



Research paper

Feature selection considering synergy between features based on soft neighborhood rough sets

Lubin Chen^a, Jinkun Chen^{a,*}, Yaojin Lin^b^a School of Mathematics and Statistics, Minnan Normal University, Zhangzhou, Fujian, 363000, PR China^b School of Computer Science, Minnan Normal University, Zhangzhou, Fujian, 363000, PR China

ARTICLE INFO

Keywords:

Feature selection

Uncertainty measures

Hybrid data

Soft neighborhood rough sets

ABSTRACT

Neighborhood entropy-based measures provide a powerful framework for feature selection to select features that are more useful for classification. However, most of these feature selection methods do not pay attention to the complementarities and synergies between features, as well as the interactions between them. In addition, most existing neighborhood rough sets are subjective in the determination of neighborhood radius when dealing with classification problems, which may lead to the omission of useful information. To solve these problems, a soft neighborhood rough set model-based feature selection method (SNCMI) is proposed. Firstly, the method dynamically adjusts the neighborhood radius, significantly minimizing its influence on the uncertainty measurement. Secondly, it comprehensively considers the correlation, redundancy, complementarity, and synergy between features through soft neighborhood uncertainty measures. Thirdly, an innovative objective evaluation function is introduced to evaluate the interactions between features. Finally, we compare the proposed SNCMI algorithm with several well-known feature selection algorithms on twenty public datasets and demonstrate the effectiveness of SNCMI.

1. Introduction

In real life, many datasets are filled with noise, irrelevant or redundant information. When processing these datasets, this increases both the required storage space and computational load, potentially lowering the classification performance of classifiers as well. This undoubtedly increases the difficulty of knowledge discovery (Cekik and Uysal, 2020), data mining (Ding et al., 2020) and pattern recognition (Taskin et al., 2017). To cope with this challenge, feature selection techniques have emerged as one of the effective ways to address this problem. Their core goal is to filter the most representative and discriminative subset from all features, optimizing the model's accuracy and efficiency.

Commonly used feature selection approaches consist of filtering methods (Forman, 2003; Zhang et al., 2013), wrapper methods (Hastie et al., 2001), and embedding methods (Zare et al., 2013). Compared with the traditional method, the rough set based method has some advantages in the feature selection task, especially it can define the neighborhood or granules without prior knowledge (Pawlak, 1982). The classical rough set theory is a mathematical method for approximating decision concepts based on equivalence relations. As a result, rough set theory has been widely used in feature selection (Raza and Qamar, 2017; Ibrahim et al., 2020; Chen and Zhu, 2024; Velayutham and

Thangavel, 2011; Yuan et al., 2021), hyperspectral classification (Yu et al., 2019; Liu et al., 2016), image labeling (Yu et al., 2013), credit rating (Yao and Lu, 2011) and incremental learning (Liu and Liang, 2014; Nandhini and Thangadurai, 2023; Xu and Liu, 2024). Nevertheless, the classical rough set depends on equivalence relations and equivalence classes to categorize the domain, limiting its ability to deal directly with numerical data. When facing numerical data, discretization is often required, which may lead to information loss. To allow for more flexible handling of continuous data, some scholars combined the principles of fuzzy set with rough set, proposing the fuzzy rough sets (Dubois and Prade, 1990; Wang et al., 2017, 2024). This approach enhances the accuracy in describing the boundaries and uncertainty of data. Furthermore, Lin (1988) introduced the neighborhood rough sets (NRS) by the idea of neighborhood granularity. Hu et al. (2008) applied the NRS to solve the problem of feature subset selection for heterogeneous data, broadening its application. Therefore, numerous scholars have extended the NRS, resulting in the development of many variants, such as neighborhood-based decision-theoretic rough sets model (Li et al., 2016), fuzzy NRS (Wang et al., 2016), local NRS (Wang et al., 2018), pseudo-labeled NRS (Yang et al., 2019), K -nearest NRS (Wang et al., 2019), multi-label NRS (Sun et al., 2020), weighted NRS (Hu et al., 2021; Wang and Zhao, 2024), neighborhood multigranulation

* Corresponding author.

E-mail addresses: cjb2304@163.com (L. Chen), cjk0132@mnnu.edu.cn (J. Chen), zzlinyaojin@163.com (Y. Lin).

rough sets (Shu et al., 2024b), and so on. However, in practice, the two existing neighborhood rough sets, i.e., NRS and kNNS, are not good enough to correctly classify the samples into the classes they belong to when facing a very dense or sparse sample space, resulting in poor classification accuracy. This is mainly due to the fact that the somewhat subjective fixing of the neighborhood radius, and the set parameter values have an impact on the model. As a result, some researchers began to consider the adaptive neighborhood radius, so as to reduce the impact of the neighborhood radius (Zhou et al., 2019; An et al., 2023; Song et al., 2022; Shu et al., 2024a). For example, An et al. (2023) combine the soft margin theory and the *KNN*, structuring a new model that allows each sample to choose an optimal neighborhood radius adaptively and define its neighborhood granularity. This approach ensures that samples within the neighborhood granularity consistently belong to the same class.

On the other hand, uncertainty measures are applicable in feature selection (Sun et al., 2019; Chen et al., 2017; Jiang and Zhang, 2024; Ma et al., 2024). Among these measures, mutual information has been widely utilized in various feature selection algorithms (Lewis, 1992; Yang and Ong, 2012; Sulaiman and Labadin, 2015; Yuan et al., 2023; Gong et al., 2024), with Lewis's maximum mutual information algorithm (Lewis, 1992) being the earliest to present the concept of mutual information. However, it lacks the consideration of relationships between the features. For this reason, Peng et al. (2005) further took into account the redundant relationships between features, building upon the original foundation. This is known as the minimum redundancy maximum correlation feature selection method. Nevertheless, in practical problems, there exist complex interactions among features. In Hagar's (Hagar et al., 2015) study, it was observed that nearly 90% of software faults in a web server application were triggered by three or fewer parameters, including 40% by a single parameter and 30% by interactions of two parameters. To reveal latent information contained within these interaction features and avoid omitting critical information, Jakulin (2005) first considered interactions between features and used Bayesian confidence intervals to evaluate the probability distribution of the expected loss assuming no interactions exists. Tang et al. (2019) considered the use of higher-order interactions to measure and mitigate information loss when facing the high-dimensional problems in the application of text mining. In the context of biological research, variations in the physiological and pathological processes occurring in organisms are often influenced by interactions between molecules. Therefore, considering interactions between features can better assess the importance of features in biological data analysis, which is crucial for disease diagnosis, treatment, and prevention (Lin et al., 2019). Regarding the interactions between features and selected features, Bennasar et al. (2013) explored it by using information interaction gain, while Nakariyakul (2018) designed an algorithm guided by mutual information. Hu et al. (2010), Wan et al. (2021), Xu et al. (2024), Wang et al. (2023) combined mutual information with neighborhood rough sets to establish algorithm that utilizes neighborhood mutual information, which defines correlation, redundancy, and interaction in the neighborhood uncertainty metric. Chen et al. (2015), Wan et al. (2023), Li et al. (2020), Ma and Ju (2022) assessed the complementarity among features and performed feature selection through the application of mutual information. Unlike complementarity, synergy refers to the fact that two features united can bring more complete information to the classification than they provide on their own (Cheng et al., 2011; Sosa-Cabrera et al., 2024). However, existing feature selection algorithms that utilize mutual information do not simultaneously consider these conditions, resulting in inadequate representation of the concepts of complementarity and synergy.

Inspired by these problems, this paper introduces the concept of conditional mutual information as a means to quantify interactions between features. Furthermore, it introduces an innovative feature selection algorithm called soft neighborhood conditional mutual information (SNCMI), aimed at selecting useful attributes in hybrid data types. The main contributions of this study include:

(1) A new approach for calculating the soft neighborhood radius in hybrid data is introduced, allowing the model to dynamically adjust the neighborhood radius, minimizing its effect on the uncertainty measure.

(2) A novel feature objective evaluation function is constructed to comprehensively discuss and redefine the neighborhood uncertainty metric, taking into account the correlation and redundancy among features, as well as the complementary and synergistic effects between features. This approach aims to enhance the completeness of the theory for evaluating feature importance and reduce the effects of underestimating the interactions between features on the construction of the actual model.

(3) Comparative experiments between the proposed SNCMI algorithm and nine other feature selection algorithms on public datasets reveal that SNCMI achieves better classification performance. Furthermore, parameter sensitivity analysis and statistical significance analysis validate the algorithm's robustness and effectiveness.

The paper is organized as follows: Section 2 discusses the fundamentals of NRS and soft interval theory and introduces an uncertainty measure for neighborhood decision systems. In Section 3, the soft neighborhood radius is computed for each sample, and an uncertainty metric is proposed to handle hybrid data. In Section 4, feature interactions under soft neighborhood uncertainty measures are introduced in detail, and a new feature evaluation function and feature selection algorithm (SNCMI) are proposed. Section 5 offers experimental comparisons, discusses the results, and provides visualizations. Lastly, Section 6 concludes the paper and suggests future work.

2. Preliminaries

A brief review of the core concepts of NRS, soft-margin theory, and uncertainty measures will be presented in this section.

2.1. δ -NRS

Given a neighborhood information system $NIS = (U, FD, f, \delta)$, where $U = \{u_1, u_2, \dots, u_m\}$ is the set of all samples in NIS , and this set is non-empty. $FD = F \cup D$ denotes conditional and decision attribute sets, where $F = \{a_i, i = 1, 2, \dots, n\}$ and $D = \{d\}$, f is a mapping that maps the Cartesian product of samples and features onto the value domain of features. Finally, δ is a parameter in the range of $[0, 1]$ that determines the neighborhood radius of the sample. In NIS , to measure the similarity between two samples, we often employ the Euclidean distance, which is computed as shown below:

$$d_F(u_i, u_j) = \sqrt{\sum_{a_i \in F} |f(u_i, a_i) - f(u_j, a_i)|^2}. \quad (1)$$

After obtaining the similarity between all samples, the neighborhood relation on feature F is defined as:

$$R_\delta(F) = \{(u_i, u_j) \mid d_F(u_i, u_j) \leq \delta\}. \quad (2)$$

Having obtained the similarity relation, the δ -neighborhood classes of u are defined as:

$$[u]_{R_\delta(F)} = \{u_j \mid d_F(u, u_j) \leq \delta\}. \quad (3)$$

Sample u is consistent if all samples in $[u]_{R_\delta(F)}$ are of the same type of decision, and vice versa for inconsistent.

2.2. k NNS

Given an NIS , for any $u_i, u_j \in U$, the k samples set closest to the sample u_i is denoted by:

$$\kappa_F(u_i) = \{u_v^v \mid d_F(u_v^v, u_i) < d_F(u_j, u_i), u_j \neq u_i \neq u_v^v, v = 1, 2, \dots, k\}. \quad (4)$$

Clearly, for any $u_i \in U$, the union of all k sample sets can cover U . Similarly, a binary relation $R_k(F)$ is defined as:

$$R_k(F) = \{(u_i, u_j) \mid u_j \in \kappa_F(u_i)\}. \quad (5)$$

2.3. Soft-margin theory

Soft-margin theory is an important part of the SVM algorithm and an extension of hard margin theory. Soft-margin relaxes the strict requirements of hard margin by introducing slack variables that allow some data points to slightly violate the limits of the margin. This means that even if some data points are not exactly classified correctly, the model can still find a valid classification hyperplane. This notion is a compromise that balances the goals of maximizing the classification margin and minimizing the classification loss, which is mathematically modeled as:

$$\min \frac{1}{2} \omega^T \omega + \lambda \sum_{i=1}^{|U|} \xi_i, \quad s.t. \quad y_i(b + \omega^T x_i) \geq 1 - \xi_i, \xi_i \geq 0, \quad (6)$$

where $\xi_i = [1 - y_i(\omega x_i + b)]_+$, λ is a penalty term that weighs the importance of counting the width of the separation surface after the classification is correct and the number of misclassified samples.

2.4. Information metrics in neighborhood decision systems

Various information entropies have been widely used for feature selection. Entropy measures the uncertainty of the information, and as the subset of features increases, the sample division becomes more precise, providing more information for classification, and vice versa. In this paper, we introduce the neighborhood entropy. It calculates the information from the neighborhood granularity that is formed by feature subsets, which is formalized as follows.

Definition 2.1 (Hu et al., 2010). Given an NIS, the neighborhood entropy of the dataset with respect to F , is defined as:

$$NH_{[u]}(F) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|[u_i]_{R_\delta(F)}|}{|U|}, \quad (7)$$

where $|[u_i]_{R_\delta(F)}|$ is the cardinality of the neighborhood granule induced by $R_\delta(F)$.

3. Soft neighborhood-based rough sets approach to hybrid data

In this section, we first describe the approach for handling hybrid data. Then, we present a computational example of the proposed algorithm to better illustrate its process. Finally, we introduce information measures in soft neighborhood decision information systems.

