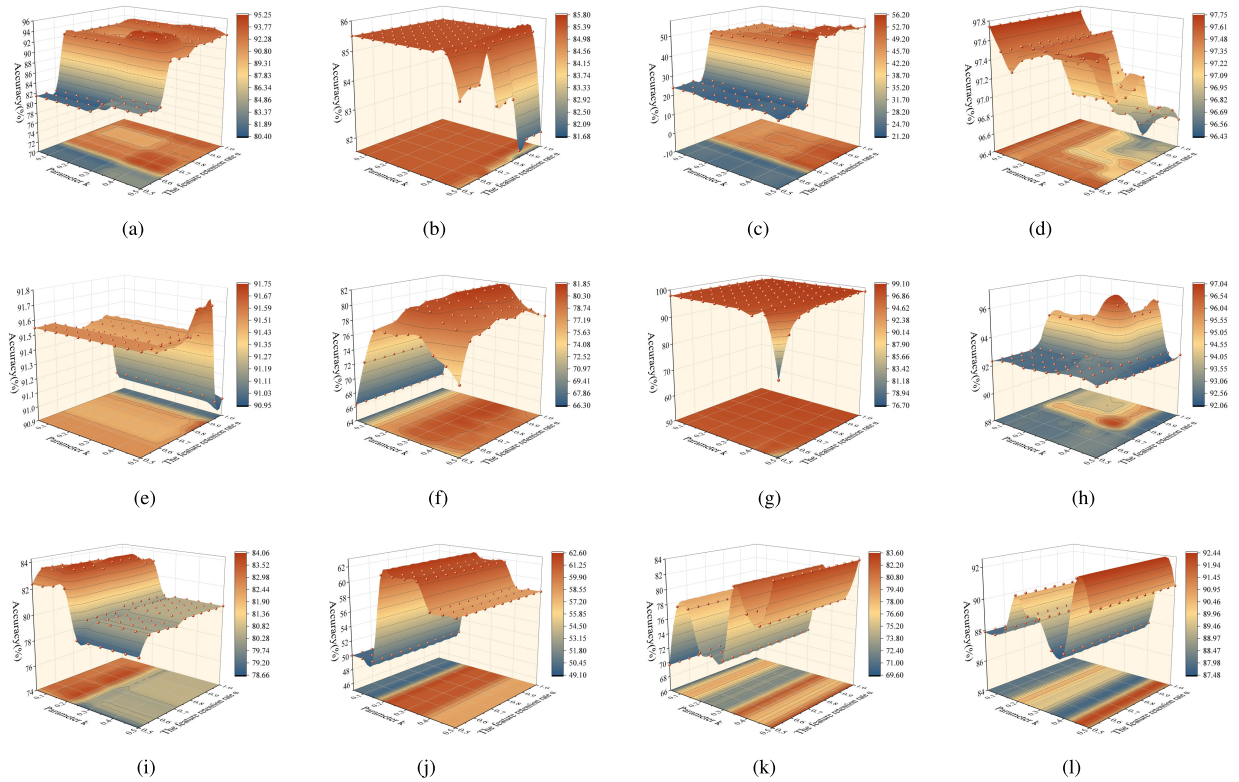


Fig. 2. Classification accuracy of different combinations of α and k in KNN classifier. (a) Seeds. (b) Australian. (c) Concrete. (d) Cardiotocography. (e) Rice. (f) Waveform. (g) Texture. (h) Thyroid. (i) Magic. (j) nervousSystem. (k) Leukemia. (l) lungCancer.

classification accuracy of the KND-UFS algorithm is less than or equal to the classification accuracy of the other eight comparative algorithms. By evaluating the p-values obtained from the Wilcoxon test, it was observed that the majority of the p-values were less than the predetermined significance level of 0.05. Therefore, the null hypothesis was refuted in these cases. This implies that our KND-UFS algorithm consistently exhibits higher classification accuracy compared with the other comparative models, as supported by the statistical analysis. The p-value results for the KNN and SVM classifiers can be found in Tables IX and X, respectively. In conclusion, based on the KNN and SVM classifiers, the KND-UFS algorithm proves to be effective for unsupervised feature selection.

