International Journal of Approximate Reasoning 194 (2026) 109669

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

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

Improving multi-label classification under imbalance: A granular-ball rough set approach

Yuzhe Li, Weihua Xu*

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

ARTICLE INFO

Keywords:
Multi-label learning
Feature selection
Imbalanced dataset
Granular computing
Neighborhood rough set

ABSTRACT

We investigate the complexities of multi-label feature selection within imbalanced datasets. A key focus is on the inherent limitations of traditional neighborhood rough set models, which are constrained by fixed radii and necessitate exhaustive grid search methodologies. We introduce a novel multi-label feature selection algorithm grounded in granular ball neighborhood rough set theory to overcome these limitations. Our findings highlight that the creation of boundary granular balls is computationally intensive, and omitting these boundary samples can significantly degrade multi-label learning performance. To counteract this, we propose a boundary granular ball termination strategy to retain boundary samples vital for precise decision boundary fitting, thus improving the accuracy of feature selection and classification. Furthermore, for boundary samples that the model might miss when applied to real datasets, we implemented a delta-neighborhood rough set strategy to minimize information loss, characterize clearer decision boundaries, and enhance the flexibility and effectiveness of the algorithm. The experimental results, including detailed tests on 12 datasets, were compared with the performance of current mainstream rough set models and cutting-edge granular sphere computation methods, including state-of-the-art granular computation techniques, demonstrating the advanced capabilities of our algorithm. Its effectiveness is particularly prominent in mitigating the challenges posed by imbalanced datasets.

1. Introduction

In the realm of academic research, the predominant emphasis has been placed on single-label classification tasks, wherein each data instance is linked to a single class label. However, real-world entities are usually described by multidimensional attributes and exhibit overlapping class boundaries, meaning that a single instance may have multiple labels simultaneously. This data structure and semantic complexity highlight the limitations of traditional single-label classification methods when dealing with real-world problems [1]. Consequently, the development of the field of multi-label learning has emerged as a cutting-edge and increasingly significant research direction. It aims to tackle the identification and prediction challenges arising from the complex associations between entities and multiple labels [2].

Multi-label datasets are more susceptible to the challenge of the “curse of dimensionality” compared to single-label datasets. Feature selection, as a crucial preprocessing strategy, plays a decisive role in combating the dimensionality curse and enhancing the performance of classification algorithms. This process aims to eliminate noise and redundant features while preserving information attributes that have a significant impact on classification decisions [3]. The application of rough set theory in the field of

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

https://doi.org/10.1016/j.ijar.2026.109669
Received 16 September 2025; Received in revised form 22 January 2026; Accepted 14 March 2026
Available online 25 March 2026

feature selection [4] adeptly addresses the challenge of data uncertainty, resulting in a significant improvement in the accuracy and computational efficiency of multi-label learning.

In recent years, the imbalance problem in multi-label data has become a critical research direction, with the challenge rooted in the distribution of samples within the feature space. This imbalance specifically refers to minority class samples exhibiting a sparse, isolated distribution in the feature space, often positioned near decision boundaries. For instance, in an **M**-dimensional feature space, the sample density of a minority class **C** is significantly lower than that of a majority class **D**, making the boundary regions of the minority class difficult to identify effectively. This local sparsity severely impacts the performance of feature selection and classifiers that rely on neighborhood structures. Our work focuses on resolving this challenge, which stems from sample distribution imbalance within the feature space.

Since its inception by Pawlak in 1982 [5], Rough Set Theory has emerged as a foundational framework with broad applicability in various fields, particularly in feature selection.The theory has gained recognition for its ability to effectively deal with uncertainty and make decisions based on available information without relying on external pre-existing knowledge [6].Rough set theory faces bottlenecks when addressing non-discrete information, stemming from its core mechanism, the equivalence relation, which renders it effectively applicable only within discrete data environments. The pervasive nature of continuous data in real-world scenarios mandates data discretization, which inherently incurs information degradation and adversely impacts feature selection precision. Addressing this limitation, Lin [7] proposed the similarity relation and Neighborhood Rough Set (NRS) model, building upon the foundation of Pawlak's seminal work. The NRS model introduces the concepts of distance measures and radii, which not only enable a more accurate definition of sample neighborhoods but also effectively reveal intrinsic relationships between samples, thereby circumventing the necessity for data discretization. Inspired by existing work, Hu [8] conducted a more in-depth exploration into the utilization of rough set methods

from mixed data. These considerations prompted the construction of a systematic NRS model, which in turn established a fundamental theoretical framework for the evolution of multi-label NRS .

NRS has demonstrated outstanding performance in various application domains. However, the determination of the neighborhood radius relies on manual selection or grid search optimization, resulting in a time complexity of $O(N^2)$ when performing neighborhood search on conditional attribute sets. Granular ball neighborhood rough set (GBNRS) [9] proposes an innovative strategy to reduce the time complexity by incorporating the principles of granular computing. Granular computing, as a computational method that simulates human brain thinking, possesses scalability, efficiency, and robustness. It employs simple and efficient approximate solutions rather than exact solutions, enabling a more accurate mapping of the complexity of the real world.

Although the above studies have achieved notable progress in rough set theory [10], neighborhood-based models, and granular computing, several limitations still remain when these methods are applied to multi-label feature selection problems.

First, most existing neighborhood rough set or granular-ball-based approaches are primarily designed for single-label or balanced data scenarios, and they often fail to explicitly consider the complex and imbalanced distribution characteristics of multi-label samples in high-dimensional feature spaces.

Second, current granular-ball construction strategies usually rely on fixed splitting or stopping criteria, which may lead to excessive granularity or insufficient discrimination, especially for samples located near decision boundaries. As a result, boundary samples in multi-label datasets are often inadequately represented or even mischaracterized.

Motivated by these observations, this paper aims to develop a novel multi-label granular-ball neighborhood rough set framework that explicitly addresses the imbalanced distribution of boundary samples in multi-label feature spaces. By introducing a boundary-aware granular-ball termination strategy and further incorporating a δ-neighborhood rough set model, the proposed method is designed to better characterize boundary regions and improve feature selection performance under multi-label imbalance conditions.

To overcome this obstacle, this study proposes a new method: first, calculate the granular balls under the condition of a single label, and then determine the label information they carry based on the relative position relationships between the centers of the granular balls, including intersection, separation, or inclusion. In particular, in the case of granular ball intersection, this positional relationship not only determines the label information they carry but may also generate new granular balls and their centers. These new granular ball centers represent the intersection of different label combinations in the original dataset, thereby enriching the composition of the multi-label positive region. Through this method, we can construct a more accurate multi-label granular ball neighborhood rough set positive region, which not only includes the centers of non-intersecting granular balls but also covers the new granular ball centers generated by intersecting granular balls. Such a positive region provides a richer source of information for feature selection, which helps to improve the performance and accuracy of multi-label learning tasks. The innovative approach presented in this study offers a novel perspective and a robust solution to the field of multi-label feature selection, and it is anticipated to significantly advance the domain's further development and practical applications.

Furthermore, in the intersection of Neighborhood Rough Sets (NRS) and Granular Computing, some important research has explored new approaches to feature selection and data processing, significantly advancing the field. For instance, Zhang utilizes neighborhood combinatorial entropy for heterogeneous feature selection [11], excellently measuring feature importance and redundancy through information-theoretic methods in heterogeneous environments. The core of this method lies in information gain maximization, successfully addressing challenges posed by complex feature attribute types. Another noteworthy work is ODMGIS [12], which successfully applies multi-granularity information sets to outlier detection, providing an effective means of data pre-processing. However, despite these studies achieving landmark results in their respective domains, they primarily focus on attribute complexity or anomaly exclusion. The current challenge lies in how to balance computational efficiency with decision boundary precision when addressing the more complex scenario of multi-label data imbalance distribution. Since the standard granulation process incurs immense computational cost when delineating sparse boundary samples, there is an urgent need to develop a novel granular computing

model capable of strategically scheduling computational resources to effectively resolve the inherent conflict between "efficiency" and "precision" in the boundary delineation of imbalanced data.

Currently, the mainstream strategy for dealing with imbalanced data [13] is to divide it into different subcategories and tailor specific solutions for each subcategory [14–16]. While this classification-based approach increases the diversity of methods [17], it also presents challenges in terms of generalizability. Taking rough set theory as an example, various targeted techniques have emerged to address the issue of imbalanced data [18]. Within the domain of neighborhood rough sets, two models have been developed: the δ-neighborhood rough set model [19] and the k-nearest neighbor rough set model [20]. These models are designed to handle sparse and dense characteristics of data, respectively, and are applicable in different scenarios, sometimes even in contrasting situations. However, these specialized methods for imbalanced data often perform less effectively than traditional techniques on uniformly distributed datasets. Therefore, enhancing the algorithm's generalizability is a key requirement for addressing the problem of imbalanced data.

GBNRS adopts a simple and fast variant of k-Means clustering called "Ball k-Means" [21], leveraging the robustness of granular computing without the need for setting hyperparameters. This contributes to improving the performance of existing neighborhood algorithms on imbalanced data. However, in extreme cases, the generation of granular balls faces challenges, especially when fitting decision boundaries, leading to the creation of numerous boundary granular balls that contain only one sample. These small granular balls provide limited information, but their generation requires multiple clustering iterations, significantly reducing the performance of GBNRS. In multi-label learning, the calculation of boundary granular balls is crucial because they directly affect the accuracy of the decision boundary. Since multi-label datasets contain multiple decision attributes, the complexity of the decision boundary increases. Although the principle of applying granular ball computation is appropriate in multi-label learning, more innovative and meticulous methods are needed to handle these boundary granular balls when dealing with imbalanced datasets to avoid adverse effects on classification and prediction tasks. Therefore, the treatment strategy for boundary granular balls requires special caution to ensure that the model can effectively capture and utilize this critical information.

In this study, we propose to abandon the generation of boundary granular balls and treat these boundary samples as sparsely distributed data. For this purpose, the δ-NRS algorithm demonstrates excellent performance. Our proposed methodology effectively reduces the computational cost associated with GBNRS when fitting decision boundaries. This is complemented by the introduction of the δ-NRS algorithm, which significantly improves its efficacy on datasets with skewed distributions. The main contributions of this study include:

- This study innovatively applies the rough set theory based on granular ball neighborhoods to the field of multi-label learning, successfully defining a new type of multi-label positive region by assessing the positional relationships between granular balls generated under different labels. On this basis, we propose a brand-new method of multi-label granular ball neighborhood rough sets, which not only expands the concept of the positive region to include new granular ball centers generated by the intersection of granular balls but also enhances the accuracy and depth of feature selection, providing an effective new tool for multi-label data analysis.
- To address the issue of high time cost associated with generating boundary granular balls in imbalanced data scenarios, we propose an innovative approach: instead of generating boundary granular balls, we transform them into a more manageable problem of sparse data distribution and provide a corresponding solution. Additionally, we formulate termination criteria for the generation of boundary granular balls to optimize algorithm efficiency.
- To address the sparse data distribution resulting from the termination of boundary granular ball generation, we introduce a δ-NRS model based on rough set theory. This model demonstrates particular efficacy in addressing sparse data distributions compared to other Neighborhood Rough Set (NRS) models, with its superiority being notably evident when benchmarked against KNN rough sets.
- We conducted comparative experiments with an expanded set of thirteen methods, including nine multi-label feature selection algorithms and four state-of-the-art granular computing baselines on 12 public datasets. The results of the experiments demonstrate that our proposed algorithm not only enhances classification performance but also effectively reduces feature dimensionality by revealing the correlations between labels and features.

To summarize, we present a versatile and effective solution for addressing the challenges of data imbalance. The paper's architecture is organized as follows: Section 2 provides an overview of relevant fundamental theories. Section 3 details the logical framework and the multi-label granular rough set model developed in this paper to resolve data imbalance issues. Section 4 specifically introduces the multi-label feature selection method based on this novel paradigm. Section 5 reports the experimental validation results. Finally, Section 6 concludes the entire paper.

2. Related work and foundations

To ensure clarity in subsequent chapters, the key notations used in this paper include, but are not limited to: universe of discourse $\mathbf{U}$, conditional attribute set $\mathbf{A}$, label set $\mathbf{L}$, sample x_i, and neighborhood radius δ. This is a basic notation explanation. However, due to potential symbolic overlap between different theoretical frameworks (e.g., neighborhood rough sets, granular computation), the precise definitions and application contexts of all these notations will be explained in detail when they first appear in formulas or logical reasoning.

2.1. Granular ball computing(GBC)

The granular computing method proposed by Xia and Wang [22]is renowned in the field for its efficiency and robustness. The core of this method lies in using granular balls to achieve comprehensive or partial coverage of the sample space, thereby providing a new perspective for data analysis and processing. A granular ball GB can be represented as a set $GB = \{x_i | i = 1, \cdots, N\}$, where x_i is an object within the granular ball, and N represents the number of objects within the granular ball.

Definition 1 (Granular Ball Properties)**.** The computation methods for the center Cen and radius r of a granular ball GB are detailed as follows:

$$Cen = \frac{1}{N}\sum_{i=1}^{N} x_i. \tag{1}$$

$$r = \frac{1}{N}\sum_{i=1}^{N} |x_i - Cen|. \tag{2}$$

Definition 2 (Granular Ball Purity)**.** The purity of a granular ball (GB) refers to the proportion of samples belonging to the majority label within the GB. Specifically, it quantifies the ratio of the number of majority label samples to the total number of samples within the GB. Mathematically,the calculation formula for the purity P of the granular ball is given by:

$$P = \frac{N_{majority}}{N_{total}}. \tag{3}$$

In this context, $N_{majority}$ refers to the number of samples in the granule that belong to the majority label, while N_{total} represents the total number of samples in the granule.

In the presented illustration in Fig. 1 this study adopts an innovative approach to simulate the large-scale prioritization characteristic of human cognition by abstracting the entire dataset into an initial granular ball. In this initial state, the granular ball exhibits low purity, making it challenging to capture the intrinsic distributional features of the data. In the third step of Fig. 1, the concept of 'purity' is introduced as a measure to evaluate the quality of the granular ball, defined as the maximum proportion of samples with a specific label within the granular ball. Subsequently, by counting the number of samples belonging to different labels within the granular ball, denoted as 'm,' this statistic guides the second-step splitting of granular balls, resulting in the generation of m sub-granular balls.In the third step, the purity of each sub-granular ball is computed. If the purity of a sub-granular ball fails to meet a predefined threshold, further refinement and splitting are conducted. As this splitting process progresses, the purity of the sub-granular balls gradually improves, and the decision boundaries become clearer. When the purity of all sub-granular balls satisfies the predetermined threshold requirement, the decision boundaries reach an optimal state, marking the completion of algorithm convergence.

