Information Sciences 715 (2025) 122220

Contents lists available at ScienceDirect

Information Sciences

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

Multi-label feature selection for imbalanced data via KNN-based multi-label rough set theory

Weihua Xu *, Yuzhe Li

College of Artificial Intelligence, Southwest University, Chongqing, 400715, PR China

ARTICLE INFO

Keywords:
Multi-label feature selection
k-nearest neighborhood
Neighborhood rough set
Imbalanced data

ABSTRACT

In the realm of multi-label feature selection, the intricacy of data structures and semantics has been escalating, rendering traditional single-label feature selection methodologies inadequate for contemporary demands to meet contemporary demands. This manuscript introduces an innovative neighborhood rough set model that integrates δ-neighborhood rough sets with k-nearest neighbor techniques, facilitating a transition from single-label to multi-label learning frameworks. The study delves into the attribute dependency concept within rough set theory and proposes a novel importance function based thereon, which can effectively quantify the significance of features within multi-label decision-making contexts. Building on this theoretical foundation, we have crafted a feature selection algorithm specifically tailored for imbalanced datasets. Extensive experiments conducted on 12 datasets, coupled with comparative analyses with 10 cutting-edge methods, have substantiated the superior performance of our algorithm in managing imbalanced datasets. This research not only offers a fresh theoretical perspective but also has significant practical implications, particularly in scenarios involving imbalanced datasets with multiple labels.

1. Introduction

Currently, the preponderance of research is centered on single-label problems, where each data sample is linked to a solitary class label. However, in multi-label learning [1], each sample may be associated with multiple labels simultaneously, adding complexity to the problem. Multi-label learning needs to consider not only the relationship between features and individual labels but also capture the interactions between features and multiple labels [2]. Moreover, This complexity is further exacerbated in the presence of data imbalance.

In traditional single-label learning, feature selection methods typically assume a uniform distribution of samples across the feature space. However, in multi-label learning, sample distribution may be highly imbalanced, with some regions being overly dense while others are sparse. This imbalance is further exacerbated by the coexistence of multiple labels, which increases the complexity of the feature space. Traditional single-label methods perform poorly in multi-label scenarios primarily because global feature importance-based selection may ignore local features that are crucial for multi-label classification. Moreover, the imbalanced feature distribution makes it difficult to select meaningful features in sparse regions. The coexistence of multiple labels complicates the feature-label interaction [3], which traditional methods fail to capture effectively. Additionally, sparse regions may contain noisy features, which traditional methods struggle to distinguish from important features. To address these challenges, this paper proposes a multi-label

* Corresponding author.
E-mail addresses: chxuwh@gmail.com (W. Xu), liyuzhe503503@gmail.com (Y. Li).

https://doi.org/10.1016/j.ins.2025.122220
Received 23 January 2025; Received in revised form 17 April 2025; Accepted 17 April 2025

Available online 23 April 2025

feature selection method based on neighborhood rough sets. By characterizing new neighborhood structures, the method captures key features in sparse regions and reduces noise interference, thereby improving classification accuracy. Additionally, by integrating label dependencies and neighborhood granularity optimization mechanisms, the method enhances the adaptability of feature selection for multi-label dataset.

Pawlak's rough set theory, proposed in 1982 [4], has been widely applied in various domains, including feature selection, owing to its advantages in addressing uncertainty and its ability to operate without external prior knowledge [5]. However, traditional rough set theory has limitations in handling diverse data types because it primarily relies on equivalence relations, thus making it suitable only for discrete data. In real-world applications, where most data are continuous, the data often needs to be discretized. However, discretization step may lead to information loss, thereby impacting the effectiveness of feature selection. To overcome this limitation, Lin introduced similarity relations [6] and the neighborhood rough set model [7], which extends and expands Pawlak's classical rough set theory. By incorporating distance and neighborhood radius, the neighborhood rough set model refines the definition domain of each sample and establishes interconnections among samples [8]. This approach comprehensively captures the data by considering sample proximity and neighborhood structure, overcoming the constraint of rough set theory's applicability to discrete data [9]. Hu conducted further research on the application of rough set theory in discovering knowledge from mixed data and proposed a more systematic neighborhood rough set model, providing strong theoretical support for the development of multi-label neighborhood rough set methods [10]. In recent years, the extended research of the neighborhood rough set model has received widespread attention, including neighborhood-based decision-theoretic rough sets [11], local neighborhood rough sets [12], fuzzy neighborhood rough sets [13], neighborhood multi-granulation rough sets [14], pseudo-labeled neighborhood rough sets [15], soft neighborhood rough sets [16], noise-resistant multi-label fuzzy neighborhood rough sets [17]. Researchers have also explored different ways to measure uncertainty [18], further broadening the applicability of rough set theory. For example, [19] used a Zentropy-based uncertainty measure for multi-granularity data analysis, demonstrating its value in robust feature selection. This work emphasizes the importance of considering multiple perspectives in data. Following this, [20] developed Ze-HFS, a Zentropy-based method for heterogeneous feature selection and knowledge discovery, demonstrating its ability to handle diverse data types. These studies highlight the potential of Zentropy for feature selection. Additionally, [21] further explored Zentropy-based uncertainty measurement for feature selection. These investigations contribute to the expanding research on uncertainty measures in feature selection. These studies greatly enrich and improve the theoretical framework of neighborhood rough set theory [22]and provide a solid theoretical foundation and methodological support for solving practical problems in various domains [23].

Recent multi-label feature selection advances include diverse strategies. Ebrahimi et al. [24] used deep learning to integrate feature selection by extracting meta-label-specific features and capturing inter-label relationships. This data-driven approach demonstrates deep learning's effectiveness. Cai et al. [25] developed a fuzzy rough set-based algorithm with metric learning and label enhancement to capture label dependencies and optimizing distance measures. This knowledge-driven approach refines fuzzy rough set methodology, improving accuracy and efficiency. These studies highlight the potential of both data-driven (deep learning [26]) and knowledge-driven (fuzzy rough sets) paradigms.

In real-world multi-label learning scenarios, the idealized assumption of a uniformly distributed sample space rarely holds true. Data often exhibits significant spatial and label imbalances, where some regions densely populated while others are sparsely populated, and some labels being significantly less frequent than others or appearing in rare combinations. This combined spatial and label imbalance poses substantial challenges for multi-label learning algorithms. Specifically, the spatial imbalance can lead to biased models that are overly influenced by dense regions, resulting in poor performance in sparse regions. Simultaneously, the label imbalance complicates learning due to the complex interplay of label dependencies and varying label frequencies. Traditional methods for addressing class imbalance, such as oversampling and undersampling, primarily focus on balancing the number of instances or labels independently and are not well-suited to handle the combined spatial and label imbalances inherent in multi-label data. These methods often fail to capture the complex relationships among spatial distribution, label co-occurrence, and feature dependencies. For example, oversampling rare label combinations in sparse regions might amplify noise, whereas undersampling frequent labels in dense regions might discard crucial information about the local data structure. Consequently, these methods may not effectively address the challenges posed by spatially and label-imbalanced multi-label data, thus motivating the need for more specialized techniques that explicitly consider both spatial context and label dependencies.

This non-uniformity in spatial distribution presents a challenge for traditional rough set theories such as δ-neighborhood rough sets [10] and k-nearest neighbor rough sets [27], as these methods may not perform well in classification tasks when dealing with imbalanced data. However, research findings indicate that when dealing with certain types of imbalanced data distributions, δ neighborhood rough sets and k-nearest neighborhood rough sets demonstrate their respective strengths. δ neighborhood rough sets perform exceptionally well in sparse data distributions because in sparse data, the distances between data points are large, and using a fixed distance δ value to define neighborhoods ensures that data points within each neighborhood are relatively independent, thereby reducing overlap between neighborhoods. This independence and discrimination are crucial for sparse data because it helps models to more accurately identify data boundaries, thereby enhancing classification accuracy.

In contrast, k-nearest neighborhood rough sets are more suitable for dense data distributions. In dense data, data points are close to each other, and defining neighborhoods based on a fixed number k of the nearest neighbors can better capture the local structure of the data. Since the data points are dense, even considering only the nearest k points can contain enough information to describe the characteristics of the neighborhood of a point. This neighborhood definition is more stable when dealing with dense data because it does not change dramatically with minor fluctuations in data points.

δ neighborhood rough set model is based on the neighborhood definition with a fixed radius, determining the neighborhood relationships between objects by setting a threshold (δ). This model is suitable for dense data distribution because in such datasets,

the distances between objects are usually small, and a fixed radius can effectively capture the similarity between objects. The δ neighborhood rough set model can simplify the data structure of large-scale datasets, reduce computational complexity, and maintain the local characteristics of objects in dense datasets, thus making it suitable for data mining and pattern recognition.

In short, for sparse data with the large intervals between points, δ neighborhood rough sets are more suitable for maintaining the clarity and discrimination of neighborhoods; for dense data, due to the small intervals between points, k-nearest neighborhood rough sets are more suitable for ensuring the richness and stability of information within neighborhoods. This choice reflects the adaptability and flexibility of rough set models under different data distribution characteristics. δ neighborhood rough sets and k-nearest neighborhood rough sets each exhibit unique advantages under different data distribution characteristics, and their features are complementary. Integrating these two models can create a new model that dynamically adjusts the definition of neighborhoods based on the local density and distribution characteristics of the data, thereby better adapting to different data distribution scenarios. This integrated model will combine the fixed-radius advantage of δ neighborhood rough sets and the dynamic neighborhood advantage of k-nearest neighborhood rough sets, providing a new, more flexible, and effective method for dealing with imbalanced datasets.

In an information system, there is often inconsistency between the information conveyed by features and labels, which reveals the limitations of feature descriptive power for labels and also reflects the uncertainty present in the data. To quantitatively analyze this uncertainty, the concept of positive region was introduced within the framework of rough set theory. However, traditional feature selection theories based on neighborhood rough sets primarily rely on positive regions. These theories assess the importance of features in the information system by observing variations in the positive region's scope through feature addition or removal. The positive region represents all deterministic rules in an abstract form, with the count of positive regions effectively denoting the collective count of deterministic rules, accounting for duplicated rules. Nevertheless, the efficacy of positive regions as a deterministic criterion is not significantly pronounced in algorithmic applications [28]. To address this limitation, this study introduces the notion of dependency function [29], which quantifies the proportion of deterministic rules within the comprehensive rule set, thereby numerically measuring the consistency between labels and feature sets. Building upon the theoretical foundations of positive regions and dependency, this research further advances a measure of feature importance and applies it to multi-label information systems. This measure employs a forward greedy algorithm to sequentially evaluate the relevance of individual features. Through this measure, we are able to quantitatively gauge the capacity of features to depict labels and employ it as a basis for feature selection. In conclusion, the introduction of the feature importance measure provides an effective strategy for conducting feature selection on multi-label datasets. This approach not only addresses the incongruity between features and labels but also furnishes a valuable framework for enhancing the performance and accuracy of information systems.

The practical significance of this integration lies in its ability to simultaneously capture key features in local neighborhoods and adapt to the distribution characteristics of sparse regions, thereby enhancing the robustness and accuracy of feature selection. In real-world applications, imbalanced data are ubiquitous. For instance, in text classification, the number of documents for some topics is significantly lower than others, leading to an uneven distribution in the feature space. In bioinformatics [30], gene expression data often exhibit high-dimensional sparsity, with sample distributions being highly imbalanced across different classes. The model proposed in this paper, by combining δ-neighborhood rough sets and k-nearest neighbor techniques, can effectively address these issues. In text classification, the model can identify key features of sparse topics to improve classification performance. In bioinformatics, it can select features related to minority classes from high-dimensional sparse data, aiding in disease diagnosis or gene function prediction. This integration not only enhances the theoretical value of the model but also provides strong support for its widespread application in real-world scenarios.

As a cognitive process, Concept Cognition Learning (CCL) emphasizes the structured expression of data and systematic cognitive understanding, aiming to extract valuable information from data [31]. Although CCL has achieved significant success in single-label classification tasks, its application in multi-label learning faces substantial challenges. The main difficulties lie in the imbalanced distribution of samples in the feature space and the failure of existing methods to effectively utilize the structural relationships between feature concepts and multi-label concepts. Key features in sparse regions are often overlooked, making it difficult for models to fully capture the complex structure of multi-label data. The method proposed in this paper addresses the issue of imbalanced sample distribution by integrating δ-neighborhood rough sets and k-nearest neighbor techniques, thereby indirectly alleviating the limitations of CCL in multi-label learning. The feature selection capability under imbalanced data provides potential support for the application of CCL in multi-label learning and lays the foundation for solving the problem of structural relationships between feature concepts and multi-label concepts.

Given the challenges posed by imbalanced data distribution in multi-label classification tasks, this study aims to propose a novel multi-label neighborhood rough set model and develop a dedicated feature selection algorithm based on this model to address this crucial problem. The main contributions of this paper are as follows:

1) A novel k-nearest neighbor multi-label rough set model is proposed, integrating the multi-label δ-neighborhood approach with the k-nearest neighbor algorithm to specifically address challenges associated with imbalanced multi-label data.

2) To enhance the descriptive capability of positive regions in handling uncertainty, the concept of dependency is introduced. Furthermore, a crucial feature importance evaluation function is developed for accurate characterization of features, demonstrating superior performance in selecting high-performing features for multi-label classification compared to certain existing techniques.

3) Building upon prior research, this paper develops an innovative multi-label feature selection algorithm tailored for imbalanced datasets. The proposed algorithm significantly improves adaptability and accuracy, particularly in scenarios with extreme class distribution imbalances.