3.1. Hybrid data handling

In actual applications, data often contain two types: numeric and nominal. Therefore, we introduce the Heterogeneous Overlapping Euclidean Metric (HOEM) to process the hybrid data.

Definition 3.1 (Xu et al., 2024). For hybrid data, the HOEM distance function is given as:

$$HOEM_{a_i}(u_i, u_j) = \begin{cases} d_{a_i}(u_i, u_j), & a_i \text{ is the numerical feature} \\ 1, & a_i \text{ is the nominal feature, and } f(u_i, a_i) \neq f(u_j, a_i) \\ 0, & a_i \text{ is the nominal feature, and } f(u_i, a_i) = f(u_j, a_i) \end{cases} \quad (8)$$

Therefore, when given an NIS, the neighborhood class for u_i is denoted by:

$$[u_i]_{R_\delta(a_i)} = \{u_j | HOEM_{a_i}(u_i, u_j) \leq \delta\}. \quad (9)$$

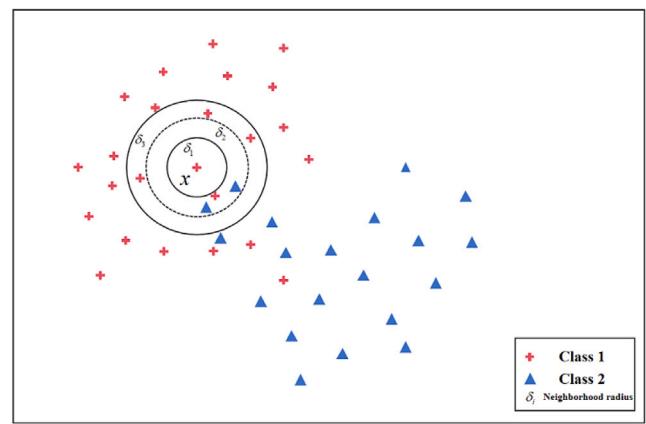


Fig. 1. The classification results of different granules.

3.2. Benefits of the soft neighborhood

In the classification task, we seek to minimize the differences within a single class while maximizing the differences between different classes, facilitating better classification. Feature selection is done to identify features that aid in classification. However, in practice, there are samples from different classes that exhibit very small differences in their eigenvalues, yet they are distributed across different classes, which leads to misclassification.

In NRS, the extent of δ represents its tolerance level for noisy data, i.e., the degree of quantification error allowed. In Fig. 1, the inconsistency in classification for different radius granularities is visualized. In this problem of binary classification, given different neighborhood radii $\delta_1, \delta_2, \delta_3$ such that $\delta_1 < \delta_2 < \delta_3$, it is clear from Fig. 1 at that granularity $\delta_1(x)$, the classification is consistent, and the sample x belongs to class 1. However, the classification at granularity $\delta_2(x)$, which has a larger error margin, is inconsistent, making it difficult to determine which class the sample x belongs to. In contrast, at granularity $\delta_3(x)$, the sample x is classified as belonging to class 1. Therefore, the impact of the neighborhood radius is considerable. For two classical NRS methods, determining the neighborhood radius involves subjectivity, making it challenging to choose a specific radius for point x , which increases the risk of misclassification. In contrast, an adaptive neighborhood radius can flexibly adjust the radius according to data density, avoiding the limitations of fixed methods and improving classification accuracy and robustness. Thus, an adaptive neighborhood radius is crucial for classification tasks.

3.3. Soft neighborhood rough sets

Although the δ -neighborhood introduced in Section 2 is capable of handling continuous data, the parameter δ needs to be set by humans, which is somewhat subjective. Additionally, when the neighborhood radius is uniform across samples, it may lead to biased results in the learning task, especially when there are differences in the distributions of feature values. Therefore, adapting the neighborhood radius for each sample is an effective approach to solving this issue.

Given a soft neighborhood information system $SNIS = (U, FD, f, K, \alpha)$, where the components U , FD and f are defined similarly as in the earlier NIS. Here, K and α are parameters for determining the neighborhood radius. For any $u, y_i \in U$, $A \subseteq F$, the K nearest samples of u are obtained by calculating the distance between u and y_i from the Euclidean distance. And arranging them in ascending order $Y = \{y_1, y_2, \dots, y_K\}$, they meet $d_A(u, y_1) \leq d_A(u, y_2) \leq \dots \leq d_A(u, y_K)$. The resulting soft neighborhood radius of u is fixed as follows:

$$d_S = \arg \max_{d_A(u, y_i)} \{d_A(u, y_i) - \alpha \cdot p_i, i = 1, \dots, K\}, \quad (10)$$

where α denotes the penalty factor, and p_i indicates the count of samples that belong to a different class than u , among those with a distance no greater than $d_A(u, y_i)$. Then the closed sphere centered on u with radius δ_S is called the soft neighborhood of u . The soft neighborhood relation on A is denoted by:

$$SNR_{\delta_S}(A) = \{(u, y_i) | d_A(u, y_i) \leq \delta_S\}. \quad (11)$$

A similarity relation $SNR_{\delta_S}(A)$ is determined by the distance measure d_A , and is represented in the universe as a relationship matrix $M(d_A)$.

Definition 3.2. Given a SNIS, for any $u_i \in U, \forall A \subseteq F$, the soft neighborhood class of u_i is defined as:

$$\aleph_{SNR_{\delta_S}(A)}(u_i) = \{u_j | d_A(u_i, u_j) \leq \delta_S, u_j \in Y\}. \quad (12)$$

Adaptively determining the neighborhood radius is key to effectively handling complex datasets. $\aleph_{SNR_{\delta_S}(A)}(u_i)$ is able to adaptively find the neighborhood radius, avoiding inappropriate neighborhood radius settings caused by data size and dimensionality, thus overcoming the drawbacks of δ -neighborhood and KNN when describing samples with uneven density. For example, in the classification problem of x discussed in Section 3.1, the value of δ_S enables the samples to obtain the optimal classification hyperplane under misclassification conditions, enhancing robustness and preventing failures in classification due to inappropriate neighborhood radius α or improper setting of K -value.

Definition 3.3. Given a SNIS, in terms of any $A \subseteq F$, the soft neighborhood entropy is defined as:

$$SNH_N(A) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|\aleph_{SNR_{\delta_S}(A)}(u_i)|}{|U|}. \quad (13)$$

Definition 3.4. Given a SNIS, in terms of any $A, B \subseteq F$, the soft neighborhood joint entropy is defined as:

$$SNJH_N(A, B) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|\aleph_{SNR_{\delta_S}(A \cup B)}(u_i)|}{|U|}. \quad (14)$$

Corollary 1. Given a SNIS, in terms of any $A_1, A_2, \dots, A_m \subseteq F$, the soft neighborhood joint entropy is defined as:

$$SNJH_N(A_1, A_2, \dots, A_m) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|\aleph_{SNR_{\delta_S}(A_1 \cup A_2 \cup \dots \cup A_m)}(u_i)|}{|U|}. \quad (15)$$

Definition 3.5. Given a SNIS, in terms of any $A, B \subseteq F$, the soft neighborhood conditional entropy is defined as:

$$SNCI_N(B|A) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|\aleph_{SNR_{\delta_S}(A \cup B)}(u_i)|}{|\aleph_{SNR_{\delta_S}(A)}(u_i)|}. \quad (16)$$

Proposition 3.1. Given a SNIS, then $SNCI_N(B|A) = SNJH_N(A, B) - SNH_N(A)$ holds.

From Proposition 3.1, the soft neighborhood conditional entropy is computed using the soft neighborhood joint entropy of A and B , along with the soft neighborhood information entropy of A .

Definition 3.6. Given a SNIS, in terms of any $A, B \subseteq F$, the soft neighborhood mutual information is defined as:

$$SNMI_N(A; B) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|\aleph_{SNR_{\delta_S}(A)}(u_i)| \cdot |\aleph_{SNR_{\delta_S}(B)}(u_i)|}{|U| \cdot |\aleph_{SNR_{\delta_S}(A \cup B)}(u_i)|}. \quad (17)$$

By Definitions 3.3–3.6 and Proposition 3.1, we can identify the relationship between soft neighborhood entropy. The following proposition explain the calculation of soft neighborhood conditional entropy and soft neighborhood mutual information in detail.

Proposition 3.2. Given a SNIS, in terms of any $A, B \subseteq F$, we have

- (1) $SNMI_N(A; B) = SNMI_N(B; A)$;
- (2) $SNMI_N(A; B) = SNH_N(A) + SNH_N(B) - SNJH_N(A, B)$;
- (3) $SNMI_N(A; B) = SNH_N(A) - SNCI_N(A|B) = SNH_N(B) - SNCI_N(B|A)$.

In Proposition 3.2, (1) indicates the symmetrical properties of soft neighborhood mutual information, and (2) and (3) show the correlation between soft neighborhood information measures.

Definition 3.7. Given a SNIS, in terms of any $A, B, E \subseteq F$, the soft neighborhood conditional mutual information is defined as:

$$SNCMI_N(A; B|E) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|\aleph_{SNR_{\delta_S}(A \cup E)}(u_i)| \cdot |\aleph_{SNR_{\delta_S}(B \cup E)}(u_i)|}{|\aleph_{SNR_{\delta_S}(E)}(u_i)| \cdot |\aleph_{SNR_{\delta_S}(A \cup B \cup E)}(u_i)|}. \quad (18)$$

Definition 3.8. Given a SNIS, in terms of any $A, B, E \subseteq F$, the soft neighborhood synergy mutual information is defined as:

$$SNSMI_N(E; A, B) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|\aleph_{SNR_{\delta_S}(A \cup B)}(u_i)| \cdot |\aleph_{SNR_{\delta_S}(A \cup E)}(u_i)| \cdot |\aleph_{SNR_{\delta_S}(B \cup E)}(u_i)|}{|U| \cdot |\aleph_{SNR_{\delta_S}(A \cup B \cup E)}(u_i)|^2}. \quad (19)$$

When there are synergies between features, the amount of information provided by the synergies can be measured by soft neighborhood synergy mutual information.

Proposition 3.3. Given a SNIS, in terms of any $A, B, E \subseteq F$, we have

- (1) $SNCMI_N(A; B|E) = SNCMI_N(B; A|E)$;
- (2) $SNCMI_N(A; B|E) = SNJH_N(A, E) + SNJH_N(B, E) - SNJH_N(A, B, E) - SNH_N(E)$;
- (3) Given E , if A and B are mutually independent, then $SNCMI_N(A; B|E) = 0$;
- (4) $SNSMI_N(E; A, B) = SNMI_N(A; E) + SNCMI_N(A; B|E) + SNCMI_N(B; E|A)$;
- (5) $SNSMI_N(E; A, B) = SNMI_N(B; E) + SNCMI_N(B; A|E) + SNCMI_N(A; E|B)$.

Similarly to Proposition 3.2, Proposition 3.3 (1) embodies the idea that given a subset E , the amount of information about A from B is equal to the amount of information about B from A , i.e., the information they provide is mutual. Proposition 3.3 (2) shows the correlation between soft neighborhood information measures, and Proposition 3.3 (3) shows that when E is known, if A and B are mutually independent, soft neighborhood conditional mutual information has a value of 0. Proposition 3.3 (4) and (5) express the relationship between soft neighborhood synergy mutual information, soft neighborhood mutual information and soft neighborhood conditional mutual information.

4. Feature importance assessment with the soft neighborhood rough set

In this section, we will present a method for feature selection that uses a soft neighborhood uncertainty measure. This method incorporates feature correlation, redundancy, complementarity, and synergy to provide a more comprehensive evaluation.

4.1. Exploring feature multiple correlations in feature selection

Features that demonstrate meaningful interactions with categories contribute valuable information to classification, with mutual information commonly used to quantify these interactions. Therefore, mutual information is introduced into the soft neighborhood rough set.

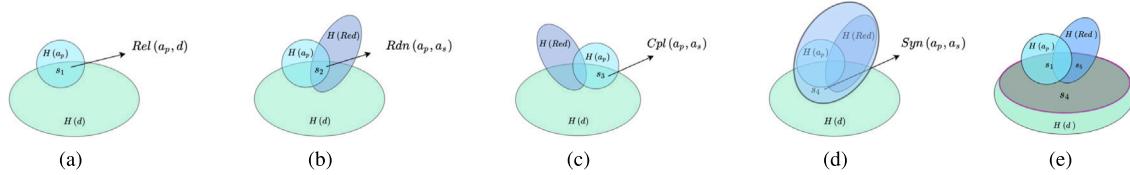


Fig. 2. Diagram of the interactions between features.

