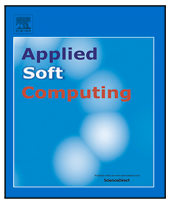
Applied Soft Computing Journal 169 (2025) 112572

Contents lists available at ScienceDirect

Applied Soft Computing

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



Fuzzy multi-neighborhood entropy-based interactive feature selection for unsupervised outlier detection

Siyu Yang [a], Zhong Yuan [a,*], Chuan Luo [a], Hongmei Chen [b], Dezhong Peng [a,c,d]

[a] College of Computer Science, Sichuan University, Chengdu 610065, China
[b] School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China
[c] Sichuan National Innovation New Vision UHD Video Technology Co., Ltd, Chengdu 610095, China
[d] National Innovation Center for UHD Video Technology, Chengdu 610095, China

ARTICLE INFO

Dataset link: https://github.com/BELLoney/Outlier-detection

Keywords:
Unsupervised feature selection
Fuzzy multi-neighborhood uncertainty measures
Feature interactivity
Outlier detection
Heterogeneous data

ABSTRACT

Unsupervised feature selection is one of the important techniques for unsupervised knowledge discovery, which aims to reduce the dimensionality of conditional feature sets as much as possible to improve the efficiency and accuracy of the algorithm. However, existing methods have the following two challenges: (1) They are mainly applicable to select numerical or nominal features and cannot effectively select heterogeneous features; (2) The relevance and redundancy are primarily considered to construct feature evaluation indexes, ignoring the interaction information of heterogeneous features. To solve the challenges mentioned above, this paper proposes an unsupervised heterogeneous feature selection method based on fuzzy multi-neighborhood entropy, which also considers the multi-correlation of features to select heterogeneous features. First, the fuzzy multi-neighborhood granule is constructed by considering the distribution characteristics of the data. Then, the concept of fuzzy entropy is introduced to define the fuzzy multi-neighborhood entropy and its associated uncertainty measures, and the relationship between them is discussed. Next, the relevance, redundancy, and interactivity among attributes are defined, and the idea of maximum relevance-minimum redundancy-maximum interactivity is used to construct the evaluation indexes of heterogeneous features. Finally, experiments are conducted on several publicly unbalanced datasets, and the results are in comparison with existing algorithms. The experimental results show that the proposed algorithm can select fewer heterogeneous features to improve the efficiency of outlier detection tasks. The code is publicly available online at https://github.com/BELLoney/MNIFS.

1. Introduction

In recent years, the explosive growth of data has led to a significant increase in the number of features in machine learning and data mining problems. As a result, the dimensional disasters and overfitting problems associated with high-dimensional data have posed great challenges for real-world applications. As an important dimensionality reduction technique, unsupervised feature selection does not consider label information. It reduces the number of features by selecting the most representative features from the original data. Also, it can lower the data processing cost and improve the performance of the learning model. Unlike supervised feature selection methods, unsupervised feature selection does not rely on label information and can be performed on large-scale datasets. Due to the ability of unsupervised feature selection to handle complex datasets without labels, it has higher flexibility and versatility. Currently, unsupervised feature selection has been widely used in a variety of learning tasks, such as clustering analysis [1], outlier detection [2], and medical analysis [3]. When performing outlier detection, redundant features and complex interactions between features in high-dimensional datasets may make outlier detection algorithms inefficient, thus making it difficult to correctly determine anomalies and seriously increasing time costs. Therefore, exploring more effective unsupervised feature selection techniques is of great importance in the field of outlier detection and other fields.

From the data perspective, representative unsupervised feature selection methods can be divided into four main categories [4]: similarity-based [5,6], information-theoretic-based [7,8], sparse learning-based [9], and statistical-based methods [10]. The similarity-based methods can identify highly correlated features. The information-theoretic-based methods adopt the quantitative information amount to assess the correlation between features, which can effectively discover the nonlinear relationship as well as the dependency relationship between features.

* Corresponding author.
E-mail address: yuanzhong@scu.edu.cn (Z. Yuan).

https://doi.org/10.1016/j.asoc.2024.112572
Received 1 June 2023; Received in revised form 6 September 2024; Accepted 24 November 2024
Available online 3 December 2024

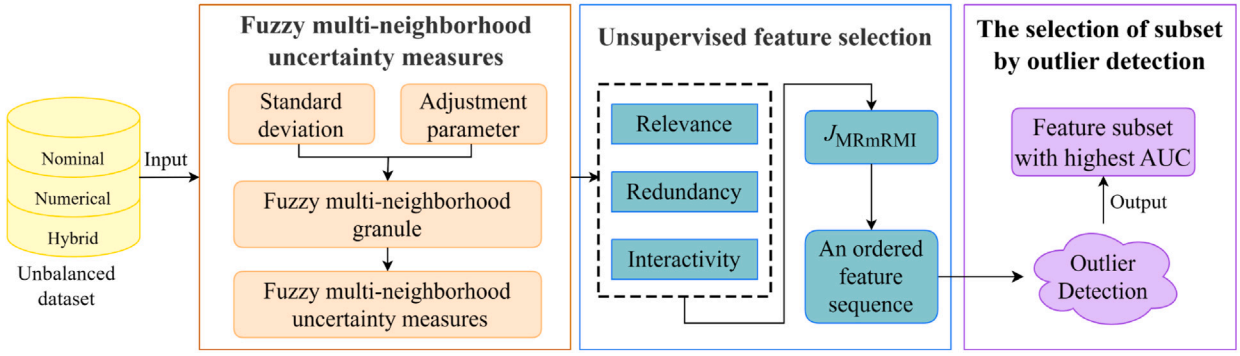


Fig. 1. Overall framework of this paper.

The sparse learning-based methods can improve the generalization ability of the model by introducing regularization terms to obtain a sparse feature weight matrix. The statistical-based methods classify important features based on the statistical significance to compare the differences between variables. However, these methods have the following shortcomings: (1) Limited applicable data types. Most of the methods based on similarity and information theory apply to numerical or nominal data and are not suitable for handling heterogeneous data. The processing of a single type of data makes it difficult to meet the demand [11]. (2) Difficult to deal with uncertain information. The complexity and variability of the environment will lead to a large amount of uncertain information in the data. The methods mentioned above are difficult to handle uncertain information which will affect the efficiency of the learning model.

In order to effectively deal with uncertain information such as random and fuzzy, while preserving the distribution structure of the data, Dubios and Prade [12] proposed a fuzzy rough set (FRS) model. This model combines rough set theory and fuzzy set theory. The degree of similarity between data objects is measured by replacing equivalence relations with fuzzy similarity relations. It makes up for the deficiency that rough sets can only be used to deal with nominal attributes. Feature selection is a major application of FRS theory. Current feature selection methods on the basis of FRSs are mainly classified into methods based on fuzzy distinguishing matrix [13], methods based on fuzzy dependency [14], and methods based on fuzzy uncertainty measure [15]. On the basis of these methods, many scholars have further extended them [16–19]. It should be noted that most of the existing fuzzy rough feature selection methods work for supervised learning scenarios. In this scenario, only labeled data is applicable. If there is a lack of decision information directing the selection process, unsupervised feature selection methods are used. In real-world applications, most of the data is unlabeled. While manually labeling data can solve part of the problem, obtaining labeled information is time-consuming and challenging. Hence, unsupervised feature selection becomes more attractive in practical applications [20]. Currently, there are still few research works on unsupervised feature selection based on FRSs [21–24]. Besides, in these methods, relevance and redundancy between features are mainly considered, while the interactivity between features is rarely taken into account. The interactivity of features refers to the interaction and dependency between features, where individual features may not describe the classification of data well and the influence of other features must be regarded [25]. If we ignore the interactivity between features, it may result in some features that can provide important discriminative information being lost or incorrectly removed.

Based on the above discussion, an unsupervised heterogeneous feature selection method based on fuzzy multi-neighborhood entropy is proposed in this paper. This method considers the distribution structure of each feature to construct the fuzzy multi-neighborhood granule. The fuzzy multi-neighborhood entropy and its extension concepts are used to better measure the uncertainty of features. First, fuzzy mutual information is used to represent the fuzzy relevance of each feature, selecting the feature with the maximum relevance. Second, for the candidate features, their fuzzy redundancy is measured by defining fuzzy conditional relevance, and fuzzy multi-neighborhood conditional mutual information is used to represent their fuzzy interactivity. Third, a fuzzy Multi-Neighborhood entropy-based Interactive unsupervised Feature Selection (MNIFS) algorithm is constructed to derive a feature sequence. Finally, a comparative analysis is performed with some existing feature selection algorithms on 24 publicly unbalanced datasets. These experiments verify the ability of this algorithm in improving the outlier detection performance of heterogeneous data by using a smaller number of features. In addition, the hypothesis statistical tests further verify the statistically significant differences of the proposed algorithm. The overall research framework of this paper is shown in Fig. 1. To summarize, our main contributions in this paper are as follows.

(1) Fuzzy multi-neighborhood entropy is utilized for efficient processing of uncertain information.
(2) In assessing the significance of features, the relevance, redundancy, and interactivity of features are comprehensively considered.
(3) No discretization of numerical attributes is required, which reduces preprocessing time while maintaining the distributional structure of data.
(4) As an unsupervised heterogeneous feature selection algorithm, MNIFS requires no label information that can be efficiently applied to nominal, numerical, and heterogeneous features.
(5) Experimental results show that MNIFS markedly outperforms other existing feature selection algorithms in terms of flexibility and performance.

The rest of the paper is described below. Section 2 briefly describes the related work of existing feature selection methods. Section 3 reviews the preliminary knowledge of FRS models. Section 4 describes in detail the construction of the fuzzy multi-neighborhood granule and fuzzy multi-neighborhood uncertainty measures. Section 5 proposes the multi-correlation in the feature selection process and designs the algorithm MNIFS. Section 6 shows the experimental results and their analysis. Section 7 concludes this paper.

2. Related work

This section introduces some classical feature selection methods, including unsupervised feature selection methods and fuzzy rough-based feature selection methods.

2.1. Unsupervised feature selection

As mentioned above, unsupervised feature selection methods can be broadly classified into four major categories based on the strategies used [4], including similarity-based, information-theoretic-based, sparse learning-based, and statistical-based methods.

Similarity-based methods construct feature importance evaluation functions by using the similarity between features as a way to select features. For example, He et al. [5] proposed a feature selection algorithm that uses Laplacian scores to evaluate feature importance. Yao et al. [6] applied the idea of locally linear embedding to feature selection based on filters to measure the difference between the local structures of each feature and the original data. The information-theoretic-based approach considers the dependencies among all features and evaluates the importance of features by calculating information entropy and information gain. Lim et al. [7] were the first to apply pairwise dependency information of features to feature selection, combining dependency between features and unsupervised feature selection based on regression. Sparse learning-based methods are sparse in the feature weights and feature selection. The data is compressed by optimizing the objective function to set the weights of some features to zero. For example, Cai et al. [9] selected the features that best preserve the multicluster structure of the data, involving only the sparse feature problem and the L1-regularized least squares problem. The statistical-based approach introduces statistical properties of the data distribution. It calculates the significant differences between different characteristics and the target variables under statistical significance. For example, He et al. [10] used traces and determinants to measure the magnitude of the covariance matrix and selected features that minimize the parametric covariance matrix of the regularized regression model.

2.2. Fuzzy rough-based feature selection

With the development carried out by combining fuzzy sets and rough sets, FRSs can effectively solve the problem that classical rough sets cannot handle heterogeneous data. According to different simplification rules, the feature selection methods can be broadly classified into three categories: the method based on fuzzy discernibility matrix [13], the method based on fuzzy dependence [14], and the method based on fuzzy uncertainty measure [15].

In the fuzzy discernibility matrix-based approach, Tsang et al. [26] extended the idea of classical distinguishing matrices to FRSs, using a discernibility matrix to compute all attribute reductions in an algorithm for attribute approximation. Wei et al. [27] proposed two classes of fuzzy rough lower and upper approximations and corresponding positive region reducts for set-valued data, based on which two classes of discriminative matrices and functions are constructed to acquire these proposed reducts.

For fuzzy dependence-based approaches, Zhao et al. [28] developed an FRS model for hierarchical structures and calculated lower and upper approximations for classes organized with class hierarchies. Ni et al. [29] proposed an incremental feature selection method based on FRSs by analyzing the basic concept of FRSs on incremental datasets.

For methods based on fuzzy uncertainty measures, Hu et al. [30] first proposed fuzzy information entropy and information index for computing the discernibility of fuzzy relations, and constructed two greedy reduction algorithms for data dimensionality reduction. Wang et al. [31] proposed the concepts of joint fuzzy entropy, conditional fuzzy entropy, and fuzzy mutual information to calculate the uncertainty information of fuzzy binary relations and applied them to feature selection.