4) The proposed multi-label feature selection algorithm was extensively compared with fourteen mainstream algorithms using twelve datasets from the *Mulan* repository. A comprehensive evaluation framework, comprising metrics for reduction capability, classification accuracy, and effectiveness, was designed to assess the algorithm.

5) Experimental investigations were conducted to analyze the impact of neighborhood granularity parameters δ and k on the algorithm's performance. The experimental results clearly indicate the excellent performance of the proposed algorithm in addressing multi-label feature selection problems.

This paper is organized as follows:In the second part of this paper, we provide an overview of the definitions of δ neighborhood rough set and k-nearest neighbor rough set. We then delve into the concepts of neighborhood rough set in the context of multi-label frameworks, with a detailed comparison to the single-label information system. The third part introduces an improved multi-label k-nearest neighbor rough set model that aims to address the issue of data imbalance. We further derive the construction of the approximate space for this new model and propose a novel measure of feature importance. In the fourth part, we present the design of a specialized multi-label feature selection algorithm called $KNMRS$. The fifth part showcases the experimental comparisons and performance results of the proposed algorithm with existing algorithms. Finally, the sixth part summarizes the research findings of this paper and provides an outlook on future research directions.

2. Related work and foundations

In this section, we will introduce the basic theory of δ neighborhood rough sets and k-nearest neighbor rough sets in detail. To overcome the inconvenience encountered by classical rough sets in processing numerical attribute datasets, we introduce the neighborhood rough set model. The core of this model lies in the different definitions of the concept of the neighborhood concept [13], from which a variety of distinctive neighborhood rough set models are derived.

2.1. Neighborhood rough set (NRS)

Neighborhood rough sets use the distance formulas to evaluate the similarity between different samples. The Euclidean distance is a classic metric used to assess the dissimilarity between samples. When dealing with two sets of random samples, each characterized by continuous real-valued attributes, the Euclidean distance between them can be determined using the following mathematical formula:

$$Dis(x_i, x_j) = \sqrt{\sum_{a \in C} (a_i - a_j)^2}.$$

Here, C is a subset of the conditional attribute set, and a is one of the conditional attributes. a_i corresponds to the specific value of x_i on attribute a.

We commonly use $NDS = (U, A, D, Dis, \delta)$ to describe the Neighborhood Decision System. In this system, U represents the set consisting of all sample, the set A corresponds to the conditional attributes, and the set D contains the decision attributes, which are crucial labels used for classification or prediction of the sample points. The distance metric, denoted as Dis, is calculated based on the relation R_C and is commonly computed using the Euclidean distance. Additionally, the neighborhood radius parameter δ determines the size of the neighborhood, thereby influencing the definition of the sample point's neighborhood and the accuracy of the decision-making process.

In the Neighborhood Decision System, we select a subset of features, denoted as C, and define the similarity relation on C as follows:

$$NR_\delta(C) = \{(x_i, x_j) \in U \times U \,|\, Dis_C(x_i, x_j) \leq \delta\}.$$

Accordingly, the definition of the neighborhood class for x_i can be determined as follows:

$$\delta_C(x_i) = \{x_j \in U \,|\, Dis(x_i, x_j) \leq \delta\}. \tag{1}$$

2.2. K-nearest neighborhood rough set (KNNRS)

In the Neighborhood Decision System (*NDS*), given a subset of conditional attributes C, the k-nearest neighborhood [32] of sample x_i based on attribute subset C can be defined as follows:

$$\begin{aligned} \kappa(x_i) = \{x_i^1, x_i^2, \ldots, x_i^k \,|\, & Dis_C(x_j, x_i) > Dis_C(x_i, x_i^t), \\ & x_j \neq x_i^s, s, t = 1, 2, \ldots, k\}. \end{aligned} \tag{2}$$

$\kappa_C(x)$ is the set of k samples that are nearest neighbors to sample x. The concept of "nearest neighbor" mentioned here refers to proximity in terms of distance. In other words, for any given sample x_i, its corresponding set $\kappa_C(x_i)$ contains exactly k samples, which are the closest to x_i in a certain metric space.

Therefore, we can define the relationship between samples in the k-nearest neighborhood as

$$Knn_C = \{(x_i, x_j) \in U \times U \,|\, x_j \in \kappa(x_i)\}.$$

Table 1
A toy example of single-label task.

U	a_1	a_2	a_3	D
x_1	-2	3	0	l_1
x_2	1	0	1	l_1
x_3	1	1	1	l_2
x_4	4	2	-1	l_3

Table 2
A toy example of multi-label task.

U	y_i
x_1	l_1, l_2
x_2	l_2
x_3	l_1, l_3
x_4	l_2, l_3

Table 3
Binary representation of multi-label task.

U	l_1	l_2	l_3
x_1	1	1	0
x_2	0	1	0
x_3	1	0	1
x_4	0	1	1

the Knn_C relation evidently possesses reflexivity. However, this relation does not exhibit symmetry or transitivity properties.

2.3. Multi-label neighborhood rough set (MNRS)

The paper delves into an in-depth analysis of how the theoretical transition from single-label to multi-label learning frameworks can be achieved within the framework of rough set theory. This theoretical shift not only forms the focal point of this section but is also crucial for theoretical construction of the multi-label feature selection algorithm proposed in this paper.

Within the context of multi-label learning, we utilize a triplet (U, A, D) to denote the multi-label decision information system [33], where U corresponds to the set of samples, A represents the set of features, and $D = \{l_1, l_2, \ldots, l_m\}$ signifies the set of labels. In *MNRS*, if x_i corresponds to a value of 1 for a decision attribute, it signifies that x_i has the corresponding label, whereas a value of 0 indicates that x_i does not possess that particular label. For sample x_1, we use y_1 to represent its corresponding label set.

As depicted in Tables 2 and 3, considering the fact that the first two decision attribute values of x_1 are 1, and the third value is 0, it indicates that x_1 possesses the labels l_1 and l_2, while not having the label l_3. Mathematically, this can be expressed as $y_1 = \{l_1, l_2\}$, representing the labels that x_1 possesses.

In the domain of single-label learning, the neighborhood decision-making system employs the lower approximation of the neighborhood rough set to precisely represent the classification ability of the attribute set towards the samples. In multi-label learning, we define the attribute set as being able to clearly classify samples across different labels. The ability in each label category is expressed in the form of a lower approximation of neighborhood rough sets in multi-label learning.

Definition 2.1 *(Approximate space of MNRS).* In the multi-label neighborhood decision-making system $MNDT = (U, A, D)$, the label set D^p represents the set of all samples with labels l_p, while D_i refers to the set of labels owned by sample x_i. When the feature subset $C \subseteq A$, the upper approximation and lower approximation of *MNRS* [34] can be defined as follows:

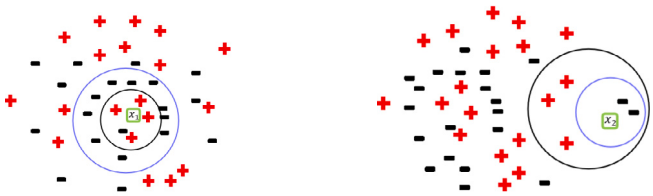
$$\underline{N_C} D = \{x_i \in U \mid \forall l_p \in D_i, N_C^{\delta}(x_i) \in D^p\},$$
$$\overline{N_C} D = \{x_i \in U \mid \forall l_p \in D_i, N_C^{\delta}(x_i) \cap D^p \neq \emptyset\}.$$

In the single-label information system illustrated in Table 1, the universe of discourse U comprises four sample objects, and their neighborhood classes based on the conditional attribute subset C are presented in Table 4. Following the classical rough set theory, we can compute lower approximation set based on C as

$$\underline{N_C} D = \{x_2\} \cup \{x_4\} = \{x_2, x_4\}.$$

Table 4
A toy example of neighborhoods about C.

x_i	$\delta_C(x_i)$
x_1	x_1, x_2, x_3
x_2	x_1, x_2
x_3	x_1, x_3
x_4	x_4



(a) high-density region (b) sparse-density region

Fig. 1. The performance of δ-neighborhood and k-nearest neighborhood under imbalanced data distribution.

But it is different in multi-label learning (as shown in Table 3). Taking x_1 as an example, its neighborhood class is

$$\delta_C(x_1) = \{x_1, x_2, x_3\}.$$

And the label owned by the neighborhood class is

$$y_1 = \{l_1, l_2\}, y_2 = \{l_2\}, y_3 = \{l_1, l_3\}.$$

The target set formed by the labels involved is

$$D^1 = \{x_1, x_3\}, D^2 = \{x_1, x_2, x_4\}, D^3 = \{x_3, x_4\}.$$

Based on $\delta_C(x_1) \not\subset D^1 \wedge \delta_C(x_1) \not\subset D^2$, we can conclude that x_1 is not an element of $\underline{N_C} D$.

Under the given evaluation method, we can determine whether other sample objects belong to the lower approximation set. This can be expressed as follows:

$$\delta_C(x_2) \subset D^2 \Rightarrow x_2 \in \underline{N_C} D,$$
$$\delta_C(x_3) \not\subset D^1 \wedge \delta_C(x_3) \not\subset D^3 \Rightarrow x_3 \notin \underline{N_C} D,$$
$$\delta_C(x_4) \subset D^2 \wedge \delta_C(x_4) \subset D^3 \Rightarrow x_4 \in \underline{N_C} D.$$

Through the above derivation, we draw the conclusion be represented as $\underline{N_C} D = \{x_2, x_4\}$. This implies that x_2, x_4 can be considered as members of the lower approximation set of D

$\underline{N_C} D$ can also be referred to as the positive region of multi-label classification under the knowledge level given by attribute C, and is recorded as $POS(D)$.

Definition 2.2 *(Information measurement of MNRS).* In the framework of multi-label learning, we commonly use dependency to measure the importance of feature subsets. For $C \subseteq A$, the dependence is defined as

$$\gamma_C(D) = \frac{|POS_C(D)|}{|U|} = \frac{|N_C(D)|}{|U|}.$$

3. KNN-based multi-label rough set theory for imbalanced data

3.1. Balancing δ-neighbors and k-nearest neighbors: a comparative analysis

Taking a simple binary classification problem as an example, let's assume the sample space consists of two classes, represented as '+' and '-', corresponding to positive and negative samples, respectively. Here, the δ-neighborhood is depicted by blue circles, while the k-nearest neighbors of the samples are indicated by black circles surrounding them.

In Fig. 1a, it can be observed that sample x_1 is situated within a region of high density, and its δ-neighborhood, denoted as $\delta(x_1)$, encompasses some samples from the positive and negative classes. According to the theory of δ-neighborhood rough sets, sample x_1 may be classified into the boundary region based on the majority voting principle. However, in practice, if we consider the five nearest neighbors, we may be able to identify the class label of x_1 more accurately. The k-nearest neighbor strategy reduces the selection

of samples in high-density regions, thereby narrowing the range of the neighborhood. This promotes greater consistency among the samples within the neighborhood, effectively reducing the risk of misclassification and ensuring the accuracy of the classification results.

Considering another scenario, the k-nearest neighbor strategy also exhibits similar limitations when dealing with samples in sparsely distributed regions. For example, as illustrated in Fig. 1b, if we set the number of nearest neighbors for x_2 to be 5 ($5 - NN$), its neighborhood will contain five samples, including three positive samples that are distant from x_2 and two closer negative samples. According to the majority voting principle, this may result in incorrectly assigning x_2 a '+' label. However, in the case, the $\delta(x_2)$ can better classify x_2 as a negative class label and overcome this limitation of the k-nearest neighbor strategy.

The KNMRS algorithm achieves an effective balance between dense and sparse distributions when handling imbalanced data by integrating δ-neighborhood and k-nearest neighbor techniques. Specifically, the k-nearest neighbor technique is suitable for dense regions, where a fixed number of neighboring points (k) is used to capture local features, ensuring stable selection of key features in densely distributed areas. However, in sparse regions, k-nearest neighbor may fail due to a lack of neighboring points, leading to the neglect of important features. In such cases, the δ-neighborhood dynamically adjusts the neighborhood range by setting a distance threshold (δ), ensuring that key features can still be captured even in sparsely distributed areas. This dual-mechanism design enables KNMRS to automatically adjust its strategy based on the density of data distribution: relying on k-nearest neighbor for stability and efficiency in feature selection in dense regions, and on δ-neighborhood for enhanced adaptability and robustness in sparse regions. For example, in the Computer and Health datasets, KNMRS effectively addressed the imbalance in feature distribution across different regions by dynamically switching between k-nearest neighbor and δ-neighborhood strategies. This trade-off not only resolves the issue of poor performance of traditional methods in sparse regions but also avoids the limitations of a single strategy when dealing with complex distributions. By combining the strengths of both techniques, KNMRS achieves more comprehensive and accurate feature selection in imbalanced data, providing deeper insights into algorithm design choices and enhancing its flexibility and reliability in practical applications.

3.2. K-nearest neighborhood multi-label rough set

Building upon a thorough analysis of Fig. 1a and 1b, this research uncovers the inherent limitations of the δ neighborhood and KNN methods when addressing the challenge of imbalanced sample distribution within the sample space [35]. Furthermore, it highlights their respective advantages in certain extreme scenarios. Based on these findings, we expanded the traditional neighborhood rough set theory and proposed a dedicated multi-label neighborhood rough set model [36]. The primary objective of this model is to effectively mitigate the issue of data imbalance by comprehensively considering the characteristics of sample distribution, optimizing classification performance, and ensuring robust and accurate prediction results across diverse data distribution conditions.