Definition 4.1 (Rel). Given a SNIS, let $Red \subseteq F$ represent the subset of selected features, and let $a_p \in F - Red$ represents the current candidate feature, the correlation of a_p with d can be expressed as:

$$Rel(a_p, d) = SNM I_R(a_p; d). \quad (20)$$

Region s_1 in Fig. 2(a) indicates the information relevant to classification. For any $a_{p_1}, a_{p_2} \in F$, if the correlation between a_{p_2} and d is greater than the correlation between a_{p_1} and d , then $SNM I_R(a_{p_2}; d) > SNM I_R(a_{p_1}; d)$, indicating that a_{p_2} provides more discriminative information for classification. Therefore, we choose the $\max_{a_p \in F - Red} Rel(a_p, d)$ as the initial feature.

To determine whether the current candidate feature can join Red feature subset, it is necessary to investigate the redundancy between a_p and the selected features in Red , as well as the information provided by the complementarity and synergy between a_p and Red for d .

Definition 4.2 (Rdn). Given a SNIS, $a_s \in Red \subseteq F$ are the features that have been selected, and $a_p \in F - Red$ represents the current candidate feature. If the classification information provided by a_p can be deduced from the classification information provided by Red , then a_p is a redundant feature, and the redundancy of a_p and Red is defined as:

$$Rdn(a_p, a_s) = SNM I_R(a_p; a_s). \quad (21)$$

Redundancy means that a_p and Red provide duplicate or redundant classification information, the redundancy of a_p and Red can be seen in Fig. 2(b), and the redundancy between a_p and Red provides redundant information about classification in the region s_2 .

Definition 4.3 (Cpl). Given a SNIS, $a_s \in Red \subseteq F$ are the features that have been selected, $a_p \in F - Red$ represents the current candidate feature, with respect to d , the complementarity between a_p and Red is defined as:

$$Cpl(a_p, a_s) = SNCM I_R(a_p; d | a_s). \quad (22)$$

Complementarity indicates that while a_p may overlap with Red in terms of redundancy, it can also contribute additional complementary information to Red , as illustrated in Fig. 2(c). The region s_3 reflects the complementary information between a_p and Red regarding d .

In contrast to complementarity, synergy refers to the phenomenon where features cooperate to provide more useful information than any single feature alone. For example, a_p and Red synergize to provide information that they cannot provide on their own. The region s_4 , as indicated in Fig. 2(d), shows the synergistic information provided by both a_p and Red .

In order to further quantify the amount of information provided by synergy, it may be assumed that the area shaded in gray is the amount of information that a_p and Red synergistically provide to d , as shown in Fig. 2(e). This part contains the information provided by a_p for d , i.e., s_1 , as well as the complementary information of a_p and Red , i.e., s_5 . Thus the effective information provided by a_p and Red synergistically for d is calculated as follows:

$$\begin{aligned} Syn(a_p, a_s) &= SNSM I_R(d; a_p, a_s) - SNM I_R(a_p; d) - SNCM I_R(a_s; d | a_p) \\ &= SNCM I_R(a_p; a_s | d). \end{aligned} \quad (23)$$

Definition 4.4 (Syn). Given a SNIS, $a_s \in Red \subseteq F$ are the features that have been selected, $a_p \in F - Red$ represents the current candidate feature, with respect to d , the synergy between a_p and Red is defined as follows:

$$Syn(a_p, a_s) = SNCM I_R(a_p; a_s | d). \quad (24)$$

Therefore, when determining feature importance, the features with greater synergy with the approximation set Red are selected, and the criterion for maximizing synergy is formalized as follows:

$$\max_{a_p \in F - Red, a_s \in Red} Syn(a_p, a_s). \quad (25)$$

4.2. Objective evaluation functions for features based on SNCMI

Evaluation criteria are often used to measure the contribution of features to information categorization. Some researchers have considered improving feature relevance to categories and reducing feature redundancy to enhance the classification accuracy of the algorithms. However, in fact, there may be complex interactions between features, such as complementarity and synergism, which can provide more useful information but are rarely mentioned. Therefore, based on the discussion in Section 4.1, this section establishes an objective evaluation function of the original features \mathcal{L}_{SNCMI} , as shown below:

$$\mathcal{L}_{SNCMI} = \arg \max_{a_p \in F - Red} J_{sig}(a_p), \quad (26)$$

where

$$\begin{aligned} J_{sig}(a_p) &= SNM I_R(a_p; d) - \frac{1}{|Red|} \sum_{s=1}^{|Red|} SNM I_R(a_p; a_s) \\ &\quad + \frac{1}{|Red|} \sum_{s=1}^{|Red|} (SNCM I_R(a_p; d | a_s) + SNCM I_R(a_p; a_s | d)). \end{aligned} \quad (27)$$

In this evaluation function, $SNM I_R(a_p; d)$ measures the feature relevance to categories, assessing the importance of the feature for classification tasks. $SNM I_R(a_p; a_s)$ measures the redundancy of a_p and Red , ensuring that the selected feature is not overly repetitive with existing features. Through $SNCM I_R(a_p; d | a_s)$ and $SNCM I_R(a_p; a_s | d)$, we can evaluate the additional information between the feature and Red , thus helping to better select complementary feature combinations.

4.3. Algorithms and frameworks

Based on the above detailed discussion and visualization, a suitable candidate feature can supply classification information for the selected features while also providing valuable extra information through synergy. Therefore, a new SNCMI algorithm is proposed, with the corresponding feature selection framework displayed in Fig. 3.

The feature selection framework, as shown in Fig. 3, comprises three main stages. The first stage calculates the soft neighborhood radius for each sample, which establishes the soft neighborhood granularity. The second stage assesses the correlation, redundancy, complementarity, and synergy among features, resulting in a new feature selection evaluation function. In the third stage, features with the highest importance scores are progressively added to the Red and removed from the remaining set, continuing this process until all features are selected. Each

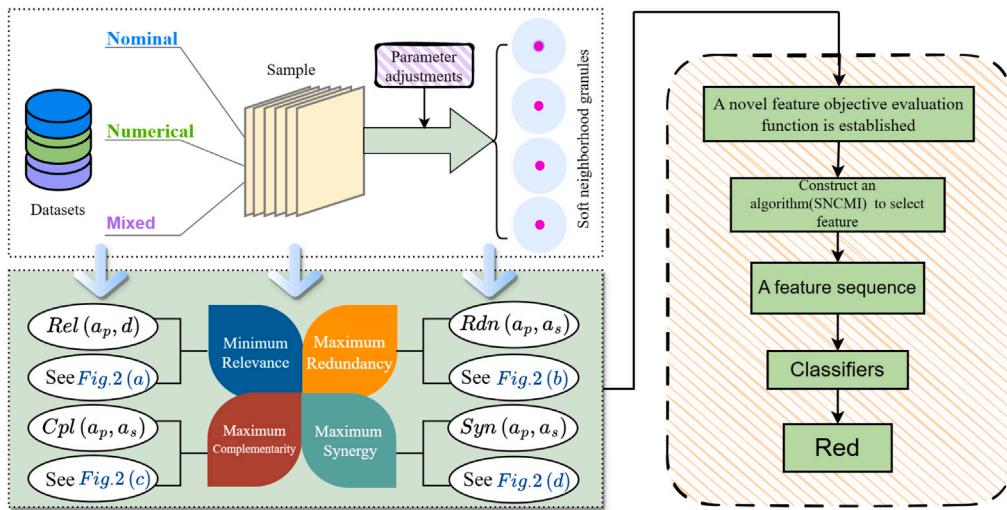


Fig. 3. The proposed feature selection framework.

sequence of features is subsequently tested on a designated classifier, with the subset yielding the highest average classification accuracy being identified as the optimal one.

Algorithm 1: SNCMI

```

Input: A SNIS = ( $U, FD, f, K, \alpha$ ),  $U = \{u_1, u_2, \dots, u_m\}$  and
 $F = \{a_i, i = 1, \dots, n\}$ ; parameter  $K$  for neighborhood
adjustment has a range of 1 to 21 in steps of 2, while
parameter  $\alpha$  for neighborhood adjustment has a range
of 0.1 to 0.3 in steps of 0.05.

Output: Reduced set  $Red$ .
1  $Red \leftarrow \emptyset$ .
2 for each  $u_i \in U$  do
3   Calculate  $N_{SNR_{\delta_S}(F)}(u_i)$ ; // The adaptive neighborhood of
   each sample is obtained.
4 end
5 for each  $a_p \in F$  do
6   Calculate  $Rel(a_p, d)$ ;
7 end
8  $Red \leftarrow a_s ; a_s = \max_{a_p \in F - Red} Rel(a_p, d) ; F \leftarrow F - \{a_s\}$ ;
9 for each  $a_q \in F$  do
10  for each  $a_s \in Red$  do
11    Calculate  $Rdn(a_q, a_s)$ ;
12  end
13  for each  $a_s \in Red$  do
14    Calculate  $Cpl(a_q, a_s)$ ;
15    Calculate  $Syn(a_q, a_s)$ ;
16  end
17  Compute  $J_{sig}(a_q)$ ;
18  Select the feature  $a_q$  with the maximum objective function
  values( $\max_{a_q \in A - Red} J_{sig}(a_q)$ );
19   $Red \leftarrow Red \cup \{a_q\}$ ;
20   $F \leftarrow F - a_q$ ;
21 end
22 Different classifiers are employed to select the  $Red$  set that
  achieves the highest classification accuracy.
23 return  $Red$ 
```

This algorithm is divided into three key parts. The first part is to calculate the soft neighborhood of each sample under different features according to the data distribution (Steps 2 to 4). The second part first uses soft neighborhood mutual information to assess the relevance of features to categories (Steps 5 to 7). It further evaluates redundancy,

complementarity, and synergy between features and classes using both soft neighborhood mutual information and soft neighborhood conditional mutual information (Steps 9 to 16). Finally, we use the feature objective function to calculate the feature importance and select the features that are more favorable for classification according to the evaluation value in order (Steps 17 to 20). In the third part, we use the packaging feature selection algorithm to determine the feature subset to improve classification performance (Step 22).

The time and space complexity analysis of the SNCMI algorithm is as follows: the time complexity and space complexity of initializing Red are both $O(1)$ (Step 1). The time complexity for calculating the soft neighborhood radius of each sample is $O(m)$, and the soft neighborhood relationship is represented by the neighborhood matrix, so the space complexity is $O(m^2)$ (Steps 2 to 4). Iterate the feature set F and calculate the associativity of features and labels has a time complexity of $O(n)$ and space complexity of $O(n)$ (Steps 5 to 7). Selecting the maximum relevance feature has a time complexity of $O(1)$ and space complexity of $O(1)$ (Step 8). Iteratively calculate redundancy, complementarity, synergy, and objective function values, and update Red and F , results in a time complexity of $O(n^2)$, and the computation results between all features need to be stored, so the space complexity is $O(n)$ (Steps 9 to 21). Finally, selecting the optimal Red using different classifiers requires $O(m)$ time complexity and $O(mn)$ space complexity (Step 22). In summary, the overall time complexity is $O(n^2)$ and the overall space complexity is $O(m^2)$.

4.4. Examples

This subsection gives an example to explain the proposed algorithm.

Example 1. On the left side of Table 1 is an initial data table, $U = \{u_1, u_2, \dots, u_6\}$ and $F = \{a_1, a_2, a_3, a_4\}$, where a_1, a_2 are numerical attributes and a_3, a_4 are categorical attributes. The raw numerical data underwent min-max normalization, with the resulting values shown on the right side of Table 1.

For any $a_j \in F$, here, it may be useful to take $K = 3$ and $\alpha = 0.15$. Take the object u_1 under feature a_1 as an example, note that a_1 is a numerical feature, we can calculate the distance between each object and u_1 using the Euclidean distance. After the computation is completed, these distance values are sorted from smallest to largest to get u_2, u_3, u_4, u_6, u_5 . Then according to Eq. (10) and the obtained parameter values, we can get $\delta_S = \arg \max \{0.125 - 1 \times 0.15, 0.125 - 2 \times 0.15, 0.375 - 2 \times 0.15\} = 0.075$, therefore, the neighborhood radius is $d_{a_1}(u_1, u_4)$,

Table 1

Initial and standardized decision information systems.