However, most of the above fuzzy rough feature selection methods are applicable to supervised learning scenarios. Existing methods to solve the unsupervised feature selection problem are still relatively few. Mac Parthaláin et al. [22] proposed an unsupervised feature selection algorithm that does not require threshold information and can manipulate real-valued data without discretization. Ganivada et al. [24] integrated granular computing and neural networks within a soft computing framework, and proposed a granular neural network for identifying salient features of data. Wang et al. [32] combined FRSs with granular loading variable accuracy and proposed a variable accuracy FRS model to compensate for the limitation of relative error limit on FRSs.

Nevertheless, most unsupervised feature selection methods are only suitable for single data types dealing with numerical or nominal attributes. Only a few attempts have been made for feature selection problems with mixed attributes. Solorio-Fernández et al. [33] were inspired by spectral feature selection and used both kernel and new spectral-based feature evaluation measures to quantify and rank the relevance of features. Chaudhuri et al. [34] utilized entropy and mutual information to produce maximally informative and non-redundant ranked features from a high-dimensional mixed dataset. Yuan et al. [23] defined the importance description of candidate attributes as evaluation criteria for feature selection, and used FRSs for unsupervised mixed data.

To address the limitations of unsupervised feature selection in mixed-attribute datasets, this paper constructs an unsupervised heterogeneous feature selection method based on fuzzy multi-neighborhood entropy. The method not only applies to heterogeneous data but also makes up for the existing unsupervised feature selection method that does not consider interactivity.

3. Preliminaries

This section will review the basic concepts of FRS theory and fuzzy uncertainty measures [35,36].

3.1. Fuzzy rough set

Let the quaternion $FIS = (U, A, V, f)$ be a fuzzy information system, where $U = \{x_1, x_2, \dots, x_n\}$ is a nonempty finite set of objects; $A = \{a_1, a_2, \dots, a_m\}$ is a nonempty finite set of features, describing the features of the objects in the theoretical domain information. And V is the union of feature value domains, i.e., $V = \bigcup_{a\in A} V_a$, where V_a is the value domain of attribute a. $f : U \times A \to V$ is the information function about the relation between the objects in U and their attribute values, satisfying when $\forall a \in A$ and $x \in U$, $f_a(x) \in V_a$.

$\widetilde{R}$ is a fuzzy relation on $U \times U$, which is defined as $\widetilde{R} : U \times U \to [0, 1]$. $\forall (x, y) \in U \times U$, $\widetilde{R}(x, y)$ shows the degree of similarity between x and y. $\widetilde{F}(U \times U)$ denotes the set of all fuzzy relations on $U \times U$.

Assume $\widetilde{R} \in \widetilde{F}(U \times U)$, $\forall x, y, z \in U$, if it satisfies the following conditions:

(1) reflexivity: $\widetilde{R}(x, x) = 1$;
(2) symmetry: $\widetilde{R}(x, y) = \widetilde{R}(y, x)$;
(3) transferability: $\sup_{y\in U} \min\{\widetilde{R}(x, y), \widetilde{R}(y, z)\} \le \widetilde{R}(x, z)$,

then $\widetilde{R}$ is a fuzzy equivalence relation on U. If $\widetilde{R}$ only satisfies reflexivity and symmetry, then $\widetilde{R}$ is said to be a fuzzy similarity relation on U. For $\widetilde{R}_1, \widetilde{R}_2 \in \widetilde{F}(U)$, we have the following operations:

(1) $\widetilde{R}_1(x, y) \le \widetilde{R}_2(x, y) \Rightarrow \widetilde{R}_1 \subseteq \widetilde{R}_2$;
(2) $(\widetilde{R}_1 \cap \widetilde{R}_2)(x, y) = \min\{\widetilde{R}_1(x, y), \widetilde{R}_2(x, y)\}$;
(3) $(\widetilde{R}_1 \cup \widetilde{R}_2)(x, y) = \max\{\widetilde{R}_1(x, y), \widetilde{R}_2(x, y)\}$.

FRS model was originally proposed by Dubois and Prade [12,37] and is defined as follows.

Definition 1. Let $\widetilde{R}$ be a fuzzy equivalence relation on U. $\forall \chi \in \widetilde{F}(U)$, the lower approximation $\underline{\widetilde{R}}\chi$ and the upper approximation $\overline{\widetilde{R}}\chi$ of χ are a set of fuzzy sets on U with membership functions respectively

$$\underline{\widetilde{R}}\chi(x) = \inf_{y\in U} \max\left\{1 - \widetilde{R}(x,y), \chi(y)\right\}; \tag{1}$$

$$\overline{\widetilde{R}}\chi(x) = \sup_{y\in U} \min\left\{\widetilde{R}(x,y), \chi(y)\right\}, \tag{2}$$

where $\overline{\widetilde{R}}\chi(x)$ denotes the possible degree of object x belonging to the fuzzy set χ, $\underline{\widetilde{R}}\chi(x)$ denotes the degree of certainty of object x belonging to the fuzzy set χ, and $(\underline{\widetilde{R}}\chi, \overline{\widetilde{R}}\chi)$ is called the FRS of χ.

3.2. Fuzzy uncertainty measures

In FRS, the uncertainty of fuzzy approximation spaces is measured by introducing the concept of entropy. The definition of entropy is an uncertainty measure in the whole system which varies monotonically with features. The fuzzy entropy is defined as follows [38].

Definition 2. Given an $FIS = (U, C, V, f)$, $\forall E \subseteq C$, the fuzzy entropy of E is calculated as

$$FE(E) = -\frac{1}{n}\sum_{i=1}^{n} \log_2 \frac{1}{|[x_i]_{\widetilde{R}_E}|}, \tag{3}$$

where $|[x_i]_{\widetilde{R}_E}| = \sum_{j=1}^{n} \widetilde{R}_E(x_i, x_j)$ is the cardinal number of the fuzzy granule derived from E.

From Definition 2, it can be seen that the fuzzy entropy decreases as the feature subset grows, and increases as the feature subset shrinks.

4. Fuzzy multi-neighborhood uncertainty measures

4.1. Fuzzy multi-neighborhood granule

FRSs are sensitive to noise [39]. As a significant component to measure the uncertainty in fuzzy approximation spaces, fuzzy granules are neighborhoodized by controlling the similarity between objects, thus avoiding disturbance from data noise.

Let $E \subseteq C$, $\forall x_i, x_j \in U$, the fuzzy granule of x_i is derived from the fuzzy relation $\widetilde{R}_E$, denoted as $[x_i]_{\widetilde{R}_E}$. $\widetilde{R}_E(x_i, x_j)$ reflects not only the similarity between object x_i and object x_j, but also the membership of object x_j to the fuzzy granule $[x_i]_{\widetilde{R}_E}$, i.e., $\widetilde{R}_E(x_i, x_j) = [x_i]_{\widetilde{R}_E}(x_j)$.

In the actual data-driven model construction, there are differences between features and different features have different distribution characteristics. Therefore, it is unreasonable to set the neighborhood value uniformly for a single fuzzy neighborhood granule [25]. Different types of data need different neighborhood values. Setting neighborhood values globally cannot be adjusted for different types of data and ignores the characteristics of the dataset.

Based on this issue, a multi-neighborhood radius set is employed. Compared with a uniform neighborhood value, this approach can better take into account the distribution characteristics of the features.

Definition 3. For $FIS = (U, C, V, f)$, where $U = \{x_1, x_2, \ldots, x_n\}$, and $C = \{c_1, c_2, \ldots, c_m\}$. $\forall x_i, x_j \in U, E \subseteq C$, and $E = \{c_{t_1}, \ldots, c_{t_l}\} = \{c_{t_k} | k \in \{1, \ldots, l\}\} (l \in \{1, \ldots, m\})$, the multi-neighborhood radius set of E is defined as

$$MnrS_E = \left\{\zeta_k | \zeta_k = \frac{std_{\{c_{t_k}\}}}{\lambda}, c_{t_k} \in E\right\}, \tag{4}$$

where std is the standard deviation used to measure the distribution characteristics of the feature c_{t_k}, and λ is the adjustment parameter for the neighborhood radius.

According to this, the fuzzy multi-neighborhood granule is defined below.

Definition 4. The fuzzy multi-neighborhood granule is derived from adaptively regulated fuzzy similarity relations, the definition of which is

$$FMnG_E = \left\{[x_i]^{\zeta_1}_{\widetilde{R}_{\{c_{t_1}\}}}, [x_i]^{\zeta_2}_{\widetilde{R}_{\{c_{t_2}\}}}, \ldots, [x_i]^{\zeta_l}_{\widetilde{R}_{\{c_{t_l}\}}}\right\}, \tag{5}$$

where,

$$[x_i]^{\zeta_k}_{\widetilde{R}_{\{c_{t_k}\}}}(x_j) = \begin{cases} \widetilde{R}^{\zeta_k}_{\{c_{t_k}\}}(x_i, x_j), & 1 - \widetilde{R}_{\{c_{t_k}\}}(x_i, x_j) \le \zeta_k; \\ 0, & 1 - \widetilde{R}_{\{c_{t_k}\}}(x_i, x_j) > \zeta_k. \end{cases} \tag{6}$$

The fuzzy similarity relation $\widetilde{R}_{\{c_{t_k}\}}(x_i, x_j)$ between x_i and x_j on the attribute c_{t_k} is computed below.

$$\widetilde{R}_{\{c_{t_k}\}}(x_i, x_j) = \begin{cases} 1 - |f_{c_{t_k}}(x_i) - f_{c_{t_k}}(x_j)|, & c_{t_k} \text{ is numerical;} \\ 1, & c_{t_k} \text{ is nominal and } f_{c_{t_k}}(x_i) = f_{c_{t_k}}(x_j); \\ 0, & c_{t_k} \text{ is nominal and } f_{c_{t_k}}(x_i) \ne f_{c_{t_k}}(x_j). \end{cases} \tag{7}$$

For convenience, $[x_i]^{\zeta_k}_{\widetilde{R}_{\{c_{t_k}\}}}(x_j)$ is shortened to $\zeta_{\{c_{t_k}\}}(x_i)$. $\widetilde{R}^{\zeta_k}_{\{c_{t_k}\}}(x_i, x_j)$ denotes a parametric fuzzy similarity relation between x_i and x_j on the attribute c_{t_k}, i.e., the fuzzy neighborhood similarity relation. $\forall B \subseteq C$, the fuzzy similarity relation $\widetilde{R}_B$ can be induced by B. And it can be expressed as the fuzzy relation matrix $M_{\widetilde{R}_B} = (r^B_{ij})_{n\times n}$, where $r^B_{ij} = \widetilde{R}_B(x_i, x_j)$. The values of objects over attribute c_{t_k} are denoted as $f_{c_{t_k}}(x_i)$ and $f_{c_{t_k}}(x_j)$.

The main advantage of multi-granularity over single-granularity is that feature extraction and indexes using multi-scale information can characterize the object more comprehensively. It ensures the comprehensiveness, flexibility, and robustness of feature multi-correlation indexes calculated using fuzzy multi-neighborhood uncertainty measures. The concept of the fuzzy multi-neighborhood granule was presented by Wan et al. [25].

Definition 5. The definition of fuzzy multi-neighborhood granule is shown as

$$[x_i]^{\zeta}_{\widetilde{R}_E} = \bigcap_{c_{t_k}\in E} [x_i]^{\zeta_k}_{\widetilde{R}_{\{c_{t_k}\}}}. \tag{8}$$

Following the strategy in [25], each feature distribution $std_{\{c_{t_k}\}}$, feature subset E, and parameter λ determine the size of the fuzzy multi-neighborhood granule.

4.2. Fuzzy multi-neighborhood entropy and its development measures

In this paper, the information available from the fuzzy multi-neighborhood granule is calculated using entropy. The fuzzy multi-neighborhood entropy is defined through generalization based on fuzzy entropy [38]. Given $E, H \subseteq C$, there are several definitions of fuzzy multi-neighborhood uncertainty measures.

Definition 6. The fuzzy multi-neighborhood entropy of the feature subset E is defined as

$$FME(E) = -\frac{1}{|U|}\sum_{i=1}^{|U|} \log_2 \frac{1}{|\zeta_E(x_i)|}. \tag{9}$$

From Definition 6, the fuzzy multi-neighborhood entropy varies monotonically with features, i.e., a decrease or increase in features in FIS leads to a change in fuzzy multi-neighborhood entropy and uncertainty.

The following is a few expansions to measure the uncertainty on the basis of fuzzy multi-neighborhood entropy.

Definition 7. The fuzzy multi-neighborhood joint entropy of the feature subset E and the feature subset H is defined as

$$FME(E,H) = -\frac{1}{|U|}\sum_{i=1}^{|U|}\log_2\frac{1}{|\zeta_E(x_i)\cap\zeta_H(x_i)|}. \tag{10}$$

After knowing E, the information provided by H is calculated using the fuzzy multi-neighborhood joint entropy. Thus, we quantify the uncertainty of multiple features for the reliability assessment of the fusion results.

Definition 8. Under the condition of a given feature subset H, the fuzzy multi-neighborhood conditional entropy of E is given by

$$FME(E|H) = -\frac{1}{|U|}\sum_{i=1}^{|U|}\log_2\frac{|\zeta_H(x_i)|}{|\zeta_E(x_i)\cap\zeta_H(x_i)|}. \tag{11}$$

Property 1. $FME(E|H) = FME(E,H) - FME(H)$.