Definition 3.1 *(K-nearest neighbor multi-label decision system).* In traditional multi-label neighborhood decision systems, the influence of δ on decision-making is typically considered. In contrast, in the k-nearest neighbor multi-label decision system $KNMDS = (U, A, D, \delta, k)$, we introduce a key parameter k, which is used to determine the k nearest sample points to a given point x_i.

Definition 3.2 *(K-nearest multi-label neighborhood).* In *KNMDS*, if we arbitrarily select a subset of conditional attributes C, then this new neighborhood class of sample x_i can be described as follows:

$$\psi_C(x_i) = \{x_j | x_j \in \delta_C(xi) \cap \kappa_C(x_i)\}. \tag{3}$$

k-nearest multi-label neighborhood class $\psi_C(x_i)$ is defined as a combination of the δ-neighborhood and the k-nearest-neighborhood of the data point x_i. Therefore, the k-nearest multi-label neighborhood successfully combines the advantages of both approaches while avoiding their inherent limitations. This approach is particularly effective in handling sample spaces with uneven sample densities. It overcomes the challenges faced by traditional classification algorithms in such cases, thereby improving the accuracy and efficiency of classification tasks.

Definition 3.3 *(Approximate space of KNMDS).* The KMNRS algorithm is based on the theory of neighborhood rough sets, whose core concepts include the upper approximation set and the lower approximation set. These two sets describe the boundary of a concept from different perspectives.

In *KNMDS*, where $\forall C \subseteq A$. Within this framework, the neighborhood relation on C is denoted as NC_C^{ψ}. For any sample $x_i \in U$, $NC_C^{\psi}(x_i)$ characterizes the neighborhood of x_i determined by C. Consequently, the upper and lower approximations of the domain U based on the C can be defined as:

$$\underline{N_C^{\psi}} D = \{x_i \in U | \forall l_p \in D_i, N_C^{\psi}(x_i) \in D^p\},$$

$$\overline{N_C^{\psi}} D = \{x_i \in U | \forall l_p \in D_i, N_C^{\psi}(x_i) \cap D^p \neq \emptyset\}.$$

By employing the aforementioned definitions, we can derive the positive region of KNMDS:

$$POS_C^{\psi}(D) = \underline{N_C^{\psi}} D.$$

Definition 3.4 *(Information measurement of KNMDS).* The dependency function is an important concept in rough set theory, used to measure the significance of features. The KNMRS algorithm introduces the dependency function to select the most relevant feature subset.

$$\gamma_C^{\psi}(D) = \frac{|POS_C^{\psi}(D)|}{|U|} = \frac{|N_C^{\psi}(D)|}{|U|}. \tag{4}$$

The higher the value of the dependency function $\gamma_C^{\psi}(D)$, the greater the dependency of the feature subset C on the decision attribute D, indicating that the features in C are more important. The KNMRS algorithm calculates the dependency function values for different feature subsets and selects the subset with the maximum dependency function value as the final feature selection result. Based on this function, the importance of features can be further quantified, providing a systematic approach to feature selection in multi-label classification tasks.

The importance of conditional attribute $a \in A\text{-}C$ relative to decision attribute set $D_q = \{l_1, \ldots, l_q\}$ based on conditional attribute set C is

$$sig_{\gamma}^{\psi}(a_i, C, D) = \gamma_{C\cup\{a_i\}}^{\psi}(D) - \gamma_C^{\psi}(D). \tag{5}$$

If $sig(a_i, C, D) = 0$, it indicates that the attribute a has no statistically significant correlation with the studied information system. In such cases, we can exclude this attribute from *KNMDS* to simplify the model and improve its computational efficiency.

4. Multi-label feature selection algorithm in *KNMDS*

Algorithm 1 The feature selection algorithm using K-Nearest Neighborhood Multi-label Rough Set (KNMRS).

Input: $KNMDS = (U, A, D, \delta, k)$, δ and k are adjustable hyperparameters;
Output: A feature selection subset $reduct$.
Initialize $C \leftarrow \emptyset$, $reduct \leftarrow \emptyset$;
for a_k to $A - reduct$ **do**
  $reduct \leftarrow reduct \cup a_k$
  for x_i to U **do**
    Calculate the $\delta(x_i), \kappa(x_i)$ by Equation (1) and (2)
    Calculate the $\psi(x_i)$ by Equation (3)
    if $\forall l_p \in D_i, \delta(x_j) \subset D^p$ **then**
      $POS^{\psi}_{reduct\cup a_k}(D) \leftarrow POS^{\psi}_{reduct\cup a_k}(D) \cup x_j$
    end if
  end for
  Calculate $\gamma_C^{\psi}(D)$ by Equation (4)
  Calculate $sig_{\gamma}^{\psi}(a_k, C, D)$ by Equation (5)
  $sig(a_l, C, D) \leftarrow max(sig_{\gamma}^{\psi}(a_k, C, D))$
  if $sig_{\gamma}^{\psi}(a_l, C, D) > 0$ **then**
    $reduct \leftarrow reduct \cup a_l$
  else
    return $reduct$
  end if
end for
return $reduct$

4.1. *KNMDS algorithm flow description*

The algorithm comprises three main stages. The first stage initializes the feature subset C and the final feature subset $reduct$. The second stage consists of two steps. First, the method identifies the most relevant features from the multiple attributes to construct a temporary decision information system. Within this system, δ-neighborhood classes and κ-neighborhood classes are calculated for all samples, subsequently determining the derived ψ-neighborhood classes and their corresponding positive regions. This process is specifically designed to mitigate data imbalance issues by combining the strengths of δ neighborhood rough sets and k-nearest neighborhood rough sets, thus offering a novel approach to handling imbalanced datasets. Second, the algorithm calculates the feature dependency, monitors changes in feature importance, and ultimately selects a feature subset with high importance and strong discriminative power as the final result. The KNMRS algorithm's independence from data distribution, high accuracy, adaptability, and effective feature selection capabilities make it well-suited for handling multi-label data.

In a multi-label information decision system containing N instances, L labels, and M feature attributes, the time complexity analysis of the KNMRS algorithm involves three key steps. First, calculating the δ neighborhood and k-nearest neighborhood for each sample involves comparing distances with all other samples, resulting in a time complexity of O(N). Second, iterating through different subsets of features adds a time complexity of O(M). Finally, assessing the consistency of sample instances within the neighborhood for each label results in the final positive region, giving a time complexity of O(NL). Combining these steps, the overall time complexity of the KNMRS algorithm is $O(MN^2L)$.

4.2. Comparison of KNMDS with traditional methods for addressing data imbalance

The KNMRS algorithm proposed in this paper demonstrates several significant advantages over traditional oversampling and undersampling methods in addressing imbalanced data:

1) **Dynamic Adaptation to Data Density**: Traditional oversampling and undersampling methods are typically effective for specific data densities but may perform poorly for others. In contrast, the KNMRS algorithm combines the strengths of δ-neighborhood and k-nearest neighbor techniques to dynamically adjust the neighborhood structure based on the local density and distribution characteristics of the data. This flexibility enhances the model's adaptability and accuracy in handling extreme data distributions, effectively mitigating the impact of data imbalance.
2) **Reduction of Information Loss and Enhancement of Classification Accuracy**: Oversampling and undersampling methods often involve modifications to the data, which can introduce noise or lead to information loss, thereby affecting classification accuracy. The KNMRS algorithm directly utilizes the original data without modification, thereby minimizing information loss and improving classification accuracy.
3) **Avoidance of Overfitting**: Oversampling methods can increase the number of samples, potentially leading to overfitting and reduced generalization ability. The KNMRS algorithm avoids this issue by not increasing the sample size, thereby enhancing the model's generalization capability. Traditional oversampling and undersampling methods balance datasets by increasing the number of minority class samples or reducing the number of majority class samples, which can lead to information loss or the introduction of noise. The KNMRS algorithm performs feature selection directly on the original data, avoiding the potential issues associated with data resampling.
4) **Consideration of Feature and Label Dependencies**: Traditional oversampling and undersampling methods primarily focus on the distribution of sample numbers and often overlook the dependencies between features and labels. The KNMRS algorithm introduces a dependency function that effectively quantifies the significance of features in a multi-label decision environment, enabling more accurate feature selection and improved classification performance.
5) **Applicability to Multi-Label Learning Scenarios**: Traditional oversampling and undersampling methods are mainly designed for single-label learning scenarios. In contrast, multi-label learning scenarios are more complex because samples may belong to multiple classes simultaneously. The KNMRS algorithm is specifically designed for multi-label learning scenarios, eliminating the need to transform the problem into a single-label one. By directly handling multi-label data, KNMRS preserves the correlations between labels, resulting in a more accurate model.
6) **Rough set theory for feature selection**: The KNMRS algorithm leverages rough set theory for feature selection, which enables it to handle uncertainties and ambiguities in the data. Through rough set theory, the algorithm selects features most relevant to the decision attributes while eliminating redundant and noisy features. This not only enhances the model's performance but also increases its robustness.

5. Experimental analysis

In this section, to demonstrate the superiority of the KNMRS algorithm in the domain of feature selection, we devised a series of comparative experiments.

5.1. Experimental preparation

This paper selected 12 publicly available multi-label datasets, covering multiple domains including text, biology, music, and images. From a theoretical perspective, Data imbalance is common in these domains. For example, in text datasets, the number of documents for some topics may be significantly lower than others. In biological datasets, certain gene function labels may be rare. Additionally, these datasets vary in sample size, feature dimensions, and label quantities to ensure that the experiments cover data from different fields and scales.

Experimental data show that the LC (average label cardinality) values of the datasets vary significantly, ranging from 1.014 to 4.237, indicating substantial differences in the average number of labels per sample. The LC value reflects the distribution of samples in the feature space to some extent. Samples with higher LC values may require more features to represent multiple labels, leading to a more dispersed distribution in the feature space. Moreover, some datasets have low LD (label density) values. For example, the LD value of the Birds dataset is only 0.053, indicating that the proportion of labels per sample is low relative to the total number of labels, which may also imply a sparser distribution of samples in the feature space.

More importantly, the imbalance rate (avgIR) of the datasets varies significantly, ranging from 1.254 to 653.531, which further corroborates the imbalance in the distribution of samples in the feature space. For instance, Dataset A has an avgIR as high as 653.531, indicating an extremely uneven distribution of samples in the feature space. Some samples may occupy the majority of the feature space, while others are confined to very limited regions. In contrast, Dataset Scene has a lower avgIR of 1.254, suggesting a relatively more uniform distribution of samples in the feature space. However, even datasets with lower avgIR values may still face issues of imbalance in feature space distribution if they have high LC or LD values. For example, Dataset Yeast has an LC value of 4.237, indicating that samples on average possess a larger number of labels. If the feature expressions of these labels are unevenly distributed in the feature space, it can still lead to difficulties for models in learning effective feature representations.

Table 5
Description of the twelve multi-label datasets.

No.	Datasets	S	F	L	LC	LD	avgIR	Domain
1	Art	5000	462	26	1.63	0.27	94.7	text
2	Computer	5000	681	159	1.50	0.04	176.6	text
3	Yeast	2417	103	14	4.23	0.30	7.197	Biology
4	Health	5000	612	32	1.63	0.05	653.5	text
5	Emotions	593	72	6	1.86	0.31	1.478	Music
6	Reference	5000	793	33	1.16	0.03	461.8	text
7	Scene	2407	294	6	1.07	0.17	1.254	Image
8	Recreation	5000	606	22	1.42	0.06	12.20	text
9	Birds	645	260	19	1.01	0.05	5.407	text
10	Education	5000	550	33	1.46	0.04	168.1	text
11	Science	5000	743	40	1.45	0.03	52.63	text
12	Social	5000	1047	39	1.27	0.03	257.70	text

We believe that this imbalance in feature space distribution poses challenges for multi-label learning, such as difficulties in learning effective feature representations or susceptibility to noise features. Therefore, studying the imbalance of multi-label data distribution in the feature space is of significant theoretical and practical importance. Detailed information on the datasets is provided in Table 5. For each dataset we provide a short description as well as some characterization metrics. It includes the number of samples (S), number of features (F), number of labels (L), label cardinality (LC), label density (LD), and average Imbalance Ratio per label (avgIR).

To objectively evaluate the results of feature selection, this study adopted the widely recognized multi-label k-nearest neighbor (MLKNN) classifier in the realm of multi-label learning. In the experimental setup of MLKNN, we used the default smoothing parameter $\sigma = 1$ and set the k-nearest neighbors parameter to 10. In the evaluation process, we employed a technique called 10-fold cross-validation. In our experiments, the primary focus was on comparing the relative performance of different feature selection methods in multi-label classification tasks, rather than achieving the absolute optimal performance of MLKNN under specific parameter settings. Therefore, we opted to use the commonly recommended default parameter setting of $k = 10$. While the choice of k does influence the performance of MLKNN, our emphasis was on comparing the performance of different feature selection methods under the same classifier and parameter configuration. To ensure consistency and fairness in our experiments, all tests were conducted using the unified parameter setting of $k = 10$. Additionally, researchers experimentally observed the impact of varying k within a certain range (e.g., $k = 5,8,12,15,20$) on the results and found that the relative ranking of the performance of different feature selection methods remained largely unchanged. In each iteration, based on the results of the feature selection algorithm, we trained the classifier using the training set. After training, we evaluated the performance of the trained classifier on the corresponding test set. By averaging the evaluation results across all test sets, we obtained the final classification performance metrics.