U	a_1	a_2	a_3	a_4	d	a_1	a_2	a_3	a_4	d
u_1	6	0.6	1	2	1	0.625	1	1	2	1
u_2	5	0.2	1	1	0	0.5	0	1	1	0
u_3	5	0.4	0	1	0	0.5	0.5	0	1	0
u_4	3	0.5	1	2	1	0.25	0.75	1	2	1
u_5	1	0.3	0	0	0	0	0.25	0	0	0
u_6	9	0.2	0	1	1	0	0	0	1	1

so we get the neighborhood class of u_1 as $\{u_1, u_2, u_3, u_4\}$. Then the neighborhood relationship matrix under each feature can be calculated

$$\text{as: } M(a_1) = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, M(a_2) = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 \end{pmatrix},$$

$$M(a_3) = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}, M(a_4) = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{pmatrix},$$

$$M(d) = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

The soft neighborhood entropy of feature a_1 is calculated by [Definition 3.3](#) as:

$$SNH_{\mathbb{N}}(a_1) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|\mathbb{N}_{SNR_{\delta_S}(a_1)}(u_i)|}{|U|} \approx 1.3208. \quad (28)$$

Similarly, we can calculate the neighborhood entropy for the remaining features as:

$$SNH_{\mathbb{N}}(a_2) \approx 0.7516, SNH_{\mathbb{N}}(a_3) \approx 0.5850, \\ SNH_{\mathbb{N}}(a_4) \approx 0.8900, SNH_{\mathbb{N}}(d) \approx 0.5850. \quad (29)$$

The correlation between a_1 and d is calculated by [Definition 3.6](#) as follows:

$$Rel(a_1, d) = SNCMI_{\mathbb{N}}(a_1; d) \\ = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log_2 \frac{|\mathbb{N}_{SNR_{\delta_S}(a_1)}(u_i)| \cdot |\mathbb{N}_{SNR_{\delta_S}(d)}(u_i)|}{|U| \cdot |\mathbb{N}_{SNR_{\delta_S}(a_1 \cup d)}(u_i)|} \approx 0.4183. \quad (30)$$

Similarly, we can compute the correlation between the remaining features and the decision.

$$Rel(a_2, d) \approx -0.0534, Rel(a_3, d) \approx 0.0724, Rel(a_4, d) \approx 0.3774. \quad (31)$$

Next, select the features that have the greatest relevance measure and add them to Red , i.e., a_1 is added to Red , then $Red = \{a_1\}$, and the subset of candidate features is $\{a_2, a_3, a_4\}$.

The soft neighborhood joint entropy of each pair of features is calculated by [Definition 3.4](#), and the soft neighborhood joint entropy of each feature and the decision category is shown in [Tables 2](#) and [3](#), respectively.

According to [Proposition 3.2](#), for selected feature a_1 and current candidate feature a_2 , the redundancy between them is calculated as:

$$Rdn(a_1, a_2) = SNH_{\mathbb{N}}(a_1) + SNH_{\mathbb{N}}(a_2) - SNJH_{\mathbb{N}}(a_1, a_2) \approx 0.0849.$$

Similarly, the redundancy of a_1 with the remaining candidate features is $Rdn(a_1, a_3) \approx 0.1541$, $Rdn(a_1, a_4) \approx 0.5566$.

Table 2

Soft neighborhood joint entropy between features.

$SNJH_{\mathbb{N}}(a_i, a_j)$	a_1	a_2	a_3	a_4
a_1	1.3208	1.9875	1.7516	1.6541
a_2	1.9875	0.7516	1.3617	1.5566
a_3	1.7516	1.3617	0.5850	1.0975
a_4	1.6541	1.5566	1.0975	0.8900

Table 3

Soft neighborhood joint entropy of features and decision class.

$SNJH_{\mathbb{N}}(a_i, d)$	a_1	a_2	a_3	a_4
d	1.4875	1.3900	1.0975	1.0975

For inter-feature interactions, according to [Proposition 3.3](#), for the selected feature a_1 and the current candidate feature a_2 , the complementarity and synergy measures between these two features can be computed as:

$$SNCMI_{\mathbb{N}}(a_2; a_1 | d) = SNJH_{\mathbb{N}}(a_2, d) + SNJH_{\mathbb{N}}(a_1, d) \\ - SNJH_{\mathbb{N}}(a_1, a_2, d) - SNH_{\mathbb{N}}(d) \approx 0.2075. \quad (32)$$

$$SNCMI_{\mathbb{N}}(a_2; d | a_1) = SNJH_{\mathbb{N}}(a_1, a_2) + SNJH_{\mathbb{N}}(a_1, d) \\ - SNJH_{\mathbb{N}}(a_1, a_2, d) - SNH_{\mathbb{N}}(a_1) \approx 0.0692. \quad (33)$$

Similarly, the complementarity and synergy between Red and the remaining candidate features can be calculated as: $SNCMI_{\mathbb{N}}(a_3; a_1 | d) \approx 0.1792$, $SNCMI_{\mathbb{N}}(a_4; a_1 | d) \approx 0.3459$, $SNCMI_{\mathbb{N}}(a_3; d | a_1) \approx 0.0975$, $SNCMI_{\mathbb{N}}(a_4; d | a_1) \approx 0.1667$.

Finally, according to the characteristics of the proposed evaluation function, the importance of all candidate features is calculated.

$$J(a_2) = Rel(a_2, d) - \frac{1}{|Red|} Rdn(a_2, a_1) + \frac{1}{|Red|} (Cpl(a_2, a_1) + Syn(a_2, a_1)) \approx 0.1384, \\ J(a_3) \approx 0.1950, J(a_4) \approx 0.3333.$$

Feature a_4 has a maximum value under the evaluation function, so it becomes the second feature to added, i.e., $Red = \{a_1, a_4\}$. The subset of candidate features is $\{a_2, a_3\}$. A similar strategy is employed to select features individually, resulting in the feature selection sequence $Red = \{a_1, a_4, a_3, a_2\}$.

5. Experiments and analysis of results

In this section, we first present some detailed experimental preparations and design different comparative experiments to verify the superiority and effectiveness of the proposed algorithm. Finally, we verify the stability of the proposed algorithm through parameter sensitivity experimental analysis.

5.1. Datasets

To validate the efficiency of the SNCMI algorithm, we use publicly available datasets. A total of twenty datasets are considered, including twelve numerical, three categorical, and five hybrid datasets. [Table 4](#) presents their detailed characteristics, including high-dimensional gene datasets used in cancer mechanism studies (Colon, Lung), leukemia subtype classification (ALLML), and lymphoma-associated gene expression analysis (LYMPHOMA), as well as other datasets commonly used for classification tasks.

5.2. Comparison setup

This subsection compares $SNCMI$ with several well-established feature selection algorithms. These algorithms are described as follows:

- mRMR ([Peng et al., 2005](#)): It is a classical filtering feature selection algorithm that uses mutual information to determine the importance of variables by calculating the correlation between the

Table 4
Datasets.

ID	Data set	# Samples	# Features	Decision class	Data type
1	Glass	214	9	7	Numerical
2	Wine	178	13	3	Numerical
3	Drug	1885	12	7	Numerical
4	Rice	3810	7	2	Numerical
5	Waveform	5000	21	3	Numerical
6	Wdbc	569	30	2	Numerical
7	Sonar	208	60	2	Numerical
8	Segment	2310	19	7	Numerical
9	Colon	62	2000	2	Numerical
10	Lung	203	3312	5	Numerical
11	LYMPHOMA	66	4026	3	Numerical
12	ALLAMI	72	7129	2	Numerical
13	Soybean	307	35	19	Categorical
14	Zoo	101	16	7	Categorical
15	NPHA	714	14	2	Categorical
16	HCV	615	12	5	Hybrid
17	German	1000	20	2	Hybrid
18	Abalone	4174	8	2	Hybrid
19	Autos	159	25	6	Hybrid
20	Contraceptive	1473	9	3	Hybrid

feature variable and the target variable, as well as the redundancy among features variables and other feature variables.

- NRS (Lin, 1988): It is a extends the classical rough set model by incorporating neighborhood relations. This model assesses the importance of heterogeneous feature subsets and uses neighborhood dependencies to develop a forward feature selection algorithm.
- NCMI_IFS (Wan et al., 2021): It is an algorithm that combines *NRS* as well as information-theoretic uncertainty measures and proposes an objective evaluation function to assess and maximize relevance and interaction while minimizing redundancy.
- KNCMI (Xu et al., 2024): It is an algorithm that combines *KNN* as well as information-theoretic uncertainty measures. It leverages the benefits of both *NRS* and *kNNRS* strategies to iteratively assess the importance of each feature in the decision-making process.
- SNRS (An et al., 2023): It combines soft boundary theory and *NRS* to propose a *SNRS* model. This model introduces a dependency-based feature evaluation metric and employs a forward greedy strategy for feature selection.
- NFRS (Wang et al., 2017): It leverages the concept of fuzzy neighborhoods to define a sample's fuzzy decision and introduces parametric fuzzy relations to portray fuzzy information grains. It uses fuzzy relations to reconstruct the fuzzy upper and lower approximation of the decision. It defines an importance criterion for candidate attributes, along with develops a greedy forward algorithm.
- WAIFS (Wang et al., 2024): It proposes a new weighted fuzzy rough set and evaluates the importance of candidate features by defining the dependency between decision variables and conditional attributes, and then proposes a new algorithm based on the weighted fuzzy rough approximation operator.
- CFS (Moran and Gordon, 2019): It combines the concept of intrinsically motivated autonomous agents with data structures to propose a feature selection method that autonomously optimizes the model. This method iteratively reduces the error in model learning by continuously selecting new features based on the previously selected ones, thereby improving the model's accuracy.
- FIRDL (Wojtas and Chen, 2020): It introduces a dual-network architecture, where two networks are jointly trained to build an importance ranking model. The networks alternate during training and incorporate a stochastic local search process into learning to solve combinatorial optimization challenges.

To ensure the reliability of the experiments, the experimental study in this paper adopts the ten-fold cross-validation method. For data

with missing values, the maximum probability value method is used to estimate the missing values, and then the dataset is randomly divided into ten equal disjoint subsets. In each iteration, nine subsets are used for training and one subset is used for testing. Initially, all numerical features were normalized in the range [0, 1] to ensure consistency and comparability. Each subset is used for testing and after ten iterations, the classification accuracy is averaged to measure the performance of feature selection.

It is important to note that in these comparison algorithms, mRMR, KNCMI, NCMI_IFS, and FIRDL are sorting the features rather than getting a subset of features directly. Therefore, in this paper, we evaluate the classification accuracy of the algorithms by selecting the top 's' features from the sorted feature set for each data subset. The average classification accuracy is then calculated using different classifiers. The optimal number of selected features, 's', corresponds to the point at which the average classification accuracy reaches its maximum.

In terms of algorithm parameters, the proposed algorithm was evaluated using fixed parameters $K = 5$ and $\alpha = 0.15$. Similarly, for the comparison algorithms, a fixed set of parameters was selected and remained unchanged throughout the experiments. These algorithms include mRMR (Peng et al., 2005), NRS (Lin, 1988), NCMI_IFS (Wan et al., 2021), KNCMI (Xu et al., 2024), SNRS (An et al., 2023), NFRS (Wang et al., 2017), and WAIFS (Wang et al., 2024). The parameter values were selected based on the ranges provided in the original articles, without performing an exhaustive search to determine the optimal settings. Specifically, for CFS (Moran and Gordon, 2019) and FIRDL (Wojtas and Chen, 2020), the number of iterations was adjusted according to the dataset size to prevent memory shortages on the experimental equipment. All experimental results reported in this paper were obtained on a system running Windows 10 (AMD Ryzen 5 3550H processor with 16 GB RAM). The experiments were conducted using MATLAB 2023a, while the FIRDL algorithm was implemented in Python 3.8.

5.3. Results and analyses

In this subsection, we conduct six experiments. First, the algorithms proposed in this paper are compared with other algorithms in terms of running time, and the results are shown in Table 5. Second, the optimal number of selected features is compared for all sorting-based algorithms, and the results are shown in Table 6. Third, the classification accuracies of the different algorithms are compared using three classifiers, and the results are shown in Tables 7 to 9. Fourth, the effect of feature synergy on classification accuracy is visualized, and the results are shown in Figs. 4 and 5. Fifth, a parametric sensitivity analysis

Table 5

Running time comparison of different algorithms.