The fuzzy multi-neighborhood conditional entropy reflects the dependencies between features.

Mutual information is used to measure the amount of information that a random variable contains about another random variable, indicating the degree of information sharing between two random variables. The introduction of mutual information in feature selection is a widespread application to measure the relevance and redundancy between features [40], which is defined as follows.

Definition 9. The fuzzy multi-neighborhood mutual information of E and H is given by

$$FMmI(E;H) = -\frac{1}{|U|}\sum_{i=1}^{|U|}\log_2\frac{|\zeta_E(x_i)\cap\zeta_H(x_i)|}{|\zeta_E(x_i)|\times|\zeta_H(x_i)|}. \tag{12}$$

Property 2. *For fuzzy multi-neighborhood mutual information, the following properties hold.*

(1) $FMmI(E;H) \geq 0$;
(2) $FMmI(E;H) = FME(E) - FME(E|H)$;
(3) $FMmI(E;H) = FME(H) - FME(H|E)$;
(4) $FMmI(E;H) = FME(E) + FME(H) - FME(E,H)$.

The change in the uncertainty of the feature subset E after knowing H is represented by the fuzzy multi-neighborhood mutual information, i.e., the degree of relevance between the amount of information contained in both.

The relevance of three feature sets is usually measured using conditional mutual information and joint mutual information. Some related definitions are proposed below.

Definition 10. For $E,H,Q\subseteq C$, under the condition that Q is known, the fuzzy multi-neighborhood conditional mutual information of E and H is defined as

$$FMmI(E;H|Q) = -\frac{1}{|U|}\sum_{i=1}^{|U|}\log_2\frac{|\zeta_E(x_i)\cap\zeta_H(x_i)\cap\zeta_Q(x_i)|\times|\zeta_Q(x_i)|}{|\zeta_E(x_i)\cap\zeta_Q(x_i)|\times|\zeta_H(x_i)\cap\zeta_Q(x_i)|}. \tag{13}$$

The above definition is used in Section 5.1 as a measure of the fuzzy interactivity among features.

Definition 11. For $E,H,Q\subseteq C$, the fuzzy multi-neighborhood joint mutual information of feature E, feature H, and feature Q is defined as

$$FMmI(E,H;Q) = -\frac{1}{|U|}\sum_{i=1}^{|U|}\log_2\frac{|\zeta_E(x_i)\cap\zeta_H(x_i)\cap\zeta_Q(x_i)|}{|\zeta_E(x_i)\cap\zeta_H(x_i)|\times|\zeta_Q(x_i)|}. \tag{14}$$

The fuzzy multi-neighborhood joint mutual information measures how much information the joint features provide.

Property 3. *The following relationships exist between these indexes.*

(1) $FMmI(E,H;Q) = FMmI(E;Q) + FMmI(H;Q|E)$;
(2) $FMmI(E,H;Q) = FMmI(H;Q) + FMmI(E;Q|H)$.

From Definitions 6–11, fuzzy multi-neighborhood uncertainty measures in this paper consider the different distribution characteristics of different features by setting the adaptive neighborhood radius. With the use of the fuzzy multi-neighborhood granule and extended concepts to measure the information uncertainty, the distinguishing ability of features can be better described. Moreover, in methods that use uncertainty measures to evaluate the importance of features, interactivity is also a very important measure in addition to relevance and redundancy between features. However, interactivity is rarely considered in the feature evaluation function. Ignoring the interactivity among features not only fails to comprehensively assess the importance of features but also may select redundant features.

To handle the mentioned problems, this paper proposes a fuzzy multi-neighborhood entropy-based Interactive unsupervised Feature Selection (MNIFS) algorithm for heterogeneous data.

5. Feature multi-correlation analysis based on MNIFS

Under the theoretical framework of the fuzzy multi-neighborhood entropy and its extended model, a comprehensive definition and analysis of feature multi-correlation is presented in Section 5.1. In Section 5.2, the MRmRMI objective evaluation function is constructed, which considers the relevance, redundancy, and interactivity among features. In Section 5.3, an unsupervised feature selection algorithm called MNIFS is devised, and its corresponding flow and complexity are analyzed.

5.1. Feature multi-correlation in the feature selection process

This subsection divides the multi-correlation calculation of features into relevance, redundancy, and interactivity. Their definitions are given below by fuzzy multi-neighborhood uncertainty measures.

(1) Relevance: The degree of relationship or dependence that exists between features is described by relevance. It is calculated using fuzzy multi-neighborhood mutual information.

Definition 12. The fuzzy relevance between the current candidate feature c_t and the remaining candidate feature c_l is defined as

$$FReL_{FMmI}(c_t) = \frac{1}{m}\sum_{l=1}^{m} FMmI(c_t;c_l). \tag{15}$$

Property 4. *Obviously, the following equation holds.*

$$FReL_{FMmI}(c_t) = \frac{1}{m}(FME(c_t) + \sum_{l=1,l\neq t}^{m} FMmI(c_t;c_l)). \tag{16}$$

In Property 4, $FME(c_t)$ denotes the fuzzy multi-neighborhood entropy of feature c_t. $\sum_{l=1,l\neq t}^{m} FMmI(c_t;c_l)$ denotes the quantity of information that both c_t and the remaining features can provide. A larger value of $\sum_{l=1,l\neq t}^{m} FMmI(c_t;c_l)$ will decrease the importance of the remaining features. Therefore, selecting the feature that has the greatest fuzzy relevance minimizes information loss.

From this, the approach proposed by this paper is to first set an empty set S for feature selection. After each iteration, a new feature is selected to join S. In the first selection, the feature should have the maximum fuzzy relevance. It meets the requirements below.

$$c_{k_1} = arg\max_{c\in C}\left\{FReL_{FMmI}(c)\right\}. \tag{17}$$

Due to the greatest fuzzy relevance of the feature c_{k_1}, it maximally reduces the uncertain information of the remaining features. That is, if only one feature can be selected, c_{k_1} is capable of providing the most information.

After selecting the feature with maximum fuzzy relevance, it is necessary to judge whether the remaining features satisfy the condition of being appended to the selected feature subset S. In each iteration, the importance of the current candidate features $c_t \in C - S$ is calculated. Therefore, the relevance of c_t, the redundancy of c_t with the selected feature subset S, and the interactivity of c_t and the remaining unselected features $c_l \in C - S - \{c_t\}$ about S need to be considered.

(2) Redundancy: When selecting new features c_t, adding new features might introduce duplicate or redundant information to the already selected feature subset S.

For the current candidate feature $c_t \in C - S$, $\forall c_{k_s} \in S$, if the candidate feature c_t is added to S, the information amount supplied by the feature c_{k_s} will change as a result of the inclusion of c_t. Thus, its fuzzy relevance will change as well. Based on this, the fuzzy conditional relevance is defined as follows.

Definition 13. The fuzzy conditional relevance of the current candidate feature c_t in relation to the feature c_{k_s} is given as

$$FReL_{FMmI}(c_{k_s}|c_t) = \frac{FME(c_{k_s}, c_t)}{FME(c_{k_s})} FReL_{FMmI}(c_{k_s}). \tag{18}$$

Further, the discrepancy between them can be described by fuzzy redundancy.

Definition 14. The fuzzy redundancy between the current candidate feature c_t and the feature c_{k_s} is defined as

$$FReD_{FMmI}(c_t, c_{k_s}) = FReL_{FMmI}(c_{k_s}) - FReL_{FMmI}(c_{k_s}|c_t). \tag{19}$$

(3) Interactivity: When only a single feature is selected, it provides less information and has weaker relevance; when combined with other features, it has increased relevance.

There may be interactivity between features. That is, two or more features together affect the target variable, while one of them alone may not have a direct effect on the target. If only individual features are considered and their interactivity is ignored, critical information may be lost. By selecting features with significant interactivity, patterns, and relationships in the data can be captured more comprehensively. Furthermore, by introducing interactivity between features, the model can reduce the selection of features that are not actually optimal or redundant. Reducing redundant features can improve the efficiency of subsequent detection models. In summary, considering the interactivity between features helps to identify key information and eliminate redundancy.

Interactivity is concerned with the impact of combining the current candidate feature with the remaining unselected features. For the current candidate feature $c_t \in C - S$ and the remaining candidate feature $c_l \in C - S - \{c_t\}$, fuzzy multi-neighborhood joint mutual information and fuzzy multi-neighborhood mutual information are used to quantify their interactivity.

Definition 15. For the selected feature subset S, the fuzzy interactivity of the current candidate feature c_t and the remaining unselected feature c_l is calculated as

$$FInT_{FMmI}(c_t, c_l) = |FMmI(c_t, c_l; S) - FMmI(c_t; S)|. \tag{20}$$

The index explains that individual features possibly contribute to the learning algorithm with less discriminative information from the viewpoint of fuzzy multi-neighborhood uncertainty measures. But when the features occur as pairs, they may contribute more discriminative information to it. Thus, interactivity reflects the effect of joint feature interactions on the learning algorithm.

According to Property 3, it can be known that $FMmI(c_t, c_l; S) = FME(c_t; S) + FMmI(c_l; S|c_t)$, so Definition 15 can be simplified as follows.

$$FInT_{FMmI}(c_t, c_l) = |FMmI(c_l; S|c_t)|. \tag{21}$$

5.2. Feature assessment considering multiple correlations

When selecting features, in addition to ensuring that features have maximum relevance, redundancy is expected to be minimized and interactivity is desired to be maximized between features. In this regard, redundancy refers to features providing identical or duplicate information, whereas interactivity reflects features that deliver new discriminative information.

Accordingly, a novel assessment function J_{MRmRMI} is built to identify the significance of each feature. The features with the maximum information content can be selected according to the following equation.

$$J_{MRmRMI} = arg \max_{c_t \in C-S} \left\{ FReL_{FMmI}(c_t) - \frac{1}{|S|} \sum_{s=1}^{|S|} FReD_{FMmI}(c_t, c_{k_s}) + \frac{1}{(|C-S|-1)} \sum_{c_l \in C-S-\{c_t\}} FInT_{FMmI}(c_t, c_l) \right\}. \tag{22}$$

As shown in Eq. (22), the significance of features during the feature selection procedure is taken into account by feature multi-correlation analysis. This consists of the relevance of the current candidate feature, the redundancy between the candidate feature and selected features, and the interactivity between the candidate feature and remaining unselected features.

Following the J_{MRmRMI} evaluation index, when selecting the rth feature, if the feature c_{k_r} satisfies J_{MRmRMI}, c_{k_r} can minimize the uncertainty of other features. This not only contains little fuzzy redundant information but also gives more discriminative information in the interaction with other features.

5.3. Feature selection algorithm

An algorithm called MNIFS is designed according to the previous discussion, and its time complexity is analyzed.

In Algorithm 1, the computation is divided into three main stages as shown below.

Stage 1: First, normalize the numerical data by the $\min - \max$ normalization method. Then the ordered feature sequence S is set as $\emptyset$, and the feature set to be selected S_u is set as C. Calculate the fuzzy multi-neighborhood granule and entropy of each attribute in Steps (3)–(6). Calculate the fuzzy multi-neighborhood joint entropy and mutual information between features in Steps (7)–(12). Calculate the fuzzy relevance of each feature in Steps (13)–(17), and select the feature with the largest relevance as c_{k_1} to be added into the ordered feature sequence, while removing the feature c_{k_1} in S_u.

Stage 2: In Steps (19)–(29), select a feature in S_u as the current candidate feature in turn. Compute the fuzzy redundancy between the current candidate feature and each selected feature in the ordered feature sequence S in Steps (20)–(22). Compute the fuzzy interactivity between the current candidate feature and the remaining features in S_u in Steps (23)–(25), respectively. If the candidate feature satisfies the maximum correlation-minimum redundancy-maximum interactivity metric, it is added to the ordered feature sequence S and removed from S_u.

Stage 3: Repeat Stage 2 until S_u is empty. Finally, output the ordered feature sequence S.

