AIMS Mathematics, 11(8): 24987–25043.
DOI: 10.3934/math.20261005
Received: 18 June 2026
Revised: 29 July 2026
Accepted: 05 August 2026
Published: 13 August 2026

https://www.aimspress.com/journal/Math

Research article

Label dimensional reduction for interval-valued data based on maximal consistent block granular-ball rough sets

Shiqi Chen[1], Zhongying Suo[1,*], Yuanbo Kong[2,3], Songlei Xue[1] and Zhuoluo Wang[4]

[1] Fundamentals Department, Air Force Engineering University, Xi'an 710051, China

[2] National Science Library, Chinese Academy of Sciences, Beijing 100190, China

[3] Department of Information Resources Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China

[4] Aviation Engineering School, Air Force Engineering University, Xi'an 710038, China

* **Correspondence:** Email: suozhongying@sohu.com.

Abstract: To address the problem of dimensional reduction in single/multi-label scenarios with interval-valued data, this paper proposes a reduction method based on maximal consistent block granular-ball rough sets. First, the dimensional reduction problem is transformed into an attribute reduction problem in interval-valued decision information systems. An interval tolerance relation is designed by integrating the Hausdorff distance and the degree of interval overlap. Maximal consistent blocks are then used to substitute for the global dataset as the initial granular-ball set, and a granular-ball splitting together with a heterogeneous overlap removal mechanism tailored for interval-valued data is designed to build a single-label granular-ball rough set attribute reduction model. The model is further extended to the multi-label scenario by defining multi-label granular-ball purity and tolerance measures, thereby establishing a multi-label granular-ball rough set attribute reduction method. The theoretical properties of the proposed models are established, and enhancement strategies including a granular-ball quality function, a mutual information candidate pool, predictive performance validation, and local refinement are introduced to form a complete algorithmic workflow. Experimental results on public datasets show improved predictive performance and substantial feature reduction, with good stability against parameter perturbations and noise.

Keywords: multi-label learning; label dimensional reduction; granular-ball rough set; decision information system; maximal consistent block; attribute reduction
Mathematics Subject Classification: 68T30, 68T37

1. Introduction

In practical applications such as medical diagnosis, industrial monitoring, and unmanned aerial vehicle (UAV)-related situational awareness, feature values are usually recorded in interval forms rather than precise point values due to sensor measurement errors, environmental disturbances, and quantization uncertainty, giving rise to widespread interval-valued data [1, 2]. Meanwhile, a single sample is often associated with multiple semantic labels. For example, a remote sensing image may simultaneously indicate "vegetation coverage", "water pollution", and "building encroachment"; such tasks are referred to as multi-label learning [3]. Compared with single-label learning, multi-label learning more closely reflects the intrinsic properties of data, but its high-dimensional feature space and complex dependencies among labels can readily lead to the "curse of dimensionality" [4], resulting in inefficient model training and weakened generalization ability. The coupling of interval-valued data with multi-label semantics makes traditional single-label dimensional reduction methods designed for precise numerical values difficult to apply directly. Therefore, a new dimensional reduction paradigm is urgently needed to handle both interval uncertainty and labels' multidimensionality.

As a key technique for mitigating the "curse of dimensionality" and improving model generalization, dimensional reduction aims to remove redundant features while preserving discriminative information [5]. Existing methods are mainly divided into two categories: Feature selection and feature extraction. Feature selection is widely used in scenarios requiring interpretability because it can preserve the original physical semantics of features [6,7]. However, most existing feature selection methods are designed for point-valued data and do not fully account for endpoint uncertainty in interval-valued data. Although feature extraction [8, 9] can effectively condense dimensionality, the converted features often lack clear physical semantics, and their practical applicability to multi-label interval-valued data is limited.

Rough set theory provides a solid mathematical foundation for dimensional reduction. Among the existing rough set models, the Pawlak rough set (PRS) and the neighborhood rough set (NRS) are the two most widely used ones. The PRS partitions the universe according to equivalence relations, and its upper and lower approximations are composed of equivalence classes, which endows it with favorable knowledge interpretability. Nevertheless, the PRS is restricted by the assumption of discrete data [10]. The NRS can directly handle continuous data through neighborhood relations, yet it loses the equivalence-class-based capability for knowledge representation because its upper and lower approximations are composed of sample points [11]. Beyond these two classical models, various extended rough set models have been developed in recent years and successfully applied to real-world problems. For example, Fang et al. [12] integrated rough set theory with prospect theory for analysis of failure mode and effects, and validated the approach in the risk assessment of a steam valve system; Senthil Kumar and Inbarani [13] applied multi granulation rough set approaches to cardiac arrhythmia classification based on real electrocardiogram (ECG) records from the Massachusetts Institute of Technology-Beth Israel Hospital (MIT-BIH) arrhythmia database; Chu et al. [14] proposed a nonadditive rough set model for long-term clinical efficacy evaluation of chronic diseases in real-world clinical settings; and Fujita [15] extended the classical fuzzy VIsekriterijumska Optimizacija I Kompromisno Resenje (VIKOR, multi-criteria optimization and compromise solution) and fuzzy DEcision-MAking Trial and Evaluation Laboratory (DEMATEL) methods to the hyperfuzzy and superhyperfuzzy settings, capturing multi layered uncertainty in real-world multi criteria decision-

making. These application-oriented studies verify the practical value of rough set models; however, they are mainly designed for specific decision or classification scenarios and do not address the trade-off between knowledge interpretability and continuous data processing in attribute reduction. To balance interpretability with the ability to process continuous data, Xia et al. [16] introduced granular-ball computing into rough set theory and proposed the granular-ball rough set (GBRS). By using adaptive granular balls, this model unifies the advantageous frameworks of the PRS and NRS, enabling efficient knowledge representation and attribute reduction for continuous data. Subsequent studies further extended the theoretical boundaries of the GBRS. Lian et al. [17] integrated the GBRS with fuzzy sets and proposed the granular-ball fuzzy rough set (GBFRS) to handle fuzzy uncertainty. Ye et al. [18] constructed a multi granularity GBRS model and incorporated the Zentropy integral to drive the evolution of generalized multi granularity rough sets, thereby improving feature selection performance. Peng et al. [19] addressed label noise by using the inclusiveness property of granular balls to design a covering rough set model, which effectively mitigated the negative impact of noise on attribute reduction. These studies demonstrate the scalability of granular-ball computing. However, the abovementioned GBRS and its derived algorithms do not explicitly characterize the positional differences and range overlap degrees of interval-valued samples, which limits their applicability in interval-valued scenarios.

To address these issues, it is necessary first to return to the intrinsic representation of interval-valued data and establish a theoretical framework. An interval-valued information system characterizes attribute values as intervals rather than precise numerical values, thereby naturally preserving the range of uncertainty and measurement boundaries of raw data [1, 2, 20]. For interval-valued data, researchers have extended rough set models based on interval order, interval dominance relations, and interval neighborhood relations. For example, Ali et al. [21] proposed a decision-theoretic rough set model based on interval-valued intuitionistic fuzzy numbers; Li et al. [22] designed an interval-dominance-based feature selection method for interval-valued ordered data; and Sang et al. [23] combined fuzzy dominance NRSs to process dynamic interval-valued ordered data. Interval-valued data have also been extensively investigated in decision-making: Interval-valued intuitionistic fuzzy sets [24] have been widely adopted to express the uncertainty and hesitancy of decision makers, and various decision methods have been developed on this basis, including aggregation-based methods [25] and the recently proposed interval-valued intuitionistic fuzzy best-worst method with multiplicative consistency for group decision-making [26]. These studies have further enriched the approaches for processing interval-valued data and for decision-making under interval-valued uncertainty. However, the abovementioned methods either focus on deriving decision rules or depend on specific interval order structures, or aim at ranking alternatives under a given attribute system; they have not yet systematically addressed single-label or multi-label dimensional reduction for interval-valued data, nor have they fully exploited tolerance structures among samples for efficient granulation.