Data set	SNCMI	mRMR	KNCMI	NCMI_IFS	NRS	SNRS	WAES	NFRS	CFS	FIRDL
Glass	0.0585	0.0196	0.2291	0.2593	3.0376	2.0173	0.2092	0.1752	50.3244	265.1061
Wine	0.0750	0.0258	0.1981	0.4682	2.1632	1.3304	0.1409	0.1425	71.0149	189.0089
Drug	7.2176	0.0627	30.3930	50.5198	452.6405	38.5895	17.0896	14.1089	618.4692	119.7506
Rice	11.2488	0.0213	33.4571	43.8489	1995.7930	27.4486	37.2675	32.8378	965.7031	201.8273
Waveform	47.4688	0.2904	377.1252	1468.7830	12562.8510	222.3193	225.8120	200.2214	6204.7201	213.2394
Wdbc	3.9651	0.0954	30.9027	36.4495	76.8461	8.7960	3.2824	3.0155	650.1158	139.0958
Sonar	0.9696	0.2734	24.6556	25.0345	51.5164	5.9340	0.94	1.0025	599.9435	503.3559
Segment	23.4461	0.0872	153.2989	260.5619	1281.1906	75.4403	42.1018	34.1038	1499.4230	299.1299
Soybean	0.6582	0.0866	9.9041	9.5619	20.9868	30.9271	1.4931	0.9710	401.8807	172.7945
Zoo	0.0470	0.0313	0.8658	0.3604	0.6501	1.9853	0.0905	0.0641	56.2675	170.2639
NPHA	1.4281	0.0347	7.0393	7.4500	70.9548	5.4888	2.4007	2.2132	273.4313	118.0563
HCV	0.8531	0.0234	3.2996	3.6015	83.8058	6.5864	1.7157	1.4782	187.7324	122.3847
German	5.0592	0.0484	31.7454	36.2274	324.1713	12.5095	6.8344	6.2309	696.4972	288.6582
Abalone	16.8864	0.0244	51.4237	67.4493	5102.4240	58.3341	70.6201	64.7230	925.0402	136.1239
Autos	0.1636	0.0985	1.4729	1.7346	7.7835	4.0905	0.2720	0.2585	150.6450	235.1830
Contraceptive	2.6628	0.0109	8.3247	13.3176	162.0477	12.3455	7.2790	6.7074	343.2291	203.0799
Colon	235.3082	218.4136	10 558.1563	11 096.6316	54.7007	56.1414	55.9193	31.1400	15 142.8097	92.3025
Lung	1407.0132	655.2318	22 050.3837	24 308.996	4825.7021	570.3524	54.9161	131.7838	-	2656.2661
LYMPHOMA	755.0161	736.5330	-	-	571.8070	135.6573	315.6303	38.1646	-	2901.5733
ALLAML	3135.1203	2927.5322	-	-	351.0903	206.0094	719.5439	72.3501	-	-
Avg.	282.7333	226.9472	1854.0486	2079.5142	1400.1081	74.1152	78.1779	32.0846	1696.3087	475.1158

Table 6

The reduction number of sorting algorithm under three classifiers.

Data set	SNCMI			mRMR			FIRDL			KNCMI			NCMI_IFS		
	CART	SVM	KNN	CART	SVM	KNN	CART	SVM	KNN	CART	SVM	KNN	CART	SVM	KNN
Glass	5	8	3	9	9	6	7	8	2	4	5	4	5	5	5
Wine	8	11	12	5	13	7	3	13	11	3	13	5	4	11	6
Drug	1	1	2	1	1	4	1	1	9	1	1	1	2	1	10
Rice	3	3	3	3	7	6	3	3	3	1	1	1	5	5	5
Waveform	13	19	19	21	21	21	15	21	20	19	19	19	20	21	21
Wdbc	17	17	8	15	19	19	13	7	20	13	10	15	13	18	13
Sonar	53	49	16	14	55	23	19	21	21	41	46	58	8	58	33
Segment	9	9	10	14	13	13	4	8	4	7	10	8	11	14	13
Soybean	12	27	28	12	31	31	22	26	31	26	22	23	16	30	22
Zoo	3	10	7	13	7	7	9	9	10	12	10	12	11	11	8
NPHA	1	1	9	2	4	14	1	5	14	1	1	10	1	6	11
HCV	10	10	11	9	11	9	11	12	8	7	10	9	6	6	8
German	10	16	16	4	17	18	8	19	18	4	12	12	10	6	15
Abalone	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Autos	13	21	6	8	22	19	25	20	24	10	23	15	10	25	20
Contraceptive	8	4	5	9	8	9	6	6	5	9	9	9	9	7	7
Colon	524	80	24	6	179	88	910	1635	16	4	45	67	16	38	101
Lung	33	79	37	386	179	280	846	788	1052	4	40	897	97	46	1036
LYMPHOMA	321	1048	41	45	2988	5	1067	81	46	-	-	-	-	-	-
ALLAML	15	763	279	2	57	48	-	-	-	-	-	-	-	-	-
Avg.	53.00	108.85	26.85	28.95	182.10	31.40	156.37	141.26	69.21	9.28	15.44	64.78	13.61	17.17	74.17

is performed to observe the effects of different parameter values on classification accuracy. Finally, the effectiveness of the algorithms in this paper is verified by statistical tests.

Table 5 records the running time of ten algorithms across twenty datasets. If the execution time exceeds eight hours or the algorithm terminates due to insufficient memory, it is represented as "-". From Table 5, although the proposed algorithm is not the fastest, it remains competitive with other feature selection algorithms. These results indicate that the SNCMI algorithm is computationally efficient and serves as an effective feature selection method. Notably, it outperforms the CFS algorithm significantly. For instance, on the Colon dataset, the proposed algorithm runs 63 times faster than the CFS algorithm. However, on the Waveform and Lung datasets, the difference between the proposed algorithm and the mRMR algorithm is more pronounced due to the higher computational cost of calculating the soft neighborhood radius for each sample. To address this issue, parallelizing the computation of the soft neighborhood radius and entropy can enhance the algorithm's efficiency when handling large-scale datasets.

Table 6 records the optimal number of features selected by the five ranking algorithms on twenty datasets. When these features are

selected, adding the remaining features does not significantly improve the average classification accuracy. From Table 6, it is observed that the average number of features selected by the algorithm proposed in this paper is not optimal for the two classifiers, CART and SVM. However, under the KNN classifier, it performs better than the other algorithms. When the dataset contains more features, the average number of features selected under all three classifiers is better than that of the mRMR algorithm. This further demonstrates that the method proposed in this paper is effective.

Tables 7 to 9 present the average classification accuracies of different algorithms under the three classifiers. From the information in the tables, it is evident that the algorithm proposed in this paper achieves the highest average classification accuracy across three classifiers, even though it does not achieve the highest accuracy on some individual datasets. Using the CART classifier (as shown in Table 7), the SNCMI algorithm achieves the best classification results on thirteen datasets. Similarly, with the SVM classifier (as shown in Table 8), SNCMI achieves the highest classification accuracy on fourteen datasets. When using the KNN classifier (as shown in Table 9), the algorithm continues to perform well, achieving the best performance

Table 7Average classification accuracy (mean \pm std. dev. %) of different feature selection algorithms on the CART.

Data set	SNCMI	mRMR	KNCMI	NCMI_IFS	NRS	SNRS	WAFS	NFRS	CFS	FIRDL
Glass	74.20 \pm 9.1541	69.11 \pm 10.2965	69.17 \pm 8.2619	69.11 \pm 10.2965	69.11 \pm 9.1138	67.75 \pm 9.2796	68.20 \pm 8.9125	68.19 \pm 9.2172	67.77 \pm 8.6414	69.29 \pm 10.9037
Wine	93.86 \pm 6.1115	89.38 \pm 6.6605	90.91 \pm 6.5495	92.19 \pm 4.6687	89.35 \pm 6.0948	88.79 \pm 5.8257	90.49 \pm 6.4637	89.93 \pm 6.8442	89.93 \pm 6.8442	90.52 \pm 6.0341
Drug	51.67 \pm 3.2699	45.67 \pm 3.7437	51.46 \pm 3.3337	45.41 \pm 3.5276	41.27 \pm 2.3545	41.27 \pm 2.8945	40.58 \pm 2.8755	40.90 \pm 2.4996	40.79 \pm 3.1652	46.84 \pm 2.6045
Rice	89.87 \pm 1.0513	90.21 \pm 1.2376	89.68 \pm 1.3097	89.32 \pm 1.2560	89.37 \pm 1.2813	89.19 \pm 1.8959	89.21 \pm 1.1832	89.06 \pm 1.3213	89.08 \pm 1.1011	89.27 \pm 1.2277
Waveform	77.08 \pm 1.6871	75.30 \pm 1.1785	75.06 \pm 1.4206	75.36 \pm 1.3786	75.34 \pm 1.0793	75.38 \pm 1.1448	75.42 \pm 1.0560	75.52 \pm 1.1933	75.14 \pm 1.2963	75.92 \pm 1.8743
Wdbc	93.85 \pm 2.2187	92.45 \pm 4.2967	92.28 \pm 4.6195	92.10 \pm 3.8933	91.39 \pm 4.4034	92.45 \pm 3.0935	92.10 \pm 4.1484	92.10 \pm 4.1484	92.63 \pm 2.9498	92.63 \pm 2.3013
Sonar	76.52 \pm 10.5045	72.14 \pm 12.7874	73.11 \pm 10.9974	73.10 \pm 7.4628	71.21 \pm 6.1028	66.81 \pm 12.4212	68.64 \pm 11.3877	66.26 \pm 11.0109	70.62 \pm 7.5394	74.05 \pm 10.2991
Segment	95.80 \pm 1.5945	95.37 \pm 1.4864	95.37 \pm 1.3233	95.50 \pm 1.2107	95.32 \pm 1.7884	95.50 \pm 1.6094	95.45 \pm 1.4321	95.41 \pm 1.4602	95.24 \pm 1.4138	95.93 \pm 1.6853
Soybean	80.15 \pm 7.2028	78.84 \pm 8.3865	78.51 \pm 6.4938	78.20 \pm 4.4750	76.18 \pm 8.7576	76.26 \pm 7.3847	79.78 \pm 5.7547	80.46 \pm 6.6059	79.16 \pm 5.3126	78.19 \pm 6.9478
Zoo	93.00 \pm 8.2327	92.00 \pm 9.1894	92.00 \pm 7.8881	92.00 \pm 7.8881	92.00 \pm 21.4189	93.00 \pm 6.7495	87.20 \pm 13.3749	91.00 \pm 8.7560	92.00 \pm 7.8881	94.00 \pm 6.9921
NPHA	55.74 \pm 4.0731	54.49 \pm 5.3709	55.61 \pm 5.0741	55.18 \pm 3.2326	49.16 \pm 5.2298	53.78 \pm 4.0041	49.30 \pm 6.9566	49.30 \pm 7.5917	49.44 \pm 7.2647	53.22 \pm 7.1895
HCV	92.52 \pm 3.1799	91.70 \pm 3.9563	89.74 \pm 6.3365	91.38 \pm 3.9863	90.07 \pm 4.9574	89.91 \pm 5.1577	90.88 \pm 4.8323	90.23 \pm 5.3574	90.40 \pm 4.8917	90.72 \pm 4.3562
German	70.40 \pm 4.2374	72.40 \pm 3.7476	70.40 \pm 4.2216	72.10 \pm 3.4464	69.90 \pm 3.2813	63.40 \pm 6.0037	67.60 \pm 3.1340	68.60 \pm 4.2999	68.80 \pm 3.9665	71.90 \pm 5.0211
Abalone	99.23 \pm 0.3349	99.19 \pm 0.3603	98.99 \pm 0.3703	99.19 \pm 0.3603	98.83 \pm 0.4978	99.14 \pm 0.3942	99.19 \pm 0.3603	98.83 \pm 0.4978	98.83 \pm 0.5230	99.23 \pm 0.3349
Autos	79.83 \pm 8.3569	75.42 \pm 10.1493	71.00 \pm 9.2384	76.04 \pm 7.9331	71.63 \pm 8.8004	72.25 \pm 10.5869	71.63 \pm 10.5910	70.38 \pm 7.5932	72.88 \pm 10.1217	72.88 \pm 11.3353
Contraceptive	52.61 \pm 3.3125	51.19 \pm 2.7818	51.19 \pm 2.7745	51.26 \pm 3.0155	43.1 \pm 4.1843	51.26 \pm 2.8773	51.33 \pm 2.8330	51.26 \pm 2.7596	51.59 \pm 3.0334	
Colon	88.57 \pm 15.7455	82.95 \pm 13.4312	87.14 \pm 12.4074	82.38 \pm 16.1951	83.80 \pm 13.0313	53.57 \pm 23.8161	79.05 \pm 15.5847	83.81 \pm 7.9206	62.86 \pm 7.6290	84.04 \pm 13.0434
Lung	93.55 \pm 5.3908	89.55 \pm 8.9928	94.05 \pm 5.1752	90.14 \pm 5.2263	81.69 \pm 10.3397	55.57 \pm 9.7759	81.69 \pm 9.2306	–	88.59 \pm 7.4961	
LYMPHOMA	97.14 \pm 9.1429	92.85 \pm 12.1405	–	–	88.33 \pm 9.1944	89.24 \pm 9.3412	95.58 \pm 9.9412	94.05 \pm 10.2409	–	97.14 \pm 9.0351
ALLAML	91.96 \pm 11.3330	89.11 \pm 13.9900	–	–	90.36 \pm 11.6399	93.04 \pm 9.9762	88.93 \pm 12.6493	87.68 \pm 12.0365	–	–
Avg.	82.37 \pm 5.8067	79.97 \pm 6.7092	79.20 \pm 5.4331	78.88 \pm 4.9562	77.13 \pm 6.6169	74.97 \pm 6.7770	76.79 \pm 6.6646	78.24 \pm 6.0329	75.70 \pm 4.9005	79.79 \pm 5.8513

Table 8Average classification accuracy (mean \pm std. dev. %) of different feature selection algorithms on the SVM.