In Algorithm 1, suppose the number of numerical attributes is j ($0 \le j \le m$). The $\min - \max$ normalization in the preprocessing part will

Algorithm 1: MNIFS algorithm

Input: $FIS = (U, C, V, f)$, where $U = \{x_i, i = 1, 2, \ldots, n\}$ and the threshold λ is in the range [0.1,2.0] with a step size of 0.1

Output: An ordered sequence of features

1 Normalize the numerical data by the min – max normalization;
2 $S \leftarrow \emptyset, S_u \leftarrow C$;
3 **for** $t \leftarrow 1$ *to* m **do**
4 Compute the fuzzy multi-neighborhood granule $[x_i]^{\zeta_t}_{\widetilde{R}_{c_t}}$;
5 Compute the fuzzy multi-neighborhood entropy $FME(c_t)$;
6 **end**
7 **for** $t \leftarrow 1$ *to* m **do**
8 **for** $l \leftarrow 1$ *to* m **do**
9 Compute the fuzzy multi-neighborhood joint entropy $FME(c_t, c_l)$;
10 Compute the fuzzy multi-neighborhood mutual information $FMmI(c_t; c_l)$;
11 **end**
12 **end**
13 **for** $t \leftarrow 1$ *to* m **do**
14 Compute the fuzzy relevance $FReL_{FMmI}(c_t)$;
15 **end**
16 Choose c_{k_1} which satisfies $FReL_{FMmI}(c_{k_1})$ takes the largest value;
17 $S \leftarrow S \cup \{c_{k_1}\}, S_u \leftarrow S_u - \{c_{k_1}\}$;
18 **while** $|S_u| \neq 0$ **do**
19 **for** $k \leftarrow 1$ *to* $|S_u|$ **do**
20 **for** $s \leftarrow 1$ *to* $|S|$ **do**
21 Compute the fuzzy redundancy $FReD_{FMmI}(c_k, c_{k_s})$;
22 **end**
23 **for** $l \leftarrow 1$ *to* $|S_u|$ **do**
24 Compute the fuzzy interactivity $FInT_{FMmI}(c_k, c_l)$;
25 **end**
26 **end**
27 Choose the feature c_{k_r}, satisfying
28 $FReL_{FMmI}(c_{k_r}) - \frac{1}{|S|}\sum_{s=1}^{|S|} FReD_{FMmI}(c_{k_r}, c_{k_s}) + \frac{1}{|S_u|-1}\sum_{c_k \in S_u - \{c_{k_r}\}} FInT_{FMmI}(c_{k_r}, c_k)$ has the maximum value;
29 $S \leftarrow S \cup \{c_{k_r}\}, S_u \leftarrow S_u - \{c_{k_r}\}$;
30 **end**
31 **return** S.

cost $n \times j$ loops. In Steps (3)–(6) has m loops, Step (3) has $n \times n$ loops, Step (5) has n loops, Steps (7)–(12) has $m \times m$ loops, Step (9) has n loops, Steps (13)–(15) has m loops, the loop counts of Steps (19)–(26) are $|S_u|$, the loop counts of Steps (20)–(22) are $|S|$, and the loop counts of Steps (23)–(25) are $|S_u|$. Thereby, the overall number of cycles to Algorithm 1 is $n \times j + m \times (n \times n + n) + m \times m \times n + m + |S_u| \times (|S| + |S_u|)$. As a result, for the worst situation, the time complexity of the Algorithm 1 is $O(mn(m + n))$.

5.4. An example

In this section, a fuzzy information system example that explains the above study is shown.

Example 1. Let $FIS = (U, C, V, f)$ as shown in the left side of Table 1, mainly involving heterogeneous feature data, where $U = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ and $C = \{c_1, c_2, c_3, c_4\}$. Among these attributes, c_1 and c_2 are numerical attributes, and c_3 and c_4 are nominal attributes.

Table 1
Initial and standardized fuzzy information system.

U	c_1	c_2	c_3	c_4	c_1	c_2	c_3	c_4
x_1	0.8	3	B	b	0.8571	0.2500	B	b
x_2	0.6	9	D	a	0.5714	1	D	a
x_3	0.5	5	C	c	0.4286	0.5000	C	c
x_4	0.9	1	A	c	1	0	A	c
x_5	0.4	7	A	b	0.2857	0.7500	A	b
x_6	0.2	4	B	a	0	0.3750	B	a

Preprocessing is first performed to normalize the numerical data by the min – max normalization method. The resulting data after processing are displayed on the right side of Table 1.

The standard deviations of the numerical attributes c_1 and c_2 respectively are $std(c_1) \approx 0.3367$, $std(c_2) \approx 0.3261$. Let $\lambda = 1$, from Definition 3, the fuzzy radius of the numerical attributes can be calculated as $\zeta_{c_1} \approx 0.3367$, $\zeta_{c_2} \approx 0.3261$ respectively.

The fuzzy relation matrix for every single attribute in the attribute set C is calculated as follows.

$$M_{\widetilde{R}_{c_1}} = \begin{bmatrix} 1 & 0.7143 & 0 & 0.8571 & 0 & 0 \\ 0.7143 & 1 & 0.8571 & 0 & 0.7143 & 0 \\ 0 & 0.8571 & 1 & 0 & 0.8571 & 0 \\ 0.8571 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0.7143 & 0.8571 & 0 & 1 & 0.7143 \\ 0 & 0 & 0 & 0 & 0.7143 & 1 \end{bmatrix}$$

$$M_{\widetilde{R}_{c_2}} = \begin{bmatrix} 1 & 0 & 0.7500 & 0.7500 & 0 & 0.8750 \\ 0 & 1 & 0 & 0 & 0.7500 & 0 \\ 0.7500 & 0 & 1 & 0 & 0.7500 & 0.8750 \\ 0.7500 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0.7500 & 0.7500 & 0 & 1 & 0 \\ 0.8750 & 0 & 0.8750 & 0 & 0 & 1 \end{bmatrix}$$

$$M_{\widetilde{R}_{c_3}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$M_{\widetilde{R}_{c_4}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

According to Definition 12, the fuzzy relevance of each attribute is calculated as follows.

$FReL_{FMmI}(c_1) \approx 1.8762$; $FReL_{FMmI}(c_2) \approx 1.8006$;
$FReL_{FMmI}(c_3) \approx 1.4999$; $FReL_{FMmI}(c_4) \approx 1.8255$.

Then, select the feature with maximum fuzzy relevance and add it to the selected feature subset S. Thus, add feature c_1 to S. In this way, $S_1 = \langle c_1 \rangle$ is obtained.

Based on Definition 13, the fuzzy conditional relevance between the remaining unselected features with respect to c_1 are calculated as follows, respectively.

$FReL_{FMmI}(c_1|c_2) \approx 1.0681$; $FReL_{FMmI}(c_1|c_3) = 0$;
$FReL_{FMmI}(c_1|c_4) = 0$.

From Definition 14, we can calculate the fuzzy redundancy between the remaining unselected features and c_1 separately as

$FReD_{FMmI}(c_2, c_1) = 0.8081$; $FReD_{FMmI}(c_3, c_1) = 1.8762$;
$FReD_{FMmI}(c_4, c_1) = 1.8762$.

According to Definition 15, the interactivity between the current candidate feature c_t and the remaining unselected feature c_l is calculated as

$FInT_{FMmI}(c_2, c_3) \approx 0.2652$; $FInT_{FMmI}(c_2, c_4) \approx 0.5675$;

$FInT_{FMmI}(c_3, c_2) \approx 0.3644$; $FInT_{FMmI}(c_3, c_4) \approx 0.6667$;

$FInT_{FMmI}(c_4, c_2) \approx 1.0000$; $FInT_{FMmI}(c_4, c_3) \approx 1.0000$.

From this, it can be calculated that

$$FReL_{FMmI}(c_2) - FReD_{FMmI}(c_2, c_1) + \frac{1}{2}(FInT_{FMmI}(c_2, c_3) + FInT_{FMmI}(c_2, c_4)) = 1.4089;$$

$$FReL_{FMmI}(c_3) - FReD_{FMmI}(c_3, c_1) + \frac{1}{2}(FInT_{FMmI}(c_3, c_2) + FInT_{FMmI}(c_3, c_4)) = 0.1392;$$

$$FReL_{FMmI}(c_4) - FReD_{FMmI}(c_4, c_1) + \frac{1}{2}(FInT_{FMmI}(c_4, c_2) + FInT_{FMmI}(c_4, c_3)) = 0.9493.$$

Then, select the feature c_2 and add it to S_1 to obtain $S_2 = \langle c_1, c_2 \rangle$ and $S_u = \{c_3, c_4\}$.

The same strategy is used to select the remaining features in turn. Eventually, a sequence of ordered features $S = \langle c_1, c_2, c_4, c_3 \rangle$ is obtained.

6. Experiments

This section applies MNIFS to the outlier detection task. The experimental results are analyzed to obtain the performance of MNIFS. For this purpose, 24 unbalanced datasets are chosen from the webpage.[1] The relevant information about the datasets can be found in Table 2.

From Table 2, 6 datasets contain only nominal attributes, 9 datasets contain only numerical attributes, and the other 9 datasets contain both nominal and numerical attributes, i.e., mixed attributes. To ensure that the experiment can be repeated, the datasets used are available through the webpage.

6.1. Experiment preparation

We will discuss the comparison algorithms, the experimental settings, and evaluation indexes used in this study in the experimental preparation.

6.1.1. Compare algorithms

In the outlier detection experiments, k-Nearest Neighbor Outlier Detection (kNNOD) [46], Histogram-based (HBOS) [47], and Empirical-Cumulative-distribution-based (ECOD) [48] outlier detection algorithms are employed for comparing the performance between algorithms. In these experiments, specific information on all comparison algorithms is given in Table 3.

[1] https://github.com/BElloney/Outlier-detection.

6.1.2. Experimental settings

The kNNOD, HBOS, and ECOD algorithms are provided by Python's PyOD library [49], and the experiments configure parameters as default values.

Both the comparison algorithms and MNIFS proposed in this paper rank the features of each dataset in descending order according to their significance. The output is a feature sequence of m features. We select the top $k(k = 1, 2, \ldots, m-1)$ features in each outlier detection to assess the efficiency of the algorithms in order to make a more reasonable comparison. Eventually, there are $(m-1)$ results available, from which the optimal result is selected for comparison. For MNIFS, set the step size to 0.1 to calculate the optimal feature subset of the parameter λ in the range of $[0.1, 2]$.

6.1.3. Evaluation indexes

The effectiveness of outlier detection is evaluated using the Receiver Operating Characteristic (ROC) curve and the Area Under Curve (AUC) [50].

Outlier detection methods generally end up with an outlier score for each sample in U. The larger its value, the more likely the object is to be an outlier. Therefore, scholars generally sort all samples in descending order according to the outlier scores obtained by each outlier detection algorithm. Then take a positive integer q that is not larger than the number of samples in U, and the objects in the top q of the ordering will be determined to be outliers.

The ROC curves are plotted with FPR (False Positive Rate) as the horizontal coordinate and TPR (True Positive Rate) as the vertical coordinate, respectively. FPR denotes the percentage of samples that are predicted as outliers but are actually normal as a percentage of all true normal points. And TPR denotes the percentage of all samples correctly recognized as outlier points to all true outliers.

AUC is the area under the ROC curve. The larger the AUC, the greater the outcome for outlier detection. Its formula is as follows.

$$\mathrm{AUC} = \underset{o_i \in OS_{true}, o_j \in U - OS_{true}}{\mathrm{Mean}} \begin{cases} 1, & score(o_i) > score(o_j) \\ 0.5, & score(o_i) = score(o_j) \\ 0, & score(o_i) < score(o_j) \end{cases}, \tag{23}$$

where OS_{true} denotes the set of samples that are truly outliers, and $score(o_i)$ denotes the outlier score of a sample o_i.

6.2. Analysis of mixed-attribute datasets

Table 4 gives the amount of best feature subsets and the corresponding parameters λ obtained using MNIFS over kNNOD, HBOS, and ECOD on the mixed-attribute datasets. Table 5 gives a comparison of the AUCs obtained on kNNOD, HBOS, and ECOD after feature selection using different comparison algorithms for the mixed-attribute datasets. Fig. 2 shows the average ROC curves of the mixed-attribute datasets on the three algorithms.

The experimental results from the mixed-attribute datasets show that MNIFS achieves a better performance on the mixed-attribute datasets. The main aspects of the analysis are as follows.

(1) From Table 4, we know that MNIFS can effectively remove the candidate features. Among them, HBOS has the lowest number of candidate features, with an average number of 15.7. On kNNOD and ECOD, their respective mean numbers of selected features are 17.2 and 18.7. This illustrates that it is possible to obtain better anomaly detection outcomes using fewer features on each dataset.

(2) According to the analysis of AUC results in Table 5, when using HBOS, MNIFS achieves the best AUC on 7 datasets, while LS, USFSM, SPEC, UFSFS, FR-FS, RSR, FMIUFS, and EUIAR only obtain the best AUC results on 0, 1, 0, 0, 0, 0, 1, 0 datasets, respectively. When using kNNOD, MNIFS achieved the best AUC on 6 datasets. The AUC distribution of each algorithm is more

Table 2
Description of the relevant datasets.

No	Datasets	Abbreviation	Number of conditional features		Number of objects	Number of outliers
			Numerical	Nominal		
1	Abalone_variant1	Abalone	7	1	4177	79
2	Annealing_variant1	Annealing	10	28	798	42
3	Autos_variant1	Autos	15	10	205	25
4	Bands_band_6_variant1	Bands1	24	15	318	6
5	Bands_band_34_variant1	Bands2	24	15	346	34
6	Bands_band_42_variant1	Bands3	24	15	354	42
7	German_1_14_variant1	German	7	13	714	14
8	Heart270_2_16_variant1	Heart	6	7	166	16
9	Horse_1_12_variant1	Horse	7	20	256	12
10	Breast_cancer_variant1	Breast	0	9	286	85
11	Chess_nowin_16_variant1	Chess1	0	36	1685	16
12	Chess_nowin_87_variant1	Chess2	0	36	1756	87
13	Chess_nowin_145_variant1	Chess3	0	36	1814	145
14	Chess_nowin_185_variant1	Chess4	0	36	1854	185
15	Chess_nowin_227_variant1	Chess5	0	36	1896	227
16	Cardio	Cardio	21	0	1831	176
17	Ecoli	Ecoli	7	0	336	9
18	Glass	Glass	9	0	214	9
19	Letter	Letter	32	0	1600	100
20	Pageblocks_1_258_variant1	Pageblocks	10	0	5171	258
21	Pima_TRUE_55_variant1	Pima	9	0	555	55
22	Spambase_spam_56_variant1	Spambase	57	0	2788	56
23	Thyroid	Thyroid	6	0	3772	93
24	Wine	Wine	13	0	129	10

Table 3
Description of the comparison algorithms.