VI. Conclusion

This study introduces a novel feature selection method that combines the FCM with k-nearest neighbor rough set. The proposed method introduces a unique ranking approach to assess the importance of features and selects the most significant ones to build the feature selection algorithm. To evaluate its effectiveness, the proposed algorithm is compared against eight existing feature selection algorithms using 12 public datasets. The experimental results validate the superiority and efficacy of the proposed method as a feature selection technique.

The research methodology in this study involves two parameters, namely, α (the proportion of selected features) and k (the number of neighbors). The performance of feature selection relies on selecting appropriate combinations of these parameters to achieve relative optimality. However, there is currently no automatic mechanism to determine the ideal values for α and k. Therefore, a research direction for future work is to investigate methods for adapting these parameters automatically. Furthermore, the feature selection algorithm studied in this article focuses on static datasets. Consequently, developing an efficient mechanism for dynamic datasets' feature selection is a potential area for future research. In addition, it is important to situate this study within a specific application context, such as the selection of hyperspectral bands [42], [43]. Therefore, identifying a suitable application scenario is also a research direction for future work.

References

[1] Y. Fan, C. Liu, and J. Wang, "Integrating multi-granularity model and similarity measurement for transforming process data into different granularity knowledge," *Adv. Eng. Informat.*, vol. 37, pp. 88–102, Aug. 2018.

[2] M. Li, H. Wang, and J. Li, "Mining conditional functional dependency rules on big data," *Big Data Mining Anal.*, vol. 3, no. 1, pp. 68–84, Mar. 2020.

[3] J.-T. Yao, A. V. Vasilakos, and W. Pedrycz, "Granular computing: Perspectives and challenges," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1977–1989, Dec. 2013.

[4] X. Hu, Y. Shen, W. Pedrycz, Y. Li, and G. Wu, "Granular fuzzy rule-based modeling with incomplete data representation," *IEEE Trans. Cybern.*, vol. 52, no. 7, pp. 6420–6433, Jul. 2022.

[5] L. Cardelli and P. Wegner, "On understanding types, data abstraction, and polymorphism," *ACM Comput. Surv.*, vol. 17, no. 4, pp. 471–523, Dec. 1985.

[6] T. Y. Lin, "Data mining and machine oriented modeling: A granular computing approach," *Appl. Intell.*, vol. 13, pp. 113–124, Sep. 2000.

[7] W. H. Xu, D. D. Guo, Y. H. Qian, and W. P. 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.

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

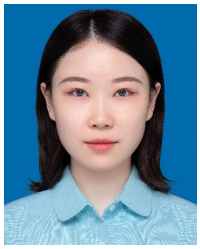
[9] L. Chen, L. Zhao, Z. Xiao, Y. Liu, and J. Wang, "A granular computing based classification method from algebraic granule structure," *IEEE Access*, vol. 9, pp. 68118–68126, 2021.