In the granulation techniques of rough sets, the maximal consistent block (MCB) serves as the largest knowledge granule under tolerance relations. It provides a reliable basis for attribute reduction by fully preserving the consistent structures among objects while avoiding information redundancy [27]. Leung and Li [27] first proposed the MCB technique for handling incomplete information systems, thereby laying its theoretical foundation. On this basis, subsequent studies have extended it to multiple application realms: Li et al. [28] applied it to telecommunications fraud detection; Sun et al. [29] proposed an MCB-based optimal scale selection method for incomplete

multi-scale data; and Sun et al. [30] further combined the MCB with the variable precision rough set, which enhances the fault-tolerant reduction capability of the model in noisy data scenarios. However, these studies all focus on incomplete data with missing values and rely on the logic of "missing-value tolerance", making them not directly transferable to interval-valued data. In particular, in multi-label learning tasks, maintaining correlations among labels under interval uncertainty while achieving efficient granulation and reduction remains an open theoretical and algorithmic challenge.

From a theoretical perspective, the dimensional reduction problem for multi-label interval-valued data can be naturally incorporated into the frame of decision information systems and transformed into the attribute reduction problem in rough set theory, namely, finding the minimal subset of conditional attributes (features) while preserving the decision (label) information. Although the GBRS [16] effectively integrates the advantages of the PRS and NRS through adaptive granular-ball generation, its current framework still has the following limitations. First, granular-ball initialization treats the entire dataset as a single granular ball and does not fully exploit the tolerance structure among samples, resulting in low purity of the initial granular balls. Second, its distance metric and purity evaluation are both designed for point-valued data, using Euclidean distance and the numerical mean of samples; consequently, they cannot explicitly characterize the length uncertainty and the degree of range overlap of interval values and are difficult to adapt directly to interval-valued data. Therefore, it is necessary to integrate the MCB with the GBRS, generate initial granular balls starting from the high consistency among samples within each consistent block, and redesign the distance metric and splitting mechanism for interval-valued data.

The main motivations of this paper originate from the following facts: (1) Interval-valued data are ubiquitous in real-world applications due to measurement errors, privacy protection, and data perturbation; however, the existing GBRS and its derived models do not explicitly characterize the positional differences and the degrees of range overlap of interval-valued samples, which calls for an interval tolerance relation that jointly captures these two aspects. (2) Dimensional reduction arises in both single-label and multi-label scenarios, whereas the existing GBRS-based reduction methods are designed only for single-label point-valued data, so a unified interval-valued GBRS framework covering both scenarios is desirable. (3) In the existing granular-ball generation mechanism, the whole dataset is taken as a single initial granular ball, which ignores the intrinsic similarity structure among samples at the coarsest granulation level; the MCBs of a tolerance relation characterize the maximal groups of mutually similar samples and can therefore provide a structure-aware initialization for granular-ball generation, so that the subsequent splitting proceeds on already consistent granules rather than on the undifferentiated universe.

In view of this, starting from the intrinsic connection between label learning and decision information systems, this paper transforms the single-label and multi-label dimensional reduction problems for interval-valued data into the attribute reduction problems in interval-valued decision information systems, and proposes an interval-valued GBRS method based on the MCB. The main contributions are summarized as follows. (1) Core mathematical model: Within the general framework of interval-valued decision information systems, an interval tolerance relation integrating the Hausdorff distance and the interval overlap degree is designed; MCBs are used in place of the global dataset as the initial granular balls; and the splitting rules, the heterogeneous overlap removal mechanism, and the purity measures for interval-valued granular balls are established, together with the theoretical properties proved in Lemma 2.1 and Theorems 2.1–3.4, forming a GBRS model for interval-valued

data. This constitutes the primary theoretical contribution of this paper. (2) Main attribute reduction algorithms: Based on the proposed model, attribute reduction algorithms for single-label and multi-label scenarios are designed, and their computational complexity is analyzed in an output-sensitive manner. (3) Optional heuristic enhancements: Enhancement strategies, including the granular-ball quality function, the mutual information candidate pool, predictive performance validation, and local refinement, etc., are integrated to form a complete implementation. These strategies are engineering-oriented improvements that are orthogonal to the core model and do not constitute the theoretical novelty of this paper. (4) Comparative experiments are conducted on 10 single-label and 6 multi-label public datasets against the full-feature baseline, the GBRS, the maximal GBRS (MaxGBRS), and various representative feature selection methods. The results show that the proposed method achieves superior performance in terms of its classification accuracy and reduction rate. Statistical tests further verify its significance, and its robustness is further examined under both attribute and decision noise.

The remainder of this paper is organized as follows: Section 2 introduces the preliminaries, including interval-valued decision information systems, interval tolerance relations, the MCB, and the GBRS. Section 3 details the single-label and multi-label granular-ball generation and attribute reduction algorithms for interval-valued data, and analyzes their computational complexity. Section 4 experimentally validates the effectiveness of the proposed method, and Section 5 concludes the paper and outlines future work. For clarity, the main notations used throughout this paper are summarized in Table 19 in the Appendix.

2. Preliminaries

To establish a theoretical framework for attribute reduction in interval-valued decision information systems, this section systematically reviews the foundational concepts and existing studies adopted in this research. First, we define interval-valued decision information systems and their single-label/multi-label forms; second, we introduce the interval tolerance relation and the MCB; finally, we revisit the basic model of the GBRS, which serves as the theoretical foundation of the proposed method in this paper.

2.1. Interval-valued decision information systems

To describe single-label and multi-label decision scenarios within a unified framework, this paper adopts the general definition of interval-valued decision information systems [20] and separates it into specific cases according to the structure of the decision attributes [31].

Definition 2.1. *[20] Let $DT = \langle U, AT, V, f\rangle$ be a quadruple for an interval-valued decision information system, where*

- *$U = \{x_1, x_2, \ldots, x_n\}$ is a non-empty finite set of objects, called the universe;*
- *$AT = A \cup D$ is a non-empty finite set of attributes, in which $A = \{a_1, a_2, \ldots, a_m\}$ is the set of conditional attributes and D is the set of decision attributes;*
- *$V = \bigcup_{a\in AT} V_a$ is the domain of attributes. For any conditional attribute $a \in A$, its domain $V_a = \{[l, r] \mid l, r \in \mathbb{R}, l \le r\}$ is a set of interval values;*
- *$f : U \times AT \to V$ is an information function such that $f(x, a) = [l_{x,a}, r_{x,a}] \in V_a$ holds for arbitrary $x \in U$ and $a \in A$.*

The general framework above can be transformed into the following two scenarios according to the structure of the set of decision attributes D [31].

Single-label scenario: $D = \{d\}$ *where* $d \notin A$ *is a single decision attribute whose domain* $V_D = \{d_1, d_2, \ldots, d_k\}$ *is a finite set of class labels. The information function satisfies* $f(x, d) = d_x \in V_D$*, meaning that each object corresponds to exactly one decision class label.*

Multi-label scenario: $D = L = \{l_1, l_2, \ldots, l_q\}$ *where* $A \cap L = \emptyset$ *and* $V_{l_p} = \{0, 1\}$ *for each* $l_p \in L$*. For an object* $x \in U$*,* $f(x, l_p) = 1$ *indicates that x is associated with the label* l_p*, while* $f(x, l_p) = 0$ *indicates no association. Each object is associated with at least one label, and each label is associated with at least one object. The label subset of object x is denoted as* $L_x = \{l_p \in L \mid f(x, l_p) = 1\}$*.*

For data whose original attributes take point values, let a perturbation radius ρ *be adopted to construct interval-valued representations. Let* $z_{x,a}$ *be the point value of object x on conditional attribute a. Its corresponding interval-valued form is defined as*

$$f(x, a) = [z_{x,a} - \rho, z_{x,a} + \rho], \quad \rho \geq 0. \tag{2.1}$$

Here, ρ *refers to the interval perturbation radius, which quantifies the range of uncertainty when extending point-valued attributes to interval-valued attributes. When* $\rho = 0$*,* $f(x, a) = [z_{x,a}, z_{x,a}]$*, which degenerates into a point-valued interval.*

2.2. Interval tolerance relation and the MCB

On interval-valued data, the similarity between samples needs to simultaneously take the positional difference of the intervals and the degree of interval overlap into account, so as to construct the interval tolerance relation and generate the MCB. To this end, the interval Hausdorff distance [1] and the interval overlap degree [2] are introduced as fundamental metrics. To simultaneously characterize the positional difference of intervals and the overlap degree between intervals, this paper draws on the idea of interval similarity measurement and defines the following normalized interval overlap degree.