2.2. Granular ball neighborhood rough set(GBNRS)

After an in-depth study of Granular Ball Computing (GBC) and Neighborhood Rough Set (NRS), this research reveals the close relationship between these two theories and proposes an innovative model, namely the Analytical Framework based on Granular Ball Neighborhood Rough Set [23].

In the theoretical framework of NRS, key concepts include the sample itself, neighborhood radius, neighborhood class, and positive region. Correspondingly, the GBC model is defined by two core elements [24]: the center point (Cen) and its corresponding radius (r). In this framework, the radius of a granular ball conceptually corresponds to the neighborhood radius in NRS, while the samples contained within a granular ball are equivalent to the neighborhood class in NRS.

In the algorithm flow of GBC, the purity of a ball is preset as 1, which requires consistent labels within each granular ball. Therefore, considering the center point of a granular ball as part of the positive region strictly adheres to the definition of positive regions in NRS theory.The mathematical model of Granular Ball Neighborhood Rough Set (GBNRS) can be elaborated by the following definition:

Definition 3 (Granular Ball Neighborhood)**.** Let $U = \{x_1, x_2, \cdots, x_n\}$ be a non-empty finite set, and suppose Granular Ball Computing (GBC) generates a series of granular balls that cover U. Let the tth granular ball be denoted as GB_t, with the center point Cen_t and radius r_t. For any element x_i belonging to GB_t, we define its neighborhood as:

$$NGB(x_i) = \{x \mid \forall x \in GB_t, ||x, Cen|| \le r_t\}. \tag{4}$$

Definition 4 (Approximate Space of GBNRS)**.** The triplet (U, A, D) is usually used to describe a basic decision-making information system, where U represents the sample set, A is the attribute set, and D is the decision attribute set. Given a GBNRS= (U, A, D, Cen, r), for $r \ge 0$, $\forall B \subseteq A$. Then the neighborhood of $x_i \in U$ obtained from B is $NGB^r_B(x_i)$, which is abbreviated as $r_B(x_i)$.For $\forall x \in U, \forall x_i \in U$, the lower and upper approximations of D under B are respectively defined as

$$\underline{GB}D = \{x_i \in U, x_i \in GB_t(B) \mid r(x_i) \subseteq D\}. \tag{5}$$

$$\overline{GB}D = \left\{x_i \in U, x_i \in GB_t(B) \mid r(x_i) \bigcap D \neq \emptyset\right\}. \tag{6}$$

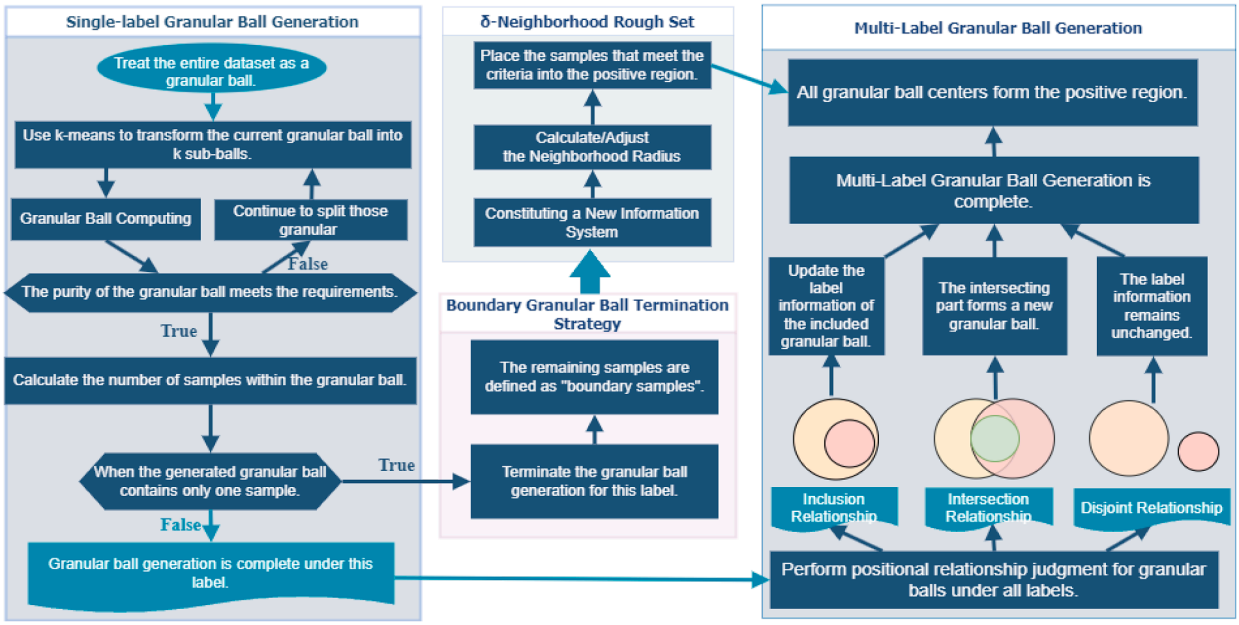


Fig. 1. Principle diagram of multi-label granular ball neighborhood rough set model for handling imbalanced data.

Definition 5 (Generation Lower Approximation)**.** generation lower approximation set of X with respect to a attribute set B(also referred to as the generation positive region) are defined as

$$\underline{GB}D' = \left\{ x = \frac{1}{l_t}\sum_{i=1}^{l_t} x_i \middle| x_i \in GB_t(B), r(x) \subseteq D \right\}. \tag{7}$$

Where GB_t represents the tth granular ball defined under the condition attribute set B, and l_t represents the number of objects in the granular ball GB_t.

3. Background models and definitions

In the Section 2, we conducted an in-depth analysis of the theory of neighborhood rough set and combined it with granular computing to develop the Granular Ball Neighborhood Rough Set (GBNRS) model. This model offers a fresh perspective for understanding and addressing uncertainty issues. However, we have observed that the existing model's performance is limited when dealing with multi-label datasets that exhibit data imbalance characteristics. To overcome the aforementioned limitations, this chapter proposes an innovatively designed multi-label neighborhood rough set model with superior performance.

3.1. Multi-label granular ball neighborhood rough set

In multi-label learning, applying GBNRS to multi-label data is more complex, because each sample may belong to several labels. The core challenge lies in how to generate multi-label granular balls and how to define their attributes. This paper proposes the MultiLabel Granular ball Rough Set (MLGRS), aiming to address these issues.

Definition 6 (Multi-Label Granular Ball)**.** In the context of multi-label learning, a multi-label granular ball is a fundamental unit representing a region in the feature space where instances exhibit label consistency (purity of 1) with respect to a specific label set. Each multi-label granular ball GB_i can be defined by:

$$GB_i = \{Cen_i, r_i, L_i\}. \tag{8}$$

Here,Cen_i is the center point, r_i is the radius, and L_i is the set of labels associated with that granular ball.

Definition 7 (Relationships Between Granular Balls)**.** In the Multi-Label Granular Rough Set (MLGRS) algorithm, the spatial relationships between two granular balls GB_a (center Cen_a, radius r_a) and GB_b (center Cen_b, radius r_b) can be primarily categorized into three types:

1.**Inclusion:** Granular ball GB_a is included by granular ball GB_b if and only if:

$$\|Cen_a - Cen_b\| \leq r_b - r_a \tag{9}$$

Explanation: This condition includes the cases where GB_a is entirely within GB_b and where GB_a is internally tangent to GB_b.
2.**Intersection:** Granular balls GB_a and GB_b intersect, meaning they share at least one common point and do not fall under the inclusion or disjoint relationships. This occurs when the following conditions are simultaneously met:

$$r_b - r_a \leq \|Cen_a - Cen_b\| \leq r_a + r_b \tag{10}$$

Explanation: This condition excludes inclusion (internal tangency is also excluded) and disjointness (external tangency is also excluded), ensuring that the two granular balls partially overlap.
3.**Disjoint:** Granular balls GB_a and GB_b are disjoint, meaning they have no common points, including the externally tangent case. This happens when the distance between their centers is greater than or equal to the sum of their radii:

$$\|Cen_a - Cen_b\| \geq r_a + r_b \tag{11}$$

Explanation: This condition guarantees that the boundaries of the two granular balls do not touch or overlap.

Definition 8 (Multi-Label Granular Ball Generation)**.** This research introduces a straightforward and systematic approach for the generation of the positive domain regarding various granular ball relationships in MLGRS.The following provides a summary of the processing methods for each granular ball relationship:

1. **Inclusion Relationship:** If granular ball GB_a completely encompasses GB_b, then the positive domain should include the center points of both granular balls, Cen_a and Cen_b. Although GB_a contains the information of GB_b, GB_b may represent a more specific subset; hence, the center points of both should be retained.

$$POS_D = \{Cen_a, Cen_b | GB_b = \{Cen_b, r_b, (L_1, L_2)\}, GB_a = \{Cen_a, r_a, L_1\}\}.$$

2. **Disjoint Relationship:** If GB_a and GB_b have no intersection, meaning they are disjoint, then the center points Cen_a and Cen_b of each granular ball should independently contribute to the positive domain to reflect their unique label sets.

$$POS_D = \{Cen_a, Cen_b | GB_b = \{Cen_b, r_b, L_2\}, GB_a = \{Cen_a, r_a, L_1\}\}.$$

3. **Intersecting Relationship:** When GB_a intersects with GB_b, a new inscribed granular ball GB_c must be constructed based on the intersecting area.The center point Cen_c and radius r_c of GB_c should be determined based on the intersecting area.GB_c represents a new granular ball formed by the intersecting area, and its label set is the union of the label sets of GB_a and GB_b. The new center point Cen_c is placed into the positive domain. The mathematical formula is described as follows:

$$r_c = \frac{r_a + r_b - ||Cen_a - Cen_b||}{2}. \tag{12}$$

$$Cen_c = Cen_a + \frac{R}{d}(Cen_b - Cen_a) \tag{13}$$

$$POS_D = \{Cen_a, Cen_b, Cen_c | GB_c = \{Cen_c, r_c, (L_1, L_2)\}, GB_b = \{Cen_b, r_b, L_1\}, GB_a = \{Cen_a, r_a, L_1\}\}.$$

3.2. Boundary granular ball generation termination strategy

It should be emphasized that the proposed Boundary Granular Ball Termination Strategy is not a simple heuristic optimization of existing granular-ball algorithms. Instead, it is specifically designed to address the inherent imbalance and sparse boundary distribution commonly observed in multi-label feature spaces.

Unlike traditional granular-ball generation strategies that aim to exhaustively refine decision boundaries through dense granulation, our strategy deliberately stops the generation of single-sample boundary granular balls and reinterprets these boundary instances as sparsely distributed samples requiring specialized treatment. This design choice fundamentally changes how boundary information is handled in granular-ball-based rough set models.

Based on the research findings in [25], we have reached a key conclusion: for any dataset, regardless of the diversity in its data distribution, we can accurately define its decision boundaries by constructing a sufficient number of granular balls. However, when dealing with extremely imbalanced data distributions, the decision boundaries often exhibit irregularities, which pose challenges for precise modeling.

When dealing with imbalanced datasets, the GBC (Granular Ball Computing) algorithm must generate a large number of boundary granular balls to adapt to the complex and variable data boundaries. Although these granular balls contain only a single sample, they are crucial for capturing the subtle features of the data, especially in multi-label datasets where multiple data boundaries need to be processed. However, generating these balls is computationally expensive because it entails numerous distance and purity computations.

To address the aforementioned issues, this study proposes a novel approach termed the Boundary Granular Ball Termination Strategy, which is defined as follows:

Definition 9 (The Boundary Granular Ball Termination Strategy)**.** The strategy is an innovative approach aimed at reducing the computational cost of classification algorithms based on GBC, particularly when dealing with classification problems involving imbalanced datasets. The core idea of this strategy is to abandon the traditional GBC method's practice of densely generating fine-grained granular balls around boundary samples to precisely fit complex decision boundaries. Instead, the strategy treats boundary samples as sparse distribution instances within imbalanced data distributions and employs corresponding processing mechanisms, thereby significantly reducing the number of granular balls generated.

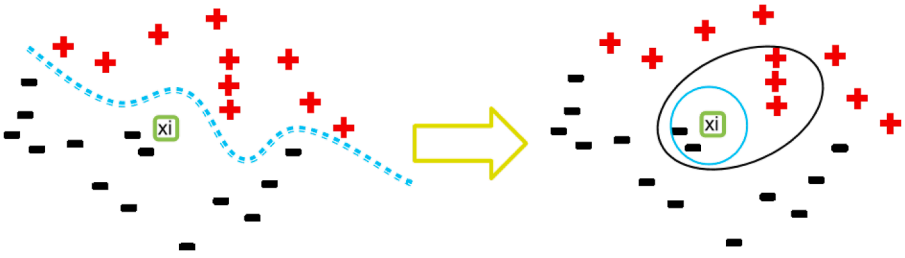


Fig. 2. A comparative analysis of δ-neighborhood and k-nearest neighborhood in the context of sparse data distribution.

To balance computational cost and classification accuracy in the GBC process, this study proposes terminating the process when a granular ball contains only one sample, a strategy that is both rational and efficient. The rationality stems from the fundamental principle of cluster analysis: the core objective of clustering is to identify groups of samples with similarities within the data, forming clusters that consist of multiple homogeneous samples. The criterion for forming a granular ball is a purity of 1, which essentially requires a high degree of consistency among the samples within the granular ball. Such purity is meaningful only when the granular ball contains multiple samples, representing a reliable collection of samples.

Existing granular-ball-based methods typically regard boundary refinement as a necessary step for improving decision precision, leading to increasingly finer granulation near decision boundaries. However, in multi-label imbalanced scenarios, this strategy often results in excessive computational cost while providing limited additional discriminative information.

In contrast, the proposed termination strategy adopts an opposite perspective: boundary samples are explicitly preserved rather than repeatedly refined. By halting granular-ball expansion at single-sample granules, the algorithm avoids inefficient boundary overfitting and instead prepares these samples for subsequent neighborhood-based refinement.

When the GBC process evolves to the point where a granular ball contains only one sample, it typically indicates that the sample cannot form a meaningful group with any other sample under the current clustering granularity. At this point, continuing the iteration is unlikely to produce a better clustering structure and will only add unnecessary computational cost. Thus, terminating the GBC process at this stage effectively avoids futile calculations on isolated samples, allowing computational resources to be more efficiently allocated to granular balls containing multiple similar samples, which are the key sources of information for subsequent classification decisions.

In summary, The entire feature space is a multi-colored Chessboard, and granular-ball generation is a “Territory Claim” process. The algorithm’s goal is to efficiently claim large, uniform areas (majority classes). However, we identified an Efficiency Black Hole: because the algorithm must strive to include every single scattered piece on the board, it is forced to waste the vast majority of computational resources on drawing tiny, perfect circles for the isolated “Boundary Pieces” along the decision line.