In the context of multi-label classification problems, the evaluation metric system [37] comprehensively reflects the performance of algorithms from multiple dimensions. This paper adopts Hamming loss as a sample-based metric for evaluating the performance of multi-label classifiers, which measures the effectiveness of classifiers by calculating the difference between the predicted label set and the actual label set. In terms of label-based metrics, this paper decomposes the multi-label problem into single-label measurements, uses all metrics applicable to single-label problems, and averages the measurement results for each label, specifically using macro-averaging and micro-averaging to achieve this process. In addition, this paper also adopts metrics based on ranking and prediction, including Average Precision (AP), Ranking Loss (OE), Coverage (CV), and Ranking Loss (RL), which focus more on evaluating the accuracy and ranking quality of the predicted label set. Additionally, the number of selected features (N) serves as an indicator reflecting the level of feature reduction. Among these metrics, higher scores in MacF1, MicF1 and AP indicate more remarkable algorithm performance, while lower scores in CV, OE, RL, and HL also signify superior algorithm performance. These evaluation metrics reveal the comprehensive performance of multi-label classification algorithms from different perspectives, providing a strong basis for algorithm optimization and selection. In the table below, "↑" indicates metrics where higher values are better, while "↓" indicates metrics where lower values are better. The optimal performance for each evaluation metric will be presented in bold format.

5.2. The comparative results between KNMRS and other multi-label feature selection algorithms

In the domain of multi-label learning, each sample can be associated with multiple categories simultaneously, in contrast to single label learning where a sample is assigned to a single category. A key challenge in multi-label learning arises from the potential influence of label ordering on model performance evaluation. Hence, disregarding the label ordering can result in misleading assessments of the model's performance. The primary objective of the first part of this section is to assess the performance of our proposed KNMRS algorithm in relation to ranking-based multi-label classification metrics, namely AP, CV, OE, RL, and N. To accomplish this, we conduct a thorough comparison between the KNMRS algorithm and cutting-edge methods in multi-label classification, which include:

We compared our proposed KNMRS algorithm against several existing multi-label feature selection methods to assess its effectiveness. The comparison included:

- **Multi-Label Naive Bayes (MLNB [38])**: MLNB, a straightforward multi-label classification method, uses the Naive Bayes assumption. While computationally efficient, it can struggle with complex label relationships and may not perform well with imbalanced data due to its sensitivity to how features are distributed.

Table 6
The experimental results of ten algorithms on the Computer dataset.

Methods	N↓	AP↑	CV↓	OE↓	RL↓
MLKNN	681	0.633	4.416	0.437	0.092
MLNB	345	0.635	4.553	0.434	0.095
MDDM_proj	**18**	0.598	4.88	0.481	0.105
MDDM_spc	23	0.599	4.847	0.48	0.103
ARMLNRS_imp	132	0.633	4.419	0.443	0.091
ARMLNRS_opt	132	0.623	4.421	**0.422**	0.089
ARMLNRS_pes	102	0.622	4.311	0.472	0.089
ML-ReliefF	387	0.6305	4.467	0.4392	0.0929
NMIFS	200	0.6318	4.466	0.447	0.0915
KNMRS	155	**0.6541**	**4.27**	**0.422**	**0.0824**
Average	217.5	0.6259	4.505	0.4477	0.0931

- **Multi-label Dimensionality Reduction via Data-Driven and Structure Preserving Projections (MDDM_proj, MDDM_spc [39]):** MDDM aims to reduce the dimensionality of multi-label data by learning projections that preserve both data-driven and structure-preserving information. While effective for dimensionality reduction, its performance on imbalanced multi-label data isn't thoroughly examined in the original work. Its primary focus is dimensionality reduction, not directly addressing feature selection or imbalance.
- **Adaptive Rough Set-Based Multi-Label Neighborhood Reduction (ARMLNRS [40]):** ARMLNRS uses rough set theory for multi-label feature selection. Different versions (imp, opt, pes) represent different strategies within the ARMLNRS framework. While rough sets are inherently suited for handling uncertainty, ARMLNRS's performance on imbalanced multi-label data isn't the main focus of the cited paper. Also, the computational cost of rough set-based methods can be a concern.
- **Multi-Label ReliefF (ML_ReliefF [41]):** ML-ReliefF adapts the ReliefF algorithm for multi-label data. ReliefF is known for identifying relevant features based on instance-based learning. However, like other instance-based methods, ML-ReliefF can be sensitive to noise and may not be optimal with significant class imbalance. It can also be challenged by high-dimensional data.
- **Neighborhood Mutual Information Feature Selection (NMIFS [34]):** NMIFS uses neighborhood mutual information to select relevant features. It aims to capture the relationships between features and labels. While NMIFS can be effective, its performance on imbalanced multi-label data and its computational demands are important factors to consider.

It's crucial to understand that these methods can perform differently depending on the specific dataset characteristics, such as the degree of imbalance, the number of labels, and the presence of noise or irrelevant features. Furthermore, the cited works don't always explicitly address the challenges of imbalanced multi-label data. This underscores the need for specialized techniques like KNMRS, which are specifically designed to handle the complexities of both multi-label learning and data imbalance.

To perform the performance comparison with these nine methods on the five multi-label evaluation metrics, we selected the Computer, Health, Yeast, Art, Scene, Recreation, Emotions, Reference, Birds, and Recreation datasets from Table 5 as the experimental data. Figs. 2-5 illustrate the performance variation of six algorithms across various levels of feature reduction. The X-axis represents the number of selected features, while the Y-axis displays the performance values for four specific metrics. Since the subplots within the same figure group use the same plotting methods and data encoding, the legend is universal and representative. For the sake of conciseness, we only display the legend in the last subplot.

Fig. 2 demonstrates that, except for the Art and Recreation datasets, KNMRS exhibits the best AP performance among the other seven datasets. In the Computer, Health, Reference, and Birds datasets, KNMRS consistently outperforms other algorithms. In the Recreation dataset, the performance of KNMRS alternates with ARMLNRS as the number of features increases, and both outperform the other comparative algorithms. For the Art dataset, KNMRS performs best when approximately 160 features are selected, and its performance significantly declines as the number of features increases. However, this precisely indicates that KNMRS achieves efficient feature reduction while maintaining good AP performance on the Art dataset. Fig. 3 displays the CV performance of the six algorithms. In the Computer, Health, Recreation, Emotion, and Reference datasets, KNMRS significantly outperforms other algorithms. In the Scene and Birds datasets, KNMRS performs comparably to NMIFS and ML-ReliefF, and it outperforms the other four algorithms. In the Yeast and Emotion datasets, the NMIFS algorithm performs better than our proposed algorithm, but we achieve competitive or optimal performance when selecting a larger number of features, thanks to the integration of two neighborhood principles in our algorithm. In practical applications, the appropriate algorithm can be selected based on different requirements for feature reduction effectiveness and performance. Fig. 4 illustrates the advantage of KNMRS in the OE metric on the Computer, Recreation, and Reference datasets. In the Health dataset, KNMRS is only slightly inferior to MDDM_proj. In the Yeast and Art datasets, KNMRS demonstrates no significant difference in performance compared to other algorithms and still achieves optimal performance at a lower degree of feature reduction. For the Emotion and Birds datasets, KNMRS performs equally well as ARMLNRS and ML-ReliefF, and it outperforms the other datasets. Fig. 5 describes the performance of the six algorithms in the RL metric. KNMRS surpasses other algorithms significantly in the Art, Recreation, Scene, and Computer datasets. In the Health dataset, KNMRS does not surpass MDDM_proj. Furthermore, in other datasets, our algorithm demonstrates no significant difference compared to ARMLNRS and NMIFS, and it ranks highly in the experimental results. In conclusion, the KNMRS algorithm performs exceptionally well on the nine datasets.

Table 7
The experimental results of ten algorithms on the Health dataset.

Methods	N↓	AP↑	CV↓	OE↓	RL↓
MLKNN	612	0.681	3.305	0.421	0.061
MLNB	289	0.667	3.555	0.425	0.068
MDDM_proj	15	0.704	3.234	0.381	**0.06**
MDDM_spc	**81**	0.647	3.704	0.453	0.071
ARMLNRS_imp	83	0.685	3.358	0.401	0.063
ARMLNRS_opt	81	0.6972	3.357	0.4023	0.0627
ARMLNRS_pes	87	0.7111	3.358	0.4019	0.0661
ML-ReliefF	286	0.6845	3.6213	0.4590	0.0685
NMIFS	215	0.7034	3.363	0.4215	0.0624
KNMRS	69	**0.7441**	**3.121**	**0.379**	**0.06**
Average	181.8	0.6924	3.397	0.4133	0.0642

Table 8
The experimental results of ten algorithms on the Yeast dataset.

Methods	N↓	AP↑	CV↓	OE↓	RL↓
MLKNN	103	0.751	6.809	0.25	**0.176**
MLNB	28	0.736	6.693	0.256	0.187
MDDM_proj	**7**	0.708	6.852	0.262	0.208
MDDM_spc	20	0.712	6.879	0.254	0.206
ARMLNRS_imp	23	0.734	6.605	0.253	0.190
ARMLNRS_opt	26	0.729	6.607	0.2543	0.1833
ARMLNRS_pes	31	0.731	6.611	0.2493	0.1871
ML-ReliefF	16	0.732	6.6759	0.2551	0.1919
NMIFS	76	0.741	6.5462	0.2437	0.1819
KNMRS	83	**0.7681**	**6.5132**	**0.2381**	0.1794
Average	41.3	0.73421	6.67913	0.25155	0.18906

Table 9
The experimental results of ten algorithms on the Art dataset.

Methods	N↓	AP↑	CV↓	OE↓	RL↓
MLKNN	462	0.5093	5.4453	0.6327	0.152
MLNB	228	0.5098	5.4917	0.629	0.153
MDDM_proj	120	0.4895	5.4943	0.6633	0.1543
MDDM_spc	120	0.4808	5.6612	0.6791	0.1599
ARMLNRS_imp	115	0.5156	5.3211	0.616	0.1468
ARMLNRS_opt	120	0.5155	5.3371	0.6132	0.1492
ARMLNRS_pes	**114**	0.5116	5.3561	0.6377	0.1479
ML-ReliefF	267	0.4959	5.3451	0.6982	0.1481
NMIFS	142	0.5101	5.4371	0.6257	0.1512
KNMRS	157	**0.5313**	**5.2937**	**0.6026**	**0.1417**
Average	184.5	0.50694	5.41827	0.63975	0.15041

Table 10
The experimental results of ten algorithms on the Scene dataset.

Methods	N↓	AP↑	CV↓	OE↓	RL↓
MLKNN	294	0.8494	**0.5686**	0.3928	0.0931
MLNB	200	0.8145	0.6941	0.3971	0.099
MDDM_proj	250	0.8167	0.6962	0.2988	0.119
MDDM_spc	250	0.8138	0.6991	0.3067	0.1183
ARMLNRS_imp	243	0.8312	0.6328	0.3917	0.1075
ARMLNRS_opt	243	0.8386	0.6017	0.3901	0.1025
ARMLNRS_pes	243	0.8109	0.6893	0.3857	0.177
ML-ReliefF	205	0.8155	0.6413	0.4662	0.1105
NMIFS	200	0.8433	0.6179	0.2954	0.1029
KNMRS	**79**	**0.8519**	0.5877	**0.2944**	**0.0909**
Average	220.7	0.82858	0.64287	0.36189	0.11207

Table 11
The experimental results of ten algorithms on the Recreation dataset.

Methods	N↓	AP↑	CV↓	OE↓	RL↓
MLKNN	606	0.4644	5.1483	0.7737	0.1928
MLNB	204	0.4367	5.0523	0.6924	0.1893
MDDM_proj	194	0.4361	5.2076	0.7372	0.1977
MDDM_spc	196	0.4372	5.2161	0.7212	0.1979
ARMLNRS_imp	228	0.5214	4.8702	0.6797	0.1819
ARMLNRS_opt	228	0.4915	4.8702	0.6714	0.1833
ARMLNRS_pes	245	0.4931	4.8702	0.6743	0.1837
ML-ReliefF	**109**	0.4065	5.3173	0.7466	0.2244
NMIFS	203	0.4814	5.1020	0.6637	0.1905
KNMRS	323	**0.5284**	**4.4779**	**0.6533**	**0.1661**
Average	253.6	0.46967	5.01321	0.70135	0.19076

Table 12
The experimental results of ten algorithms on the Emotion dataset.

Methods	N↓	AP↑	CV↓	OE↓	RL↓
MLKNN	72	0.781	1.92	0.332	0.173
MLNB	25	0.753	2.074	0.376	0.205
MDDM_proj	**3**	0.669	**0.649**	0.4059	0.3164
MDDM_spc	6	0.748	2.2134	0.3956	0.226
ARMLNRS_imp	30	0.79	1.964	0.2769	0.1741
ARMLNRS_opt	33	0.785	1.959	0.2777	0.1749
ARMLNRS_pes	29	0.79	2.011	0.2773	0.2095
ML-ReliefF	42	0.7512	1.9801	0.3521	0.3999
NMIFS	29	0.7415	1.9255	0.3517	0.2033
KNMRS	59	**0.7931**	1.9119	**0.276**	**0.1701**
Average	32.8	0.76018	1.86079	0.33212	0.22522

Table 13
The experimental results of ten algorithms on the Reference dataset.

Methods	N↓	AP↑	CV↓	OE↓	RL↓
MLKNN	793	0.583	3.813	0.532	0.102
MLNB	400	0.623	3.431	0.47	0.089
MDDM_proj	400	0.611	3.446	0.489	0.089
MDDM_spc	400	0.613	3.44	0.484	0.089
ARMLNRS_imp	190	0.605	3.963	0.564	0.105
ARMLNRS_opt	174	0.609	3.8763	0.564	0.106
ARMLNRS_pes	188	0.601	3.992	0.559	0.105
ML-ReliefF	**68**	0.5834	3.7497	0.5080	0.0982
NMIFS	212	0.6312	3.408	0.4653	0.0858
KNMRS	276	**0.642**	**3.224**	**0.4403**	**0.0847**
Average	310.1	0.61016	3.6343	0.50756	0.09537

Table 14
The experimental results of ten algorithms on the Birds dataset.