Definition 2.2. *[1, 2] Let* $I_1 = [l_1, r_1]$ *and* $I_2 = [l_2, r_2]$ *be any two intervals. The Hausdorff distance between them is defined as*

$$H(I_1, I_2) = \max\{|l_1 - l_2|, |r_1 - r_2|\}. \tag{2.2}$$

The normalized interval overlap degree is defined as

$$O(I_1, I_2) = \frac{|I_1 \cap I_2|}{\max\{|I_1|, |I_2|, \varepsilon_l\}}, \tag{2.3}$$

where $|I_1 \cap I_2| = \max\{0, \min(r_1, r_2) - \max(l_1, l_2)\}$*,* $|I_1| = r_1 - l_1$*,* $|I_2| = r_2 - l_2$*, and* $\varepsilon_l > 0$ *denotes a small constant to avoid division by zero caused by zero-length intervals; it is smaller than every positive interval length in* $DT = \langle U, AT, V, f\rangle$ *(DT). The Hausdorff distance is adopted to characterize the positional difference of intervals and satisfies non-negativity, symmetry, and the triangle inequality. The interval overlap degree is used to quantify the rationality of interval aggregation, with a value range of* $[0, 1]$*. A value of 1 represents complete overlap of two intervals, while a value of 0 indicates no overlap.*

In particular, when both intervals degenerate into zero-length intervals, if their endpoints are identical, their overlap degree is specified as 1; otherwise, the overlap degree is set to 0, i.e.,

$$O([x, x], [y, y]) = \begin{cases} 1, & x = y \\ 0, & x \neq y \end{cases}. \tag{2.4}$$

Lemma 2.1 (Properties of the two interval metrics)**.** *(1) H is a metric on the space of closed intervals, i.e., it satisfies the conditions of non-negativity, the identity of indiscernibles, symmetry, and the triangle inequality.*

(2) The normalized interval overlap degree satisfies: $0 \le O(I_1, I_2) \le 1$; $O(I_1, I_2) = O(I_2, I_1)$; $O(I, I) = 1$; $O(I_1, I_2) = 1 \iff I_1 = I_2$; and $O(I_1, I_2) = 0 \iff I_1$ and I_2 have disjoint interiors.

Proof. (1) Let $\varphi([l, r]) = (l, r) \in \mathbb{R}^2$. Then φ is injective and $H(I_1, I_2) = \|\varphi(I_1) - \varphi(I_2)\|_\infty$; since the L_∞ norm is a metric on $\mathbb{R}^2$, H inherits all four metric axioms.

(2) Boundedness follows from $|I_1 \cap I_2| \le \min\{|I_1|, |I_2|\} \le \max\{|I_1|, |I_2|, \varepsilon_l\}$; symmetry from $\cap$ and max. If $|I| > 0$ then $O(I, I) = |I|/\max\{|I|, \varepsilon_l\} = 1$; if $|I| = 0$, the degenerate rule gives $O(I, I) = 1$. If $O(I_1, I_2) = 1$, then $|I_1 \cap I_2| = \max\{|I_1|, |I_2|, \varepsilon_l\} \ge \max\{|I_1|, |I_2|\} \ge \min\{|I_1|, |I_2|\} \ge |I_1 \cap I_2|$, forcing $|I_1| = |I_2| = |I_1 \cap I_2|$, and two equal-length intervals whose intersections have the same length coincide. Finally, $O(I_1, I_2) = 0 \iff \min(r_1, r_2) \le \max(l_1, l_2)$.

□

Definition 2.3. *Let attribute subset $B \subseteq A$, and let $x_i, x_j \in U$ be arbitrary objects. The interval tolerance relation $T_B^{\varepsilon,\theta}$ is defined as*

$$T_B^{\varepsilon,\theta} = \{(x_i, x_j) \in U^2 \mid \max_{a \in B} H(f(x_i, a), f(x_j, a)) \le \varepsilon \wedge \min_{a \in B} O(f(x_i, a), f(x_j, a)) \ge \theta\}, \tag{2.5}$$

where $\varepsilon > 0$ denotes the distance threshold and $\theta \in [0, 1]$ denotes the overlap degree threshold ($\theta = 0$ makes the overlap condition vacuous). The interval tolerance relation on the empty attribute set satisfies $T_\emptyset^{\varepsilon,\theta} = U \times U$.

By integrating the Hausdorff distance and the interval overlap degree, this tolerance relation extends the classical neighborhood tolerance relation to interval-valued scenarios. This relation requires all interval values of objects on attribute subset B to simultaneously satisfy the distance constraint and overlap constraint for the two objects to be regarded as mutually tolerant.

Theorem 2.1. *For any $B \subseteq A$, $\varepsilon > 0$, $\theta \in [0, 1]$, we have following.*

(1) $T_B^{\varepsilon,\theta}$ is reflexive and symmetric, i.e., a tolerance relation on U;
(2) $T_B^{\varepsilon,\theta} = \bigcap_{a \in B} T_{\{a\}}^{\varepsilon,\theta}$;
(3) $B_1 \subseteq B_2 \subseteq A \implies T_{B_2}^{\varepsilon,\theta} \subseteq T_{B_1}^{\varepsilon,\theta}$;
(4) $\varepsilon_1 \le \varepsilon_2 \implies T_B^{\varepsilon_1,\theta} \subseteq T_B^{\varepsilon_2,\theta}$;
(5) $\theta_1 \le \theta_2 \implies T_B^{\varepsilon,\theta_2} \subseteq T_B^{\varepsilon,\theta_1}$.

Proof. (1) For any $x \in U$ and $a \in B$, $H(f(x, a), f(x, a)) = 0 \le \varepsilon$, and $O(f(x, a), f(x, a)) = 1 \ge \theta$ by Lemma 2.1(2); hence $(x, x) \in T_B^{\varepsilon,\theta}$. Symmetry follows from the symmetry of H and O (Lemma 2.1).

(2) $\max_{a \in B} H \le \varepsilon$ holds if and only if $H \le \varepsilon$ for every single $a \in B$, and $\min_{a \in B} O \ge \theta$ holds if and only if $O \ge \theta$ for every single $a \in B$.

(3) Taking maxima and minima over a larger index set gives $\max_{a \in B_2} H \ge \max_{a \in B_1} H$ and $\min_{a \in B_2} O \le \min_{a \in B_1} O$; hence the pair of constraints for B_2 implies those for B_1.

(4) Items 4 and 5 follow directly from the directions of the two defining inequalities.

□

Definition 2.4. *[27] Let the attribute subset $B \subseteq A$. A subset $X \subseteq U$ is called an MCB with respect to B if the following two conditions hold:*

(1) For any $x_i, x_j \in X$, $(x_i, x_j) \in T_B^{\varepsilon,\theta}$;
(2) There is not any set $Y \supset X$ with $Y \subseteq U$ such that Y satisfies Condition (1).

A set X satisfying Condition (1) is said to be consistent with respect to B (B-consistent for short). The collection of all MCBs with respect to B is denoted as $\mathcal{MCB}(B)$.

All samples within an MCB are mutually tolerant under the interval tolerance relation, and no proper superset of this block satisfies the mutual tolerance property. Therefore, MCBs naturally form high-quality local granulation units.

Theorem 2.2. *For any $B \subseteq A$:*

(1) $\mathcal{MCB}(B)$ is a covering of U, i.e., $\bigcup \mathcal{MCB}(B) = U$;
(2) if $B_1 \subseteq B_2 \subseteq A$, then every $X \in \mathcal{MCB}(B_2)$ is contained in some $X' \in \mathcal{MCB}(B_1)$;
(3) if $T_B^{\varepsilon,\theta}$ is an equivalence relation, then $\mathcal{MCB}(B) = U/T_B^{\varepsilon,\theta}$.