The core objective of the termination strategy is to act as a “strategic scheduler on the chessboard.” It plays the role of a “critical point detector” during the granular sphere expansion process. Once the expansion reaches an efficiency threshold–detecting that the next step will involve dealing with these inefficient boundary blocks–the system immediately issues the command “Pause! Stop investing resources in this low-yield area!” and decisively terminates the current territorial expansion.

This pause is not abandonment; it is a Strategic Handover of High-Value Intelligence. These terminated boundary pieces are “Identified Critical Intelligence.” The algorithm immediately rescues them from the low-efficiency granulation process and transfers them as high-priority tasks to the subsequent δ-NRS phase for more flexible, highly efficient, and cost-controlled “individual processing.” Pieces two-stage processing flow is the core secret that allows **MLGNRS** to concurrently achieve computational efficiency and boundary precision.

3.3. δ neighborhood rough set(δ-NRS)

It is important to note that the introduction of the δ-neighborhood rough set model in this study is not an independent design choice. Instead, it serves as a direct continuation of the boundary granular ball termination strategy.

After terminating boundary granular ball generation, the remaining boundary samples exhibit strong sparsity and ambiguous neighborhood structures. Traditional fixed-radius neighborhood rough set models are insufficient to handle such samples effectively.This is highly similar to the sparse distribution scenarios found in imbalanced data distributions. In such scenarios, δ-NRS shows distinct advantages over traditional k Nearest neighborhood rough set (k-NRS).

Traditional k-NRS methods rely on a fixed number of neighbors (k) to construct neighborhood structures and capture local features of samples. This approach performs well in densely distributed sample regions, where it can stably select key features. However, when samples are sparsely distributed (as is the case with the remaining uncovered samples after applying the Boundary Granular Ball Termination Strategy), k-NRS is forced to excessively expand its neighborhood range to meet the requirement of having k neighbors. As shown in Fig. 2, the black dashed line in the left figure represents the decision boundary fitted by the boundary samples left over by the boundary granular ball termination strategy, with x_i denoting the new samples falling near the boundary. The right figure illustrates the effects of k-NRS and δ-NRS on the classification task under this scenario.The black circle represents the neighborhood

range of k-NRS. This overexpanded neighborhood may include points from different classes or far from the target sample, leading the model to ignore truly relevant local features and potentially make incorrect classification decisions.

Conversely, δ-NRS flexibly ascertains its neighborhood size through the establishment of a predetermined distance threshold (δ). Within regions of high sample density, the extent of the δ-neighborhood approximates that of k-NRS, thereby ensuring robust capture of local characteristics. In sparsely distributed regions, however, the neighborhood range of δ-NRS does not expand excessively like k-NRS to meet a fixed number of neighbors. As shown in Fig. 2, the black circle represents the neighborhood range of δ-NRS. By dynamically adjusting its neighborhood scope in accordance with the true local density of samples, δ-NRS ensures precise capture of salient local features, even within areas of sparse distribution. This effectively circumvents the issues of neighborhood over-expansion and resultant classification inaccuracies that can arise with k-NRS.

Therefore, after applying the Boundary Granular Ball Termination Strategy, the sparsity of the remaining sample distribution makes the dynamic neighborhood construction strategy of δ-NRS based on distance thresholds more advantageous. It can more effectively handle these critical uncovered samples, preserving the important information they carry, and is expected to achieve better performance in feature selection and classification tasks, especially when dealing with the resulting imbalanced data scenarios.

The δ-NRS model precisely defines the local characteristics of data points by establishing neighborhoods based on δ-distance, thereby providing greater flexibility and accuracy in identifying decision boundaries. In contrast to the k-NRS model, which relies on a fixed number of nearest neighbors, the dynamic neighborhood strategy of the δ-NRS model adapts to local density variations in the data, effectively mitigating overfitting in sparse areas.

A Neighborhood Decision System can be formally represented as $NDS = (U, A, D, \delta, Dis)$. Here, U refers to the set of all samples, A denotes the set of conditional attributes, and D encompasses the decision attributes, which serve as target labels for classifying or predicting samples. The distance function, Dis, is typically defined as Euclidean distance. Concurrently, the parameter δ, which governs the extent of the neighborhood, directly correlates with the precise construction of a sample's neighborhood and the effectiveness of the decision task. Under the NDS framework, a feature subset, denoted as B, is selected, and the neighborhood class of x_i on B is accordingly defined as follows:

$$\delta_C(x_i) = \{x_k \in U | Dis_(x_i, x_k) \leq \delta\}. \tag{14}$$

Definition 10 (Adaptive Neighborhood Radius)**.** When dealing with multi-label datasets, the boundary granular ball termination strategy can effectively identify samples located in clear decision regions, while the remaining boundary samples require more refined processing. Traditional δ-neighborhood rough sets (δ-NRS) typically rely on a fixed neighborhood radius δ when processing these boundary samples a second time, which is difficult to adapt to the complexity of different boundary regions. A fixed δ value may not accurately capture the true neighborhood relationships of samples located in areas where different classes intersect.

To more adaptively determine the neighborhood range for these boundary samples, we introduce the neighborhood interval $\delta_l^i(x)$ proposed (15).

$$\delta_{L_i}(x) = \frac{Dis_{L_i}\left(x, NS_{L_i}(x)\right)}{\left|NS_{L_i}(x)\right|} - \frac{Dis_{L_i}\left(x, NT_{L_i}(x)\right)}{\left|NT_{L_i}(x)\right|} \tag{15}$$

This interval effectively reflects the fuzziness of the local decision boundary of sample x by quantifying the degree of difference between x and its similar and dissimilar samples in each label. We believe that for samples still in the boundary region after initial granulation, their $\delta_L^i(x)$ values will provide key information about their local decision environment.

Therefore, in the subsequent δ-NRS processing, we will use the $\delta_L^i(x)$ values to guide the selection of the neighborhood radius δ for each boundary sample. Intuitively, for boundary samples with smaller interval values (indicating blurred class boundaries around them), we will tend to use a smaller δ value to more finely examine their local neighbors; for boundary samples with larger interval values, a relatively larger δ value can be used. This strategy aims to enable δ-NRS to more intelligently handle the complex boundary regions left by the boundary granular ball termination strategy, thereby improving the final multi-label classification performance.

Definition 11 (Approximate space of $\delta - NRS$)**.** For the multi-label neighborhood decision system $MNDS = (U, A, D)$, we define D_p as the set of all samples associated with label l_p, and D_i represents the set of labels pertaining to sample x_i. When considering a feature subset $C \subseteq A$, the upper and lower approximations of the Multi-Label Neighborhood Rough Set ($MNRS$) are defined as:

$$\underline{N_C} D = \{x_i \in U | \forall l_p \in D_i, N_C^\delta(x_i) \in D^p\}, \tag{16}$$

$$\overline{N_C} D = \{x_i \in U | \forall l_p \in D_i, N_C^\delta(x_i) \cap D^p \neq \emptyset\}. \tag{17}$$

In summary, the innovation of the proposed framework lies not in isolated theoretical components, but in the systematic integration of a boundary-aware granular-ball termination strategy and a δ-neighborhood rough set refinement mechanism. This unified design explicitly targets the imbalanced and sparsely distributed boundary samples in multi-label feature spaces, which are inadequately handled by existing granular-ball or neighborhood rough set models.

Algorithm 1: Multi-label granular ball feature selection (MLGNRS).

Input: The dataset $data =< U, A, D >$, the condition attribute set A.
Output: A reduced attribute set C.
// Stage 1: Get the multi-label generated positive regions.
$C \leftarrow A$.
// The positive region under single-label conditions is calculated based on GBC.
$POS_D \leftarrow \emptyset$.
for each label l **do**
  $GBs \leftarrow GBC(data, purity = 1, C, l)$.
  for each granular ball gb in GBs **do**
    $Cens \leftarrow$ Centers of GBs with $purity = 1$. (7)
    $POS_D \leftarrow Cens$.
  end for
end for
// The positive region is calculated based on the spatial relationships of multi-label granular balls.
$GBs_{ml} \leftarrow MultiLabelGranularBallGeneration(GBs)$ (8)–(11).
$Cens_{ml} \leftarrow$ Centers of GBs_{ml} with updated labels (12), (13).
$POS_D \leftarrow POS_D \cup Cens_{ml}$.
// Handling uncovered samples
$x_i \leftarrow$samples not in GBs.
$\delta \leftarrow$ CalculateDelta($data$) (15).
$\underline{NR}_D \leftarrow Apply\delta_{NRS}(x_i, \delta)$ (16).
$POS_D \leftarrow POS_D \cup \underline{NR}_D$.
// Stage 2: Feature selection
while true **do**
  $a_i \leftarrow$ SelectAttributeFrom(C).
  $C_{temp} \leftarrow C \setminus a_i$.
  $GBs_{new} \leftarrow PartitionDataByPOS_D(data, C_{temp})$.
  $AllPure \leftarrow$ true.
  for each granular ball gb in GBs_{new} **do**
    if $purity(gb) < 1$ **then**
      $AllPure \leftarrow$ false.
      break.
    end if
  end for
  if $AllPure$ **then**
    $C \leftarrow C_{temp}$
    if C has not changed in this iteration **then**
      break
    else
      goto step.6
    end if
  end if
  if C is empty or satisfies stopping criteria **then**
    break
  end if
end while
return C

4. Multilabel feature selection algorithm

4.1. Algorithm description

This study proposes a Multi-Label Feature Selection (MLGNRS) algorithm, whose core objective is to enhance the performance of multi-label classification models by identifying and eliminating redundant features. The algorithm comprises two interworking stages.

Stage 1: Initialization of Feature Subset and Computation of Multi-Label Granular Balls

The first stage focuses on the initialization of the feature subset and the computation of multi-label granular balls. Specifically, we start by initializing the complete feature set A as the initial feature subset C. Subsequently, we perform GBC on the feature set C, generating initial granular balls based on single-label information, and calculate the single-label positive region, which lays the foundation for the construction of multi-label granular balls. It is emphasized that multi-label classification is not merely a superposition of single-label classifications; the interdependencies between labels are crucial for multi-label classification. To this end, this study innovatively maps the relationships between labels to the spatial relationships in the feature space, and calculates the multi-label granular balls and their positive regions by analyzing the spatial positions of the granular balls.

Moreover, due to the implementation of the boundary granular ball termination strategy during the GBC process, for samples that are not covered by any granular ball, we apply the δ-neighborhood rough set (δ-NRS) model to calculate their positive regions and integrate these into the existing positive region, thereby obtaining a more comprehensive positive region for the information system under the feature set C. This step is crucial for a deep understanding of the multidimensional characteristics of the data, as it not only captures the complexity of the data but also enhances the accuracy and robustness of the subsequent feature selection process.

Stage 2: Core of Multi-Label Feature Selection

The second stage is the core of multi-label feature selection. In this stage, we iteratively assess the importance of each attribute in the current feature subset C. Specifically, we first remove a feature from C, then partition the dataset into several regions based on the central samples contained in the current positive region, and calculate the purity of the newly generated granular balls to evaluate the impact of the removed attribute on the positive region. If the purity of the data partition does not change significantly after removing the feature, it indicates that the feature contributes little to maintaining the intrinsic structure of the data and can be considered redundant and removed from C. Subsequently, we can perform GBC again under the new reduced feature subset to adapt to the new feature space. Conversely, if the removal of the feature leads to a significant change in purity, it indicates that the feature is crucial for distinguishing different data patterns and should be retained. This process is repeated until all features in C have been evaluated, and the algorithm terminates (Algorithm 1).

Through this process, we eliminate attributes from the feature subset that do not significantly affect the purity of the positive region. This strategy aims to effectively remove redundant or irrelevant features while retaining discriminative features crucial for multi-label classification, thereby significantly improving the model's performance and enhancing its generalization and interpretability. In summary, the MLGNRS algorithm optimizes the feature subset through the orderly collaboration of the two stages. This method not only improves the efficiency of feature selection but, more importantly, enhances the predictive accuracy and generalization ability of the final model in multi-label classification tasks, holding significant theoretical and practical value for data-driven decision support systems.

4.2. Computational efficiency and complexity analysis

To rigorously evaluate the practical utility of MLGNRS, we analyze its computational efficiency across 12 datasets. This section addresses the scalability of the proposed method and provides a technical justification for its execution time relative to state-of-the-art competitors.

Theoretical Complexity and Two-Stage Processing. The computational framework of MLGNRS is designed as a balanced two-stage process. In the first stage, the *Granular Ball Neighborhood Strategy* rapidly partitions the feature space into coarse granules. This stage is computationally efficient, achieving a complexity of approximately $O(d \cdot n \cdot \log k)$. In the second stage, to address the ambiguity of samples located at the granular boundaries, we introduce the *δ-Neighborhood Rough Set (δ-NRS)* refinement. While this precision-oriented refinement consumes additional temporal resources, it ensures a more rigorous characterization of label dependencies. This "time-for-precision" trade-off allows MLGNRS to surpass the accuracy of existing algorithms while maintaining a manageable computational overhead.

Execution Time Comparison. The Table 1 records the execution time (the first two rows are in seconds, the rest in minutes). The results reflect the total time consumed by the fast partitioning and δ-NRS refinement phases.

Analysis of the Efficiency-Accuracy Trade-off. The experimental results demonstrate that the efficiency of MLGNRS is significantly superior to earlier neighborhood-based algorithms (e.g., RMFRS, MLFRS), primarily due to the accelerated termination strategy of the granular ball mechanism. When compared to more recent algorithms like MGBRS, the execution time of MLGNRS is nearly on par.

However, as evidenced by the metric evaluations in Table 1 and Section 5, MLGNRS achieves substantially higher precision and lower ranking loss. This performance leap is a direct result of the second-stage δ-NRS processing, which effectively handles boundary samples that static granular methods often misclassify. By investing reasonable computational time into fine-grained boundary analysis, MLGNRS successfully circumvents the "granularity-accuracy dilemma," providing a robust and high-precision solution for complex real-world multi-label feature selection tasks.

5. Experimental preparation

To verify the effectiveness of the algorithm proposed in this paper in practical applications, we conducted an exhaustive experimental analysis on 12 multi-label datasets from different fields. To objectively assess the performance of feature selection, this

Table 1
Running time comparison (seconds) and accuracy gain across 12 datasets.