Methods	N↓	AP↑	CV↓	OE↓	RL↓
MLKNN	260	0.695	**3.399**	0.39	0.125
MLNB	131	0.697	3.458	0.365	0.128
MDDM_proj	**13**	0.626	3.564	0.523	0.141
MDDM_spc	19	0.643	3.616	0.489	0.139
ARMLNRS_imp	104	0.719	3.421	0.344	**0.123**
ARMLNRS_opt	109	0.7201	3.422	0.3477	**0.123**
ARMLNRS_pes	122	0.7255	3.425	0.3598	0.124
ML-ReliefF	128	0.702	3.5111	0.3741	0.129
NMIFS	99	0.682	3.7462	0.3781	0.1381
KNMRS	145	**0.7364**	3.4611	**0.3241**	**0.123**
Average	113	0.6946	3.50234	0.38948	0.12931

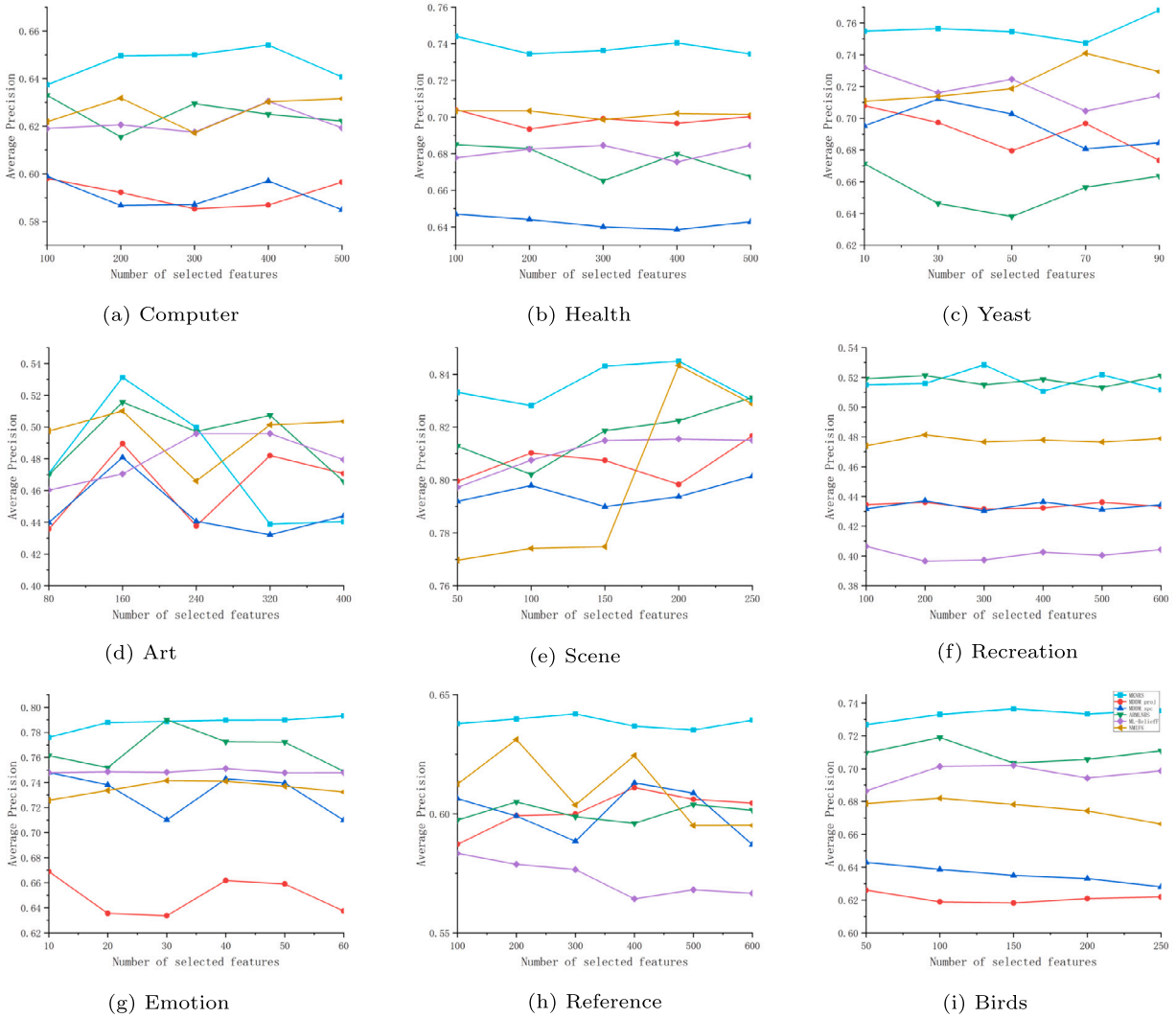


Fig. 2. The performance variation of six algorithms in terms of AP across nine datasets.

Figs. 2-5 provide an intuitive presentation of the comparative results for algorithm performance. To accurately assess the effectiveness of the KNMRS algorithm, it is necessary to comprehensively compare the various quantified evaluation metrics of all algorithms on each dataset. We evaluated the performance of various algorithms using nine different feature selection methods on the aforementioned feature selection methods, and the results are listed in Tables 6-14. MLKNN (using all features) serves as the benchmark for comparison, and KNMRS achieved significant performance improvements across several key metrics, especially in terms of Average Precision (AP), where it showed significant improvement compared to the original dataset. Across most datasets, with the exception of Scene, although the N index of KNMRS is not the lowest, its AP value is the highest. On the Computer and Yeast datasets, as shown in Tables 6 and 8, KNMRS significantly outperforms algorithms such as MLNB, ML-ReliefF, MDDM_proj, and MDDM_spc in terms of AP, and it also demonstrated superior performance compared to the other four algorithms. KNMRS also achieves the best results in terms of CV, OE, and RL metrics. On the Health dataset (see Table 7), KNMRS significantly reduces the number of selected features compared to MLNB, ML-ReliefF, and NMIFS, while slightly increasing it compared to other algorithms. KNMRS excels in AP and CV, although its OE and RL metric values are slightly lower than MDDM_proj, they are still better than all other methods. In the Art, Recreation, and Reference datasets, KNMRS achieves improvements in all evaluation metrics, outperforming all comparative algorithms. As shown in Table 6, on the Computer dataset, KNMRS performs the best in terms of AP, OE, and RL metrics, and it is the only algorithm that achieves a significant improvement in AP. Although the CV metric is slightly lower than ARMLNRS_pes, it still far surpasses other algorithms. Table 8 demonstrates that on the Yeast dataset, all feature selection algorithms result in a decrease in AP compared to the original dataset. However, our KNMRS algorithm successfully improves AP during the feature selection process, and it performs excellently in other metrics, except for the RL metric, where it achieves optimal results compared to the comparative algorithms. The performance of KNMRS in RL is also superior to the original data by a marginal margin. In Table 10, the Scene dataset demonstrates a unique situation. MLKNN, which represents the original data without feature selection, performs

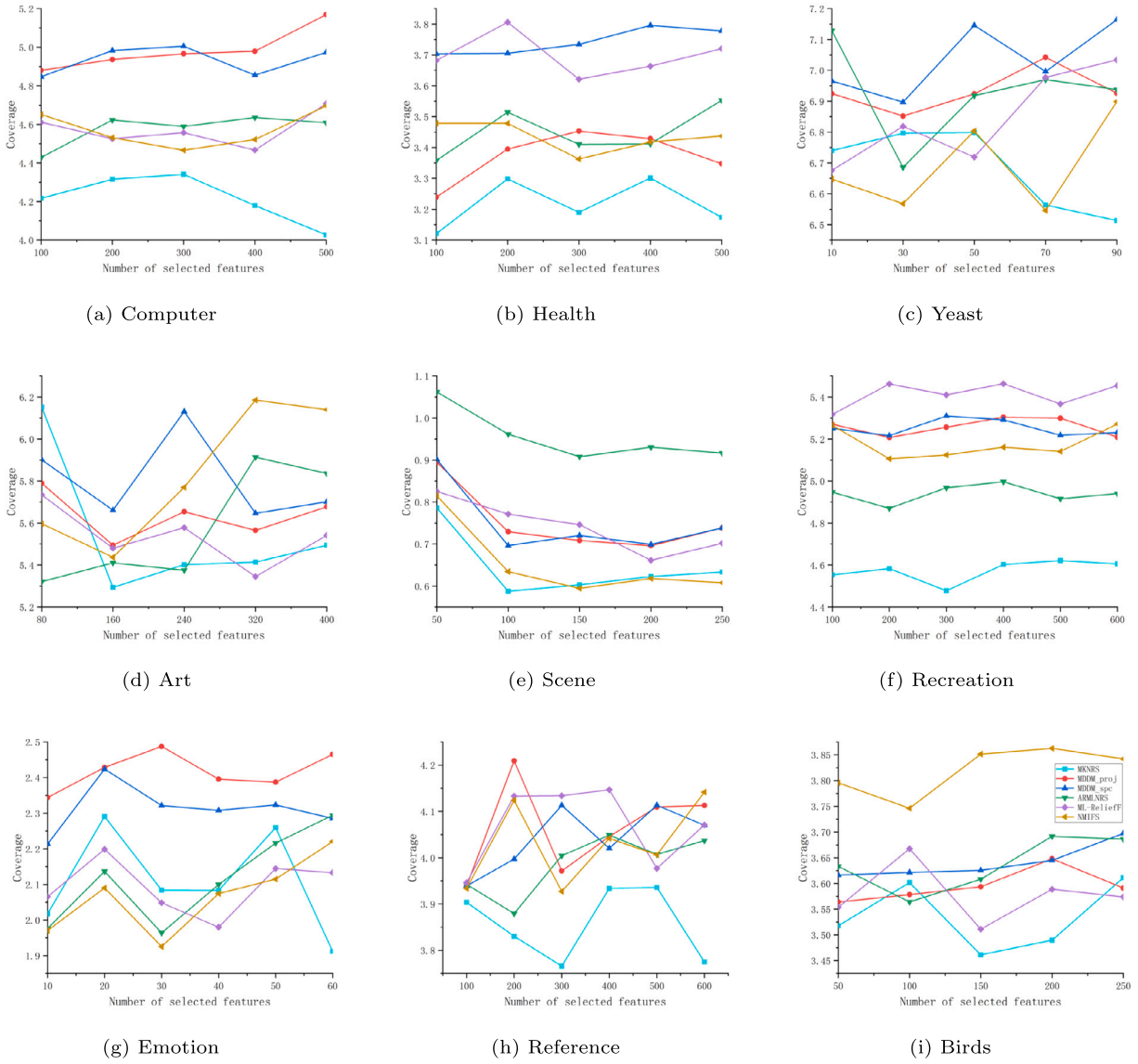


Fig. 3. The performance variation of six algorithms in terms of CV across nine datasets.

extremely well. This may be related to the characteristics of the dataset: for ideally distributed and smoothly fluctuating datasets, feature selection may reduce performance because each feature may be closely related to the stability of the information system. Nevertheless, the performance of KNMRS is almost consistent with MLKNN, and it even achieves steady improvements in OE and RL. For the remaining datasets, our algorithm not only achieves improvements in AP but also performs exceptionally well in other metrics. The comprehensive analysis indicates that the KNMRS algorithm performs exceptionally well in ranking-based evaluation metrics. However, considering that KNMRS is based on the combination of two neighborhood rough sets, the attribute reduction criteria used during feature selection are more stringent, which limits the algorithm's ability to achieve extensive attribute reduction. Additionally, the KNMRS algorithm is particularly suitable for handling imbalanced data, which makes its performance improvement particularly significant on sparse datasets such as Computer, Health, Recreation, Art, and Reference. This aspect is clearly reflected in the comparative data presented in Tables 6-14. The second part of this study aims to evaluate the performance of our proposed KNMRS algorithm on the classification performance metric called Hamming Loss (HL). In the field of multi-label learning, Hamming Loss plays a crucial role as a metric that accurately reflects the performance of the classifier. Furthermore, it assists the model in handling the complexity and sparsity inherent in multi-label data. These two aspects are the key perspectives for evaluating the performance of multi-label learning models. Fig. 6 depicts the performance of six comparison algorithms and our KNMRS algorithm on the Hamming Loss (HL) metric across nine different datasets as the number of selected features varies. The results show that none of the comparison algorithms outperforms KNMRS in all conditions. Our algorithm demonstrates excellent performance in most cases, particularly at specific values of the feature count N, where KNMRS achieves the best performance among all the algorithms. Even on