Proof. (1) For $x \in U$, $\{x\}$ is consistent by reflexivity (Theorem 2.1(1)); iteratively adding any object tolerant to all current members yields, by finiteness of U, an MCB $X \supseteq \{x\}$, hence $x \in X \in \mathcal{MCB}(B)$.

(2) Any $X \in \mathcal{MCB}(B_2)$ is B_2-consistent, hence B_1-consistent by Theorem 2.1(3); applying the same extension procedure to X gives a maximal B_1-consistent superset $X' \in \mathcal{MCB}(B_1)$.

(3) Under transitivity, consistent blocks are exactly subsets of equivalence classes, so MCBs coincide exactly with the equivalence classes.

□

2.3. Granular-ball computing and the GBRS

The GBRS [16] takes granular balls as knowledge granulation units and induces equivalence partitions via granular balls. It simultaneously inherits the interpretability of Pawlak rough set based on equivalence classes and the capability of neighborhood rough set to handle continuous data.

Definition 2.5. *[16] Let $D = \{x_i \mid i = 1, \ldots, n\}$ be a dataset, and let $GB = \{x_i \mid i = 1, \ldots, t\}$ be a granular ball; its center c and radius r are defined as*

$$c = \frac{1}{t} \sum_{i=1}^{t} x_i, \tag{2.6}$$

$$r = \max_{i=1}^{t} |x_i - c|. \tag{2.7}$$

Definition 2.6. *[16] Let $DT = \langle U, AT, V, f \rangle$ be a decision information system. For the attribute subset $B \subseteq A$, let GBR_B be the indiscernibility relation induced by granular balls, and let $U/GB(B)$ be the partition of the universe generated by GBR_B. Each equivalence class $[x]_{GB(B)} = \{y \in U \mid (x, y) \in GBR_B\}$*

forms a granular ball. For any $X \subseteq U$, the lower approximation and upper approximation of the GBRS are, respectively, defined as

$$\underline{GBR_B}(X) = \bigcup \left\{ [x]_{GB(B)} \in U/GB(B) \,\middle|\, [x]_{GB(B)} \subseteq X \right\}, \tag{2.8}$$

$$\overline{GBR_B}(X) = \bigcup \left\{ [x]_{GB(B)} \in U/GB(B) \,\middle|\, [x]_{GB(B)} \cap X \neq \emptyset \right\}. \tag{2.9}$$

Let $DT = \langle U, AT, V, f \rangle$ be a decision system. Let the set of decision attributes D partition the universe into Z equivalence classes satisfying $U/D = \{X_1, \ldots, X_Z\}$. The lower approximation, upper approximation, positive region, and boundary region of the set of decision attributes with respect to the set of conditional attributes B are, respectively, defined as

$$\underline{GBR_B}(D) = \bigcup_{i=1}^{Z} \underline{GBR_B}(X_i), \tag{2.10}$$

$$\overline{GBR_B}(D) = \bigcup_{i=1}^{Z} \overline{GBR_B}(X_i), \tag{2.11}$$

$$\mathrm{POS}_B(D) = \underline{GBR_B}(D), \tag{2.12}$$

$$\mathrm{BN}_B(D) = \overline{GBR_B}(D) - \underline{GBR_B}(D). \tag{2.13}$$

The positive region dependency and attribute significance are defined as

$$\gamma_B(D) = \frac{|\mathrm{POS}_B(D)|}{|U|}, \tag{2.14}$$

$$\mathrm{SIG}(a, B, D) = \gamma_{B\cup\{a\}}(D) - \gamma_B(D). \tag{2.15}$$

The forward attribute reduction of the GBRS selects attributes greedily based on the abovementioned dependency, until the positive region no longer expands. Compared with traditional NRSs, both the lower and upper approximations of GBRS are constructed from granular balls (equivalence classes). Thus the GBRS maintains the capability to handle continuous data while recovering the equivalence-class-based interpretability of knowledge.

3. Attribute reduction method for interval-valued decision information systems

This section is organized into two parts of different natures. The first part (Sections 3.1–3.3) constitutes the core mathematical contribution of this paper: It first presents a universal MCB generation algorithm applicable to both single-label and multi-label tasks, and then constructs GBRS attribute reduction methods for single-label and multi-label scenarios separately, including the granular-ball generation theory and algorithm as well as the attribute reduction theory and algorithm for each case. Finally, we use a simple numerical example to illustrate the computational process of the main definitions. The second part (Section 3.4) is auxiliary in nature: It introduces optional heuristic enhancement strategies and provides the complete workflow integrating the main algorithms with these strategies. These components accelerate the attribute search and improve practical performance, but they do not modify any definition or theoretical result of the core model, and removing them does not affect the validity of the method presented in the first part. In addition, Section 3.5 provides an analysis of the computational complexity of Algorithms 1– 7.

3.1. Maximal consistent block generation

The MCB serves as the core unit for granular-ball initialization in this paper. Its generation process only relies on the interval tolerance relation defined on the set of conditional attributes, independent of the single-label or multi-label structure of the set of decision attributes. Thus, it is established in advance as a general basic module.

For attribute subset $B \subseteq A$, the generation of MCBs adopts an efficient strategy: Tolerance neighborhood construction, maximality screening, and greedy clique repairing. We first calculate the tolerance class of each object x_i under B:

$$N_B(x_i) = \{x_j \in U \,\big|\, (x_i, x_j) \in T_B^{\varepsilon,\theta}\}. \tag{3.1}$$

All tolerance classes are deduplicated to form a set of candidate blocks. Maximality screening is implemented on these candidate blocks: If two candidate blocks X, Y satisfy $X \subset Y$, X is removed, and only candidate blocks without any proper superset are retained.

Let the tolerance matrix restricted to the candidate block X and the tolerance degree of an object be defined as follows:

$$M_{ij} = \mathbb{I}((x_i, x_j) \in T_B^{\varepsilon,\theta}), \tag{3.2}$$

$$\deg_X(x_i) = \sum_{x_j \in X} M_{ij}. \tag{3.3}$$

GreedyClique first orders all objects in X by non-increasing tolerance degree, with the object index used to break ties. It then scans the ordered objects once. An object is added to the repaired clique C only if it is consistent with every object already in C; otherwise, it is rejected. Since C only grows, every rejected object remains inconsistent with at least one retained object after the scan terminates. Consequently, C is an inclusion-maximal clique within X. The repaired blocks are subsequently deduplicated and filtered by set inclusion. This procedure produces an efficient deterministic approximation of the MCB collection rather than an exhaustive enumeration of all maximal cliques.

Based on the abovementioned principles, the steps of the MCB generation algorithm are shown in Algorithm 1.

Algorithm 1 Maximal consistent block generation

Input: Interval-valued decision information system $DT = \langle U, AT, V, f\rangle$, attribute subset $B \subseteq A$, distance threshold ε, overlap degree threshold θ.

Output: Approximate MCB collection $\widehat{\mathcal{MCB}}(B)$.

1: Calculate the tolerance class $N_B(x_i) = \{x_j \in U \mid (x_i, x_j) \in T_B^{\varepsilon,\theta}\}$ for each object $x_i \in U$.
2: Deduplicate all tolerance classes to form a candidate block set $Cand$.
3: Perform maximality screening on $Cand$: If $X, Y \in Cand$ satisfy $X \subset Y$, remove X; retain only blocks without any proper superset, yielding $MaxCand$.
4: Initialize the repaired-block collection "Repaired" as empty.
5: **for** each block X in $MaxCand$ **do**
6: Calculate the tolerance degree of every object and sort the objects by decreasing degree, breaking ties by increasing the object index.
7: **if** X is already a clique **then**

Set $C = X$.
else
Initialize C as empty, scan the ordered objects, and add an object to C only if it is consistent with every object already in C.
end if
Add the resulting C to Repaired.
end for
Deduplicate Repaired and remove every repaired block that is a proper subset of another repaired block.
return $\widehat{\mathcal{MCB}}(B)$.