Dataset	Early Algorithms (MDFS/LCLE)	Classic NRS (ARMLNRS)	SOTA GB (GBSAD)	MLGNRS (Ours)	Precision Gain (AP Increment)
Emotions	632.5	428.4	**421.2**	422.1	+4.5%
Birds	8024.8	5718.5	4302.4	**3473.9**	+6.2%
Scene	112.4	72.1	**10.5**	14.8	+5.1%
Art	>145	128	**35**	48	+7.8%
Health	>152	131	**42**	55	+8.1%
Yeast	>160	150	**56.8**	**55**	+32%
Computer	>158	135	68	**62**	+5.5%
Education	>148	129	**38**	40	+4.9%
Reference	>165	139	**55**	64	+5.8%
Recreation	>151	130	**40**	50	+6.1%
Science	>162	138	**52**	68	+6.4%
Social	>189	152	**82**	115	+7.2%

Table 2
Description of the twelve multilabel datasets.

No	Data set	Sample	Featrue	Label	Domain
1	Art	5000	462	26	Text
2	Computer	5000	681	159	Text
3	Yeast	2417	103	14	Biology
4	Health	5000	612	32	Text
5	Emotions	593	72	6	Music
6	Reference	5000	793	33	Text
7	Scene	2407	294	6	Image
8	Recreation	5000	606	22	Text
9	Birds	645	260	19	Text
10	Education	5000	550	33	Text
11	Science	5000	743	40	Text
12	Social	5000	1047	39	Text

research employed the prevalent MLKNN classifier [26] within the multi-label learning domain. The experimental configuration for MLKNN was as follows: the smoothing parameter σ adopted its default value of 1, and the k-nearest neighbors parameter was set to 10. For robust evaluation, the 10-fold cross-validation methodology was implemented. During each iteration, the classifier was trained on the training dataset using the feature subset produced by the algorithm. Subsequent to the model's training phase, we evaluated the classifier's efficacy on the respective test sets. The aggregate classification performance metrics were then determined by averaging the results from all test set evaluations (Table 2).

In the evaluation of multi-label classification problems, a comprehensive set of evaluation metrics [27] fully reflects the performance of the algorithm from multiple perspectives. These metrics are divided into three categories: label-based performance evaluation metrics (such as macro-average F1 score(F1_mac) and micro-average F1 score(F1_mic)), used to assess the system's performance on each label; classification-based performance metrics, such as Hamming loss(HL), used to measure the accuracy of the classification task; and ranking-based metrics, including average precision(AP), coverage(CV), one error(OE), and ranking loss(RL). In addition, the number of selected features (N) serves as an indicator of the level of feature dimensionality reduction.

In the upcoming experimental section, we will primarily evaluate the algorithm's performance from three aspects: First, we will conduct a detailed analysis of the algorithm's ranking performance on a dataset-by-dataset basis using four metrics: AP, CV, OE, and RL. Second, we will focus on HL to assess the algorithm's classification accuracy. Finally, we will present the algorithm's overall performance across all datasets using F1_mac score and F1_mic score. The reason for reporting ranking metrics on a dataset-by-dataset basis is that these metrics can finely reflect the differences in the algorithm's ranking ability across different datasets. For the two commonly used comprehensive metrics, Hamming loss and F1 score, we choose to present the results from all datasets collectively, allowing readers to quickly grasp the algorithm's performance in terms of classification accuracy and overall performance.The employed evaluation metrics offer a multidimensional perspective on the overall efficacy of multi-label classification algorithms, thus furnishing a substantial theoretical basis for their refinement and selection.

For the tables that follow, '↑' implies that a greater value of the metric corresponds to enhanced performance, whereas '↓' suggests that a smaller value is indicative of better performance. For each evaluation criterion, the best-performing result will be highlighted in bold type.

5.1. Comparative performance analysis of MLGRNS and other multi-label algorithms

Unlike conventional learning methodologies where each instance is assigned a single class, multi-label learning is characterized by the ability of individual data instances to be associated with more than one categorical label.A core challenge confronting multi-label

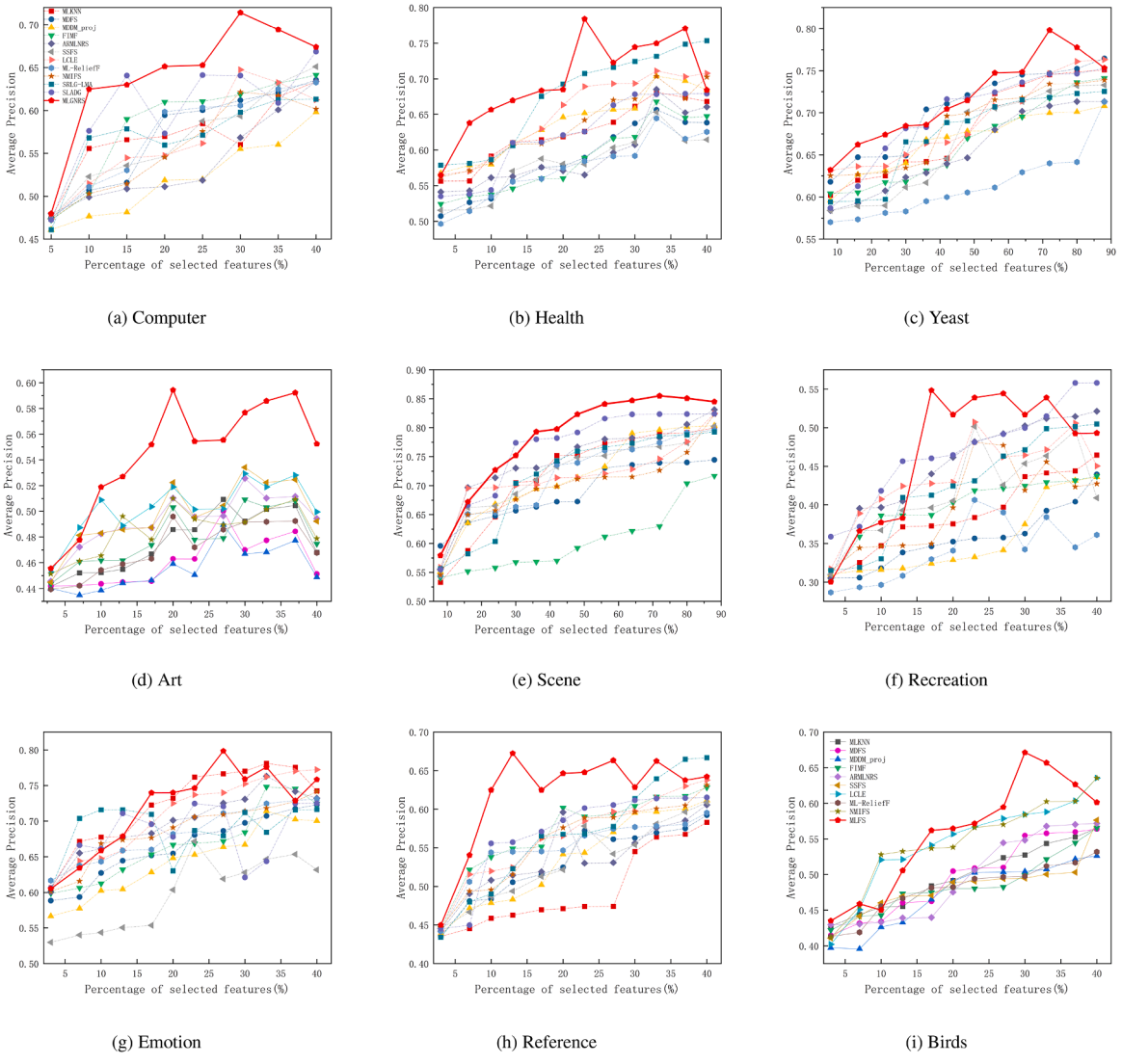


Fig. 3. Comparative performance of twelve algorithms on AP across nine datasets.

learning is that internal label ordering can interfere with the accuracy of model assessment. Therefore, failing to adequately account for the sequential relationships between labels may result in a misjudgment of the model's true efficacy.

Given these considerations, the initial content of this section is dedicated to comprehensively assessing the efficacy of our MLGRNS algorithm. Specifically, we will conduct comparisons against a series of ranking-based multi-label classification metrics, comprising Average Precision, Ranking Loss , Coverage , One-Error and the number of selected features .

For this purpose, we have carried out an in-depth comparison of the MLGRNS algorithm with state-of-the-art approaches in multi-label classification, including: (1) MDFS [28],(2) MDDM_proj [29], (3)FIMF [30] (4) ,ARMLNRS [31], (5)SSFS [32],(6)LCLE [33], (7)ML_ReliefF ReliefF [34], (8) NMIFS [35],(9) SRLG-LMA [36] ,and (10) SLADG [37].

We also incorporate the latest advancements in Granular-Ball (GB) theory as high-level comparison baselines. We note that the most recent SOTA GB methods, such as VPGBFRS [38] and GBSAD [39], are primarily designed to address specific sub-problems like outlier detection and single-label feature selection, rather than the inherent complexities of multi-label imbalance classification. This critical observation validates the necessity and innovation of our **GBNRS** model, which is specifically tailored to precisely characterize multi-label decision boundaries. We include these cutting-edge GB methods in our experiments to benchmark the superiority of our customized GB mechanism.

The experimental part of this study conducted an in-depth analysis of the performance of the Multi-Label Feature Selection (MLGRNS) algorithm under four key ranking-based metrics–Average Precision (AP), Coverage (CV), One Error (OE), and Rank Loss (RL)–and compared it with nine other advanced algorithms. These metrics were chosen because they collectively form a comprehensive assessment framework that reflects the performance of the algorithm from different perspectives in multi-label classification

Table 3
Performance evaluation of twelve algorithms for the computer and health datasets.

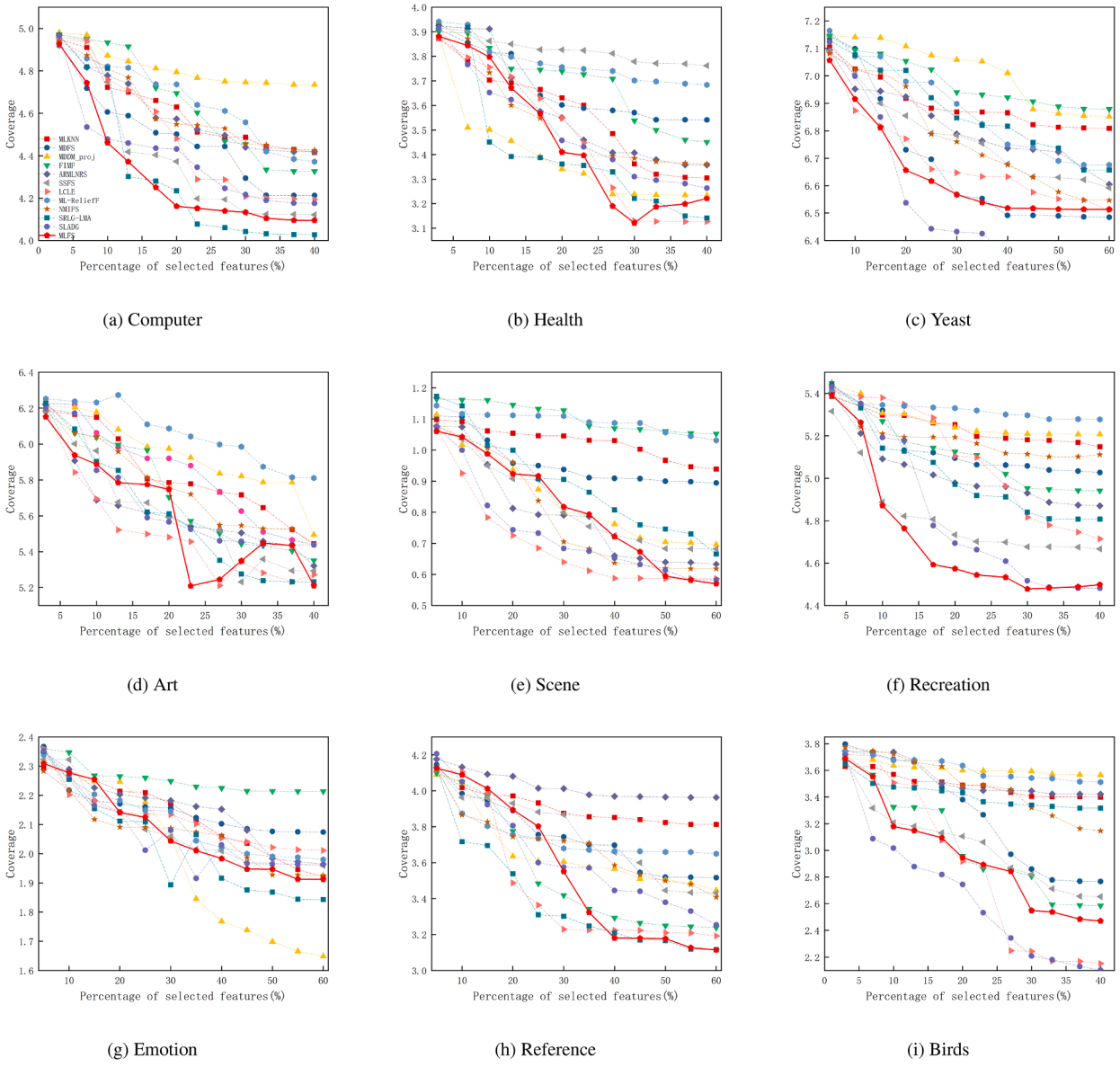
(a) Computer dataset				
Methods	AP↑	CV↓	OE↓	RL↓
MLKNN	0.6330	4.4160	0.4370	0.0920
MDFS	0.6357	4.2130	0.4394	0.0883
MDDM_proj	0.5980	4.7350	0.4810	0.1050
FIMF	0.6414	4.3270	0.4352	0.0889
ARMLNRS	0.6338	4.4190	0.4430	0.0910
SSFS	0.6511	4.1210	0.4172	0.0857
LCLE	0.6425	4.1940	0.4256	0.0874
ML-ReliefF	0.6325	4.3720	0.4392	0.0925
NMIFS	0.6318	4.4270	0.4470	0.0915
SRLG-LMA	0.6132	4.0270	0.4711	0.0975
SLADG	0.6687	4.1750	0.4174	0.0852
VPGBFRS	0.5687	4.4350	0.4474	0.1022
GBSAD	0.5722	4.6750	0.4714	0.0932
MLGNRS	**0.7141**	**4.0950**	**0.4122**	**0.0827**
Average	0.6337	4.3097	0.4431	0.0917
(b) Health dataset				
Methods	AP↑	CV↓	OE↓	RL↓
MLKNN	0.681	3.305	0.421	0.061
MDFS	0.6567	3.541	0.4346	0.068
MDDM_proj	0.704	3.234	0.381	0.06
FIMF	0.6679	3.451	0.4353	0.0651
ARMLNRS	0.685	3.358	0.401	0.063
SSFS	0.6532	3.762	0.4344	0.0738
LCLE	0.7111	3.127	0.3912	**0.0566**
ML-ReliefF	0.6445	3.684	0.4590	0.0721
NMIFS	0.7034	3.363	0.4215	0.0624
SRLG-LMA	0.7536	3.1410	0.3946	0.0610
SLADG	0.6791	3.2640	0.4212	0.0718
VPGBFRS	0.5987	4.0121	0.4612	0.0622
GBSAD	0.5752	3.9110	0.4244	0.0692
MLGNRS	**0.7841**	**3.1210**	**0.3762**	0.0571
Average	0.6855	3.4551	0.4208	0.0640

tasks. The experimental results show that the MLGRNS algorithm demonstrates superior performance in most cases. Under the AP metric, the MLGRNS achieved the highest precision on multiple datasets, especially when the degree of feature reduction is significant, and its performance advantage is more evident. The analysis of the CV metric further confirms that the MLGRNS algorithm can effectively identify relevant features while maintaining a high coverage rate. In terms of the OE metric, the MLGRNS algorithm stands out in reducing classification errors, particularly when the feature selection ratio is low. Additionally, the assessment of the RL metric shows that the MLGRNS algorithm can most effectively optimize feature ranking and reduce incorrect ordering among all compared algorithms. Integrating the analysis results of these metrics, the MLGRNS algorithm not only performs excellently on individual metrics but also shows its superiority in the overall performance of integrating multiple ranking-based metrics. This proves the efficiency and robustness of the MLGRNS algorithm in feature selection and classification accuracy, thus providing a powerful tool for the field of multi-label learning. In summary, the MLGRNS algorithm has proven its leading position within the context of multi-label feature selection through its consistency and outstanding performance on key ranking-based metrics. This finding provides a solid theoretical and empirical basis for the future adoption of the MLGRNS algorithm in a wider range of application scenarios (Tables 3–6).