[10] C. Ma, L. Zhang, W. Pedrycz, and W. Lu, "The long-term prediction of time series: A granular computing-based design approach," *IEEE Trans. Syst. Man, Cybern., Syst.*, vol. 52, no. 10, pp. 6326–6338, Oct. 2022.
[11] X. Zhu, W. Pedrycz, and Z. Li, "Development and analysis of neural networks realized in the presence of granular data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 9, pp. 3606–3619, Sep. 2020.
[12] L. A. Zadeh, "Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic," *Fuzzy Sets Syst.*, vol. 90, no. 2, pp. 111–127, Sep. 1997.
[13] Z. Pawlak, "Rough sets," *Int. J. Comput. Inf. Sci.*, vol. 11, no. 5, pp. 341–356, Oct. 1982.
[14] M. De Cock, C. Cornelis, and E. E. Kerre, "Fuzzy rough sets: The forgotten step," *IEEE Trans. Fuzzy Syst.*, vol. 15, no. 1, pp. 121–130, Feb. 2007.
[15] D. Dubois and H. Prade, "Rough fuzzy sets and fuzzy rough sets," *Int. J. General Syst.*, vol. 17, nos. 2–3, pp. 191–209, 1990.
[16] Z. Pawlak, "Rough set theory and its applications to data analysis," *Cybern. Syst.*, vol. 29, no. 7, pp. 661–688, 1998.
[17] J. W. Guan and D. A. Bell, "Rough computational methods for information systems," *Artif. Intell.*, vol. 105, nos. 1–2, pp. 77–103, Oct. 1998.
[18] W. Zheng, S. Chen, Z. Fu, F. Zhu, H. Yan, and J. Yang, "Feature selection boosted by unselected features," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 9, pp. 4562–4574, Sep. 2022.
[19] X. Chen, G. Yuan, W. Wang, F. Nie, X. Chang, and J. Z. Huang, "Local adaptive projection framework for feature selection of labeled and unlabeled data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 12, pp. 6362–6373, Dec. 2018.
[20] J. Miao and L. Niu, "A survey on feature selection," *Proc. Comput. Sci.*, vol. 91, pp. 919–926, Jan. 2016.
[21] N. M. Parthaláin and R. Jensen, "Unsupervised fuzzy-rough set-based dimensionality reduction," *Inf. Sci.*, vol. 229, pp. 106–121, Apr. 2013.
[22] Z. Yuan, H. Chen, T. Li, Z. Yu, B. Sang, and C. Luo, "Unsupervised attribute reduction for mixed data based on fuzzy rough sets," *Inf. Sci.*, vol. 572, pp. 67–87, Sep. 2021.
[23] P. Huang, Z. Kong, M. Xie, and X. Yang, "Robust unsupervised feature selection via data relationship learning," *Pattern Recognit.*, vol. 142, Oct. 2023, Art. no. 109676.
[24] X. Lin, J. Guan, B. Chen, and Y. Zeng, "Unsupervised feature selection via orthogonal basis clustering and local structure preserving," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 11, pp. 6881–6892, Nov. 2022.
[25] R. Zhang and X. Li, "Regularized regression with fuzzy membership embedding for unsupervised feature selection," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 12, pp. 3743–3753, Dec. 2021.
[26] Z. Yuan, H. Chen, P. Zhang, J. Wan, and T. Li, "A novel unsupervised approach to heterogeneous feature selection based on fuzzy mutual information," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 9, pp. 3395–3409, Sep. 2022.
[27] C. Tang, X. Zheng, W. Zhang, X. Liu, X. Zhu, and E. Zhu, "Unsupervised feature selection via multiple graph fusion and feature weight learning," *Sci. China Inf. Sci.*, vol. 66, no. 5, May 2023, Art. no. 152101.
[28] Y. Gan, X. Dong, H. Zhou, F. Gao, and J. Dong, "Learning the precise feature for cluster assignment," *IEEE Trans. Cybern.*, vol. 52, no. 8, pp. 8587–8600, Aug. 2022.
[29] Z. Li, F. Nie, D. Wu, Z. Hu, and X. Li, "Unsupervised feature selection with weighted and projected adaptive neighbors," *IEEE Trans. Cybern.*, vol. 53, no. 2, pp. 1260–1271, Feb. 2023.
[30] C. Wang, W. Pedrycz, J. Yang, M. C. Zhou, and Z. Li, "Wavelet frame-based fuzzy C-means clustering for segmenting images on graphs," *IEEE Trans. Cybern.*, vol. 50, no. 9, pp. 3938–3949, Jun. 2020.
[31] W. Li et al., "Feature selection approach based on improved fuzzy C-means with principle of refined justifiable granularity," *IEEE Trans. Fuzzy Syst.*, vol. 31, no. 7, pp. 2112–2126, Jul. 2023.
[32] 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.*, vol. 35, no. 9, pp. 12576–12589, Sep. 2024.
[33] C. Wang, Q. Hu, X. Wang, D. Chen, Y. Qian, and Z. Dong, "Feature selection based on neighborhood discrimination index," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 7, pp. 2986–2999, Jul. 2018.
[34] M. Hu, E. C. C. Tsang, Y. Guo, D. Chen, and W. Xu, "Attribute reduction based on overlap degree and k-nearest-neighbor rough sets in decision information systems," *Inf. Sci.*, vol. 584, pp. 301–324, Jan. 2022.
[35] 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, Nov. 2021, Art. no. 107398.
[36] 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.
[37] J. Dai, W. Wang, and J. Mi, "Uncertainty measurement for interval-valued information systems," *Inf. Sci.*, vol. 251, pp. 63–78, Dec. 2013.
[38] 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.
[39] X. Zhang and Z. Feng, "Feature selection based on contradictory state sequence for multi-scale interval valued decision table," *Inf. Sci.*, vol. 677, Aug. 2024, Art. no. 120926.
[40] M. G. Altarabichi, S. Nowaczyk, S. Pashami, and P. Sheikholharam Mashhadi, "Fast genetic algorithm for feature selection—A qualitative approximation approach," in *Proc. Companion Conf. Genetic Evol. Comput.*, Jul. 2023, pp. 11–12.
[41] K. Robindro, S. S. Devi, U. B. Clinton, L. Takhellambam, Y. R. Singh, and N. Hoque, "Hybrid distributed feature selection using particle swarm optimization-mutual information," *Data Sci. Manage.*, vol. 7, no. 1, pp. 64–73, Mar. 2024.
[42] J. Wang et al., "Region-aware hierarchical latent feature representation learning-guided clustering for hyperspectral band selection," *IEEE Trans. Cybern.*, vol. 53, no. 8, pp. 5250–5263, Aug. 2023.
[43] C. Tang et al., "Spatial and spectral structure preserved self-representation for unsupervised hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5531413.