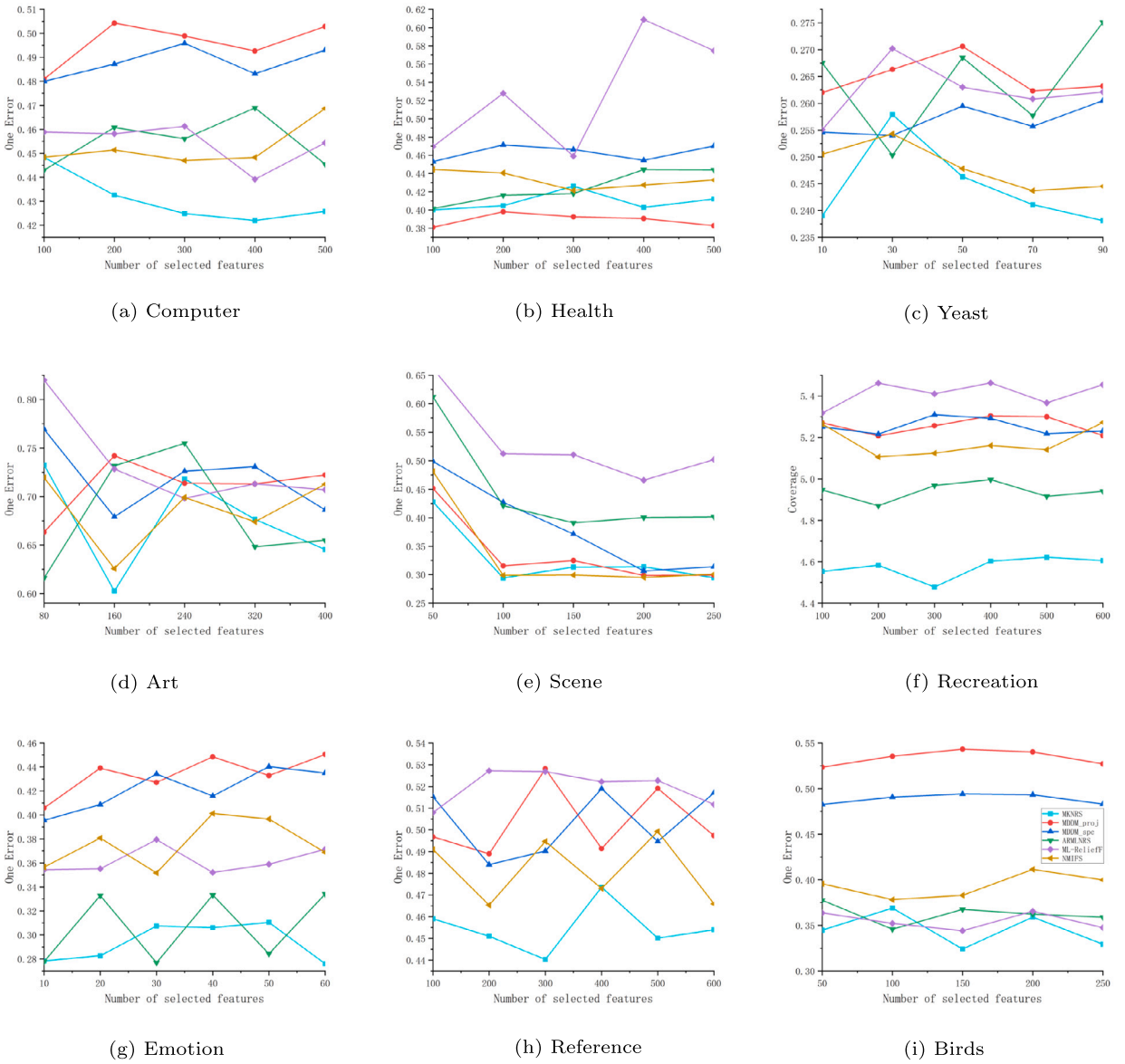


Fig. 4. The performance variation of six algorithms in terms of OE across nine datasets.

the challenging Scene dataset, which poses the greatest challenge to KNMRS, the algorithm maintains a leading position, highlighting its robustness and superiority. To objectively evaluate the performance of the algorithms, this study conducted a quantitative analysis of the performance of six different algorithms on the HL metric, as shown in Table 15. The research demonstrates that, except for the balanced Scene dataset, KNMRS algorithm consistently exhibits excellent performance on imbalanced datasets. Particularly noteworthy is its significant reduction in Hamming Loss when handling raw data, which fully demonstrates the efficiency of our algorithm in addressing imbalanced data distributions.

The third part of this study aims to evaluate the performance of our proposed KNMRS algorithm on label-based performance metrics, Macro-F1 and Micro-F1. Macro-F1 and Micro-F1 are important in multi-label learning as they provide a comprehensive evaluation of model performance from different perspectives. Macro-F1 focuses on the performance of individual labels, while Micro-F1 emphasizes the consistency in handling the overall samples. In Tables 16 and 17, we provide a detailed comparison between the KNMRS algorithm and five leading multi-label learning algorithms: PPT-CHI, MIFS, MDMR, SCLS, and LRFS. To comprehensively evaluate the performance of these algorithms, we selected six large-scale datasets with significant sample distribution imbalance and sparse characteristics, namely Enron, Art, Science, Social, Education, and Reference, as shown in Table 5. Comprehensive comparisons with existing methods indicate that KNMRS does not blindly pursue the intensity of attribute reduction in multi-label feature selection tasks, but rather focuses more on preserving local features and samples with sparse distributions. Although this may result in selecting a larger number of features, these features can more comprehensively reflect the local structure and key information of the data,

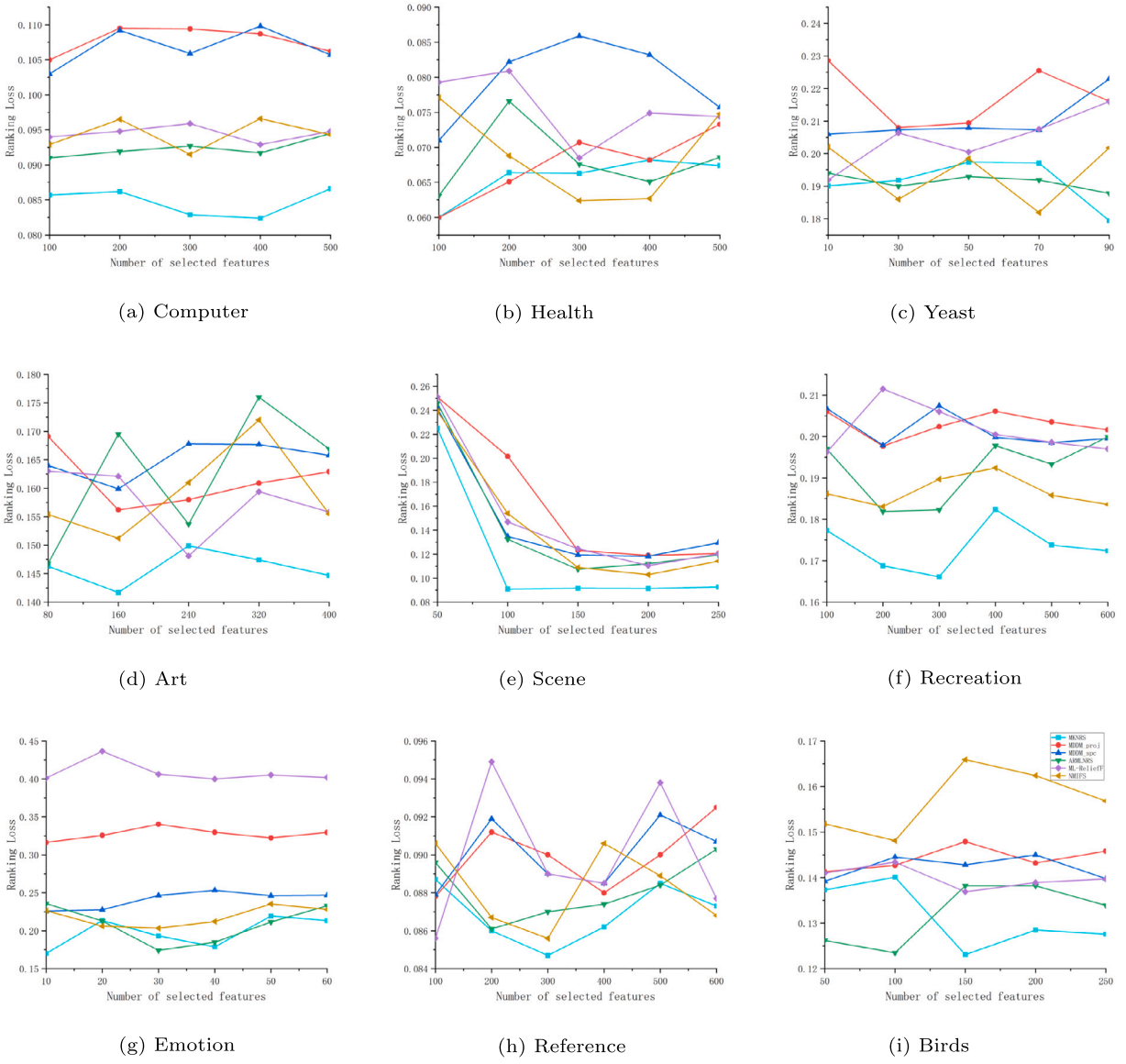


Fig. 5. The performance variation of six algorithms in terms of *RL* across nine datasets.

Table 15
Hamming loss (↓) of the ten algorithms on the nine datasets.

Datasets	MLKNN	MLNB	MDDM _proj	MDDM _spc	ARMLNRS _imp	ARMLNRS _opt	ARMLNRS _pes	ML _ReliefF	NMIFS	KNMRS	Average
Computer	0.041	0.094	0.044	0.043	0.040	0.053	0.049	0.0407	0.0401	**0.0352**	0.048
Health	0.046	0.044	0.041	0.048	0.043	0.0436	0.0426	0.0852	0.0452	**0.041**	0.0479
Yeast	0.201	0.208	0.229	0.229	0.215	0.215	0.215	0.204	0.235	**0.204**	0.2155
Art	0.0612	0.0604	0.0621	0.0614	0.0605	0.0603	0.0603	0.0616	0.0604	**0.0591**	0.0607
Scene	0.099	0.1026	0.0983	0.0968	0.1014	0.1075	0.1104	0.0953	**0.0867**	0.0989	0.0996
Recreation	0.0623	0.0644	0.0638	0.0642	0.0634	0.0641	0.0641	0.0658	0.0637	**0.0597**	0.0635
Emotions	0.214	0.245	0.3139	0.2441	0.2258	0.2347	0.2264	0.2382	0.2496	**0.214**	0.2405
Reference	0.033	0.03	0.031	0.032	0.031	0.033	0.0320	0.0344	0.033	**0.028**	0.0317
Birds	0.054	0.052	0.068	0.067	0.052	0.053	0.0510	0.0534	0.0529	**0.051**	0.0554

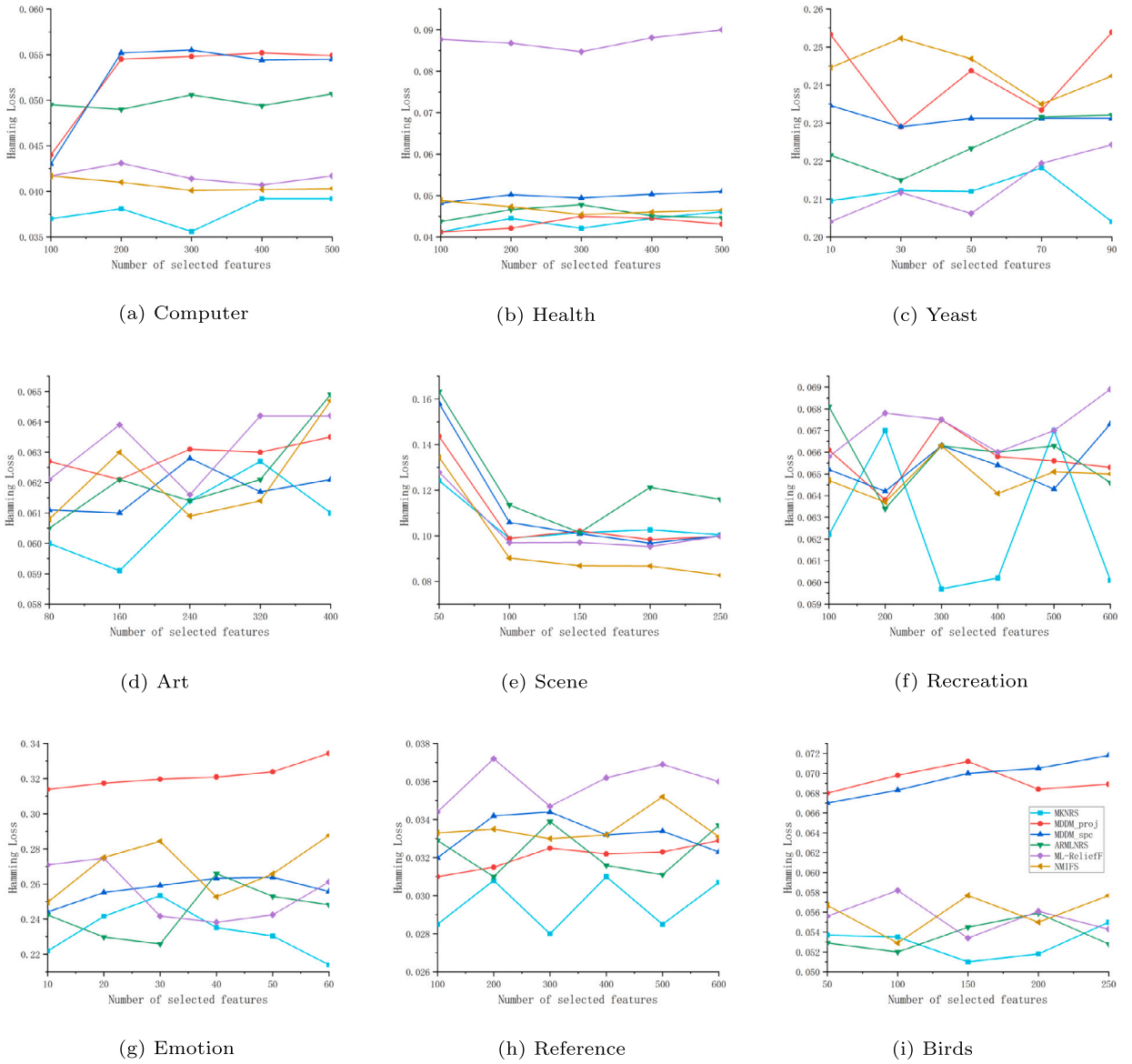


Fig. 6. The performance variation of six algorithms in terms of HL across nine datasets.