Comparison algorithm	Main idea	Application scope
LS [5]	Using Laplace scores to evaluate the significance of features	Numerical
USFSM [33]	Kernel and spectral-based feature evaluation measures	Mixed
SPEC [41]	A unified framework based on spectral graph theory	Numerical
UFSFS [42]	Introduction of maximum information compression index	Numerical
FR-FS [34]	Propose two stages of feature ranking and feature selection	Mixed
RSR [43]	Propose a regularized self-representation model	Numerical
FMIUFS [44]	Calculating relevance and redundancy to rank features	Mixed
EUIAR [45]	Based on fuzzy complementary entropy	Mixed

Table 4
Number of features and λ obtained by MNIFS for mixed-attribute datasets.

Datasets	Original feature	kNNOD	λ	HBOS	λ	ECOD	λ
Abalone	8	2	0.1	2	0.1	2	0.1
Annealing	38	32	0.6	33	0.6	32	0.6
Autos	25	6	0.5	15	0.6	4	0.6
Bands1	39	33	1.2	35	0.1	32	1.4
Bands2	39	29	1.1	12	1.4	28	1.1
Bands3	39	28	1.1	15	1.4	28	1
German	20	15	1.2	18	0.1	15	1.2
Heart	13	7	0.7	8	0.8	12	0.1
Horse	27	3	0.5	3	0.5	15	1.2
Average	27.6	17.2	0.8	15.7	0.6	18.7	0.8

scattered on ECOD, except for MNIFS and UFSFS which achieve the best AUC on 3 datasets, the rest of the algorithms have no more than two.

(3) As shown in Table 5, the average AUC of MNIFS is the highest value on both kNNOD and HBOS, and the highest AUC is obtained by EUIAR on ECOD. On all three algorithms, the AUC of MNIFS is greater as compared to the original data. This indicates that MNIFS can effectively remove uncorrelated or unnecessary features to increase the AUC of outlier detection algorithms.

From Fig. 2, we can get the effectiveness of MNIFS in a more descriptive way. For example, for the dataset Abalone, we can see from Fig. 2(a) that MNIFS is closest to the upper left corner of the first quadrant and has the largest area under the ROC curve compared to the rest of the algorithms. Thus we know that MNIFS has the highest AUC value on the dataset Abalone and achieves the best results. Similarly, from Fig. 2(e), it can be obtained that the ROC curve of MNIFS is also closest to the upper left corner of the first quadrant on the dataset Bands2 and gets the best result.

In summary, MNIFS can use fewer feature subsets to get better experimental results. Therefore, it is applicable to the mixed-attribute dataset.

6.3. Analysis of numerical-attribute datasets

The number of features and corresponding parameters λ obtained after feature selection using MNIFS on the numerical-attribute datasets are presented as Table 6. The experimental results obtained using different comparison experiments on kNNOD, HBOS, and ECOD are given in Table 7. The average ROC curves about the numerical-attribute datasets are shown in Fig. 3.

Combining the experimental results in Tables 6 and 7 shows that MNIFS achieves better performance on numerical-attribute datasets under the majority of situations. The following are specific discussions of the experimental results.

(1) From Table 6, MNIFS can remove many uncorrelated or redundant features in comparison to the original data. For different outlier detection algorithms, the same feature selection algorithm yields different feature subsets. On kNNOD, HBOS, and ECOD, the average number of features obtained by MNIFS is only 6.3, 6.4, and 4.2, respectively, which can get the optimal performance using fewer features.

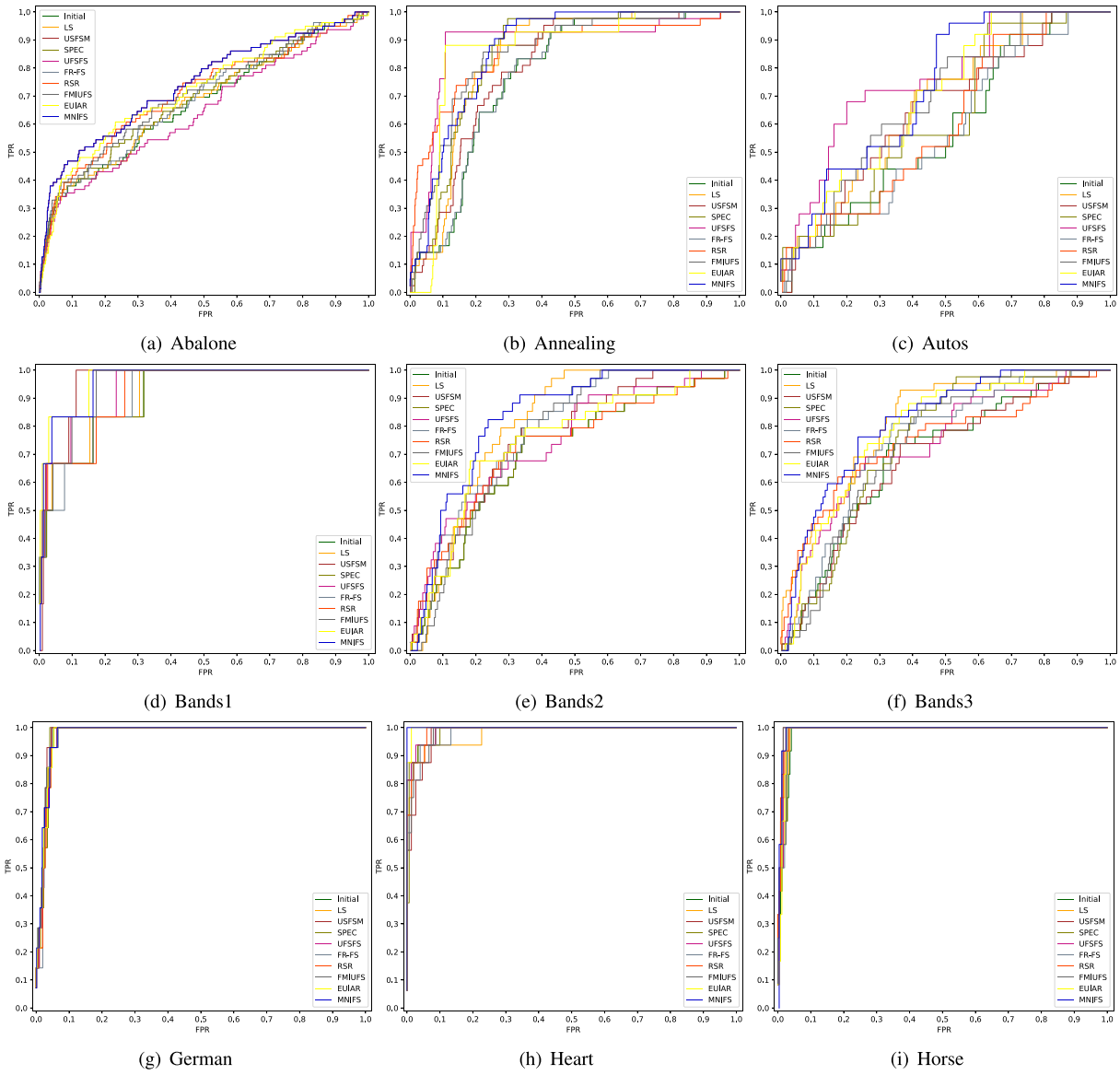


Fig. 2. ROC curves for mixed-attribute datasets.

(2) From Table 7, we can get that when using kNNOD as the anomaly detection algorithm, MNIFS can achieve the highest AUC values on 6 datasets. Meanwhile, when using HBOS, MNIFS has the highest AUC values on 4 datasets. And on ECOD, MNIFS has 6 datasets to obtain the best results, while LS, USFSM, SPEC, UFSFS, FR-FS, RSR, FMIUFS, and EUIAR obtain the best results only on 0, 0, 2, 0, 0, 0, 0, 2 datasets, respectively, where EUIAR and MNIFS obtained equal AUC values for the dataset Pima.

(3) According to Table 7, it can be seen that the average AUC of MNIFS is the highest value on all three algorithms. The AUC of MNIFS is higher than the original AUC compared with the original data. It indicates that MNIFS can effectively remove uncorrelated or redundant features and help outlier detection methods get better results using a smaller subset of features.

From Fig. 3, it can be concluded that on most datasets, such as Figs. 3(a) and 3(f), the ROC curves for MNIFS are the curves closest to the upper left corner of the first quadrant with the largest area under the curves. This shows that MNIFS performs superior to other compared algorithms. For the dataset Thyroid, the curves overlap well because each algorithm shows good performance.

Therefore, MNIFS is equally applicable to the numerical-attribute dataset in addition to the mixed-attribute dataset.

6.4. Analysis of nominal-attribute dataset

Since the different parameters λ in MNIFS do not have any effect on the feature subsets of nominal attributes, Table 8 only gives the number of feature subsets derived using MNIFS on the nominal-attribute datasets. Table 9 and Fig. 4 present the outlier detection result comparison experiments and the average ROC curves obtained by using the three algorithms for outlier detection on different datasets, respectively.

According to Tables 8 and 9, the following analysis can be performed.

(1) From Table 8, it can be concluded that MNIFS removes some of the candidate features from all datasets on all three outlier detection algorithms, where the subset of features required to use HBOS and ECOD is slightly less than that of kNNOD.

(2) The comparison of the AUC results given in Table 9 shows that MNIFS can achieve the highest AUC values on a total of 10

Table 5
Comparison of AUC of outlier detection for mixed-attributes (%).

Algorithms	Datasets	Initial	LS	USFSM	SPEC	UFSFS	FR-FS	RSR	FMIUFS	EUIAR	MNIFS
kNNOD	Abalone	70.99	70.54	74.44	70.02	70.79	73.29	72.50	72.16	73.29	**74.44**
	Annealing	74.67	77.23	87.08	81.43	78.70	77.89	83.84	86.68	84.89	**88.47**
	Autos	48.24	65.10	63.50	60.43	61.11	58.63	56.56	57.96	67.12	**71.91**
	Bands1	92.68	93.59	**97.06**	92.68	94.55	93.16	94.28	94.39	95.89	94.60
	Bands2	72.08	72.31	67.39	72.08	77.63	76.78	73.71	76.82	79.59	**79.98**
	Bands3	72.65	74.69	71.12	72.65	77.87	74.89	76.28	75.91	79.99	**80.49**
	German	95.68	94.76	**96.42**	95.88	95.67	95.12	94.76	96.33	96.30	96.24
	Heart	94.52	93.08	97.90	95.81	94.79	96.31	94.83	98.02	97.29	**98.04**
	Horse	88.47	88.64	**98.33**	88.42	90.23	90.61	97.35	88.71	97.30	98.29
	Average	78.89	81.10	83.69	81.04	82.37	81.85	82.68	83.00	85.74	**86.94**
HBOS	Abalone	66.42	68.42	67.94	66.86	64.30	67.32	69.57	**69.68**	67.36	67.94
	Annealing	74.52	82.42	75.01	82.50	85.27	74.55	84.05	83.57	80.48	**86.16**
	Autos	55.33	66.16	68.17	60.34	63.27	55.24	58.11	65.79	67.91	**73.44**
	Bands1	84.03	84.24	91.61	84.03	91.13	84.03	86.59	88.94	94.12	**94.28**
	Bands2	62.01	68.10	61.71	62.01	64.14	68.17	65.08	66.86	66.57	**74.69**
	Bands3	62.00	71.89	60.86	64.17	63.70	66.35	67.98	67.43	68.25	**75.51**
	German	96.83	96.63	**97.10**	96.97	96.90	96.40	96.63	97.01	97.08	96.88
	Heart	97.21	95.96	96.25	96.96	97.46	96.42	96.83	96.92	99.21	**99.58**
	Horse	96.65	96.89	98.26	97.37	97.47	97.30	97.75	98.43	98.39	**98.72**
	Average	77.22	81.19	79.66	79.02	80.40	78.42	80.29	81.63	82.15	**85.24**
ECOD	Abalone	65.31	63.77	74.87	66.35	61.93	66.79	68.06	68.35	73.16	**74.87**
	Annealing	78.74	83.86	80.33	85.88	**89.69**	78.81	85.23	82.63	86.87	86.29
	Autos	58.29	67.92	59.00	66.14	**74.39**	71.26	71.69	71.26	73.47	68.82
	Bands1	93.59	93.86	95.03	93.59	93.64	94.39	96.10	96.42	**97.06**	96.42
	Bands2	78.18	79.23	76.75	78.18	81.26	**84.13**	81.24	78.72	81.84	80.49
	Bands3	75.05	78.34	74.64	76.28	**80.60**	79.36	78.46	75.90	78.13	78.24
	German	96.55	96.33	96.64	96.81	96.81	95.84	96.33	96.93	97.01	**97.07**
	Heart	99.63	97.38	98.38	98.75	99.58	98.54	98.46	98.92	**100.00**	99.92
	Horse	98.05	98.36	98.43	97.64	98.60	98.33	98.36	98.36	98.36	**98.60**
	Average	82.60	84.34	83.79	84.40	86.28	85.27	85.99	85.28	**87.32**	86.75

Table 6
Number of features and λ obtained by MNIFS for numerical-attribute datasets.