Weihua Xu received the Ph.D. degree in mathematics from School of Sciences, Xian Jiaotong University, Xian, China, in 2007, and the M.Sc. degree in mathematics from the School of Mathematics and Information Sciences, Guangxi University, Nanning, China, in 2004.

He is currently a Professor with the College of Artificial Intelligence, Southwest University, Chongqing, China. He has published five monographs and more than 240 articles in international journals. His current research interests include granular computing, cognitive computing, and information fusion.

Dr. Xu also serves as a Senior Member of Chinese Association for Artificial Intelligence (CAAI). He serves on the Associate Editor of *International Journal of Machine Learning* and *Cybernetics and Journal of Intelligent and Fuzzy Systems*.



Yang Zhang received the B.S. degree from the School of Computer and Information Science, Chongqing Normal University, Chongqing, China, in 2020. She is currently pursuing the M.Sc. degree with the College of Artificial Intelligence, Southwest University, Chongqing.

Her current research interests include feature selection, granular computing, and fuzzy sets.

Yuhua Qian (Member, IEEE) received the M.S. and Ph.D. degrees in computer applications from Shanxi University, Taiyuan, China, in 2005 and 2011, respectively.

He currently holds the position of Director of the Institute of Big Data Science and Industry, Shanxi University. He is a Professor with the Key Laboratory of Computational Intelligence and Chinese Information Processing, Ministry of Education, Taiyuan. His research interests include machine learning, pattern recognition, feature selection, granular computing, and artificial intelligence. He has published more than 100 articles on these subjects in renowned international journals.

Dr. Qian has been a member of the Editorial Board of the *International Journal of Knowledge-Based Organizations* and *Journal of Artificial Intelligence Research*. He has been a PC Member at various conferences focused on machine learning and data mining. He has also taken on roles such as the Program Chair or the Special Issue Chair at conferences such as the Conference on Rough Sets and Knowledge Technology, the Joint Rough Set Symposium, and the Conference on Industrial Instrumentation and Control.