3.2. Single-label GBRS attribute reduction

This subsection focuses on single-label interval-valued decision information systems, and constructs a complete reduction framework consisting of granular-ball initialization, splitting, heterogeneous overlap removal, positive region calculation, and forward attribute selection.

3.2.1. Structural parameters and purity measure of interval-valued granular balls

The center of each granular ball in the traditional GBRS [16] is the numerical mean of all samples, and its radius is the maximum Euclidean distance from the samples to the center. For interval-valued scenarios, granular-ball parameters should be adapted to an interval-weighted form.

Definition 3.1. *Let $GB \subseteq U$ be a granular ball, with the attribute $a \in B$, and let $f(x, a) = [l_{x,a}, r_{x,a}]$ denote the value of $x \in GB$ on attribute a, and let $m_{x,a} = \frac{l_{x,a}+r_{x,a}}{2}$ denotes the midpoint of this interval. To reduce the excessive impact of high-uncertainty samples (long intervals) on the global center position and to enable the center to stably reflect the concentration trend of high-confidence samples within the granular ball, intervals with larger lengths are assigned lower weights. The interval-weighted center on attribute a is defined as*

$$c_{GB}^{a} = \frac{\sum_{x \in GB} \omega_{x,a} \cdot m_{x,a}}{\sum_{x \in GB} \omega_{x,a}}, \tag{3.4}$$

where $\omega_{x,a} = \frac{1}{r_{x,a}-l_{x,a}+10^{-6}}$ is the interval length weight. The interval length $r_{x,a} - l_{x,a}$ quantifies the uncertainty of x on attribute a: A longer interval implies a less reliable point estimate of the true value. We adopt an inverse-length weighting scheme, which is analogous to inverse-variance weighting in statistical estimation, where observations are weighted inversely to their uncertainty so that high-confidence samples dominate the estimation of the center. Compared with other decreasing functions of interval length (e.g., linear or exponential forms), the inverse form is parameter-free and monotonically decreasing, and assigns the highest weight to exact (degenerate) values. The small constant 10^{-6} is added to the denominator to keep the weight well-defined for exact values with a zero interval length.

If $B = \emptyset$, the center is an empty vector, and the center vector c_{GB} of the granular ball is composed of the center values of all attributes.

The granular-ball's radius adopts a multi attribute Hausdorff-Euclidean distance metric. The center is treated as a degenerate interval $[c_{GB}^{a}, c_{GB}^{a}]$, and the distance from sample x to the center on attribute

a is $H(f(x,a),[c_{GB}^a, c_{GB}^a])$*. The integrated multi attribute radius of the granular ball is defined as*

$$r_{GB} = \max_{x\in GB}\left\{\sqrt{\sum_{a\in B}\left[H(f(x,a),[c_{GB}^a, c_{GB}^a])\right]^2}\right\}. \tag{3.5}$$

Definition 3.2. *Let GB be a granular ball. The single-label purity is defined as*

$$P_{\text{single}}(GB) = \frac{\max_{d\in V_D} n_d^{GB}}{|GB|}, \tag{3.6}$$

where $n_d^{GB} = \left|\{x \in GB \mid f(x,D) = d\}\right|$ *for* $d \in V_D$*, which denotes the number of samples belonging to label d within granular ball GB.*

When $P_{\text{single}}(GB) = 1$, all samples inside the granular ball fall into the same decision class, satisfying the complete label consistency requirement of the positive region in traditional rough sets.

In addition, to guarantee the tolerance of samples within a granular ball in the interval sense, we define the single-label tolerance purity as

$$P_{\text{single}}^{\text{tol}}(GB) = P_{\text{single}}(GB) \times \frac{\sum_{x_i,x_j\in GB}\mathbb{I}((x_i,x_j)\in T_B^{\varepsilon,\theta})}{|GB|^2}, \tag{3.7}$$

where $\mathbb{I}(\cdot)$ represents the indicator function, which equals 1 if the tolerance relation holds and 0 otherwise. $P_{\text{single}}^{\text{tol}}(GB) \geq T_{\text{single}}^{\text{tol}}$ indicates that every pair of samples inside the granular ball satisfies the interval tolerance relation, in which $T_{\text{single}}^{\text{tol}}$ stands for the threshold of single-label tolerance purity.

3.2.2. Granular-ball splitting and heterogeneous overlap removal mechanism

The granular-ball generation process iteratively splits granular balls to make each granular ball gradually satisfy the purity and tolerance constraints, and eliminates geometric overlaps between heterogeneous granular balls. This mechanism of splitting and heterogeneous overlap removal follows the granular-ball generation framework of the GBRS [16], while replacing the distance metric with the multi-attribute Hausdorff-Euclidean distance to adapt to interval-valued data.

(1) Granular-ball splitting.

Let GB be the current granular ball, and let the attribute subset $B \subseteq A$. We first judge whether the sample size of GB exceeds the minimum size threshold $LBS(B)$. This paper adopts a dynamic threshold defined as $LBS(B) = \max(2, 2|B|)$, where $|B|$ denotes the cardinality of attribute subset B. If $|GB| \leq LBS(B)$, this granular ball is directly retained as a final granular ball without further splitting; if $|GB| > LBS(B)$, it is processed according to the following three rules.

1) If the single-label purity fails to reach the standard, i.e., $P_{\text{single}}(GB) < 1$, the granular ball needs to be split into k sub-granular balls according to label differences, where k is the number of distinct decision class labels inside the granular ball;
2) If the single-label purity meets the standard yet the tolerance purity is insufficient, i.e., $P_{\text{single}}(GB) = 1$ and $P_{\text{single}}^{\text{tol}}(GB) < T_{\text{single}}^{\text{tol}}$, the granular ball is split into two sub-granular-balls by default to refine the interval structure and improve the internal tolerance;

3) If both the single-label purity and tolerance purity reach the standard, the granular ball is retained directly.

The splitting operation is implemented by the interval k-means algorithm. The cluster centers of this algorithm take interval forms, and their left and right endpoints need to be maintained separately. Let C be a cluster, and let its interval center be $c = [c_1, c_2, \ldots, c_m]$. The interval center on attribute a is $c_a = [\bar{l}_a, \bar{r}_a]$, and let $I_{x,a} = [l_{x,a}, r_{x,a}]$ denote the interval value of sample x on attribute a, where

$$\bar{l}_a = \frac{\sum_{x \in C} \omega'_{x,a} \cdot l_{x,a}}{\sum_{x \in C} \omega'_{x,a}}, \tag{3.8}$$

$$\bar{r}_a = \frac{\sum_{x \in C} \omega'_{x,a} \cdot r_{x,a}}{\sum_{x \in C} \omega'_{x,a}}. \tag{3.9}$$

Here, the weight is defined as the interval length $\omega'_{x,a} = r_{x,a} - l_{x,a}$. Longer intervals carry larger "information volume" (i.e., the discriminability brought by uncertainty). When calculating the cluster centers, such weights can better capture the overall distribution of intervals, making the clusters after splitting both reduce the discrete degree of internal intervals and optimize the intra cluster label consistency; longer intervals are assigned higher weights. If all intervals share identical lengths, the formulas degenerate into ordinary interval averages:

$$\bar{l}_a = \frac{1}{|C|} \sum_{x \in C} l_{x,a}, \quad \bar{r}_a = \frac{1}{|C|} \sum_{x \in C} r_{x,a}. \tag{3.10}$$

The distance metric between samples adopts the multi attribute Hausdorff-Euclidean distance:

$$d(x_i, x_j) = \sqrt{\sum_{a \in B} \left[H(f(x_i, a), f(x_j, a))\right]^2}. \tag{3.11}$$

The steps above are executed iteratively until the granular-ball set no longer changes. At this point, all granular balls satisfy both the purity and size constraints.