Datasets	Original feature	kNNOD	λ	HBOS	λ	ECOD	λ
Cardio	21	5	0.7	5	0.7	10	0.8
Ecoli	7	4	1.9	2	0.1	4	1.9
Glass	9	4	1.1	4	1.1	1	0.1
Letter	32	23	1.6	2	1.6	2	1.6
Pageblocks	10	2	0.1	9	0.1	2	0.1
Pima	9	6	0.5	8	0.5	8	0.5
Spambase	57	9	0.3	22	0.3	6	0.1
Thyroid	6	2	0.1	3	1.1	2	0.1
Wine	13	2	0.6	3	1.1	3	0.6
Average	18.2	6.3	0.8	6.4	0.7	4.2	0.6

datasets on KNN, HBOS, and ECOD. Not only that, the average AUC value of MNIFS can be taken as the maximum when using both HBOS and ECOD. Compared with the original data, for all datasets on the three algorithms, MNIFS is much higher than the original AUC. It indicates that MNIFS is effective in removing uncorrelated candidate features and using a smaller subset of features to increase the outlier detection performance.

According to Fig. 4, the AUCs obtained using MNIFS on the nominal-attribute datasets have a maximum value on most of the datasets, but the difference in curve area with other comparison algorithms is small. On the dataset Chess2, as Fig. 4(c) shows, the ROC curve of MNIFS is closest to the upper left corner of the first quadrant and has the largest curve area.

The above analysis shows that MNIFS is also applicable to the nominal-attribute dataset.

6.5. Parameter sensitivity analysis

From the analysis in Section 4.1, we know that the multi-neighborhood radius comes from the standard deviation *std* and the adjustment parameter λ in MNIFS. Since the standard deviation *std* is computed from the attribute values, the influence of the value of the parameter λ and the number of selected features on the outlier detection need further exploration when different feature sequences are obtained.

Fig. 5 gives the relationship between the parameter λ, the number of selected features, and the AUC on partially mixed-attribute datasets and numerical-attribute datasets. The following analysis can be obtained from Fig. 5.

(1) On the majority of datasets, the AUC increases and then decreases with the increasing number of features like Autos, Bands1, and Letter. This indicates that the selected features are unable to offer new information for the learning algorithm when they are redundant features or uncorrelated features. On the contrary, it might give false guidance to the outlier detection algorithm.
(2) The AUCs obtained from the original data of Pageblocks and Wine are the same as, or even higher than, those of the data after partially performing feature selection. It indicates that there will be cases where the wrong selection of features leads to lower outlier detection performance instead.
(3) When using MNIFS on several datasets, such as Cardio and Ecoli, their best performance is obtained when outlier detection is performed after selecting only the first few features, which indicates that MNIFS successfully places the most representative and important features in front of the feature sequence.
(4) As shown in Fig. 5, for some datasets, such as Heart and Glass, the maximum value of AUC does not change significantly as the parameter λ increases. However, for datasets like Autos and Bands2, the maximum value of AUC has great fluctuation with the change of λ, which has some sensitivity to λ. This indicates that λ needs to be taken to the most appropriate value in order for the algorithm to be optimal.

From the preceding analyses, it is known that the obtained AUC values are different when different parameters λ are used and different numbers of feature subsets are selected. Therefore, the value of the parameter λ can be adjusted to further adapt to the distribution structure

Table 7
Comparison of AUC of outlier detection for numerical-attributes (%).

Algorithms	Datasets	Initial	LS	USFSM	SPEC	UFSFS	FR-FS	RSR	FMIUFS	EUIAR	MNIFS
kNNOD	Cardio	73.30	80.26	70.20	85.72	74.06	78.79	76.12	73.58	**87.33**	**87.33**
	Ecoli	87.99	87.87	87.12	90.35	88.92	88.70	88.58	87.12	**91.10**	90.49
	Glass	84.61	84.61	85.58	84.88	85.88	86.18	86.53	86.15	83.60	**87.24**
	Letter	89.50	89.96	90.10	90.01	89.85	90.33	90.53	91.91	**93.16**	92.75
	Pageblocks	88.12	87.73	94.76	95.21	88.77	91.01	88.72	87.73	96.02	**97.79**
	Pima	93.82	92.29	94.48	91.97	94.14	94.48	94.77	92.84	94.77	**95.18**
	Spambase	79.10	79.08	79.53	91.37	82.10	83.96	79.39	79.53	87.23	**92.07**
	Thyroid	95.08	96.69	95.45	98.77	95.55	96.69	95.25	93.12	98.89	**98.93**
	Wine	49.92	83.87	85.71	**99.87**	63.78	87.69	94.08	91.30	96.60	96.13
	Average	82.38	86.93	86.99	92.02	84.78	88.65	88.22	87.03	92.08	**93.10**
HBOS	Cardio	85.11	87.45	86.66	89.52	86.13	86.52	86.17	86.17	**91.30**	**91.30**
	Ecoli	80.84	88.70	79.88	88.48	85.63	86.70	81.58	79.88	**88.74**	88.48
	Glass	70.08	90.98	78.56	71.82	89.32	76.26	**90.98**	77.07	73.41	76.23
	Letter	59.78	60.06	62.75	63.50	60.57	61.65	66.58	61.61	68.40	**69.90**
	Pageblocks	92.85	93.23	95.35	91.26	94.87	95.35	94.51	92.84	**97.30**	95.35
	Pima	94.93	92.74	93.52	92.30	95.68	93.52	95.68	94.11	**95.68**	**95.68**
	Spambase	77.41	77.39	77.65	84.45	79.83	77.50	78.36	77.62	**85.68**	85.52
	Thyroid	95.82	96.27	96.19	96.53	96.53	96.27	94.21	96.02	96.01	**97.31**
	Wine	90.67	97.39	91.93	**99.62**	94.29	91.60	92.18	95.63	99.24	95.88
	Average	83.05	87.13	84.72	86.39	86.98	85.04	86.69	84.55	88.42	**88.41**
ECOD	Cardio	93.50	94.25	93.77	93.56	93.73	94.07	94.01	93.77	95.39	**95.78**
	Ecoli	78.08	89.23	76.93	88.48	86.68	86.36	81.28	76.93	91.35	**92.46**
	Glass	62.06	59.40	78.43	63.58	68.18	75.99	67.91	74.77	80.49	**80.98**
	Letter	57.23	56.31	61.86	56.59	59.00	58.08	65.70	57.42	65.66	**71.49**
	Pageblocks	93.76	94.52	95.56	92.52	98.04	92.75	97.56	93.05	**98.87**	98.22
	Pima	94.72	94.08	94.53	93.04	95.58	94.53	95.58	94.00	**95.58**	**95.58**
	Spambase	81.23	80.05	81.24	**93.54**	87.21	81.33	81.24	81.41	89.45	93.27
	Thyroid	97.71	98.24	98.08	98.16	98.16	98.24	96.08	96.86	98.20	**99.04**
	Wine	73.28	73.87	84.79	**95.71**	74.79	83.70	77.48	85.46	95.46	89.33
	Average	81.29	82.22	85.02	86.13	84.60	85.01	84.09	83.74	90.05	**90.68**

Table 8
Number of features obtained by MNIFS for nominal-attribute datasets.

Datasets	Original feature	kNNOD	HBOS	ECOD
Breast	9	8	2	6
Chess1	36	35	28	28
Chess2	36	35	28	28
Chess3	36	28	27	27
Chess4	36	35	27	27
Chess5	36	35	34	34
Average	31.5	29.3	24.3	25.0

of the data, so as to obtain better outlier detection performance. Not only that, when choosing the suitable λ and the number of feature subsets, MNIFS achieves favorable outcomes under most conditions, further verifying the feasibility of MNIFS in feature selection.

6.6. Statistical test

In order to assess whether the comparison algorithms are statistically significant, according to the strategy in Refs. [44,51], we first use Friedman's test to verify that the performance of all algorithms is different. Then use Nemenyi's post hoc test to distinguish them. To more visually show the differences between the compared algorithms, Nemenyi's test figures are available. In Nemenyi's test figures, the mean ordinal values of 9 feature selection algorithms are drawn at the appropriate positions of the number axis in the test plot. If a set of algorithms is linked together using horizontal line segments, then it can be assumed that there is no significant difference between the set of algorithms.

More concretely, based on the experimental results in Tables 5, 7, and 9, it can be obtained that there are 9 comparison algorithms and 24 datasets, i.e., $M = 9$ and $N = 24$. Thus, we can compute the τ_F distribution with 8 and 184 degrees of freedom. Before using Friedman's test, the AUC of each algorithm on all datasets is sorted from low to high, and the sequence number is assigned (1,2,...). Among them, if the AUC of the two algorithms is the same, the ordinal values are equally divided. The average ordinal values of different algorithms are calculated as shown in Table 10. Based on Friedman's test, the τ_F and critical values of different learning algorithms at the significance level $\alpha = 0.05$ are shown in Table 11. From Table 11, it can be seen that each value of τ_F on kNNOD, HBOS, and ECOD is greater than the critical value of 1.9890 when $\alpha = 0.05$. Therefore, the null hypothesis that "all algorithms have the same performance" does not hold for these three algorithms. This indicates that the performance of all feature selection algorithms on kNNOD, HBOS, and ECOD is significantly different. In this case, it is necessary to perform a post hoc test with the aim of further distinguishing them.

When the significance level $\alpha = 0.05$, the corresponding critical distance $CD_{0.05} = 2.4523$ can be calculated. Nemenyi's test figures for three outlier detection algorithms can be seen in Fig. 6. From Fig. 6, MNIFS is statistically significantly different from most of the other algorithms. For example, as can be seen from Fig. 6(a), MNIFS is only connected to EUIAR by horizontal line segments, which indicates that MNIFS is statistically significantly different from the other compared algorithms in terms of kNNOD. In addition, the three learning algorithms have large average ordinal values in Fig. 6 which also indicates the effectiveness of MNIFS.

7. Conclusions

For the unsupervised hybrid feature selection problem, an unsupervised heterogeneous feature selection method based on the fuzzy multi-neighborhood granule is proposed in this paper. The method makes use of fuzzy multi-neighborhood entropy to deal with uncertain information effectively, and it also takes into account the relevance, redundancy, and interactivity of each feature. Furthermore, the presented method is highly applicable to numerical, nominal, and mixed-attribute datasets by constructing the fuzzy multi-neighborhood granule. In addition, the method does not require the discretization of numerical attributes, which not only reduces the preprocessing time but also maintains the distribution structure of the data. For this method, an algorithm called MNIFS is designed and compared with existing