Table 16
Macro-f1↑ (Mean ± Std) of six algorithms on six datasets.

Datasets	PPT-CHI	MIFS	MDMR	SCLS	LRFS	KNMRS
Enron	0.062 ± 0.016	0.074 ± 0.017	0.105 ± 0.033	0.114 ± 0.041	0.091 ± 0.033	**0.119 ± 0.045**
Art	0.041 ± 0.032	0.065 ± 0.037	0.036 ± 0.012	0.035 ± 0.026	0.045 ± 0.029	**0.061 ± 0.039**
Science	0.034 ± 0.031	0.031 ± 0.026	0.037 ± 0.019	0.011 ± 0.003	0.037 ± 0.029	**0.069 ± 0.032**
Social	0.084 ± 0.031	0.034 ± 0.016	0.074 ± 0.032	0.035 ± 0.014	0.073 ± 0.031	**0.092 ± 0.019**
Education	0.032 ± 0.026	0.029 ± 0.017	0.034 ± 0.025	0.029 ± 0.015	0.051 ± 0.028	**0.061 ± 0.041**
Reference	0.042 ± 0.018	0.053 ± 0.024	0.042 ± 0.021	0.041 ± 0.017	0.052 ± 0.021	**0.068 ± 0.029**

thereby providing more stable support for subsequent classification tasks. For example, in the Computer and Health datasets, KNMRS dynamically adjusts the neighborhood granularity by integrating δ-neighborhood rough sets and k-nearest neighbor techniques, effectively capturing key features in sparse regions and significantly improving classification performance. In contrast, traditional methods, due to excessive reduction or ignoring local features, perform poorly on imbalanced data. The advantage of KNMRS lies in its ability to adapt to sparse distributions, model label dependencies, and enhance robustness through dynamic neighborhood optimization, thereby addressing the limitations of traditional methods in capturing local features and label interactions when dealing

Table 17
Micro-f1↑ (Mean ± Std) of six algorithms on six datasets.

Datasets	PPT-CHI	MIFS	MDMR	SCLS	LRFS	KNMRS
Enron	0.351 ± 0.019	0.372 ± 0.027	0.474 ± 0.049	0.488 ± 0.037	0.389 ± 0.054	**0.501 ± 0.035**
Art	0.092 ± 0.056	0.132 ± 0.079	0.097 ± 0.023	0.103 ± 0.066	0.112 ± 0.053	**0.172 ± 0.032**
Science	0.114 ± 0.055	0.125 ± 0.053	0.126 ± 0.071	0.039 ± 0.037	0.092 ± 0.031	**0.184 ± 0.072**
Social	0.443 ± 0.108	0.286 ± 0.106	0.475 ± 0.106	0.323 ± 0.12	0.142 ± 0.116	**0.517 ± 0.061**
Education	0.121 ± 0.085	0.075 ± 0.039	0.126 ± 0.061	0.077 ± 0.061	0.173 ± 0.046	**0.273 ± 0.092**
Reference	0.345 ± 0.074	0.351 ± 0.101	0.356 ± 0.069	0.264 ± 0.033	0.401 ± 0.131	**0.411 ± 0.049**

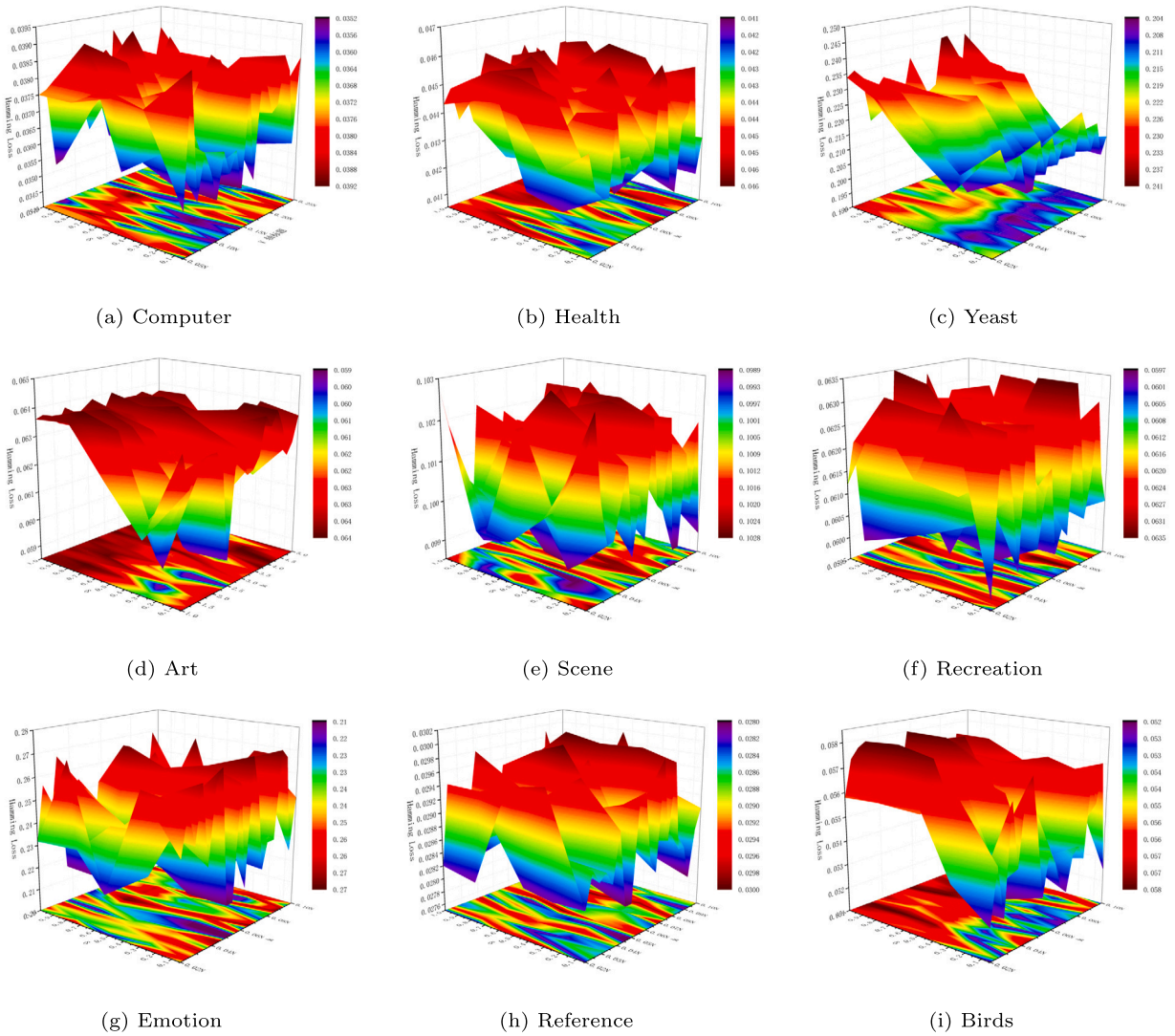


Fig. 7. The changing trend of Hamming loss affected by parameters δ and k on nine data sets.

with imbalanced data. This emphasis on local features not only improves the model's performance but also enables it to excel in practical applications where imbalanced data are common, such as text classification and bioinformatics.

5.3. Parameter classification performance analysis under MLKNN

In this paper, we propose a new algorithm called KNMRS, which includes two neighborhood parameters used to control their respective neighborhood ranges. The parameter δ is used to adjust the size of the δ-neighborhood class, while the parameter k is used to control the size of the k-nearest neighbor neighborhood class. To study the role of these two parameters in feature selection, we conducted experiments on nine datasets and analyzed their parameter sensitivity.

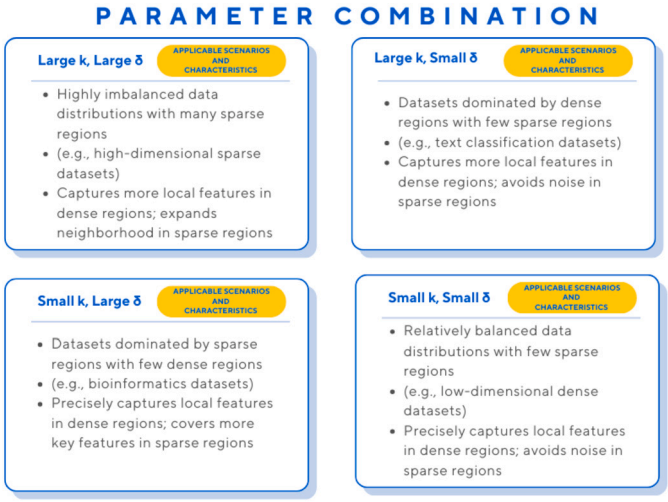


Fig. 8. Parameter combination.

In the mentioned graphical analysis, the X-axis represents the number of nearest samples to x_i. The Y-axis represents the radius size of the δ-neighborhood. The Z-axis measures the Hamming Loss. The Hamming Loss is a key metric for measuring classification performance, and its impact on algorithm performance is particularly significant in our experimental study.

In this round of experiments, the range of δ values was set to [0.05, 1], and the range of k values was set to [0.02 N, 0.1 N], where N represents the number of samples. Fig. 7 illustrates the performance variation of KNMRS in terms of Hamming Loss under different parameter combinations. We can observe that the performance does not consistently improve with increasing parameter values. Instead, it exhibits fluctuations, and we observed the best performance in specific parameter combinations. Hence, it can be inferred that the parameters have a notable impact on the performance of the proposed algorithm. The varying combinations of parameters also affect the results of feature selection.

For commonly occurring sparse datasets such as Recreation and Reference, it is more likely to encounter situations similar to those shown in Fig. 1b, where the best performance occurs with smaller values of the parameter k. However, for matrices that exhibit dense states only at certain points, such as Computer and Health, the parameter k needs to be larger to compensate for the shortcomings of the δ-neighborhood rough set. Thus, the results shown in Fig. 7 corroborate our underlying principles.

Based on the preceding parameter analysis, this paper further explores the impact of different parameter combinations (k and δ) on the performance of the KNMRS algorithm, particularly how to select appropriate parameters for different data distribution characteristics. The number of neighbors in k-nearest neighbors (k) and the distance threshold in δ-neighborhood (δ) are key parameters of the algorithm, and their combination directly affects the algorithm's performance on imbalanced data. We summarize four typical scenarios, analyzing the strengths and limitations of each in imbalanced data, as well as their impact on feature selection and classification tasks. Fig. 8 detailedly demonstrates the characteristics, advantages, and disadvantages of different parameter combinations, as well as their applicable scenarios. Through the above analysis, this paper provides more comprehensive theoretical support for the parameter selection of the KNMRS algorithm and, combined with experimental results, offers clear guidance for parameter tuning.

5.4. Statistical analysis

To rigorously evaluate the statistical performance of the comparative algorithms across various evaluation metrics, we employed the Friedman test followed by the Bonferroni-Dunn post-hoc test. The formulas for these tests are presented below.

$$\chi_F^2 = \frac{12T}{s(s+1)}\left(\sum_{i=1}^{s} R_i^2 - \frac{s(s+1)^2}{4}\right),$$

$$F_F = \frac{(T-1)\chi_F^2}{T(s-1)-\chi_F^2}$$

Let T represent the total number of datasets, s denote the number of methods, and R_i be the average ranking of method i across all datasets. The Friedman test statistic (F_F) follows an F-distribution with degrees of freedom $(s-1)$ and $(s-1)(T-1)$. If the null hypothesis is rejected based on the Friedman test, subsequent post-hoc tests, such as the Bonferroni-Dunn test, are conducted to further analyze the classification performance of the compared methods. The significant differences between methods can be described by the critical difference defined as follows:

The classification performance of each algorithm across all datasets for different classification metrics is ranked, with the best performance assigned a rank of 1, the second-best a rank of 2, and so on. Furthermore, we utilize the Critical Difference (CD) diagram to visually display the correlations between different algorithms. The average ranking of each algorithm's performance is plotted

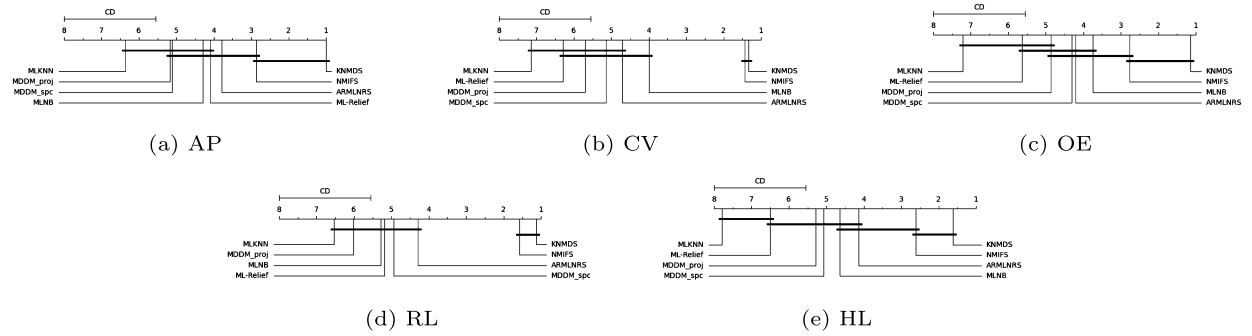


Fig. 9. Bonferroni-Dunn test results for 10 methods under MLKNN classification.