(2) Heterogeneous overlap removal

If heterogeneous granular balls with geometric overlaps exist in the granular-ball set, further processing is required to eliminate classification ambiguity. Let GB_i and GB_j be two granular balls; they are judged to be heterogeneous if their labels differ. Overlap judgment relies on the multi attribute Hausdorff-Euclidean distance between the centers of granular balls. The center distance between the two granular balls is

$$d_{\text{center}}(GB_i, GB_j) = \sqrt{\sum_{a \in B} \left[H([c^a_{GB_i}, c^a_{GB_i}], [c^a_{GB_j}, c^a_{GB_j}])\right]^2}. \tag{3.12}$$

If $d_{\text{center}}(GB_i, GB_j) < r_{GB_i} + r_{GB_j}$, the two granular balls are determined to overlap.

For any pair of heterogeneous and overlapping granular balls, split the larger one into two sub-granular-balls: If $|GB_i| > |GB_j|$, split GB_i; otherwise split GB_j. This process is executed iteratively until no overlaps remain between heterogeneous granular balls, and a non overlapping granular-ball set $\mathcal{NOLGBS}$ is finally obtained.

To intuitively demonstrate the complete execution flow of the abovementioned granular-ball splitting and heterogeneous overlap removal, Figure 1 provides a schematic diagram of the interval-valued granular-ball generation process based on MCBs. Unlike the original GBRS framework [16], which takes all samples as the sole initial granular ball, this process first takes the collection of MCBs as the initial granular-ball set, then iteratively performs single-label purity judgment, single-label tolerance purity judgment, interval k-means splitting, and heterogeneous overlap removal until all granular balls satisfy the constraints. The granular-ball generation flow for multi-label scenarios is consistent with this process, and the main differences lie in the definition of purity and the heterogeneous judgment criterion.

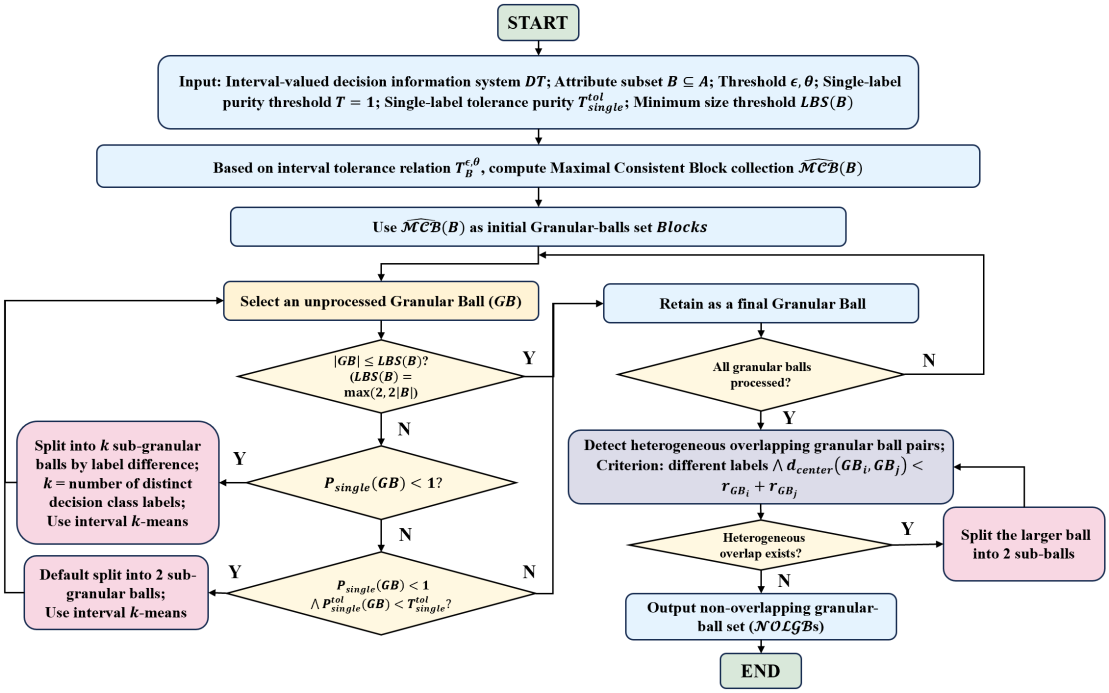

Figure 1. Simplified interval-valued granular-ball generation process based on MCBs.

3.2.3. Single-label granular-ball generation algorithm

Based on the abovementioned granular-ball structural parameters, splitting rules and heterogeneous overlap removal mechanism, the single-label granular-ball generation algorithm takes MCBs as initial granular balls. After iterative splitting and heterogeneous overlap removal, it outputs a non-overlapping granular-ball set as well as the positive region.

Definition 3.3. *Let $\mathcal{NOLGBs}$ be a non-overlapping granular-ball set and let $D_i \subseteq U$ be a decision class. The single-label tolerance lower approximation and upper approximation are, respectively, defined as*

$$\underline{apr}_B^{\text{tol}}(D_i) = \bigcup \left\{ GB \in \mathcal{NOLGBs} \,\middle|\, GB \subseteq D_i \ \wedge\ P_{\text{single}}^{\text{tol}}(GB) \geq T_{\text{single}}^{\text{tol}} \right\}. \tag{3.13}$$

$$\overline{apr}_B^{\text{tol}}(D_i) = \bigcup \left\{ GB \in \mathcal{NOLGBs} \,\middle|\, GB \cap D_i \neq \emptyset \ \wedge\ P_{\text{single}}^{\text{tol}}(GB) \geq T_{\text{single}}^{\text{tol}} \right\}. \tag{3.14}$$

The single-label tolerance positive region is the union of the lower approximations of all decision

classes:

$$\mathrm{POS}_B^{\mathrm{tol}}(D) = \bigcup_{i=1}^{k} \underline{apr}_B^{\mathrm{tol}}(D_i). \tag{3.15}$$

The steps of the single-label granular-ball generation algorithm are shown in Algorithm 2.

Algorithm 2 Single-label granular-ball generation

Input: Single-label interval-valued decision information system $DT = \langle U, A \cup \{d\}, V, f\rangle$, attribute subset $B \subseteq A$, thresholds ε, θ, single-label purity threshold $T' = 1$, tolerance purity threshold $T_{\mathrm{single}}^{\mathrm{tol}}$, minimum granular ball size $LBS(B) = \max(2, 2|B|)$.

Output: Non overlapping granular-ball set $\mathcal{NOLGBS}_B$; single-label tolerance of lower and upper approximations; positive region $\mathrm{POS}_B^{\mathrm{tol}}(D)$; dependency $\gamma_B^{\mathrm{tol}}(D)$.

Call Algorithm 1 to obtain $\widehat{\mathcal{MCB}}(B)$; let the initial block set $Blocks = \widehat{\mathcal{MCB}}(B)$.
for each $GB \in Blocks$ **do**
  Compute interval-weighted center $c_{GB}^a = \frac{\sum_{x\in GB} \omega_{x,a} \cdot m_{x,a}}{\sum_{x\in GB} \omega_{x,a}}$, where $m_{x,a} = \frac{l_{x,a}+r_{x,a}}{2}$ and $\omega_{x,a} = \frac{1}{r_{x,a}-l_{x,a}+10^{-6}}$.
  Compute radius $r_{GB} = \max_{x\in GB}\left\{\sqrt{\sum_{a\in B}\left[H(f(x,a),[c_{GB}^a, c_{GB}^a])\right]^2}\right\}$.
  Compute single-label purity $P_{\mathrm{single}}(GB) = \frac{\max_{d\in V_D} n_d^{GB}}{|GB|}$.
  Compute tolerance purity $P_{\mathrm{single}}^{\mathrm{tol}}(GB) = P_{\mathrm{single}}(GB) \cdot \frac{\sum_{x_i,x_j\in GB} \mathbb{I}((x_i,x_j)\in T_B^{\varepsilon,\theta})}{|GB|^2}$.
end for
repeat
  for each $GB \in Blocks$ **do**
    if $|GB| \le LBS(B)$ **then**
      Retain GB directly.
    else if $P_{\mathrm{single}}(GB) < 1$ **then**
      Let k be the number of distinct decision classes in GB; split GB into k sub-granular-balls via interval k-means.
    else if $P_{\mathrm{single}}(GB) = 1$ and $P_{\mathrm{single}}^{\mathrm{tol}}(GB) < T_{\mathrm{single}}^{\mathrm{tol}}$ **then**
      Split GB into two sub-granular-balls via interval k-means.
    else
      Retain GB.
    end if
  end for
  Deduplicate the split blocks and update $Blocks$.
until no new splits occur in a round
Let the candidate granular-ball set be $GBs_B = Blocks$.
repeat
  if two granular balls GB_i, GB_j have different decision labels and $d_{\mathrm{center}}(GB_i, GB_j) < r_{GB_i} + r_{GB_j}$ **then**
    Split the granular ball with the larger sample size.
  end if
  Update GBs_B.
until no geometric overlap exists between heterogeneous granular balls