Data set	SNCMI	mRMR	KNCMI	NCMI_IFS	NRS	SNRS	WAFS	NFRS	CFS	FIRDL
Glass	58.42 \pm 7.3984	57.49 \pm 7.0132	57.51 \pm 7.5534	57.49 \pm 7.0132	57.03 \pm 7.6691	57.49 \pm 7.0132	57.49 \pm 7.0132	58.40 \pm 6.7835	57.49 \pm 7.0132	58.42 \pm 7.7025
Wine	99.44 \pm 1.7568	98.89 \pm 3.5136	98.89 \pm 3.5136	98.89 \pm 3.5136	98.30 \pm 3.7886	98.89 \pm 3.5136				
Drug	51.77 \pm 3.2212									
Rice	92.94 \pm 1.1166	92.47 \pm 1.3725	92.94 \pm 1.1166	92.91 \pm 1.1340	92.94 \pm 1.1166	92.94 \pm 1.1166				
Waveform	87.00 \pm 1.5916	86.98 \pm 1.6233								
Wdbc	97.89 \pm 1.3839	97.89 \pm 1.3839	97.72 \pm 1.6644	97.89 \pm 1.6122	97.19 \pm 1.8859	97.72 \pm 1.6644	97.72 \pm 1.6644	97.54 \pm 1.4794	97.02 \pm 1.4443	97.72 \pm 1.4443
Sonar	79.86 \pm 7.9557	79.33 \pm 9.2730	78.38 \pm 5.1113	79.81 \pm 8.0483	76.95 \pm 10.1920	73.12 \pm 9.5175	76.93 \pm 7.9238	75.93 \pm 8.0738	73.10 \pm 5.0458	77.40 \pm 9.5307
Segment	93.12 \pm 1.2654	93.12 \pm 1.2654	93.12 \pm 1.2654	93.12 \pm 1.2654	93.11 \pm 1.2654	93.11 \pm 1.2654	93.03 \pm 1.3296	93.03 \pm 1.3606	93.12 \pm 1.2654	93.12 \pm 1.2654
Soybean	91.57 \pm 5.4946	89.56 \pm 6.7006	89.24 \pm 6.9407	89.07 \pm 3.4253	73.90 \pm 13.2797	89.24 \pm 6.9407	87.61 \pm 7.3490	88.24 \pm 6.6505	91.50 \pm 5.6667	
Zoo	96.00 \pm 5.1640	93.00 \pm 6.7495	94.00 \pm 5.1640	94.00 \pm 5.1640	69.10 \pm 21.4189	92.00 \pm 6.3246	94.00 \pm 5.1640	93.00 \pm 4.8305	92.00 \pm 6.3246	94.00 \pm 5.1640
NPHA	55.19 \pm 3.5682	55.33 \pm 6.2562	55.05 \pm 5.1369	55.47 \pm 4.7568	54.76 \pm 3.7797	53.50 \pm 4.0983	52.24 \pm 4.1614	52.24 \pm 4.1614	55.59 \pm 4.3146	
HCV	90.07 \pm 5.3093	90.07 \pm 5.3093	90.23 \pm 5.2329	89.91 \pm 5.6045	89.91 \pm 5.6045	89.91 \pm 5.6045	88.93 \pm 5.8202	89.58 \pm 5.5153	89.91 \pm 5.6045	89.91 \pm 5.2274
German	74.50 \pm 4.0346	74.60 \pm 4.1952	74.70 \pm 4.1647	74.70 \pm 3.5606	74.40 \pm 4.0879	71.30 \pm 4.6679	70.50 \pm 3.8370	74.50 \pm 4.0346	74.50 \pm 3.8944	
Abalone	99.23 \pm 0.3349									
Autos	87.15 \pm 16.5672	85.71 \pm 13.9737	88.81 \pm 12.9465	85.42 \pm 17.1785	78.57 \pm 15.9126	64.29 \pm 12.9441	82.14 \pm 12.4099	84.05 \pm 13.0434	80.48 \pm 15.4630	85.24 \pm 12.2849
Lung	96.05 \pm 3.9319	94.53 \pm 3.5355	97.52 \pm 3.5176	94.55 \pm 3.6979	86.24 \pm 8.9593	68.45 \pm 12.3209	68.54 \pm 12.3209	95.59 \pm 6.4026	–	96.05 \pm 3.1471
LYMPHOMA	100.00 \pm 0.0000	100.00 \pm 0.0000	–	–	89.76 \pm 9.7880	93.54 \pm 10.1032	100.00 \pm 0.0000	94.28 \pm 9.9887	–	100.00 \pm 0.0000
ALLAML	98.75 \pm 3.9528	98.75 \pm 3.9528	–	–	87.68 \pm 12.0365	79.64 \pm 15.0632	98.75 \pm 3.9528	94.64 \pm 9.1054	–	–
Avg.	83.32 \pm 4.7825	82.69 \pm 4.9648	81.14 \pm 5.0043	80.98 \pm 5.0461	78.42 \pm 7.2926	78.05 \pm 6.3278	80.57 \pm 5.0148	81.71 \pm 5.6326	78.98 \pm 5.1007	82.02 \pm 4.7179

Table 9Average classification accuracy (mean \pm std. dev. %) of different feature selection algorithms on the KNN.

Data set	SNCMI	mRMR	KNCMI	NCMI_IFS	NRS	SNRS	WAFS	NFRS	CFS	FIRDL
Glass	73.20 \pm 10.0093	69.50 \pm 9.4857	72.34 \pm 6.8240	69.98 \pm 9.4857	70.41 \pm 9.2948	69.50 \pm 9.4857	69.05 \pm 9.7555	69		

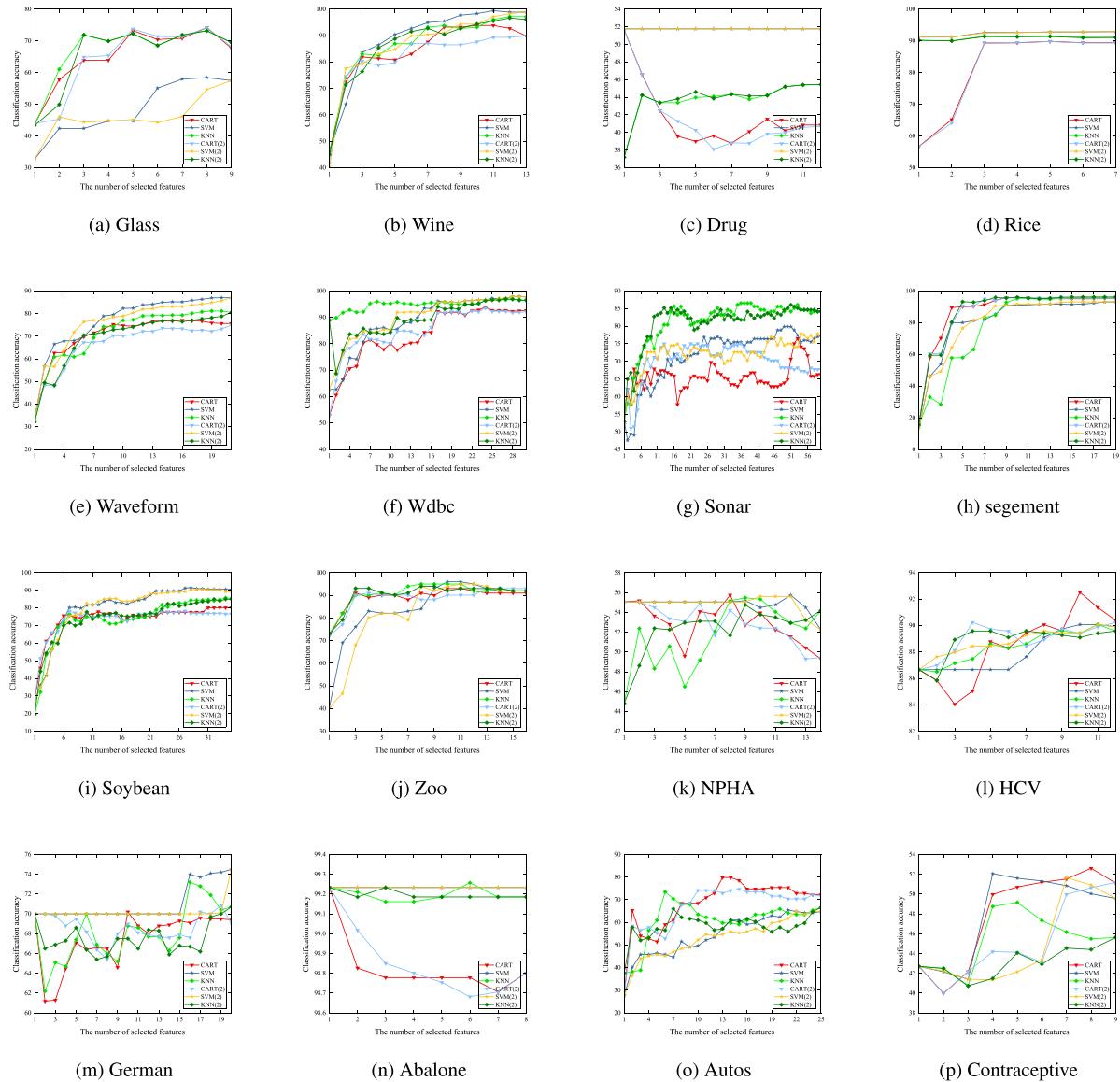


Fig. 4. Relationship between classification accuracy and the number of selected features on sixteen datasets.

Observing Figs. 4 and 5, it can be seen that the highest average classification accuracy achieved by the SNCMI algorithm is higher than that of SNCMI(2) on most datasets, indicating that the synergies considered in this paper are beneficial for feature selection. Notably, on the Sonar and ALLML datasets, the highest classification accuracy of SNCMI(2) under the CART classifier is lower than that of SNCMI. However, the average classification accuracy during the intermediate stages of feature selection is higher for SNCMI(2). On the Rice dataset, the performance of the two algorithms is similar. Both methods effectively select features, as adding more features beyond a certain threshold does not improve classification accuracy, but considering feature synergy enhances the performance better.

Further, to assess the impact of the two parameters discussed in this paper on average classification accuracy, we conducted a parameter sensitivity analysis based on the classification results obtained from the CART classifier, as illustrated in Fig. 6. In this heatmap, the x -axis represents K , the y -axis denotes α , and the z -axis indicates classification accuracy. The heatmap visualizes how varying values of these two parameters influence classification performance. In this case, the choice of parameter K is critical, as it determines the number of domains considered when calculating sample similarity. This typically depends

on data density and distribution. For low-dimensional datasets, selecting a smaller value for K allows for a more detailed capture of local features, whereas for high-dimensional datasets, a larger value for K aids in identifying global trends. Parameter α regulates the degree of penalization applied to the neighborhood radius. Smaller values permit greater variability in neighborhood delineation by allowing more significant differences in similarity measures, whereas larger values impose stricter limitations on neighborhood size. Consequently, for most datasets examined herein, the specified ranges for both parameters provide sufficient diversity and are applicable to common analytical challenges encountered in various contexts.