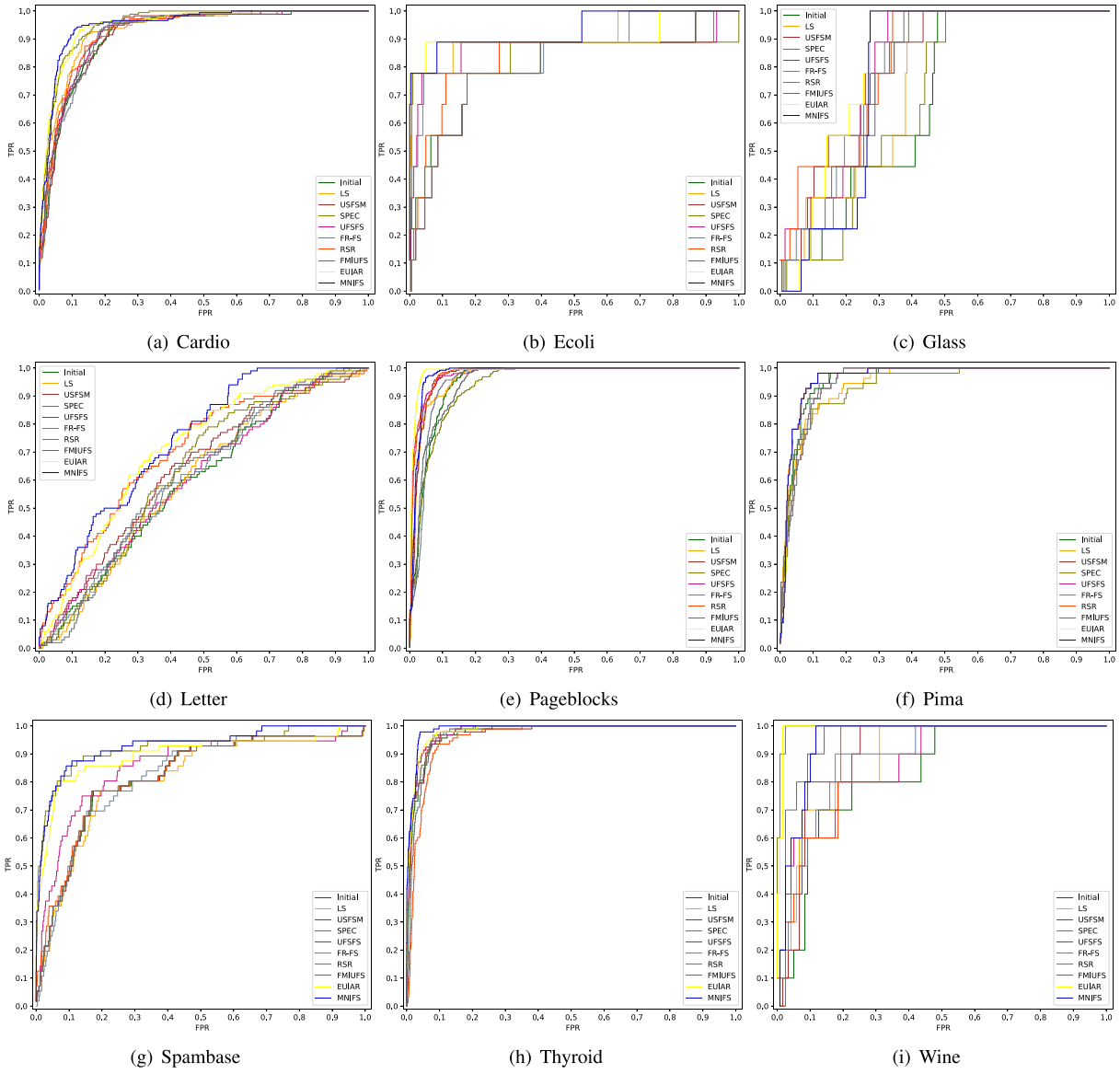


Fig. 3. ROC curves for numerical-attribute datasets.

algorithms on 24 publicly unbalanced datasets. By analyzing the experimental results, it can be concluded that MNIFS can use a smaller number of features for better performance of outlier detection. It can be effectively used for heterogeneous data outlier detection analysis. However, the complementarity between features was not considered when accessing the feature importance. It refers to the supplementary information provided by the new features relative to the selected features. Ignoring complementarity may lead to erroneous deletion of features that provide important discriminative information. For future research, increasing the complementarity between features may be included to explore unsupervised heterogeneous feature selection.

CRediT authorship contribution statement

Siyu Yang: Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Zhong Yuan:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition. **Chuan Luo:** Validation, Funding acquisition, Project administration. **Hongmei Chen:** Resources, Project administration, Funding acquisition. **Dezhong Peng:** Validation, Funding acquisition, Project administration.

Declaration of competing interest

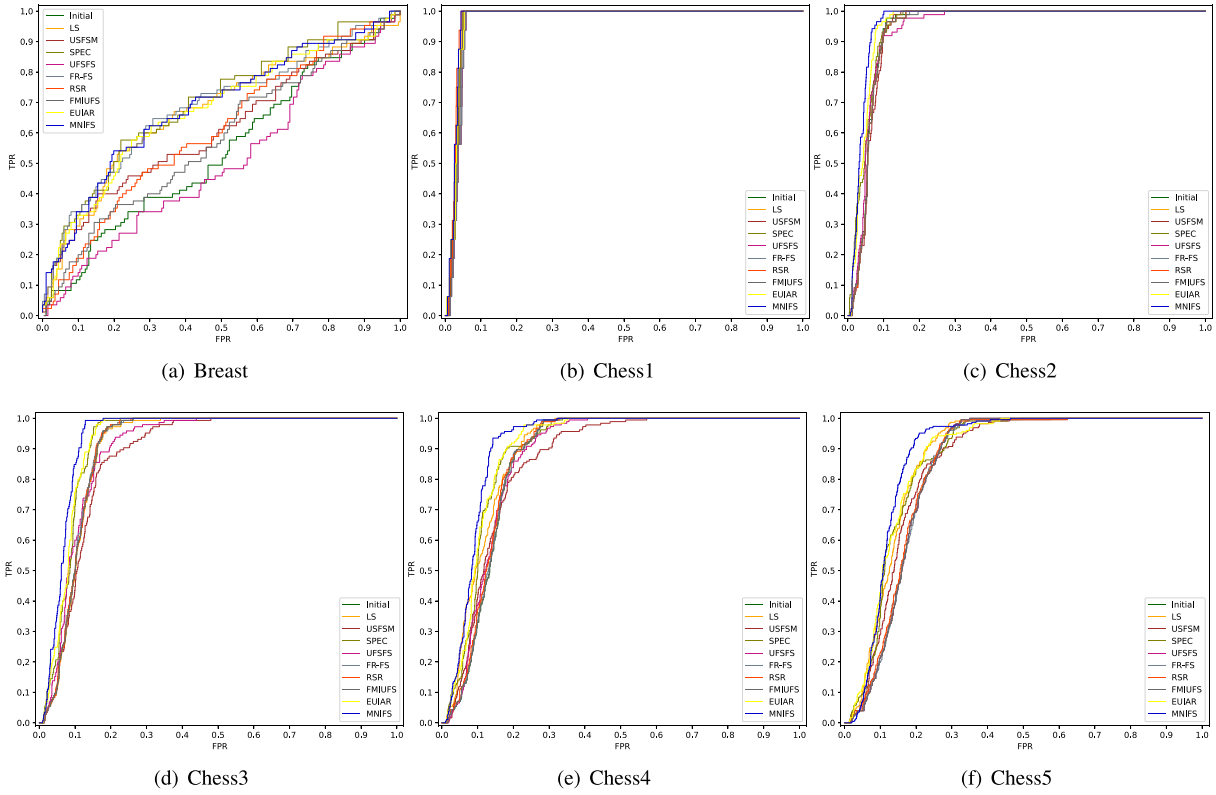
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (62306196, 62476182, and 62376230), Sichuan Science and Technology Program, China (2024NSFTD0049, 2024ZDZX0004, 2024YFHZ0144, 2024YFHZ0089, and 2024NSFSC0443), and the Fundamental Research Funds for the Central Universities, China(YJ202245).

Table 9
Comparison of AUC of outlier detection for nominal-attributes (%).

Algorithms	Datasets	Initial	LS	USFSM	SPEC	UFSFS	FR-FS	RSR	FMIUFS	EUIAR	MNIFS
kNNOD	Breast	70.23	69.13	65.91	70.21	65.90	70.07	65.91	65.26	69.13	**70.35**
	Chess1	99.08	99.13	**99.34**	99.32	99.08	99.08	99.31	99.09	99.17	99.24
	Chess2	96.18	95.60	95.61	96.49	95.30	95.60	96.14	96.14	**96.88**	96.49
	Chess3	90.20	90.18	90.93	93.08	92.49	**92.84**	90.85	90.20	92.57	91.73
	Chess4	88.16	88.17	89.41	**90.61**	88.65	89.20	88.44	88.17	90.29	88.96
	Chess5	85.98	85.98	86.95	**88.46**	85.98	87.89	85.98	85.98	88.20	86.81
	Average	88.31	88.03	88.03	**89.70**	87.90	89.11	87.77	87.47	89.37	88.93
HBOS	Breast	48.73	65.68	57.08	**68.14**	46.23	65.97	53.79	51.71	63.60	65.87
	Chess1	95.18	95.75	95.71	96.30	95.18	95.23	96.12	95.60	96.33	**97.20**
	Chess2	91.92	91.72	91.36	93.95	91.61	91.72	91.72	91.67	95.11	**96.16**
	Chess3	86.19	85.99	84.00	89.93	86.10	86.19	86.19	86.19	91.93	**93.35**
	Chess4	82.37	82.97	81.42	85.56	82.36	82.37	82.56	82.37	89.31	**91.05**
	Chess5	79.06	80.40	80.38	81.77	79.06	78.96	79.24	79.06	86.54	**87.76**
	Average	80.58	83.75	81.66	85.94	80.09	83.41	81.60	81.10	87.14	**88.57**
ECOD	Breast	65.55	66.30	63.62	**67.77**	63.62	66.86	63.62	63.84	67.76	67.69
	Chess1	96.88	97.54	97.48	96.90	96.88	96.93	**97.95**	97.09	97.12	96.94
	Chess2	95.22	95.09	95.08	95.36	95.33	95.09	95.09	95.06	95.45	**96.14**
	Chess3	91.50	91.31	89.75	91.62	91.16	91.50	91.62	91.50	91.85	**93.22**
	Chess4	88.95	88.89	88.59	89.10	88.96	88.95	89.46	88.95	89.22	**90.91**
	Chess5	86.54	86.54	87.05	86.74	86.54	86.40	87.02	86.54	86.46	**87.70**
	Average	87.44	87.61	86.93	87.92	87.08	87.62	87.46	87.16	87.98	**88.77**



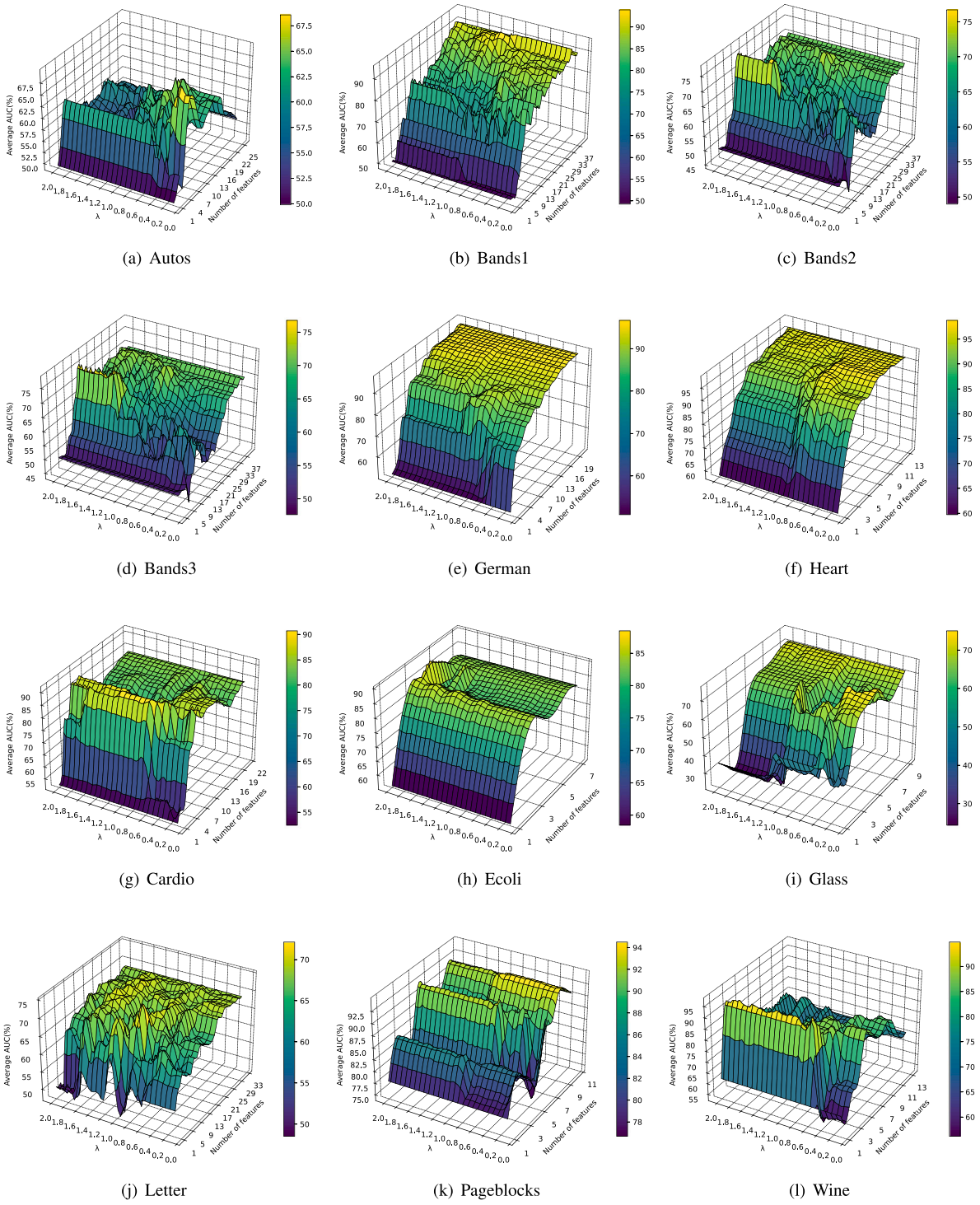
(a) Breast (b) Chess1 (c) Chess2

(d) Chess3 (e) Chess4 (f) Chess5

Fig. 4. ROC curves for nominal-attribute datasets.