29: Let $\mathcal{NOLGB}\int_B = GBs_B$.
30: **for** each decision class D_i **do**
31: Compute lower approximation $\underline{apr}_B^{\text{tol}}(D_i) = \bigcup\{GB \in \mathcal{NOLGB}s_B \mid GB \subseteq D_i \wedge P_{\text{single}}^{\text{tol}}(GB) \geq T_{\text{single}}^{\text{tol}}\}$.
32: Compute upper approximation $\overline{apr}_B^{\text{tol}}(D_i) = \bigcup\{GB \in \mathcal{NOLGB}s_B \mid GB \cap D_i \neq \emptyset \wedge P_{\text{single}}^{\text{tol}}(GB) \geq T_{\text{single}}^{\text{tol}}\}$.
33: **end for**
34: Compute positive region $\text{POS}_B^{\text{tol}}(D) = \bigcup_{i=1}^{k} \underline{apr}_B^{\text{tol}}(D_i)$.
35: Compute dependency $\gamma_B^{\text{tol}}(D) = \frac{|\text{POS}_B^{\text{tol}}(D)|}{|U|}$.
36: **return** $\mathcal{NOLGB}s_B$, approximations, $\text{POS}_B^{\text{tol}}(D)$, $\gamma_B^{\text{tol}}(D)$.

Theorem 3.1. *Let $\mathcal{NOLGB}s_B$ be the output of Algorithm 2 on the attribute subset B. We then have the following.*

(1) Every $GB \in \mathcal{NOLGB}s_B$ is contained in some $X \in \widehat{\mathcal{MCB}}(B)$; hence every generated ball is a $T_B^{\varepsilon,\theta}$-clique (a set of objects that are pairwise tolerant under $T_B^{\varepsilon,\theta}$) and satisfies $P_{\text{single}}^{\text{tol}}(GB) = P_{\text{single}}(GB)$.

(2) If $X \in \mathcal{MCB}(B)$ satisfies $X \subseteq D_i$ for some class D_i, then $X \subseteq \underline{apr}_B^{\text{tol}}(D_i)$; consequently, $\bigcup\{X \in \mathcal{MCB}(B) : \exists i,\ X \subseteq D_i\} \subseteq \text{POS}_B^{\text{tol}}(D)$, i.e., the pure consistent core of $\mathcal{MCB}(B)$ is stably retained in the positive region.

Proof. (1) By induction on the operations: The initial block set is $\widehat{\mathcal{MCB}}(B)$; each splitting operation is implemented by interval k-means on the parent's sample set and therefore partitions the parent into subsets, and the subsets of a clique are cliques; deduplication and maximality screening only delete blocks. Hence the invariant holds, and $P_{\text{single}}^{\text{tol}} = P_{\text{single}}$ follows, since the tolerant pair ratio is 1.

(2) Every descendant GB of X satisfies $GB \subseteq X \subseteq D_i$, is decision-pure (all samples within a granular ball belong to the same decision class), and, by (1), is a clique, so $P_{\text{single}}^{\text{tol}}(GB) = 1 \geq T_{\text{single}}^{\text{tol}}$; hence GB enters the union defining $\underline{apr}_B^{\text{tol}}(D_i)$ in Definition 3.3, and X equals the union of its descendants.

□

3.2.4. Single-label attribute reduction

Based on the granular-ball generation results obtained from Algorithm 2, a forward greedy strategy is adopted to perform attribute reduction.

Definition 3.4. *Let attribute subset $B \subseteq A$, and let $a \in A - B$ be a candidate attribute. The single-label tolerance attribute significance is defined as*

$$\text{SIG}_{\text{tol}}(a, B, D) = \gamma_{B\cup\{a\}}^{\text{tol}}(D) - \gamma_B^{\text{tol}}(D), \tag{3.16}$$

where the single-label tolerance dependency is defined as

$$\gamma_B^{\text{tol}}(D) = \frac{\left|\text{POS}_B^{\text{tol}}(D)\right|}{|U|}. \tag{3.17}$$

This significance metric quantifies the incremental contribution to the positive region after attribute a is added to the current subset, and it acts as the theoretical basis for forward greedy selection.

Theorem 3.2. *Let $B_1 \subseteq B_2 \subseteq A$ and suppose that $\mathcal{NOLGBS}_{B_2}$ union-refines $\mathcal{NOLGBS}_{B_1}$, i.e., every ball of $\mathcal{NOLGBS}_{B_1}$ is a union of balls of $\mathcal{NOLGBS}_{B_2}$. Then $\underline{apr}^{\mathrm{tol}}_{B_1}(D_i) \subseteq \underline{apr}^{\mathrm{tol}}_{B_2}(D_i)$ for every class D_i, and hence $\mathrm{POS}^{\mathrm{tol}}_{B_1}(D) \subseteq \mathrm{POS}^{\mathrm{tol}}_{B_2}(D)$ and $\gamma^{\mathrm{tol}}_{B_1}(D) \leq \gamma^{\mathrm{tol}}_{B_2}(D)$. Consequently, along any forward selection path on which each extension $B \cup \{a\}$ induces a union refinement,*

$$\mathrm{SIG}_{\mathrm{tol}}(a, B, D) = \gamma^{\mathrm{tol}}_{B\cup\{a\}}(D) - \gamma^{\mathrm{tol}}_{B}(D) \geq 0. \tag{3.18}$$

Proof. Let $GB_1 \in \mathcal{NOLGBS}_{B_1}$ satisfy $GB_1 \subseteq D_i$. By the assumption $GB_1 = \bigcup_j GB_j$ with $GB_j \in \mathcal{NOLGBS}_{B_2}$. Each $GB_j \subseteq D_i$, and by Theorem 3.1(1), a clique, so $P^{\mathrm{tol}}_{\mathrm{single}}(GB_j) = 1 \geq T^{\mathrm{tol}}_{\mathrm{single}}$ and $GB_j \subseteq \underline{apr}^{\mathrm{tol}}_{B_2}(D_i)$. Hence $GB_1 \subseteq \underline{apr}^{\mathrm{tol}}_{B_2}(D_i)$; taking unions over all such GB_1 and all classes proves the approximation and positive-region inclusions, and the dependency inequality follows. The significance claim is the special case $B_1 = B$, $B_2 = B \cup \{a\}$. □

The steps of the single-label attribute reduction algorithm are shown in Algorithm 3.

Algorithm 3 Single-label attribute reduction

Input: Single-label interval-valued decision information system $DT = \langle U, A \cup \{d\}, V, f\rangle$; thresholds $\varepsilon, \theta, T', T^{\mathrm{tol}}_{\mathrm{single}}$; maximum attribute count $K_{\max}$.

Output: Single-label reduction set $R \subseteq A$.