Table 18
Statistical results of five evaluation indicators.

	AP	CV	OE	RL	HL
χ_F	24.7641	21.7544	28.3843	18.0924	13.5397
F_F	7.4501	6.5562	3.7501	3.5653	2.1786

along the coordinate axis, with the best ranking value positioned on the right side of the axis. When the average rankings of different algorithms are equivalent within the error range, they are connected by a thick line; if not within the error range, it is considered that there are significant differences between these algorithms.

$$CD_\alpha = q_\alpha \sqrt{\frac{s(s+1)}{6T}}$$

q_α represents the critical value from the statistical table, which is used to determine the significance of differences between methods. α is the importance level in the Bonferroni-Dunn test.

The average rankings of the KNMRS algorithm and nine other comparative algorithms—MLKNN, MLNB, MDDMM_proj, MDDM_spc, ARMLNRS, ML_Relief and NMIFS on five evaluation metrics, corresponding to two evaluation indicators (XF and FF), are listed in Table 18 based on experimental data from Tables 6 to 15. The corresponding CD diagram is shown in Fig. 9. At $\alpha = 0.1$, $q_\alpha = 2.216$, CD = 2.45, s = 10, T = 12. Fig. 9 demonstrates RFNMIFS significantly outperforms six algorithms across all metrics. In AP, CV, and RL, RFNMIFS outperforms MLKNN, MLNB, MDDM_proj, MDDM_spc, ARMLNRS, and ML_Relief, matching NMIFS. In OE and HL, RFNMIFS outperforms MLKNN, ML_Relief, MDDM_proj, and MDDM_spc, matching ARMLNRS and NMIFS. The KNMRS algorithm demonstrates excellent performance in statistical analysis, reaffirming its superior efficacy in multi-label classification.

6. Conclusions and future work

In this study, we introduced KNMRS, a novel multi-label feature selection algorithm designed to address the challenges of high dimensionality and data sparsity prevalent in complex real-world datasets. The core innovation of KNMRS lies in its unique integration of δ-neighborhood and k-nearest neighbor concepts. This hybrid approach empowers the algorithm to dynamically adapt to the local density and distribution characteristics of the data, effectively mitigating the adverse effects of sparsity and enhancing its capacity to discern relevant features. Unlike traditional methods that often falter in the face of uneven data distributions, KNMRS's adaptive neighborhood mechanism enables it to capture nuanced relationships within the data, leading to improved feature selection accuracy.

Extensive experiments conducted on a diverse suite of 12 multi-label datasets, spanning various domains and scales, validated the effectiveness of the KNMRS algorithm. The results demonstrate that KNMRS exhibits superior performance compared to several state-of-the-art multi-label feature selection methods, particularly in handling sparse data. For instance, on the Computer dataset, KNMRS achieved a 2.1% improvement in Average Precision (AP) compared to MLKNN, while also reducing the Hamming Loss by 0.58%. Furthermore, on the Recreation dataset, KNMRS attained the best performance in terms of Coverage (CV), exhibiting a 0.67% reduction compared to MLKNN. These improvements can be attributed to KNMRS's ability to effectively identify and prioritize informative features, even in regions characterized by limited data points, and its robustness in handling noisy or irrelevant features often encountered in sparse datasets. Moreover, the algorithm's computational efficiency and stability make it a practical and valuable tool for multi-label feature selection.

Specifically, considering the challenges and opportunities presented by the increasing availability of multi-label data in bioinformatics, KNMRS can be instrumental in tasks such as gene function prediction and disease subtyping. By accurately identifying relevant features from complex gene expression profiles, KNMRS can aid in uncovering critical biological insights. Furthermore, in the realm of image classification, where images are often associated with multiple labels depicting various objects or scenes, KNMRS can facilitate the development of more precise and comprehensive image understanding systems. Future research directions include extending the proposed method to other data types such as images, audio, and time series to verify its generality. Additionally,

combining the method with techniques like ensemble learning, cost-sensitive learning, or Generative Adversarial Networks (GANs) can further enhance model performance on imbalanced data. Dynamic neighborhood granularity optimization methods could also be explored to adaptively adjust neighborhood granularity to better fit different data distribution characteristics. These directions will promote further development and application of the method in the field of imbalanced data processing.

While the KNMRS algorithm demonstrates promise for imbalanced multi-label data, certain limitations remain. Its strict attribute reduction criteria may hinder scalability in high-dimensional feature spaces. Future work will explore alternative reduction strategies, such as information-theoretic measures or ensemble approaches, to enhance efficiency. Furthermore, the algorithm's current focus on imbalanced data and specific matrix structures may limit broader applicability. Research will investigate extending its capabilities to diverse data distributions, potentially through adaptive parameter tuning. Finally, while KNMRS demonstrates improved performance, further hyperparameter optimization could yield additional gains. Future research will explore automated tuning techniques like Bayesian optimization. These efforts aim to enhance scalability, broaden applicability, and optimize performance for diverse multi-label applications.

CRediT authorship contribution statement

Weihua Xu: Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Yuzhe Li:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation.

Declaration of competing interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Nos. 62376229 and 12371465) and the Natural Science Foundation of Chongqing (Nos. CSTB2023NSCQ-LZX0027 and CSTB2023NSCQ-MSX1063).

Data availability

No data was used for the research described in the article.

References

[1] Q. Lou, Z. Deng, K. Choi, H. Shen, J. Wang, S. Wang, Robust multi-label relief feature selection based on fuzzy margin co-optimization, IEEE Trans. Emerg. Top. Comput. Intell. 6 (2021) 387–398.
[2] J. Huang, W. Qian, C. Vong, W. Ding, W. Shu, Q. Huang, Multi-label feature selection via label enhancement and analytic hierarchy process, IEEE Trans. Emerg. Top. Comput. Intell. 7 (2023) 1377–1393.
[3] Y. Lin, Z. He, L. Guo, W. Ding, Multi-label feature selection via positive or negative correlation, IEEE Trans. Emerg. Top. Comput. Intell. (2023).
[4] Z. Pawlak, Rough sets, Int. J. Comput. Inf. Sci. 11 (1982) 341–356.
[5] Q. Zhang, Q. Xie, G. Wang, A survey on rough set theory and its applications, CAAI Trans. Intell. Technol. 1 (2016) 323–333.
[6] W. Xu, K. Yuan, W. Li, W. Ding, An emerging fuzzy feature selection method using composite entropy-based uncertainty measure and data distribution, IEEE Trans. Emerg. Top. Comput. Intell. 7 (2022) 76–88.
[7] T. Lin, Neighborhood systems-a qualitative theory for fuzzy and rough sets, Adv. Mach. Intell. Soft Comput. 4 (1997) 132–155.
[8] W. Xu, Q. Bu, Feature selection using generalized multi-granulation dominance neighborhood rough set based on weight partition, IEEE Trans. Emerg. Top. Comput. Intell. (2024).
[9] S. Sun, L. Li, K. Hu, A new approach to rough set based on remote neighborhood systems, Math. Probl. Eng. 2019 (2019) 8712010.
[10] Q. Hu, D. Yu, J. Liu, C. Wu, Neighborhood rough set based heterogeneous feature subset selection, Inf. Sci. 178 (2008) 3577–3594.
[11] W. Li, Z. Huang, X. Jia, X. Cai, Neighborhood based decision-theoretic rough set models, Int. J. Approx. Reason. 69 (2016) 1–17.
[12] Q. Wang, Y. Qian, X. Liang, Q. Guo, J. Liang, Local neighborhood rough set, Knowl.-Based Syst. 153 (2018) 53–64.
[13] C. Wang, M. Shao, Q. He, Y. Qian, Y. Qi, Feature subset selection based on fuzzy neighborhood rough sets, Knowl.-Based Syst. 111 (2016) 173–179.
[14] C. Hu, L. Zhang, B. Wang, Z. Zhang, F. Li, Incremental updating knowledge in neighborhood multigranulation rough sets under dynamic granular structures, Knowl.-Based Syst. 163 (2019) 811–829.
[15] X. Yang, S. Liang, H. Yu, S. Gao, Y. Qian, Pseudo-label neighborhood rough set: measures and attribute reductions, Int. J. Approx. Reason. 105 (2019) 112–129.
[16] S. An, X. Guo, C. Wang, G. Guo, J. Dai, A soft neighborhood rough set model and its applications, Inf. Sci. 624 (2023) 185–199.
[17] T. Yin, H. Chen, Z. Yuan, T. Li, K. Liu, Noise-resistant multilabel fuzzy neighborhood rough sets for feature subset selection, Inf. Sci. 621 (2023) 200–226.
[18] W. Xu, K. Cai, D.D. Wang, A novel information fusion method using improved entropy measure in multi-source incomplete interval-valued datasets, Int. J. Approx. Reason. 164 (2024) 109081.
[19] K. Yuan, D. Miao, W. Pedrycz, H. Zhang, L. Hu, Multigranularity data analysis with zentropy uncertainty measure for efficient and robust feature selection, IEEE Trans. Cybern. (2024).
[20] K. Yuan, D. Miao, W. Pedrycz, W. Ding, H. Zhang, Ze-hfs: Zentropy-based uncertainty measure for heterogeneous feature selection and knowledge discovery, IEEE Trans. Knowl. Data Eng. (2024).
[21] K. Yuan, D. Miao, Y. Yao, H. Zhang, X. Zhao, Feature selection using zentropy-based uncertainty measure, IEEE Trans. Fuzzy Syst. 32 (2023) 2246–2260.
[22] S. Luo, D. Miao, Z. Zhang, Y. Zhang, S. Hu, A neighborhood rough set model with nominal metric embedding, Inf. Sci. 520 (2020) 373–388.
[23] W. Shu, W. Qian, Y. Xie, Incremental feature selection for dynamic hybrid data using neighborhood rough set, Knowl.-Based Syst. 194 (2020) 105516.

[24] H. Ebrahimi, K. Majidzadeh, F. Soleimanian Gharehchopogh, Integration of deep learning model and feature selection for multi-label classification, Int. J. Nonlinear Anal. Appl. 13 (2022) 2871–2883.
[25] M. Cai, M. Yan, P. Wang, F. Xu, Multi-label feature selection based on fuzzy rough sets with metric learning and label enhancement, Int. J. Approx. Reason. 168 (2024) 109149.
[26] R. Wang, J. Xiong, H. Ke, Y. Jia, D.D. Wang, Improving the adversarial robustness of deep neural networks via efficient two-stage training, in: 2023 International Conference on Machine Learning and Cybernetics (ICMLC), IEEE, 2023, pp. 43–49.
[27] Q. Hu, J. Liu, D. Yu, Mixed feature selection based on granulation and approximation, Knowl.-Based Syst. 21 (2008) 294–304.
[28] M.S. Raza, U. Qamar, Feature selection using rough set-based direct dependency calculation by avoiding the positive region, Int. J. Approx. Reason. 92 (2018) 175–197.
[29] I. Park, G. Choi, Rough set approach for clustering categorical data using information-theoretic dependency measure, Inf. Sci. 48 (2015) 289–295.
[30] D.D. Wang, W. Wu, R. Wang, Structure-based, deep-learning models for protein-ligand binding affinity prediction, J. Cheminform. 16 (2024) 2.
[31] D. Guo, W. Xu, W. Ding, Y. Yao, X. Wang, W. Pedrycz, Y. Qian, Concept-cognitive learning survey: Mining and fusing knowledge from data, Inf. Fusion 109 (2024) 102426.
[32] H. Han, B. Mao, Fuzzy-rough k-nearest neighbor algorithm for imbalanced data sets learning, 2010 seventh international conference on fuzzy systems and knowledge discovery, vol. 3, IEEE, 2010, pp. 1286–1290.
[33] Y. Lin, Y. Li, C. Wang, J. Chen, Attribute reduction for multi-label learning with fuzzy rough set, Knowl.-Based Syst. 152 (2018) 51–61.
[34] L. Sun, T. Yin, W. Ding, Y. Qian, J. Xu, Multilabel feature selection using ml-relieff and neighborhood mutual information for multilabel neighborhood decision systems, Inf. Sci. 537 (2020) 401–424.
[35] H. Chen, T. Li, X. Fan, C. Luo, Feature selection for imbalanced data based on neighborhood rough sets, Inf. Sci. 483 (2019) 1–20.
[36] W. Xu, Z. Yuan, Z. Liu, Feature selection for unbalanced distribution hybrid data based on k-nearest neighborhood rough set, IEEE Trans. Artif. Intell. 5 (2023) 229–243.
[37] S. Li, N. Li, Z. Li, Multi-label data mining: A survey, Comput. Sci. 40 (2013) 14–21.
[38] M. Zhang, Feature selection for multi-label naive Bayes classification, Inf. Sci. 179 (2009) 3218–3229.
[39] Y. Zhang, Z. Zhou, Multilabel dimensionality reduction via dependence maximization, ACM Trans. Knowl. Discov. Data 4 (2010) 1–21.
[40] L. Zhang, Q. Hu, J. Duan, X. Wang, Multi-label feature selection with fuzzy rough sets, in: Rough Sets and Knowledge Technology: 9th International Conference, RSKT 2014, Shanghai, China, October 24-26, 2014, Proceedings 9, Springer, 2014, pp. 121–128.
[41] Y. Cai, M. Yang, H. Yin, Relieff-based multi-label feature selection, Int. J. Database Theory Appl. 8 (2015) 307–318.