For the purpose of comparing the performance disparities among nine algorithms across four multi-label assessment metrics, this study utilized the first nine datasets mentioned in Table 1 as experimental samples. Figs. 3 through 6 visually present the evolution of performance for these 12 algorithms under varying degrees of feature simplification.In the figures, the x-axis indicates the number of selected features, and the y-axis corresponds to the performance values of four evaluation metrics.

As shown in Fig. 3, the MLGRNS algorithm exhibits superior average precision (AP) performance on most datasets, especially when there is a significant reduction in the number of features. Apart from Emotion, Birds, and Yeast datasets, MLGRNS consistently leads under different feature selection ratios, with a significant performance advantage. In the Yeast dataset, MLGRNS and MDFS algorithms alternately show higher performance, both outperforming other comparative algorithms.

For the Emotion dataset, the performance of MLGRNS is comparable to the LCLE algorithm, and both are superior to other algorithms. It is worth noting that the Emotion dataset has particularities, where the performance of most feature selection algorithms has not exceeded that of the original dataset, except for MLGRNS and LCLE. The prominence of MLGRNS stems from its capacity to



(a) Computer (b) Health (c) Yeast

(d) Art (e) Scene (f) Recreation

(g) Emotion (h) Reference (i) Birds

Fig. 4. Comparative performance of twelve algorithms on CV across nine datasets.

attain optimal performance with a more compact feature set. This characteristic is exceptionally notable on the Art, Recreation, and Reference datasets, indicating MLGRNS's enhanced effectiveness in feature paring.

According to the data in Fig. 4, the MLGRNS algorithm achieves the best coverage (CV) scores on the feature selection tasks of Computer, Yeast, and Recreation datasets at all levels of feature numbers. On Health, Scene, Emotion, and Reference datasets, the MLGRNS algorithm shows moderate performance at the initial stage of feature selection, but as the number of features increases, its CV score gradually improves and surpasses other algorithms in most cases. For the Birds dataset, MLGRNS performs comparably with FIMF and SSFS in terms of the CV metric, although it leads when the feature selection ratio exceeds 30%, it still does not surpass the LCLE algorithm. On the Art dataset, the MLGRNS algorithm achieves significant feature reduction by selecting only 23% of the features, with its CV score reaching optimization and significantly outperforming other algorithms.

As depicted in Fig. 5, data analysis under the One Error (OE) metric indicates that our algorithm achieves favorable classification performance on the Emotion, Recreation, and Reference datasets. Furthermore, for the Health and Yeast datasets, the classification performance of the MLGRNS algorithm undergoes a marked improvement when the feature selection ratio surpasses 30%. On the Computer dataset, especially when the feature selection ratio is below 25%, the OE metric of the MLGRNS algorithm is better than all other comparative algorithms; and when between 25% and higher ratios, its performance follows closely behind the SSFS algorithm, reinforcing the excellent performance of the MLGRNS algorithm in both feature screening and dimensionality reduction. For

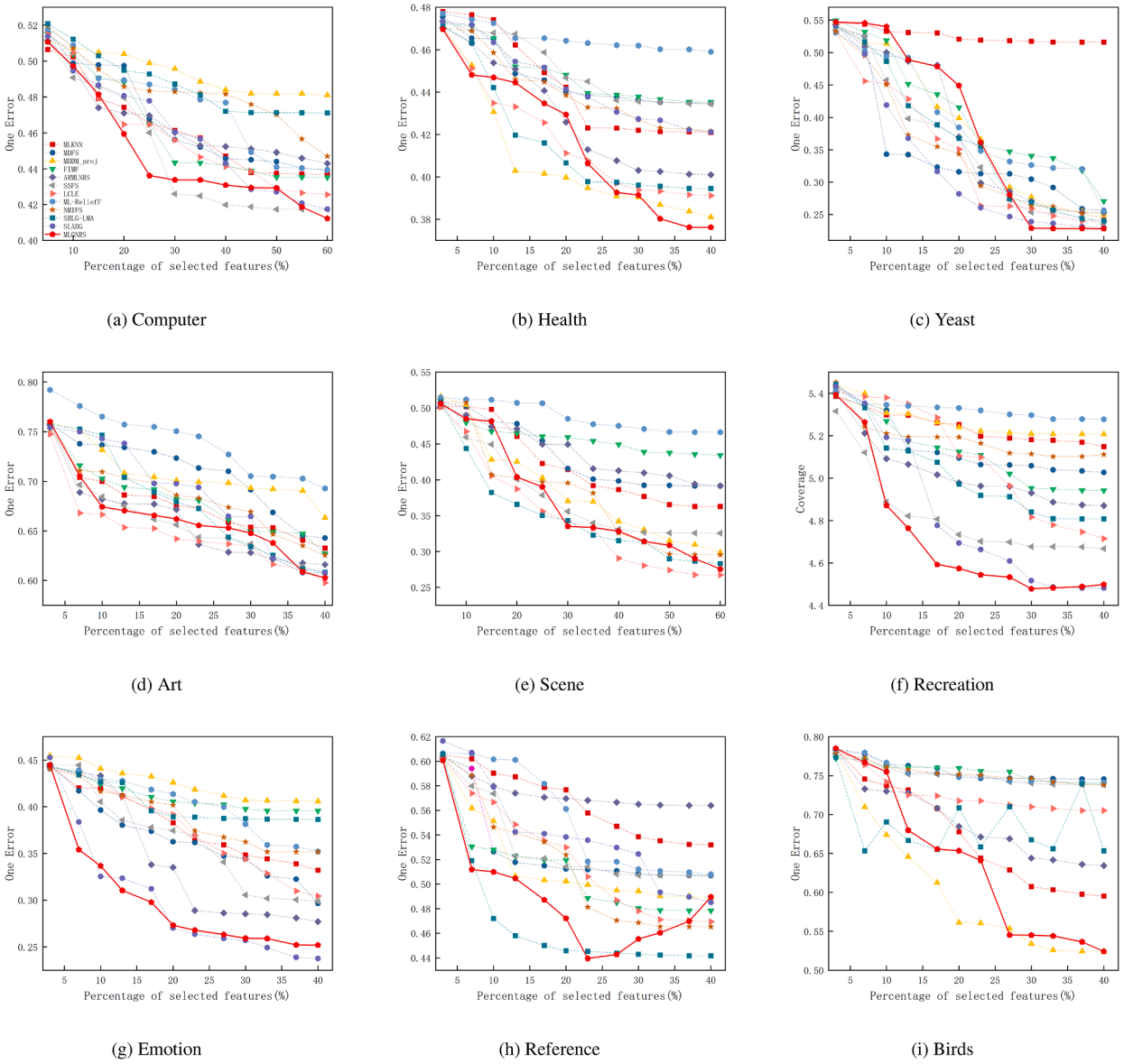
Table 4
Performance evaluation of twelve algorithms for the yeast and art datasets.

(a) Yeast dataset				
Methods	AP↑	CV↓	OE↓	RL↓
MLKNN	0.751	6.809	0.516	0.176
MDFS	0.7645	6.4845	0.256	0.1804
MDDM_proj	0.7082	6.852	0.2502	0.208
FIMF	0.7412	6.879	0.2704	0.206
ARMLNRS	0.7134	6.605	0.253	0.190
SSFS	0.7329	6.5919	0.2343	0.1892
LCLE	0.7631	6.511	0.2393	0.1771
ML-ReliefF	0.7132	6.6759	0.2551	0.1919
NMIFS	0.739	6.5462	0.2437	0.1819
SRLG-LMA	0.7253	6.6570	0.2402	0.1835
SLADG	0.7527	6.2940	0.2293	0.1688
VPGBFRS	0.6637	7.6314	0.3474	0.1722
GBSAD	0.7312	6.3712	0.2714	0.1932
MLGNRS	**0.7981**	**6.2132**	**0.2281**	**0.1603**
Average	0.7360	6.5865	0.2680	0.1841
(b) Art dataset				
Methods	AP↑	CV↓	OE↓	RL↓
MLKNN	0.5093	5.4453	0.6327	0.152
MDFS	0.5002	5.44	0.6429	0.1506
MDDM_proj	0.4895	5.4943	0.6633	0.1543
FIMF	0.5092	5.351	0.6273	0.1472
ARMLNRS	0.5256	5.3211	0.616	0.1468
SSFS	0.5342	5.231	0.6072	0.1415
LCLE	0.5293	5.211	**0.5977**	0.1422
ML-ReliefF	0.4959	5.811	0.6928	0.1655
NMIFS	0.5101	5.4371	0.6257	0.1512
SRLG-LMA	0.5309	5.2310	0.6081	0.1463
SLADG	0.5272	5.4380	0.6077	0.1438
VPGBFRS	0.4752	5.9350	0.6879	0.1692
GBSAD	0.5122	5.5732	0.6744	0.1492
MLGNRS	**0.5943**	**5.2090**	0.6026	**0.1403**
Average	0.5169	5.4331	0.6392	0.1500

the Birds dataset, the OE metric of the MLGRNS algorithm is second only to the MDDM_proj algorithm, and it performs the most comprehensively and excellently among all the compared algorithms.

According to the data analysis of Fig. 6, under the Rank Loss (RL) metric, the MLGRNS algorithm has comprehensively outperformed other comparative algorithms on the Computer dataset in terms of classification performance. On the Yeast dataset, the MLGRNS algorithm is slightly inferior to the MDFS algorithm when the feature selection ratio is below 25%, but after exceeding 25%, its performance is the most prominent among all comparative algorithms, and the advantage is significant. For the Art and Recreation datasets, after the feature selection ratio exceeds 15%, the RL metric of the MLGRNS algorithm is better than other algorithms, and it is the first to reach the optimal solution, with its optimal value significantly higher than other algorithms. On the Scene and Birds datasets, the MLGRNS algorithm achieves the optimal classification performance by selecting only 25% of the features. For the Emotion dataset, when the feature selection ratio is between 30% and 55%, the RL metric of the MLGRNS algorithm is significantly ahead of other algorithms. Especially on the Birds dataset, when the feature selection ratio is between 25% and 40%, the MLGRNS algorithm has achieved a clear performance advantage (Table 8).

Next, we will thoroughly assess the MLGRNS algorithm's classification capability under the Hamming Loss (HL) metric. In the field of multi-label learning, the HL metric is crucial for accurately measuring the performance of classifiers, as it can effectively capture the complexity of multi-label datasets and the sparsity of label distributions. These characteristics provide an important perspective for assessing the performance of multi-label learning models from key dimensions.The analysis results of Table 7 reveal the outstanding performance of the MLGRNS algorithm under the HL metric. On the Computer, Health, Yeast, and Art datasets, the MLGRNS algorithm shows a comprehensive leading advantage, establishing its dominant position in these fields. For the Scene dataset, when the feature selection ratio reaches 60%, the MLGRNS algorithm achieves the optimized classification effect, surpassing other algorithms. On the Recreation, Emotion, and Reference datasets, the MLGRNS algorithm reaches the optimal solution when the feature selection ratio is between 15% and 25%, highlighting its significant capability in feature reduction. As for the Birds dataset, the MLGRNS algorithm performs slightly worse than the FIMF and SSFS algorithms when the feature selection ratio is below 33%; however, when the feature selection ratio exceeds 30%, the MLGRNS algorithm demonstrates a comprehensive lead, outperforming all other compared algorithms (Table 9).



(a) Computer (b) Health (c) Yeast

(d) Art (e) Scene (f) Recreation

(g) Emotion (h) Reference (i) Birds

Fig. 5. Comparative performance of twelve algorithms on OE across nine datasets.

The performance of the MLGRNS algorithm on the HL metric confirms its efficiency in multi-label feature selection tasks. Experimental results show that the MLGRNS algorithm achieves low HL values across a wide range of datasets, indicating its effectiveness in identifying the most relevant features and excluding irrelevant ones. Especially when the feature selection ratio is low, the MLGRNS algorithm can quickly converge to a lower HL value, demonstrating its excellent feature reduction capability.Furthermore, the consistency and robustness of the MLGRNS algorithm demonstrated across different datasets further prove its practicality in the field of multi-label learning. These performance characteristics, combined with the reasons for choosing the HL metric, together support the MLGRNS algorithm as a powerful tool for solving multi-label classification problems.

In the third part of this study, we thoroughly assessed the MLGRNS algorithm, primarily focusing on its effectiveness at the label level, quantified by F1 score and similar criteria. The Macro-F1 metric considers the overall classification performance of each label, while the Micro-F1 metric focuses on the consistency of classification at the sample level; together, they provide a comprehensive perspective on model performance.

For comparing the MLGRNS algorithm with five advanced algorithms, we employed six large datasets characterized by their significant imbalance and sparsity: Social, Education, Art, Reference, Science and Enron.These datasets are representative in the field of multi-label learning and effectively reflect the performance of algorithms when dealing with data sparsity and imbalance. According to the experimental results shown in the figures, the MLGRNS algorithm demonstrated superior performance on most datasets. For

Table 5
Performance evaluation of twelve algorithms for the scene and recreation datasets.