By observing Fig. 6, it can be seen that when the parameter K is fixed, changes in the value of α do not necessarily affect the classification accuracy. For example, on the four datasets Glass, Rice, Waveform, and Soybean, the classification accuracy remains unchanged despite variations in α under the same K value, indicating that the algorithm is not highly dependent on precise parameter tuning. Similarly, on the Drug, NPHA, German, Abalone, and Contraceptive datasets, changes in the parameters have no impact on classification accuracy. For other datasets, the fluctuations in classification accuracy due to parameter

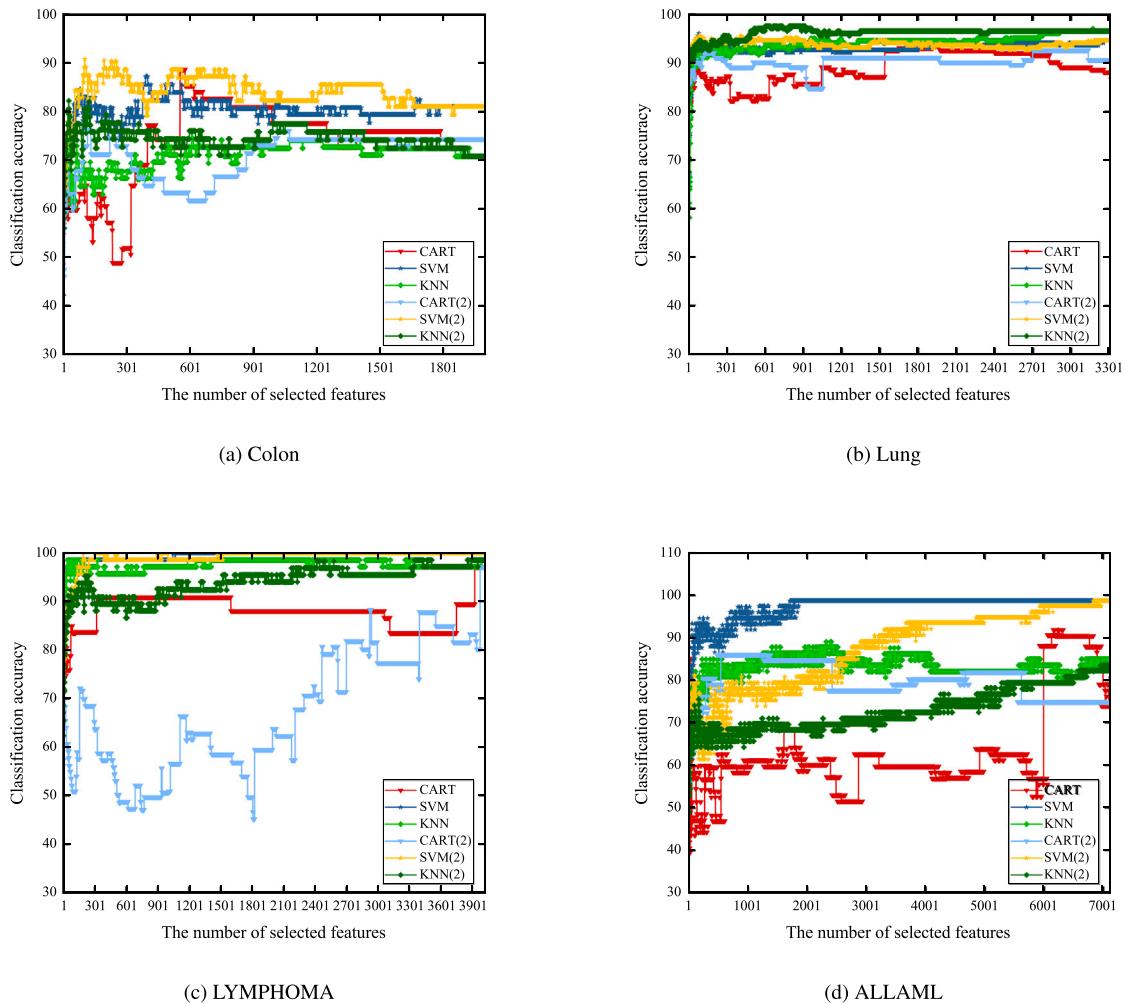


Fig. 5. Relationship between classification accuracy and the number of selected features on four high-dimensional datasets.

changes are minimal, further demonstrating the robustness of the algorithm proposed in this paper. Therefore, when selecting parameters for the algorithm, the distribution of the samples should be taken into account. If the dataset has high dimensionality, it is preferable to use a larger K value and a smaller α value. Conversely, if the dataset is sparse, a smaller K value and a larger α value are recommended.

The advantage of SNCMI algorithm can likely be attributed to its ability to adaptively determine an appropriate neighborhood radius for each sample. This adaptive approach helps mitigate the impact of a fixed neighborhood radius on the uncertainty measure, ensuring more accurate and reliable feature selection. Relying on a single fixed radius is often insufficient, as it can introduce classification bias, particularly in cases where sample distributions vary widely. Similarly, using a fixed value for K -nearest neighbors may not be adequate, as it can lead to suboptimal classification performance, especially for samples that are centrally distributed. By dynamically adjusting these parameters, SNCMI is able to better capture the local structure of the data, leading to more effective feature selection and improved classification accuracy across diverse datasets.

5.4. Statistical significance test

To further evaluate the statistical significance of the SNCMI algorithm, a Friedman test was conducted, comparing the performance of ten algorithms across twenty datasets. Before the Friedman test, the test performance of each algorithm under the three classifiers is ranked and assigned values 1, 2, ..., if the performance is same then

score the ordinal values. Then the average ordinal value under the three classifiers is obtained as shown in Table 10. If the algorithms have the same performance, the average ordinal values are the same. The calculation formula used in the Friedman test and Nemenyi's subsequent test in this paper is as follows.

$$T_{\chi^2} = \frac{12N}{h(h+1)} \left(\sum_{i=1}^h r_i^2 - \frac{h(h+1)^2}{4} \right) \quad (34)$$

$$\tau_F = \frac{(N-1)T_{\chi^2}}{N(h-1) - T_{\chi^2}} \quad (35)$$

$$CD = q_\alpha \sqrt{\frac{h(h+1)}{6N}} \quad (36)$$

According to Eqs. (34) and (35), the value of T_{χ^2} and τ_F with different classifiers are shown in Table 11. It can be seen that the τ_F value on each classifier is greater than the critical value 1.935. That is, the assumption that all algorithms have comparable classification performance was rejected, indicating that there are significant performance differences between the algorithms.

Subsequently, a Nemenyi post-hoc test was conducted to compare the algorithms pairwise. Using the critical difference (CD) hypothesis, two algorithms were considered to be significantly different in terms of mean ranking if the difference in mean ranking between the two algorithms met or exceeded the critical difference (CD). If the difference reaches or exceeds the CD value, the two algorithms are considered to be significantly different in terms of their mean rankings. For this study, the significance level was set to 0.05, which resulted in a CD value

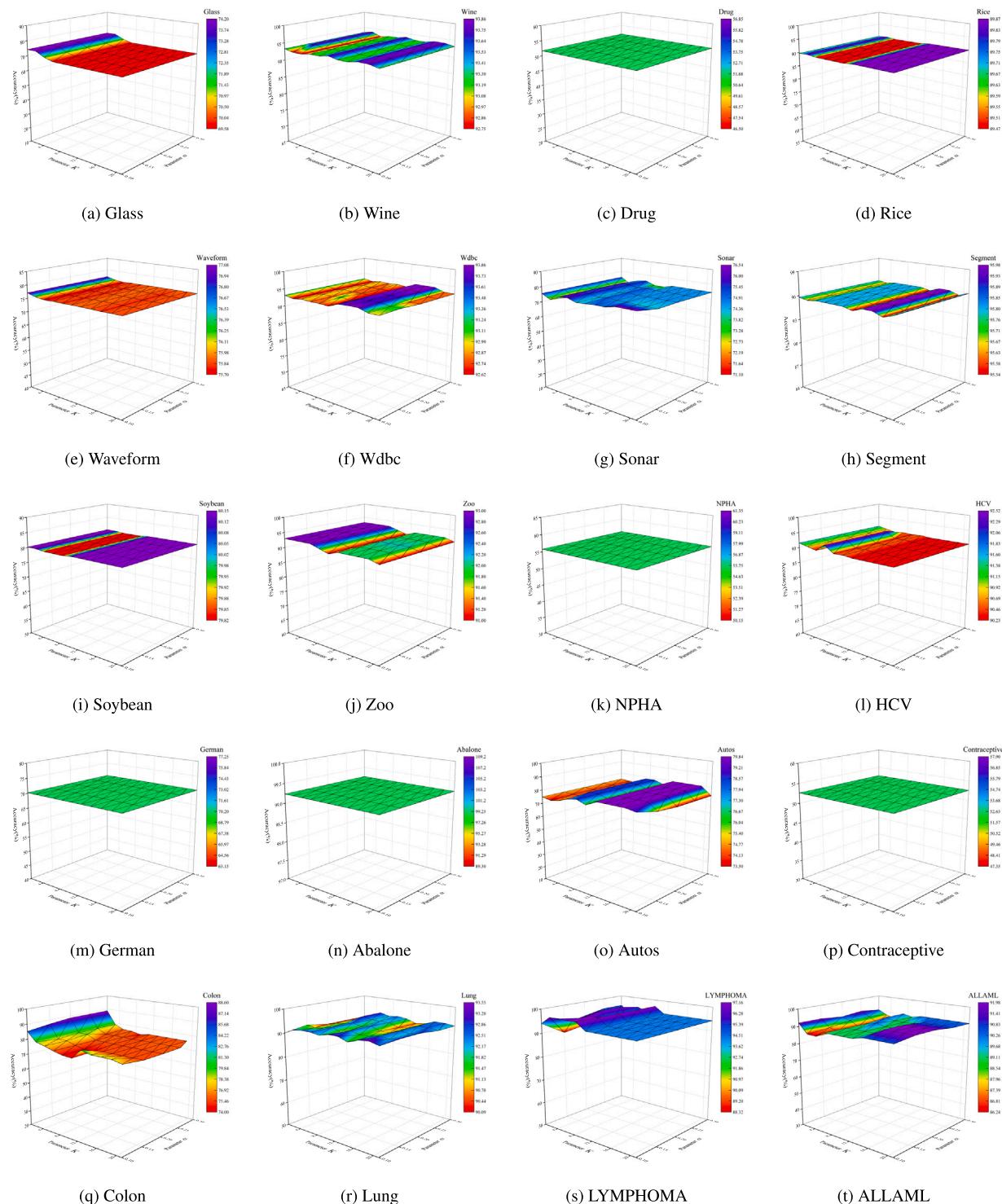
Fig. 6. Relationship between classification accuracy and parameters K and α on twenty datasets.

Table 10

The average ranks of ten algorithms over twenty datasets with three classifiers.

Classifiers	mRMR	KNMCI	NCMI_IFS	NRS	SNRS	WAFS	NFRS	CFS	FIRDL	SNCMI
CART	4.73	5.48	5.03	7.18	6.85	6.48	6.80	7.33	3.58	1.58
SVM	4.20	4.65	4.65	7.48	7.20	6.60	6.43	6.90	4.33	2.58
KNN	4.10	4.60	3.95	7.30	7.48	6.48	6.53	7.60	4.73	2.25

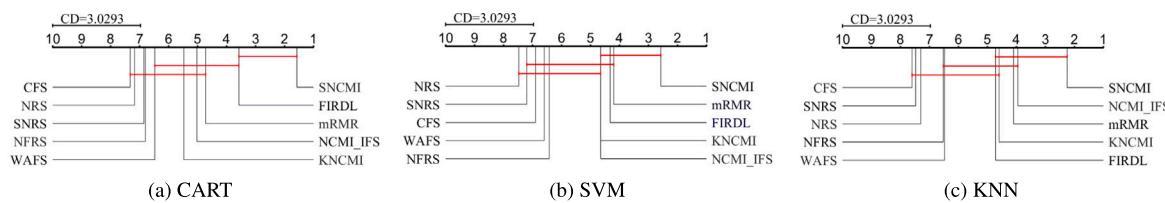


Fig. 7. CD diagrams by CART, KNN and SVM.

Table 11
The values of T_{χ^2} and τ_F with different classifiers.

Classifiers	T_{χ^2}	τ_F	Critical value ($\alpha = 0.05$)
CART	67.5299	11.4081	1.935
SVM	52.5722	7.8387	1.935
KNN	65.7591	10.9367	1.935

of 3.0293. To visually represent the relative ranking performance of the algorithms, a CD plot is presented in Fig. 7. This plot highlights the pairwise differences in algorithm rankings, offering an intuitive understanding of which algorithms outperform others and how SNCMI compares to its competitors in classification effectiveness.

The axes in Fig. 7 depict the average ranking of algorithms. A red line connects any algorithm whose average ranking is within one critical difference (CD) of SNCMI. Notably, SNCMI consistently ranks first, clearly demonstrating its superior effectiveness compared to the other algorithms. For the CART classifier, SNCMI statistically outperforms algorithms such as CFS, NRS, SNRS, NFRS, WAFS, KNCMI, NCMI_IFS and mRMR, while performing comparably to FIRDL. This highlights SNCMI's strength in feature selection for decision tree-based classification. For SVM classification, SNCMI maintains its advantage by significantly outperforming NRS, SNRS, CFS, WAFS, and NFRS, and achieving similar results to NCMI_IFS, KNCMI, FIRDL, and mRMR. For KNN classification, SNCMI demonstrates statistical superiority over CFS, SNRS, NRS, NFRS, and WAFS, while remaining competitive with FIRDL, KNCMI, mRMR, and NCMI_IFS. These results underscore the versatility and robustness of SNCMI across different classifiers, as it consistently delivers strong performance.

6. Conclusion and future work

The neighborhood radius of adaptive samples greatly affects the effectiveness and robustness of algorithms based on the neighborhood rough set theory. Furthermore, considering multiple correlations among features aids in retaining useful features and eliminating redundant ones. Based on these two crucial elements, this paper proposes SNCMI, a novel algorithm that takes into account the various correlations of features, specifically correlation, redundancy, complementarity, and synergy, as defined under the uncertainty measure. Additionally, the results of parameter sensitivity analyses confirm the good robustness of the proposed algorithm. However, by adapting the neighborhood radius for each sample and considering feature interactions, the algorithm's computational burden increases to some extent. Hence, future work will focus on optimizing the method's computational efficiency while maintaining consideration of feature interactions.

CRediT authorship contribution statement

Lubin Chen: Writing – original draft, Investigation. **Jinkun Chen:** Writing – review & editing, Supervision, Methodology. **Yaojin Lin:** Writing – review & editing, Software.