Table 10
Average ordinal values of Nemenyi's test between algorithms.

Algorithms	LS	USFSM	SPEC	UFSFS	FR-FS	RSR	FMIUFS	EUIAR	MNIFS
KNNOD	2.7500	5.1042	5.2708	3.7708	4.8333	4.3333	3.9375	7.1458	7.8542
HBOS	4.6458	3.9583	5.0625	3.9792	3.5417	4.6667	4.3542	6.8750	7.9167
ECOD	3.6875	3.7917	4.2083	5.0417	4.2708	5.0417	4.0417	7.2083	7.7083



(a) Autos (b) Bands1 (c) Bands2

(d) Bands3 (e) German (f) Heart

(g) Cardio (h) Ecoli (i) Glass

(j) Letter (k) Pageblocks (l) Wine

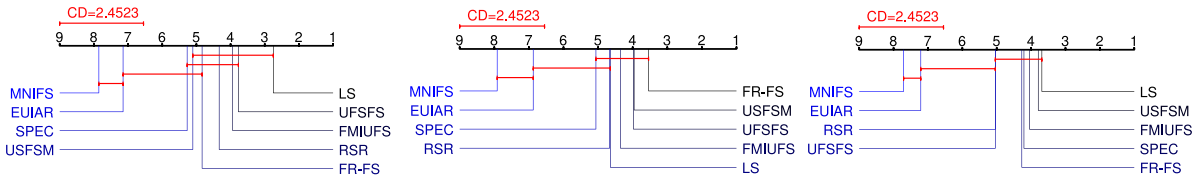
Fig. 5. AUC varies with the parameter λ and the number of features.

Table 11
τ_F of different outlier detection algorithms.

Algorithms	τ_F	Critical value ($\alpha = 0.05$)
kNNOD	12.3933	
HBOS	9.0439	1.9890
ECOD	9.4514	

Data availability

The data is publicly available online at https://github.com/BELLoney/Outlier-detection.



(a) kNNOD (b) HBOS (c) ECOD

Fig. 6. Nemenyi's test figures on three outlier detection algorithms.

References

[1] S. Vluymans, L. D'eer, Y. Saeys, C. Cornelis, Applications of fuzzy rough set theory in machine learning: A survey, Fund. Inform. 142 (1–4) (2015) 53–86.

[2] N. Aslan, G.O. Koca, M.A. Kobat, S. Dogan, Multi-classification deep CNN model for diagnosing COVID-19 using iterative neighborhood component analysis and iterative relieff feature selection techniques with X-ray images, Chemometr. Intell. Lab. Syst. 224 (2022) 104539.

[3] J.M. Xing, C. Gao, J. Zhou, Weighted fuzzy rough sets-based tri-training and its application to medical diagnosis, Appl. Soft Comput. 124 (2022) 109025.

[4] J.D. Li, K.W. Cheng, S.H. Wang, F. Morstatter, R.P. Trevino, J.L. Tang, H. Liu, Feature selection: A data perspective, ACM Comput. Surv. (CSUR) 50 (6) (2017) 1–45.

[5] X.F. He, D. Cai, P. Niyogi, Laplacian score for feature selection, Adv. Neural Inf. Process. Syst. 18 (2005).

[6] C. Yao, Y.F. Liu, B. Jiang, J.G. Han, J.W. Han, LLE score: A new filter-based unsupervised feature selection method based on nonlinear manifold embedding and its application to image recognition, IEEE Trans. Image Process. 26 (11) (2017) 5257–5269.

[7] H. Lim, D.W. Kim, Pairwise dependence-based unsupervised feature selection, Pattern Recognit. 111 (2021) 107663.

[8] D.M. Qian, K.Y. Liu, S.M. Zhang, X.B. Yang, Semi-supervised feature selection by minimum neighborhood redundancy and maximum neighborhood relevancy, Appl. Intell. 54 (2024) 7750—-7764.

[9] D. Cai, C.Y. Zhang, X.F. He, Unsupervised feature selection for multi-cluster data, in: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2010, pp. 333–342.

[10] X.F. He, M. Ji, C.Y. Zhang, H.J. Bao, A variance minimization criterion to feature selection using laplacian regularization, IEEE Trans. Pattern Anal. Mach. Intell. 33 (10) (2011) 2013–2025.

[11] L. Yang, K.Y. Qin, B.B. Sang, W.H. Xu, Dynamic fuzzy neighborhood rough set approach for interval-valued information systems with fuzzy decision, Appl. Soft Comput. 111 (2021) 107679.

[12] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, Int. J. Gen. Syst. 17 (2–3) (1990) 191–209.

[13] Z. Yuan, H.M. Chen, P. Xie, P.F. Zhang, J. Liu, T.R. Li, Attribute reduction methods in fuzzy rough set theory: An overview, comparative experiments, and new directions, Appl. Soft Comput. 107 (2021) 107353.

[14] J.Y. Zhao, Z.L. Zhang, C.Z. Han, Z.F. Zhou, Complement information entropy for uncertainty measure in fuzzy rough set and its applications, Soft Comput. 19 (2015) 1997–2010.

[15] S. An, J.Y. Liu, C.Z. Wang, S.Y. Zhao, A relative uncertainty measure for fuzzy rough feature selection, Internat. J. Approx. Reason. 139 (2021) 130–142.

[16] C.Z. Wang, Y.L. Qi, M.W. Shao, Q.H. Hu, D.G. Chen, Y.H. Qian, Y.J. Lin, A fitting model for feature selection with fuzzy rough sets, IEEE Trans. Fuzzy Syst. 25 (4) (2016) 741–753.

[17] A. Tan, W.Z. Wu, Y.H. Qian, J.Y. Liang, J.K. Chen, J.J. Li, Intuitionistic fuzzy rough set-based granular structures and attribute subset selection, IEEE Trans. Fuzzy Syst. 27 (3) (2018) 527–539.

[18] Q.H. Hu, L. Zhang, D. Zhang, W. Pan, S. An, W. Pedrycz, Measuring relevance between discrete and continuous features based on neighborhood mutual information, Expert Syst. Appl. 38 (9) (2011) 10737–10750.

[19] C.Z. Wang, Y. Huang, W.P. Ding, Z.H. Cao, Attribute reduction with fuzzy rough self-information measures, Inform. Sci. 549 (2021) 68–86.

[20] P.F. Zhu, Q. Xu, Q.H. Hu, C.Q. Zhang, Co-regularized unsupervised feature selection, Neurocomputing 275 (2018) 2855–2863.

[21] C. Velayutham, K. Thangavel, Unsupervised quick reduct algorithm using rough set theory, J. Electron. Sci. Technol. 9 (3) (2011) 193–201.

[22] N. Mac Parthaláin, R. Jensen, Unsupervised fuzzy-rough set-based dimensionality reduction, Inform. Sci. 229 (2013) 106–121.

[23] Z. Yuan, H.M. Chen, T.R. Li, Z. Yu, B.B. Sang, C. Luo, Unsupervised attribute reduction for mixed data based on fuzzy rough sets, Inform. Sci. 572 (2021) 67–87.

[24] A. Ganivada, S.S. Ray, S.K. Pal, Fuzzy rough sets, and a granular neural network for unsupervised feature selection, Neural Netw. 48 (2013) 91–108.

[25] J.H. Wan, H.M. Chen, T.R. Li, Z. Yuan, J. Liu, W. Huang, Interactive and complementary feature selection via fuzzy multigranularity uncertainty measures, IEEE Trans. Cybern. 53 (2) (2021) 1208–1221.

[26] E.C. Tsang, D.G. Chen, D.S. Yeung, X.Z. Wang, J.W. Lee, Attributes reduction using fuzzy rough sets, IEEE Trans. Fuzzy Syst. 16 (5) (2008) 1130–1141.

[27] W. Wei, J.B. Cui, J.Y. Liang, J.H. Wang, Fuzzy rough approximations for set-valued data, Inform. Sci. 360 (2016) 181–201.

[28] H. Zhao, P. Wang, Q.H. Hu, P.F. Zhu, Fuzzy rough set based feature selection for large-scale hierarchical classification, IEEE Trans. Fuzzy Syst. 27 (10) (2019) 1891–1903.

[29] P. Ni, S.Y. Zhao, X.Z. Wang, H. Chen, C.P. Li, E.C. Tsang, Incremental feature selection based on fuzzy rough sets, Inform. Sci. 536 (2020) 185–204.

[30] Q.H. Hu, D. Yu, Z.X. Xie, Information-preserving hybrid data reduction based on fuzzy-rough techniques, Pattern Recognit. Lett. 27 (5) (2006) 414–423.

[31] C.Z. Wang, Y. Huang, M.W. Shao, D.G. Chen, Uncertainty measures for general fuzzy relations, Fuzzy Sets and Systems 360 (2019) 82–96.

[32] C.Y. Wang, L.J. Wan, New results on granular variable precision fuzzy rough sets based on fuzzy (co) implications, Fuzzy Sets and Systems 423 (2021) 149–169.

[33] S. Solorio-Fernández, J.F. Martínez-Trinidad, J.A. Carrasco-Ochoa, A new unsupervised spectral feature selection method for mixed data: A filter approach, Pattern Recognit. 72 (2017) 314–326.

[34] A. Chaudhuri, D. Samanta, M. Sarma, Two-stage approach to feature set optimization for unsupervised dataset with heterogeneous attributes, Expert Syst. Appl. 172 (2021) 114563.

[35] S. An, Q.H. Hu, D. Yu, Fuzzy entropy based max-relevancy and min-redundancy feature selection, in: 2008 IEEE International Conference on Granular Computing, IEEE, 2008, pp. 101–106.

[36] Q.H. Hu, D. Yu, Z.X. Xie, J.F. Liu, Fuzzy probabilistic approximation spaces and their information measures, IEEE Trans. Fuzzy Syst. 14 (2) (2006) 191–201.

[37] D. Dubois, H. Prade, Putting rough sets and fuzzy sets together, in: Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory, vol. 11, Springer, 1992, pp. 203–232.

[38] Z.H. Wang, H.M. Chen, Z. Yuan, X.L. Yang, P.F. Zhang, T.R. Li, Exploiting fuzzy rough mutual information for feature selection, Appl. Soft Comput. 131 (2022) 109769.

[39] S. An, Q.H. Hu, C.Z. Wang, Probability granular distance-based fuzzy rough set model, Appl. Soft Comput. 102 (2021) 107064.

[40] F. Macedo, M.R. Oliveira, A. Pacheco, R. Valadas, Theoretical foundations of forward feature selection methods based on mutual information, Neurocomputing 325 (2019) 67–89.

[41] Z. Zhao, H. Liu, Spectral feature selection for supervised and unsupervised learning, in: Proceedings of the 24th International Conference on Machine Learning, 2007, pp. 1151–1157.

[42] P. Mitra, C. Murthy, S.K. Pal, Unsupervised feature selection using feature similarity, IEEE Trans. Pattern Anal. Mach. Intell. 24 (3) (2002) 301–312.

[43] P.F. Zhu, W.M. Zuo, L. Zhang, Q.H. Hu, S.C. Shiu, Unsupervised feature selection by regularized self-representation, Pattern Recognit. 48 (2) (2015) 438–446.

[44] Z. Yuan, H.M. Chen, P.F. Zhang, J.H. Wan, T.R. Li, A novel unsupervised approach to heterogeneous feature selection based on fuzzy mutual information, IEEE Trans. Fuzzy Syst. 30 (9) (2021) 3395–3409.

[45] Z. Yuan, H.M. Chen, T.R. Li, Exploring interactive attribute reduction via fuzzy complementary entropy for unlabeled mixed data, Pattern Recognit. 127 (2022) 108651.

[46] T. Cover, P. Hart, Nearest neighbor pattern classification, IEEE Trans. Inf. Theory 13 (1) (1967) 21–27.

[47] M. Goldstein, A. Dengel, Histogram-based outlier score (hbos): A fast unsupervised anomaly detection algorithm, in: KI-2012: Poster and Demo Track, vol. 1, Citeseer, 2012, pp. 59–63.

[48] Z. Li, Y. Zhao, X.Y. Hu, N. Botta, C. Ionescu, G. Chen, Ecod: Unsupervised outlier detection using empirical cumulative distribution functions, IEEE Trans. Knowl. Data Eng. 35 (12) (2023) 12181–12193.

[49] Y. Zhao, Z. Nasrullah, Z. Li, PyOD: A python toolbox for scalable outlier detection, J. Mach. Learn. Res. 20 (96) (2019) 1–7.

[50] Z. Yuan, H.M. Chen, T.R. Li, B.B. Sang, S. Wang, Outlier detection based on fuzzy rough granules in mixed attribute data, IEEE Trans. Cybern. 52 (8) (2021) 8399–8412.

[51] Z. Yuan, H.M. Chen, C. Luo, D.Z. Peng, MFGAD: Multi-fuzzy granules anomaly detection, Inf. Fusion 95 (2023) 17–25.