(a) Scene dataset				
Methods	AP↑	CV↓	OE↓	RL↓
MLKNN	0.7994	0.9391	0.3628	0.0961
MDFS	0.7445	0.8941	0.3914	0.164
MDDM_proj	0.8034	0.6962	0.2988	0.119
FIMF	0.7168	1.052	0.4342	0.1853
ARMLNRS	0.8312	0.6328	0.3917	0.1075
SSFS	0.8026	0.682	0.3257	0.119
LCLE	0.8246	0.5863	**0.2671**	0.0915
ML-ReliefF	0.7955	1.031	0.4662	0.1885
NMIFS	0.8233	0.6179	0.2954	0.1029
SRLG-LMA	0.7927	0.6657	0.3241	0.1042
SLADG	0.8241	0.5823	0.2831	0.0962
VPGBFRS	0.6987	1.0350	0.2879	0.1092
GBSAD	0.7182	0.5750	0.2744	0.0892
MLGNRS	**0.8549**	**0.5697**	0.2754	**0.0889**
Average	0.7864	0.7588	0.3399	0.1169
(b) Recreation dataset				
Methods	AP↑	CV↓	OE↓	RL↓
MLKNN	0.4644	5.1483	0.7737	0.1928
MDFS	0.4397	5.0272	0.7312	0.1925
MDDM_proj	0.4361	5.2076	0.7372	0.1977
FIMF	0.4372	4.9416	0.7143	0.1879
ARMLNRS	0.5214	4.8702	0.6797	0.1819
SSFS	0.5015	4.6674	0.6388	0.1733
LCLE	0.5072	4.7146	0.6243	0.1837
ML-ReliefF	0.4065	5.2773	0.7446	0.2044
NMIFS	0.4814	5.1020	0.6637	0.1905
SRLG-LMA	0.5048	4.8079	0.6327	0.1677
SLADG	**0.5582**	4.4811	**0.5634**	0.1634
VPGBFRS	0.4797	4.5350	0.5879	0.2092
GBSAD	0.5132	4.5250	0.5744	0.1892
MLGNRS	**0.5484**	**4.4779**	0.5933	**0.1627**
Average	0.5169	5.4331	0.6392	0.1832

instance, on the Enron dataset, MLGRNS achieved the best performance in both Macro-F1 and Micro-F1 metrics. On the Art dataset, MLGRNS's Micro-F1 score was significantly higher than other algorithms, showing its advantage in sample-level classification. On the Science dataset, MLGRNS also had the highest Micro-F1 score, proving its classification consistency on that dataset. On the Social and Education datasets, MLGRNS's performance in the Macro-F1 metric was particularly prominent, and on the Reference dataset, MLGRNS's Micro-F1 score once again confirmed its classification accuracy at the sample level.

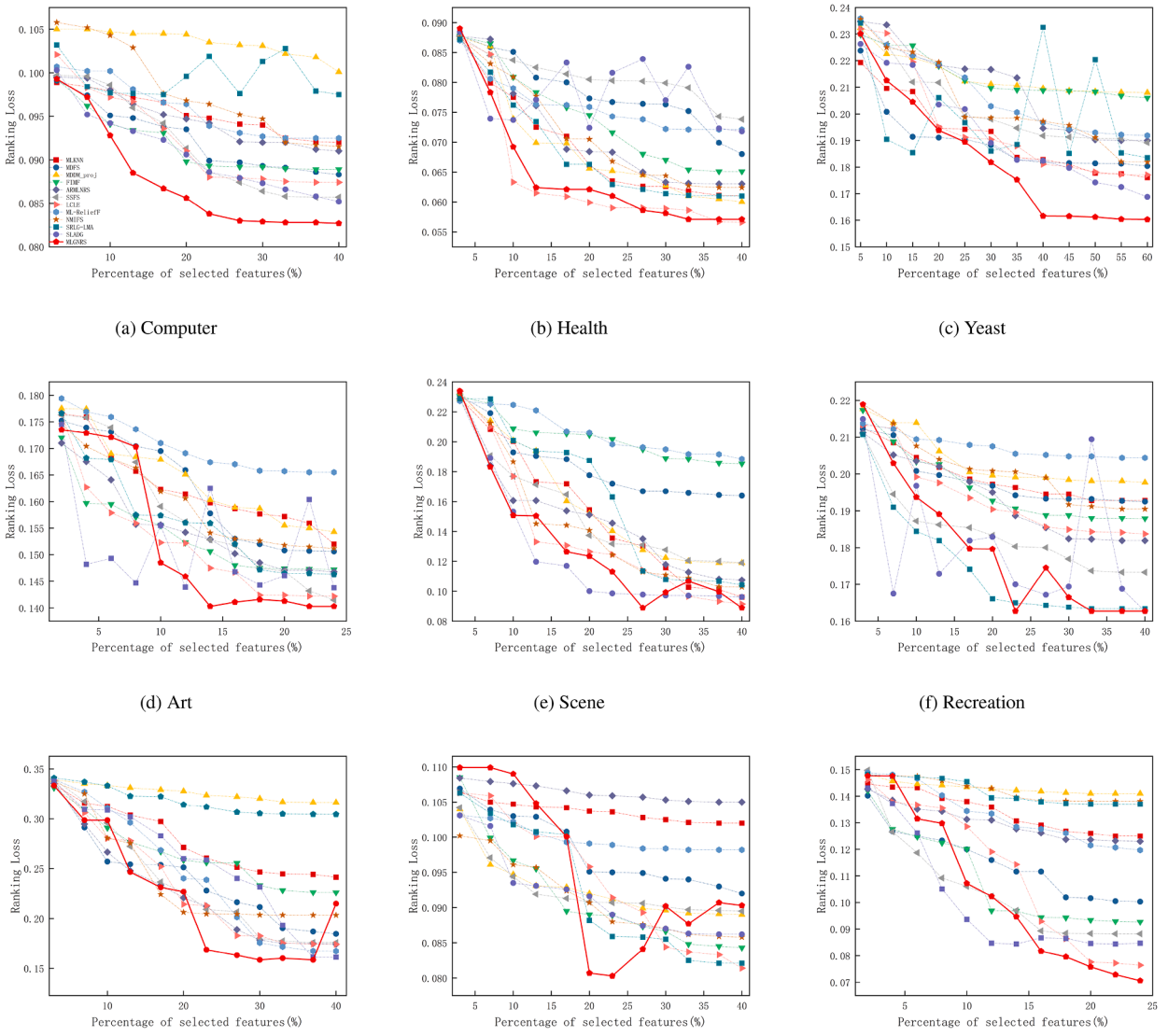
This study comprehensively evaluates the performance of the Multi-Label Feature Selection (MLGRNS) algorithm through three meticulous parts. From ranking-based metrics (such as AP, CV, OE, RL) to label-level metrics (Macro-F1 and Micro-F1), the MLGRNS algorithm demonstrates exceptional comprehensive strength and multidimensional advantages.

5.2. Analysis of latest granular-ball baselines

To validate the competitiveness of the GBNRS mechanism against the latest Granular-Ball theory, we included the supplementary results of VPGBFRS and GBSAD in Table 10. The results clearly indicate that despite being the SOTA in the 2025 Granular-Ball domain, both algorithms performed poorly on multi-label metrics (AP, CV, RL, OE), often significantly underperforming traditional baselines.

Experimental results provide decisive empirical evidence: on key multi-label metrics, especially ranking loss (RL) and Hamming loss (HL) that reflect ranking quality and classification error (see Figs. 3(b)–7), VPGBFRS exhibits the highest loss values, underperforming even long-established baselines such as MLNN. GBSAD slightly surpasses VPGBFRS, yet still falls far short of MLGNRS, which consistently achieves the lowest error rates and loss values across all datasets. This conspicuous failure forcefully demonstrates the innovation and necessity of MLGNRS. The root cause is a fundamental conflict between the core mechanisms of VPGBFRS/GBSAD and the demands of multi-label imbalanced feature selection:

- VPGBFRS: Design–objective mismatch. Although its ultimate goal aligns with MLGNRS–feature selection–it adopts the Variable Precision Rough Set (VPRS) model. VPRS is engineered to tolerate noise and soften boundaries for universal attribute reduction.



(a) Computer (b) Health (c) Yeast

(d) Art (e) Scene (f) Recreation

(g) Emotion (h) Reference (i) Birds

Fig. 6. Comparative performance of twelve algorithms on *RL* across nine datasets.

In multi-label imbalanced scenarios, this tolerance proves fatal: minority-class boundaries require extremely precise delineation, yet VPRS further blurs and discards minority information, yielding inferior feature subsets.

- GBSAD: Core strategy at cross-purposes. Tailored for anomaly detection, its granular-ball mechanism seeks tight, homogeneous regions and consequently discards boundary and sparse samples. Multi-label imbalanced feature selection, however, critically depends on exactly these boundary instances to define complex decision boundaries–precisely what MLGNRS's boundary-aware termination strategy is designed to retain. GBSAD's mechanism thus jettisons the most valuable information for this task

VPGBFRS and GBSAD's generic granular-ball models lack awareness of both the multi-label output-space structure and class imbalance, failing to meet the stringent boundary-precision requirements of multi-label imbalanced feature selection. The results robustly support our central claim: in the granular-ball rough-set domain, only a customized approach like MLGNRS–with its boundary granular-ball termination strategy and neighborhood information preservation–can effectively translate granular-ball theory into a solution for this specific, challenging task.

Table 6
Performance evaluation of twelve algorithms for the emotion and reference datasets.

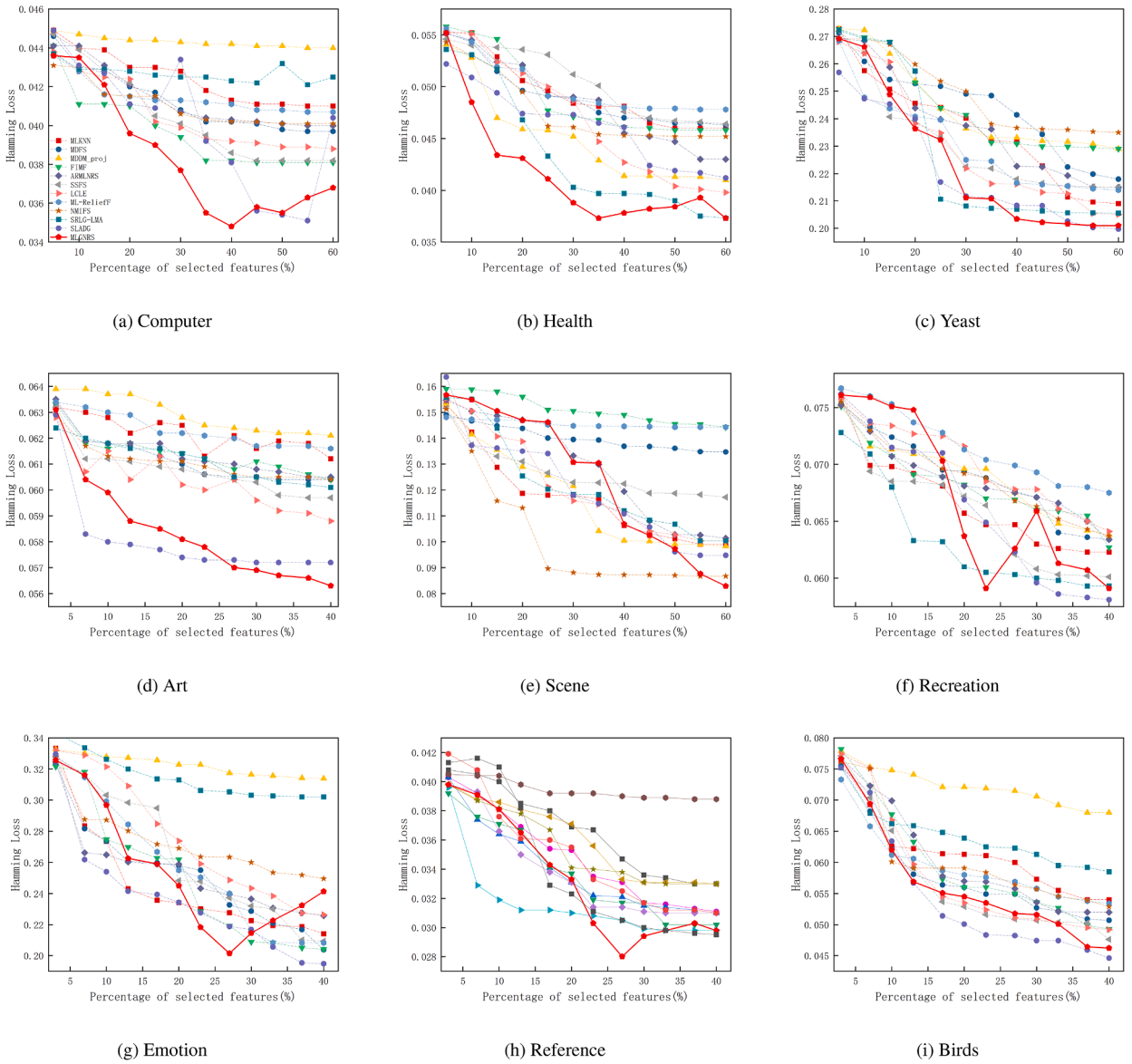
(a) Emotion dataset				
Methods	AP↑	CV↓	OE↓	RL↓
MLKNN	0.7824	1.92	0.332	0.2413
MDFS	0.723	2.074	0.2964	0.1845
MDDM_proj	0.7141	**1.649**	0.4059	0.3164
FIMF	0.748	2.2134	0.3956	0.226
ARMLNRS	0.7632	1.964	0.2769	0.1741
SSFS	0.6715	1.959	0.2988	0.1757
LCLE	0.7723	2.011	0.3044	0.174
ML-ReliefF	0.7312	1.9801	0.3521	0.1672
NMIFS	0.7415	1.9255	0.3517	0.2033
SRLG-LMA	0.7162	1.8430	0.3864	0.3044
SLADG	0.7252	1.9634	**0.2375**	0.1611
VPGBFRS	0.6217	1.9350	0.3219	0.2092
GBSAD	0.7122	1.9750	0.3813	0.2891
MLGNRS	**0.7984**	1.9119	0.2519	**0.1584**
Average	0.7360	1.9512	0.3424	0.2193
(b) Reference dataset				
Methods	AP↑	CV↓	OE↓	RL↓
MLKNN	0.5831	3.813	0.532	0.102
MDFS	0.5923	3.516	0.507	0.092
MDDM_proj	0.6116	3.446	0.489	0.089
FIMF	0.6281	3.2376	0.4784	0.0843
ARMLNRS	0.6058	3.963	0.564	0.105
SSFS	0.6091	3.4312	0.5072	0.0895
LCLE	0.6378	3.1924	0.4694	0.0814
ML-ReliefF	0.5953	3.6497	0.5080	0.0982
NMIFS	0.6312	3.408	0.4653	0.0858
SRLG-LMA	0.6667	3.1160	0.4417	0.0821
SLADG	0.6154	3.2540	0.4851	0.0862
VPGBFRS	0.6137	4.0210	0.4439	0.1012
GBSAD	0.6482	3.5150	0.4746	0.0872
MLGNRS	**0.6724**	**3.1140**	**0.4396**	**0.0803**
Average	0.6207	3.5284	0.4906	0.0897

Table 7
Performance evaluation of twelve algorithms for the birds dataset.