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 National Natural Science Foundation of China (No. 62076116), Fujian Provincial Natural Science Foundation of China (Nos. 2024J01799 and 2021J02049) and Fujian Alliance of Mathematics (2025SXLMMS05).

Data availability

The authors do not have permission to share data.

References

- An, S., Guo, X., Wang, C., Guo, G., Dai, J., 2023. A soft neighborhood rough set model and its applications. *Inform. Sci.* 624, 185–199.
- Bennasar, M., Setchi, R., Hicks, Y., 2013. Feature interaction maximisation. *Pattern Recogn. Lett.* 34 (14), 1630–1635.
- Cekik, R., Uysal, A.K., 2020. A novel filter feature selection method using rough set for short text data. *Expert Syst. Appl.* 160, 113691.
- Chen, Z., Wu, C., Zhang, Y., Huang, Z., Ran, B., Zhong, M., Lyu, N., 2015. Feature selection with redundancy-complementariness dispersion. *Knowl.-Based Syst.* 89, 203–217.
- Chen, Y., Xue, Y., Ma, Y., Xu, F., 2017. Measures of uncertainty for neighborhood rough sets. *Knowl.-Based Syst.* 120, 226–235.
- Chen, J., Zhu, P., 2024. Feature selection of dominance-based neighborhood rough set approach for processing hybrid ordered data. *Internat. J. Approx. Reason.* 167, 109134.
- Cheng, H., Qin, Z., Feng, C., Wang, Y., Li, F., 2011. Conditional mutual information-based feature selection analyzing for synergy and redundancy. *ETRI J.* 33 (2), 210–218.
- Ding, W., Lin, C., Witold, P., 2020. Multiple relevant feature ensemble selection based on multilayer co-evolutionary consensus mapreduce. *IEEE Trans. Cybern.* 50 (2), 425–439.
- Dubois, D., Prade, H., 1990. Rough fuzzy sets and fuzzy rough sets. *Int. J. Gen. Syst.* 17 (2–3), 191–209.
- Forman, G., 2003. An extensive empirical study of feature selection metrics for text classification. *J. Mach. Learn. Res.* 3, 1289–1305.
- Gong, H., Li, Y., Zhang, J., Zhang, B., Wang, X., 2024. A new filter feature selection algorithm for classification task by ensembling pearson correlation coefficient and mutual information. *Eng. Appl. Artif. Intell.* 131, 107865.
- Hagar, J.D., Wissink, T.L., Kuhn, D.R., Kacker, R.N., 2015. Introducing combinatorial testing in a large organization. *Comput.* 48 (4), 64–72.
- Hastie, T., Tibshirani, R., Friedman, J., 2001. *The Element of Statistical Learning*. Springer, Berlin.
- Hu, Q., Pan, W., An, S., Ma, P., Wei, J., 2010. An efficient gene selection technique for cancer recognition based on neighborhood mutual information. *Int. J. Mach. Learn. Cyber.* 1, 63–74.
- Hu, M., Tsang, E.C., Guo, Y., Chen, D., Xu, W., 2021. A novel approach to attribute reduction based on weighted neighborhood rough sets. *Knowl.-Based Syst.* 220, 106908.
- Hu, Q., Yu, D., Liu, J., Wu, C., 2008. Neighborhood rough set based heterogeneous feature subset selection. *Inform. Sci.* 178 (18), 3577–3594.
- Ibrahim, R.A., Elaziz, M.A., oliva, D., Lu, S., 2020. An improved runner-root algorithm for solving feature selection problems based on rough sets and neighborhood rough sets. *Appl. Soft Comput.* 97, 105517.
- Jakulin, A., 2005. Machine learning based on attribute interactions. (Ph.D. thesis). University of Ljubljana.

- Jiang, J., Zhang, X., 2024. Feature selection based on self-information combining double-quantitative class weights and three-order approximation accuracies in neighborhood rough sets. *Inform. Sci.* 657, 119945.
- Lewis, D.D., 1992. Feature selection and feature extraction for text categorization. In: *Speech and Natural Language: Proceedings of a Workshop Held at Harriman, New York*, pp. 23–26.
- Li, W., Huang, Z., Jia, X., Cai, X., 2016. Neighborhood based decision-theoretic rough set models. *Internat. J. Approx. Reason.* 69, 1–17.
- Li, C., Luo, X., Qi, Y., Gao, Z., Lin, X., 2020. A new feature selection algorithm based on relevance, redundancy and complementarity. *Comput. Biol. Med.* 119, 103667.
- Lin, T.Y., 1988. Neighborhood systems and approximation in relational databases and knowledge bases. In: *Proceedings of the 4th International Symposium on Methodologies of Intelligent Systems*. pp. 75–86.
- Lin, X., Li, C., Ren, W., Luo, X., Qi, Y., 2019. A new feature selection method based on symmetrical uncertainty and interaction gain. *Comput. Biol. Chem.* 83, 107149.
- Liu, D., Liang, D., 2014. Incremental learning researches on rough set theory: Status and future. *IJRSDA* 1 (1), 99–112.
- Liu, Y., Xie, H., Chen, Y., Tan, K., Wang, L., Xie, W., 2016. Neighborhood mutual information and its application on hyperspectral band selection for classification. *Chemom. Intell. Lab. Syst.* 157, 140–151.
- Ma, X., Ju, C., 2022. Fuzzy information-theoretic feature selection via relevance, redundancy, and complementarity criteria. *Inform. Sci.* 611, 564–590.
- Ma, X., Xu, H., Liu, Y., Zhang, J., 2024. Class-specific feature selection using fuzzy information-theoretic metrics. *Eng. Appl. Artif. Intell.* 136, 109035.
- Moran, M., Gordon, G., 2019. Curious feature selection. *Inform. Sci.* 485, 42–54.
- Nakariyakul, S., 2018. High-dimensional hybrid feature selection using interaction information-guided search. *Knowl.-Based Syst.* 145, 59–66.
- Nandhini, N., Thangadurai, K., 2023. An incremental rough set approach for faster attribute reduction. *Int. J. Inf. Technol.* 15, 1–15.
- Pawlak, Z., 1982. Rough sets. *Int. J. Comput. Inf. Sci.* 11, 341–356.
- Peng, H., Long, F., Ding, C., 2005. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (8), 1226–1238.
- Raza, M.S., Qamar, U., 2017. Redefining core preliminary concepts of classic Rough Set Theory for feature selection. *Eng. Appl. Artif. Intell.* 65, 375–387.
- Shu, T., Lin, Y., Guo, L., 2024a. Online hierarchical streaming feature selection based on adaptive neighborhood rough set. *Appl. Soft Comput.* 152, 111276.
- Shu, W., Xia, Q., Qian, W., 2024b. Neighborhood multigranulation rough sets for cost-sensitive feature selection on hybrid data. *Neurocomputing* 565, 126990.
- Song, K., Han, J., Cheng, G., Lu, J., Nie, F., 2022. Adaptive neighborhood metric learning. *IEEE Trans. Pattern Anal. Mach. Intell.* 44 (9), 4591–4604.
- Sosa-Cabrera, G., Gómez-Guerrero, S., García-Torres, M., E-Schaerer, C., 2024. Feature selection: a perspective on inter-attribute cooperation. *Int. J. Data Sci. Anal.* 17, 139–151.
- Sulaiman, M.A., Labadin, J., 2015. Feature selection based on mutual information. In: *2015 9th International Conference on IT in Asia. CITA*, pp. 1–6.
- Sun, L., Yin, T., Ding, W., Qian, Y., Xu, J., 2020. Multilabel feature selection using ML-Relieff and neighborhood mutual information for multilabel neighborhood decision systems. *Inform. Sci.* 537, 401–424.
- Sun, L., Zhang, X., Qian, Y., Xu, J., Zhang, S., 2019. Feature selection using neighborhood entropy-based uncertainty measures for gene expression data classification. *Inform. Sci.* 502, 18–41.
- Tang, X., Dai, Y., Xiang, Y., 2019. Feature selection based on feature interactions with application to text categorization. *Expert Syst. Appl.* 120, 207–216.
- Taşkin, G., Kaya, H., Bruzzone, L., 2017. Feature selection based on high dimensional model representation for hyperspectral images. *IEEE Trans. Image Process.* 26 (6), 2918–2928.
- Velayutham, C., Thangavel, K., 2011. Unsupervised quick reduct algorithm using rough set theory. *J. Electron. Sci. Technol.* 9 (3), 193–201.
- Wan, J., Chen, H., Li, T., Yuan, Z., Liu, J., Huang, W., 2023. Interactive and complementary feature selection via fuzzy multigranularity uncertainty measures. *IEEE Trans. Cybern.* 53 (2), 1208–1221.
- Wan, J., Chen, H., Yuan, Z., Li, T., Yang, X., Sang, B., 2021. A novel hybrid feature selection method considering feature interaction in neighborhood rough set. *Knowl.-Based Syst.* 227, 107167.
- Wang, W., Guo, M., Han, T., Ning, S., 2023. A novel feature selection method considering feature interaction in neighborhood rough set. *Intell. Data Anal.* 27 (2), 45–59.
- Wang, C., Qi, Y., Shao, M., Hu, Q., Chen, D., Qian, Y., Lin, Y., 2017. A fitting model for feature selection with fuzzy rough sets. *IEEE Trans. Fuzzy Syst.* 25 (4), 741–753.
- Wang, Q., Qian, Y., Liang, X., Guo, Q., Liang, J., 2018. Local neighborhood rough set. *Knowl.-Based Syst.* 153, 53–64.
- Wang, C., Shao, M., He, Q., Qian, Y., Qi, Y., 2016. Feature subset selection based on fuzzy neighborhood rough sets. *Knowl.-Based Syst.* 111, 173–179.
- Wang, C., Shi, Y., Fan, X., Shao, M., 2019. Attribute reduction based on k-nearest neighborhood rough sets. *Internat. J. Approx. Reason.* 106, 18–31.
- Wang, C., Wang, C., Qian, Y., Leng, Q., 2024. Feature selection based on weighted fuzzy rough sets. *IEEE Trans. Fuzzy Syst.* 32 (7), 4027–4037.
- Wang, N., Zhao, E., 2024. A new method for feature selection based on weighted k-nearest neighborhood rough set. *Expert Syst. Appl.* 238, 122324.
- Wojtas, M., Chen, K., 2020. Feature importance ranking for deep learning. *Adv. NeurIPS.* 33, 5105–5114.
- Xu, W., Liu, C., 2024. Incremental reduction of imbalanced distributed mixed data based on k-nearest neighbor rough set. *Internat. J. Approx. Reason.* 172, 109218.
- Xu, W., Yuan, Z., Liu, Z., 2024. Feature selection for unbalanced distribution hybrid data based on k-nearest neighborhood rough set. *IEEE Trans. Artif. Intell.* 5 (1), 229–243.
- Yang, X., Liang, S., Yu, H., Guo, S., Qian, Y., 2019. Pseudo-label neighborhood rough set: measures and attribute reductions. *Internat. J. Approx. Reason.* 105, 112–129.
- Yang, J., Ong, C., 2012. An effective feature selection method via mutual information estimation. *IEEE Trans. Syst. Man Cybern. Part B (Cybern.)* 42 (6), 1550–1559.
- Yao, P., Lu, Y., 2011. Neighborhood rough set and SVM based hybrid credit scoring classifier. *Expert Syst. Appl.* 38 (9), 11300–11304.
- Yu, Y., Pedrycz, W., Miao, D., 2013. Neighborhood rough sets based multi-label classification for automatic image annotation. *Internat. J. Approx. Reason.* 54 (9), 1373–1387.
- Yu, W., Zhang, M., Yi, S., 2019. Learning a local manifold representation based on improved neighborhood rough set and LLE for hyperspectral dimensionality reduction. *Signal Process.* 164, 20–29.
- Yuan, Z., Chen, H., Li, T., Yu, Z., Sang, B., Luo, C., 2021. Unsupervised attribute reduction for mixed data based on fuzzy rough sets. *Inform. Sci.* 572, 67–87.
- Yuan, G., Lu, L., Zhou, X., 2023. Feature selection using a sinusoidal sequence combined with mutual information. *Eng. Appl. Artif. Intell.* 126, 107168.
- Zare, H., Haffari, G., Gupta, A., Brinkman, R., 2013. Scoring relevancy of features based on combinatorial analysis of Lasso with application to lymphoma diagnosis. *BMC Genomics* 14 (1), 1471–2164.
- Zhang, Y., Li, S., Wang, T., Zhang, Z., 2013. Divergence-based feature selection for separate classes. *Neurocomputing* 101, 32–42.
- Zhou, P., Hu, X., Li, P., Wu, X., 2019. Online streaming feature selection using adapted neighborhood rough set. *Inform. Sci.* 481, 258–279.