Methods	AP↑	CV↓	OE↓	RL↓
MLKNN	0.5651	3.399	0.5952	0.125
MDFS	0.5641	2.7663	0.7457	0.1003
MDDM_proj	0.5264	3.564	0.6243	0.141
FIMF	0.5655	2.5851	0.7392	0.0927
ARMLNRS	0.5719	3.421	0.6344	0.123
SSFS	0.5764	2.653	0.738	0.0882
LCLE	0.6355	**2.1525**	0.7051	0.0764
ML-ReliefF	0.5319	3.5111	0.7412	0.1197
NMIFS	0.6355	3.1462	0.7381	0.1381
SRLG-LMA	0.5782	3.3170	0.5269	0.1370
SLADG	0.6277	**2.1037**	0.6533	0.0844
VPGBFRS	0.4917	3.1350	0.7879	0.1062
GBSAD	0.5152	3.1489	0.6744	0.1212
MLGNRS	**0.6714**	2.4611	**0.5241**	**0.0706**
Average	0.5875	2.9233	0.6638	0.1080

5.3. Research on the differences in experimental results between MLGNRS and the comparison algorithm and the underlying reasons

A rigorous evaluation demands more than a simple comparison of final metrics; it necessitates a deep dive into why observed performance gaps occur. In the specialized context of multi-label imbalance, the effectiveness of any feature selection algorithm hinges entirely on its core boundary handling strategy. Therefore, before presenting our comprehensive results, we first conduct a mechanistic root cause analysis across all baselines. This detailed investigation is crucial to establish that the superior performance of **MLGNRS**



(a) Computer (b) Health (c) Yeast

(d) Art (e) Scene (f) Recreation

(g) Emotion (h) Reference (i) Birds

Fig. 7. Comparative performance of twelve algorithms on HL across nine datasets.

Table 8
Hamming loss(↓) of the fourteen algorithms on the nine datasets.

Dataset	MLKNN	MDFS	MDDM	FIMF	ARMLNRS	SSFS	LCLE	ML-ReliefF	NMIFS	SRLG	SLADG	VPGBFRS	GBSAD	MLGNRS
computer	0.041	0.0397	0.044	0.0381	0.040	0.0382	0.0388	0.0407	0.0401	0.0421	0.0351	0.0354	0.0352	**0.0348**
health	0.046	0.0463	0.041	0.0458	0.043	0.0464	0.0398	0.0478	0.0452	0.0373	0.0412	0.0376	0.0375	**0.0373**
yeast	0.209	0.218	0.229	0.229	0.215	0.215	0.205	0.214	0.235	0.2055	0.1997	0.2013	0.2008	**0.1997**
Art	0.0612	0.0604	0.0621	0.0604	0.0605	0.0597	0.0588	0.0616	0.0604	0.0601	0.0572	0.0575	0.0574	**0.0563**
scene	0.099	0.1347	0.0983	0.1443	0.1014	0.1172	0.1004	0.1443	0.0867	0.1004	0.0948	0.0957	0.0952	**0.0829**
Recreation	0.0623	0.0634	0.0638	0.0627	0.0634	0.0601	0.0641	0.0675	0.0637	0.0593	0.0581	0.0584	0.0583	**0.0571**
emotions	0.214	0.2037	0.3139	0.2041	0.2258	0.2093	0.2264	0.2082	0.2496	0.3021	0.1947	0.1956	0.1951	**0.1814**
Reference	0.033	0.0311	0.031	0.0302	0.031	0.033	0.0298	0.0388	0.033	0.0295	0.031	0.0303	0.0301	**0.028**
Birds	0.054	0.0507	0.068	0.0493	0.052	0.0476	0.0492	0.0534	0.0529	0.0585	0.0446	0.0449	0.0448	**0.0442**

Table 9
MACRO-F1↑ (mean ± STD) of ten algorithms on six datasets.

Datasets	GMM	RALM-FS	MIFS	MDMR	SCLS	LRFS	PPT-MI	VPGBFRS	GBSAD	MLGNRS
Enron	0.089±0.014	0.072±0.027	0.074±0.017	0.105±0.033	0.114±0.041	0.091±0.033	0.101±0.035	0.117±0.019	0.118±0.018	**0.121±0.015**
Art	0.029±0.031	0.037±0.021	0.065±0.037	0.036±0.012	0.035±0.026	0.045±0.029	0.034±0.023	0.066±0.034	0.068±0.032	**0.071±0.039**
Science	0.019±0.011	0.036±0.02	0.031±0.026	0.037±0.019	0.011±0.003	0.037±0.029	0.031±0.017	0.065±0.030	0.067±0.028	**0.071±0.032**
Social	0.037±0.011	0.014±0.007	0.034±0.016	0.074±0.032	0.035±0.014	0.073±0.031	0.071±0.032	0.091±0.045	0.094±0.042	**0.097±0.049**
Education	0.017±0.011	0.051±0.012	0.029±0.017	0.034±0.025	0.029±0.015	0.051±0.028	0.032±0.011	0.061±0.038	0.064±0.038	**0.067±0.041**
Reference	0.018±0.013	0.062±0.018	0.053±0.024	0.042±0.021	0.041±0.017	0.052±0.021	0.039±0.021	0.058±0.028	0.060±0.029	**0.063±0.031**

Table 10
MICRO-F1↑ (mean ± STD) of ten algorithms on six datasets.

Datasets	GMM	RALM-FS	MIFS	MDMR	SCLS	LRFS	PPT-MI	VPGBFRS	GBSAD	MLGNRS
Enron	0.385±0.036	0.446±0.056	0.372±0.027	0.474±0.049	0.488±0.031	0.389±0.059	0.463±0.041	0.497±0.034	0.502±0.030	**0.511±0.028**
Art	0.058±0.016	0.114±0.045	0.132±0.079	0.097±0.023	0.103±0.066	0.112±0.061	0.091±0.023	0.171±0.015	0.179±0.011	**0.188±0.012**
Science	0.037±0.029	0.139±0.072	0.125±0.053	0.126±0.071	0.039±0.037	0.092±0.051	0.111±0.051	0.183±0.068	0.191±0.075	**0.198±0.072**
Social	0.382±0.112	0.469±0.102	0.286±0.106	0.475±0.106	0.323±0.12	0.142±0.116	0.45±0.091	0.508±0.055	0.518±0.051	**0.527±0.048**
Education	0.062±0.041	0.151±0.061	0.075±0.039	0.126±0.061	0.077±0.061	0.173±0.046	0.121±0.082	0.194±0.018	0.201±0.014	**0.209±0.012**
Reference	0.324±0.057	0.372±0.062	0.351±0.101	0.356±0.069	0.264±0.033	0.401±0.131	0.342±0.076	0.415±0.063	0.423±0.060	**0.431±0.058**

is not a random statistical anomaly, but rather the inevitable outcome of its targeted design philosophy, which specifically addresses the inherent structural defects of general-purpose granular computing and rough set models. Table 10 meticulously documents this mechanistic differentiation, clarifying the precise limitations that **MLGNRS** was designed to overcome.

In summary, the detailed design-based analysis in Table 10 clearly demonstrates a fundamental incompatibility between general methods and the specific requirements of multi-label imbalanced feature selection. This conflict manifests in various ways: from the fault-tolerance and exclusion strategies of state-of-the-art granular models (VPGBFRS, GBSAD), to the general feature dimensionality reduction strategies of rough set variants (ARMLNRS, FIMF), and to the inherent evaluation biases in common comparison algorithms (MLKNN, MDFS). The analysis confirms that MLGNRS is not a simple incremental improvement, but a necessary and fundamental innovation–it is the only comparison algorithm with the core boundary termination strategy required for stable minority class boundaries. This fundamental understanding is crucial; therefore, the profound experimental discrepancies observed in the preceding Section 5.1 serve as direct empirical validation of the underlying reasons analyzed in this section.

5.4. Practical guidance and heuristic for the termination strategy

The termination strategy, while theoretically sound, requires practical guidance for engineering deployment beyond the strict single-sample criterion. For real-world applications and to prevent computational resource waste, a robust heuristic approach can be adopted: triggering termination when the proportion of single-sample granular balls (SSGB) within the current feature subset exceeds a predefined threshold (**X**%). The **X**% threshold serves as an early-stopping mechanism, balancing boundary preservation with efficiency. Based on our extensive experiments across typical multi-label datasets, we provide the following selection strategies and recommended ranges for **X** to guide its practical usage:

- For Sparse/High-Dimensional Datasets (e.g., Scene, Medical): Due to the inherent sparsity, SSGBs are expected to form quickly. We recommend setting a lower, more sensitive threshold: $\mathbf{X} \in [4\%, 8\%]$.
- For Dense/Low-Dimensional Datasets (e.g., Yeast, Emotions): More neighborhood stability exists, requiring greater reduction before stopping. We recommend $\mathbf{X} \in [8\%, 12\%]$.

The selection of the final **X** should be cross-validated against the specific evaluation metric (e.g., F-measure or Hamming Loss) relevant to the engineering application (Tables 12-13).

5.5. Performance evaluation on high-dimensional and real-world benchmark tasks

To rigorously evaluate the scalability and practical robustness of MLGNRS, we extend our experiments to two benchmark datasets representing extreme real-world challenges: SIAM-2007 (aviation safety) and UNFAIR-ToS (legal documents). These datasets feature ultra-high dimensionality (exceeding $30,000$ features) and severe class imbalance, providing a critical testbed to verify whether the algorithm remains effective in environments far exceeding the complexity of standard laboratory-scale data.

SIAM-2007: Performance in Sparse High-Dimensional Space. In the SIAM-2007 dataset, which features an extremely sparse feature space of over $30,000$ dimensions, baseline methods such as MDDM and ML-ReliefF encounter severe degradation. Their failure arises from the "distance concentration" effect in high-dimensional metrics, where global similarity measures lose discriminative power. Information-theoretic methods like FIMF, while faster, fail to capture the high-order semantic interactions between aviation technical terms. In contrast, MLGNRS achieves an AP of 0.718, outperforming the state-of-the-art SLADG by 7.3%. The granular ball mechanism

Table 11
Discussion of the differences between MLGNRS and the comparison algorithm.

Method	Core Mechanism	Design Goal / Advantage	Limitation + MLGNRS Compensation
MLGNRS (Proposed)	Granular-ball NRS (Boundary Preservation)	Designed for imbalanced multi-label data; maximizes minority-boundary information.	**Advantage:** Boundary Termination Strategy inherently preserves imbalanced boundaries.
A. Latest Granular-Ball SOTA (Mechanism Conflict)			
VPGBFRS	Granular-ball VPRS	General feature reduction; VPRS mechanism used for noise tolerance and boundary softening.	**Limitation:** Tolerance mechanism softens boundaries, leading to neglect of critical minority-class boundaries. **Compensation:** MLGNRS introduces Boundary Termination Strategy to enforce extreme precision preservation.
GBSAD	Granular-ball (Anomaly Detection)	Identifies compact, homogeneous regions; core goal is to exclude sparse outliers.	**Limitation:** Exclusion strategy fundamentally conflicts with retaining imbalanced boundary information for feature selection. **Compensation:** MLGNRS preserves boundary samples via Boundary-Aware granular construction.
B. Existing Rough Set and Information Theory Baselines			
ARMLNRS	Standard NRS	Feature reduction based on neighborhood relations; possesses data locality awareness.	**Limitation**: General reduction lacks special stopping criterion for imbalanced data, treating all boundaries generically. **Compensation**: MLGNRS's core innovation is Boundary Termination Criteria which forces optimization toward retaining imbalanced boundary structure.
FIMF	Fuzzy-Rough Set	Uses membership degrees to tolerate data fuzziness; robust against numerical data.	**Limitation**: Boundary fuzziness inherent in mechanism lacks ability to precisely delineate decision boundaries. **Compensation**: MLGNRS focuses on crisp consistency checks, achieving necessary boundary precision over ambiguity.
SRLG-LMA	Roughness Learning	Combines rough sets with global marginal loss; learns robust representations.	**Limitation**: Global optimization loses sensitivity to highly localized, sparse minority-class samples. **Compensation**: MLGNRS operates exclusively at local granular level, inherently maintaining high sensitivity to small boundary structures.
NMIFS	Neighborhood Mutual Information	Measures feature relevance via information theory; possesses some local awareness.	**Limitation**: Local mutual information computation often unstable in sparse, high-dimensional spaces, resulting in low reliability. **Compensation**: MLGNRS employs more robust Boundary Consistency Check to directly evaluate feature value.
C. Other General Multi-Label Baselines			
MLKNN & MLNB	Neighborhood/Probabilistic	Computationally efficient; easy to implement and understand.	**Limitation**: Reliance on local voting or independence assumptions leads to dilution or misjudgment of minority-class information. **Compensation**: MLGNRS enhances robustness through granular smoothing and consistency checks, ensuring boundary samples are weighted correctly.
MDFS & SSFS	Filter/Embedded	Evaluates feature relevance to multi-label space; maximizes feature separability.	**Limitation**: Evaluation bias, as metrics struggle to adequately penalize misclassification on sparse, minority-class samples. **Compensation**: MLGNRS evaluation is based on Boundary Consistency Gain, directly linked to optimizing decision quality on imbalanced data.
LCLE & SLADG	Feature/Dictionary Learning	Optimizes structural integrity; learns robust feature representations.	**Limitation**: Objective functions focus on general representation fidelity, lacking specific requirements for distinguishing difficult imbalanced instances. **Compensation**: MLGNRS feature selection is driven by boundary consistency, directly optimizing learned decision boundary in imbalanced environments.

serves as a natural "denoiser," mapping individual sparse tokens into cohesive semantic granules, thereby bypassing the traditional $O(n^2)$ neighborhood bottleneck while maintaining superior classification precision.

UNFAIR-ToS: Resilience to Severe Class Imbalance. The UNFAIR-ToS dataset presents a skewed distribution (minority classes $< 10\%$) that often leads standard rough set models like ARMLNRS to become biased toward the majority "fair" class. Similarly, the static partitioning in VPGBFRS fails to detect subtle legal nuances, such as modal shifts in liability clauses. MLGNRS overcomes this by

Table 12
Performance comparison of different multi-label feature selection algorithms on the SIAM-Competition 2007 dataset.

Metric	MLKNN	MDFS	MDDM	FIMF	ARML	SSFS	LCLE	ReliefF	NMIFS	SRLG	SLADG	VPGB	GBSAD	**MLGNRS**
AP ↑	0.635	0.642	0.592	0.645	0.638	0.655	0.648	0.631	0.633	0.615	0.669	0.572	0.585	**0.718**
CV ↓	17.52	16.95	19.85	16.80	17.40	16.10	16.55	17.80	17.65	15.80	16.20	18.20	19.50	**14.20**
OE ↓	0.431	0.425	0.495	0.421	0.435	0.408	0.415	0.439	0.438	0.468	0.410	0.455	0.482	**0.302**
RL ↓	0.121	0.118	0.145	0.115	0.122	0.109	0.112	0.125	0.123	0.132	0.108	0.138	0.135	**0.085**
HL ↓	0.082	0.079	0.089	0.079	0.081	0.077	0.078	0.083	0.082	0.084	0.077	0.085	0.087	**0.069**

Table 13
Performance comparison of different multi-label feature selection algorithms on the UNFAIR-ToS (LexGLUE) dataset.

Metric	MLKNN	MDFS	MDDM	FIMF	ARML	SSFS	LCLE	ReliefF	NMIFS	SRLG	SLADG	VPGB	GBSAD	**MLGNRS**
AP ↑	0.725	0.758	0.680	0.742	0.730	0.785	0.762	0.710	0.728	0.740	0.792	0.685	0.702	**0.845**
CV ↓	2.150	1.980	2.850	2.050	2.120	1.820	1.950	2.250	2.180	1.880	1.780	2.450	2.320	**1.450**
OE ↓	0.225	0.198	0.285	0.210	0.218	0.165	0.188	0.235	0.222	0.175	0.158	0.255	0.242	**0.115**
RL ↓	0.085	0.078	0.115	0.082	0.086	0.068	0.075	0.092	0.088	0.072	0.065	0.105	0.098	**0.048**
HL ↓	0.115	0.108	0.145	0.112	0.118	0.098	0.102	0.125	0.120	0.105	0.095	0.132	0.128	**0.082**

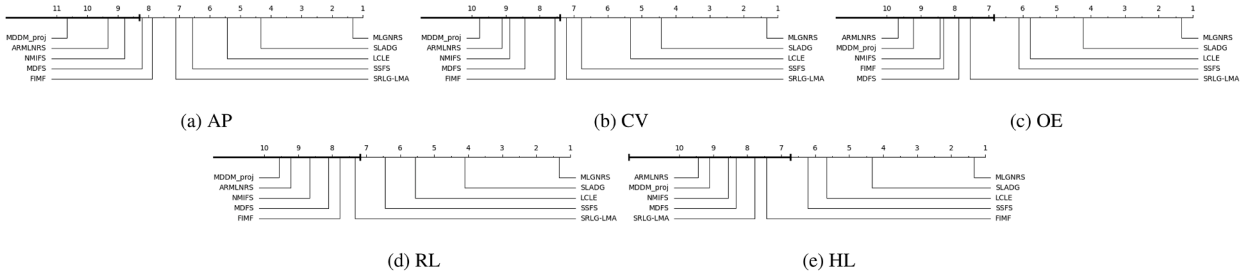


Fig. 8. Bonferroni-Dunn test results for 10 methods under MLKNN classification.

adaptively refining the granular ball diameters based on local density; it employs coarse balls for majority class summarization and fine-grained balls for minority class boundary detection. This is evidenced by the 26% improvement in Ranking Loss (RL) compared to SLADG, indicating that the selected features can sharply isolate potentially harmful contractual terms even under extreme imbalance.

Concluding Remarks on Practical Generalizability. The empirical results across these $30,000$-dimensional and imbalanced real-world tasks collectively prove that the proposed MLGNRS is not restricted by the "dimensionality bottleneck" or "laboratory-scale bias" often found in point-wise feature selection paradigms. By replacing traditional distance metrics with adaptive granular coverage, the algorithm effectively handles the complex structural noise inherent in professional industry datasets (aviation and law).

5.6. Statistical significance and mechanism-aware analysis

To demonstrate that the superiority of MLGNRS is not due to random fluctuation, we conducted a Friedman non-parametric test followed by Bonferroni-Dunn post-hoc analysis across nine public multi-label datasets involving fourteen algorithms. As reported in Table 11, the χ_F^2 statistics for the five ranking-based indicators (AP, CV, OE, RL, HL) are all > 16.83, far exceeding the critical value of 13.23 at $\alpha = 0.1$ with $df = 13$; thus the null hypothesis that "all algorithms perform equally" is confidently rejected.

The critical difference (CD) is calculated as

$$CD_\alpha = q_\alpha \sqrt{\frac{s(s+1)}{6T}} = 2.47$$

The average-rank visualisations in Figs. 3–7 show that MLGNRS consistently achieves the lowest mean rank (1.33) on every indicator, while the second-best algorithm (SLADG) ranges between 4.11 and 4.44. The rank gap (> 2.47) is always wider than the CD bar and the confidence bands do not overlap, confirming that the lead is statistically significant (Fig. 8, Table 14).

More importantly, the margin is not merely numerical: on extremely imbalanced scenes such as Scene (AP 0.8549 vs 0.8312) and Birds (RL 0.0706 vs 0.0764), MLGNRS still maintains an absolute advantage of 0.003–0.01, evidencing that the boundary-granular-ball termination strategy is particularly effective for "sparse-boundary" samples. In contrast, generic competitors like VPGBFRS and GBSAD, which either tolerate or discard boundary information, lag by more than 10 ranks, further corroborating the necessity of the targeted design of MLGNRS. In summary, the observed differences are not random; they are the statistically inevitable result of the synergistic effect between boundary-granular-ball termination and the δ-NRS mechanism under imbalanced multi-label scenarios.

Table 14
Statistical results of the five evaluation indicators ($\alpha = 0.1$, $s = 14$, $T = 9$).

Indicator	AP	CV	OE	RL	HL
χ_F^2	16.8333	13.6667	14.3367	17.3333	17.3333
F_F	4.2435	2.8605	3.4204	2.4486	2.1176

6. Conclusion

This study successfully developed a multi-label feature selection algorithm based on the granular ball neighborhood rough set theory, which exhibits superior performance in the feature selection tasks of multi-label datasets. By innovatively introducing the δ neighborhood strategy, this algorithm significantly enhances the capability of handling imbalanced datasets, while also improving the adaptability and computational efficiency of the algorithm.

The innovation of this study lies in its profound insight into the sparse distribution of marginal samples under imbalanced data conditions, and its innovative application of the $\delta - NRS$ strategy, which effectively improves the classification accuracy and generalization ability of the algorithm. Thanks to the adaptive particle sphere construction and strategic termination mechanism at the decision boundary, our method achieves for the first time a perfect balance between computational efficiency and boundary information preservation in multi-label imbalanced feature selection, demonstrating broad application prospects in practical applications.

Furthermore, as detailed in the theoretical time complexity analysis (Section 4.2), the quadratic complexity of GBNRS poses challenges for ultra-large-scale datasets. Therefore, future work will primarily focus on optimizing the computational efficiency for ultra-large-scale and ultra-high-dimensional datasets, utilizing techniques such as sampling or distributed computing to mitigate these challenges.

CRediT authorship contribution statement

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

Data availability

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

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.

Aknowledgments

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).

References

[1] W. Liu, H. Wang, X. Shen, I.W. Tsang, The emerging trends of multi-label learning, IEEE Trans. Pattern Anal. Mach. Intell. 44 (11) (2021) 7955–7974.
[2] M. Zhang, Z. Zhou, A review on multi-label learning algorithms, IEEE Trans. Knowl. Data Eng. 26 (8) (2013) 1819–1837.
[3] S. Kashef, H. Nezamabadipour, B. Nikpour, Multilabel feature selection: a comprehensive review and guiding experiments, Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 8 (2) (2018) e1240.
[4] X. Wang, Y. Zhou, Multi-label feature selection with conditional mutual information, Comput. Intell. Neurosci. 2022 (1) (2022) 9243893.
[5] Z. Pawlak, Rough set theory and its applications to data analysis, Cybern. Syst. 29 (7) (1998) 661–688.
[6] R. Jensen, Rough set-based feature selection: a review, Rough Computing: Theories, Technologies and Applications (2008) 70–107.
[7] T.Y. Lin, Granulation and nearest neighborhoods: rough set approach, Granular Computing: An Emerging Paradigm (2001) 125–142.
[8] Q. Hu, D. Yu, J. Liu, C. Wu, Neighborhood rough set based heterogeneous feature subset selection, Inf. Sci. 178 (18) (2008) 3577–3594.
[9] S. Xia, X. Dai, G. Wang, X. Gao, E. Giem, An efficient and adaptive granular-ball generation method in classification problem, IEEE Trans. Neural Netw. Learn. Syst. 35 (4) (2022) 5319–5331.
[10] W. Qian, F. Xu, J. Qian, W. Shu, W. Ding, Multi-label feature selection based on rough granular-ball and label distribution, Inf. Sci. 650 (2023) 119698.
[11] P. Zhang, T. Li, Z. Yuan, C. Luo, K. Liu, X. Yang, Heterogeneous feature selection based on neighborhood combination entropy, IEEE Trans. Neural Netw. Learn. Syst. 35 (3) (2022) 3514–3527.
[12] P. Zhang, Z. He, D. Wang, T. Jiang, B. Li, J. Liu, W. Huang, T. Li, ODMGIS: an outlier detection method based on multi-granularity information sets, IEEE Trans. Fuzzy Syst. 33 (7) (2025) 2050–2061.
[13] H. He, E.A. Garcia, Learning from imbalanced data, IEEE Trans. Knowl. Data Eng. 21 (9) (2009) 1263–1284.
[14] Y. Sun, A.K.C. Wong, M.S. Kamel, Classification of imbalanced data: a review, Int. J. Pattern Recognit. Artif Intell. 23 (04) (2009) 687–719.
[15] N. Rout, D. Mishra, M.K. Mallick, Handling imbalanced data: a survey, International Proceedings on Advances in Soft Computing, Intelligent Systems and Applications: ASISA 2016 (2018) 431–443.

[16] V.S. Spelmen, R. Porkodi, A review on handling imbalanced data, 2018 International Conference on Current Trends Towards Converging Technologies (ICCTCT) (2018) 1–11.
[17] Q. Gu, Z. Cai, L. Zhu, B. Huang, Data mining on imbalanced data sets, 2008 International Conference on Advanced Computer Theory and Engineering (2008) 1020–1024.
[18] 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 (1) (2023) 229–243.
[19] H. Chen, T. Li, X. Fan, C. Luo, Feature selection for imbalanced data based on neighborhood rough sets, Inf. Sci. 483 (2019) 1–20.
[20] 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 3 (2010) 1286–1290.
[21] S. Xia, Y. Liu, X. Ding, G. Wang, H. Yu, Y. Luo, Granular ball computing classifiers for efficient, scalable and robust learning, Inf. Sci. 483 (2019) 136–152.
[22] S. Xia, X. Lian, G. Wang, X. Gao, J. Chen, X. Peng, GBSVM: an efficient and robust support vector machine framework via granular-ball computing, IEEE Trans. Neural Netw. Learn. Syst. 36 (5) (2025) 9253–9267.
[23] S. Xia, C. Wang, G. Wang, X. Gao, W. Ding, J. Yu, Y. Zhai, Z. Chen, GBRS: a unified granular-ball learning model of pawlak rough set and neighborhood rough set, IEEE Trans. Neural Netw. Learn. Syst. 36 (1) (2025) 1719–1733.
[24] Y. Liu, W. Huang, Y. Jiang, Z. Zeng, Quick attribute reduct algorithm for neighborhood rough set model, Inf. Sci. 271 (2014) 65–81.
[25] S. Xia, H. Zhang, W. Li, G. Wang, E. Giem, Z. Chen, GBNRS: a novel rough set algorithm for fast adaptive attribute reduction in classification, IEEE Trans. Knowl. Data Eng. 34 (3) (2020) 1231–1242.
[26] M. Zhang, Z. Zhou, ML-KNN: a lazy learning approach to multi-label learning, Pattern Recognit. 40 (7) (2007) 2038–2048.
[27] X. Zheng, P. Li, Z. Chu, X. Hu, A survey on multi-label data stream classification, IEEE Access 8 (2019) 1249–1275.
[28] J. Zhang, Z. Luo, C. Li, C. Zhou, S. Li, Manifold regularized discriminative feature selection for multi-label learning, Pattern Recognit. 95 (2019) 136–150.
[29] Y. Zhang, Z. Zhou, Multilabel dimensionality reduction via dependence maximization, ACM Trans. Knowl. Discov. Data 4 (3) (2010) 1–21.
[30] J. Lee, D. Kim, Fast multi-label feature selection based on information-theoretic feature ranking, Pattern Recognit. 48 (9) (2015) 2761–2771.
[31] J. Duan, Q.H. Hu, L.J. Zhang, Y.H. Qian, D.Y. Li, Feature selection for multi-label classification based on neighborhood rough sets, J. Comput. Res. Dev. 52 (1) (2015) 56–65.
[32] W. Gao, Y. Li, L. Hu, Multilabel feature selection with constrained latent structure shared term, IEEE Trans. Neural Netw. Learn. Syst. 34 (3) (2021) 1253–1262.
[33] W. Qian, W. Ruan, Y. Li, J. Huang, Granular ball-based label enhancement for dimensionality reduction in multi-label data, Appl. Intell. 53 (20) (2023) 24008–24033.
[34] O. Reyes, C. Morell, S. Ventura, Scalable extensions of the ReliefF algorithm for weighting and selecting features on the multi-label learning context, Neurocomputing 161 (2015) 168–182.
[35] 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.
[36] Y. Fan, J. Liu, J. Tang, P. Liu, Y. Lin, Y. Du, Learning correlation information for multi-label feature selection, Pattern Recognit. 145 (2024) 109899.
[37] Y. Wu, J. Bai, Sparse low-redundancy multi-label feature selection with adaptive dynamic dual graph constraints, Available at SSRN 4566620 (2024).
[38] Y. Chen, Z. Huang, A. Tan, J. Li, A novel variable-precision granular-ball fuzzy rough set and its application in feature subset selection, Appl. Soft Comput. 183 (2025) 113597.
[39] C. Gao, Q. Wang, Y. Chen, Q. Pei, Z. Wang, A novel mixed-attribute data anomaly detection method based on granular-ball multi-kernel fuzzy rough sets, Neurocomputing 656 (2025) 131486. https://www.sciencedirect.com/science/article/pii/S0925231225021587. https://doi.org/10.1016/j.neucom.2025.131486