Initialize $R = \emptyset$, *Remaining* $= A$, $\gamma_{\mathrm{old}} = 0$.
while *Remaining* $\neq \emptyset$ and $|R| < K_{\max}$ **do**
  for each candidate attribute $a \in$ *Remaining* **do**
    Let $B = R \cup \{a\}$.
    Call Algorithm 2 to generate granular balls on B and compute $\mathrm{POS}^{\mathrm{tol}}_B(D)$, obtaining $\gamma^{\mathrm{tol}}_B(D)$.
  end for
  Compute attribute significance $\mathrm{SIG}_{\mathrm{tol}}(a, R, D) = \gamma^{\mathrm{tol}}_{R\cup\{a\}}(D) - \gamma^{\mathrm{tol}}_R(D)$ for each a; select a^* with maximum significance.
  if $\gamma^{\mathrm{tol}}_{R\cup\{a^*\}}(D) > \gamma_{\mathrm{old}}$ **then**
    Update $R \leftarrow R \cup \{a^*\}$, *Remaining* $\leftarrow$ *Remaining* $\setminus \{a^*\}$, $\gamma_{\mathrm{old}} \leftarrow \gamma^{\mathrm{tol}}_R(D)$.
  else
    Stop forward selection.
  end if
end while
if $R = \emptyset$ **then**
  As a fallback, select the single attribute a maximizing $\gamma^{\mathrm{tol}}_{\{a\}}(D)$.
end if
return R.

Theorem 3.3 (Relationship with existing rough set models)**.** *(1) PRs: On point-valued data with $\theta \in (0, 1]$, $T^{\varepsilon,\theta}_B$ coincides with the indiscernibility relation $\mathrm{IND}(B)$ for every $B \subseteq A$ and every $\varepsilon > 0$, and $\mathcal{MCB}(B) = U/\mathrm{IND}(B)$; if the equivalence classes are retained intact as balls, then the lower approximations, positive regions, and dependencies coincide with those of the Pawlak*

model, while the upper approximation coincides, provided that $T^{\mathrm{tol}}_{\mathrm{single}}$ *does not exceed the purity of any mixed class; in the general pipeline with label-driven splitting,* $\mathrm{POS}_B(D) \subseteq \mathrm{POS}^{\mathrm{tol}}_B(D)$ *and* $\gamma_B(D) \le \gamma^{\mathrm{tol}}_B(D)$.

(2) *GBRS: On point-valued data, formally setting* $\theta = 0$ *and initializing with the single ball* U *instead of* $\mathcal{MCB}(A)$*, Definitions 3.1– 3.3 reduce to GBRS [16] (the center degenerates to the arithmetic mean, the radius to the maximum Euclidean distance, the purity coincides, and the interval k-means degenerates to standard k-means); hence, the proposed model is a strict generalization of the GBRS, and MaxGBRS is its point-metric special case.*

(3) *NRS: On point-valued data with* $\theta = 0$*,* $N_B(x)$ *is exactly the* L_∞ *neighborhood of [11], and the NRS-positive region satisfies* $\mathrm{POS}^{\mathrm{NRS}}_B(D) \subseteq \mathrm{POS}^{\mathrm{tol}}_B(D)$*, and hence* $\gamma^{\mathrm{NRS}}_B(D) \le \gamma^{\mathrm{tol}}_B(D)$.

Proof. (1) For point values, $O \in \{0, 1\}$ and equals 1 if and only if the two values are equal; since $\theta > 0$, $\min_{a\in B} O \ge \theta$ forces equality on every attribute, and $H = 0 \le \varepsilon$ is then automatic, so $T^{\varepsilon,\theta}_B = \mathrm{IND}(B)$. Theorem 2.2(3) gives $\mathcal{MCB}(B) = U/\mathrm{IND}(B)$. Pure classes have $P^{\mathrm{tol}}_{\mathrm{single}} = 1$ and enter the lower approximation exactly as in Pawlak's model; mixed classes are excluded from the lower approximations in both models, hence the lower approximations, positive regions, and dependencies coincide. With label-driven splitting, every pure class $E \subseteq D_i$ is retained in $\mathrm{POS}^{\mathrm{tol}}_B(D)$ by Theorem 3.1(3), giving the stated inclusions.

(2) With zero interval lengths, the weights $\omega_{x,a} = 1/(0 + 10^{-6})$ are constant, so the weighted center equals the arithmetic mean of the midpoints, i.e., of the point values; $H(f(x, a), [c_a, c_a]) = |x_a - c_a|$, so the radius is $\max_{x\in GB} \sqrt{\sum_{a\in B}(x_a - c_a)^2}$; identical zero lengths make the interval k-means centers degenerate to ordinary averages, i.e., standard k-means.

(3) On point data with $\theta = 0$, $N_B(x) = \{y \mid \max_{a\in B} |x_a - y_a| \le \varepsilon\}$, which is the L_∞ neighborhood. If x belongs to the NRS positive region, i.e., $N_B(x) \subseteq D_{d_x}$, take $X \in \mathcal{MCB}(B)$ with $x \in X$ (Theorem 2.2(1)); since X is a clique containing x, every $y \in X$ is tolerant to x, so $X \subseteq N_B(x) \subseteq D_{d_x}$, and Theorem 3.1(3) yields $x \in \mathrm{POS}^{\mathrm{tol}}_B(D)$.

□

3.3. Multi-label GBRS attribute reduction

This section extends the abovementioned framework to multi-label scenarios. The multi-label pipeline maintains consistency with the single-label case in the granular-ball generation and reduction workflow; the main differences lie in the definition of purity, the heterogeneous judgment rules, and the definition of the positive region.

3.3.1. Multi-label granular-ball purity and tolerance metric

The structural parameters of multi-label granular balls (interval-weighted center, radius) are consistent with those for the single-label case, as stated in Definition 3.1. This subsection focuses on constructing the granular-ball purity metric under multi-label conditions.

For multi-label tasks, the Jaccard similarity metric is adopted to measure the consistency of label subsets.

Definition 3.5. *Let* x_i, x_j *be two samples inside granular ball GB. The Jaccard similarity of their label*

subsets is defined as

$$J(L_{x_i}, L_{x_j}) = \frac{|L_{x_i} \cap L_{x_j}|}{|L_{x_i} \cup L_{x_j}|}. \tag{3.19}$$

The multi-label purity of a granular ball is defined as the average Jaccard similarity over all sample pairs within the ball

$$P_{\text{multi}}(GB) = \frac{1}{|GB|^2} \sum_{x_i, x_j \in GB} J(L_{x_i}, L_{x_j}). \tag{3.20}$$

$P_{\text{multi}}(GB)$ *takes values in the range* $[0, 1]$*; larger values indicate higher consistency of the label subsets among samples inside the granular ball.*

The multi-label tolerance purity is defined as:

$$P^{\text{tol}}_{\text{multi}}(GB) = \frac{1}{|GB|^2} \sum_{x_i, x_j \in GB} \left[J(L_{x_i}, L_{x_j}) \times \mathbb{I}((x_i, x_j) \in T_B^{\varepsilon,\theta}) \right]. \tag{3.21}$$

$P^{\text{tol}}_{\text{multi}}(GB) \geq T^{\text{tol}}_{\text{multi}}$ *means that the granular ball satisfies interval tolerance, where* $T^{\text{tol}}_{\text{multi}}$ *denotes the threshold of multi-label tolerance purity.*

3.3.2. Multi-label granular-ball splitting and heterogeneous overlap removal mechanism

Multi-label granular-ball generation also iteratively splits granular balls to make each granular ball gradually satisfy the purity and tolerance constraints, and eliminates geometric overlaps between heterogeneous granular balls.

(1) Granular-ball splitting

Let GB be the current granular ball, and let the attribute subset $B \subseteq A$. We first judge whether the sample size of GB exceeds the minimum size threshold $LBS(B)$. If $|GB| \leq LBS(B)$, this granular ball is directly retained as a final granular ball without further splitting; if $|GB| > LBS(B)$, it is processed according to the following three rules:

1) If the multi-label purity fails to reach the standard, i.e., $P_{\text{multi}}(GB) < T$ (where T denotes the threshold of multi-label purity), the granular ball needs to be split into k sub-granular-balls according to the label differences, where k is the number of distinct label subsets contained in the granular ball;
2) If the multi-label purity meets the standard yet the tolerance purity is insufficient, i.e., $P_{\text{multi}}(GB) \geq T$ and $P^{\text{tol}}_{\text{multi}}(GB) < T^{\text{tol}}_{\text{multi}}$, the granular ball is split into two sub-granular-balls by default to refine the interval structure and improve internal tolerance;
3) If both purity metrics meet their respective standards, the granular ball is retained directly.

The splitting operation also uses the interval k-means algorithm. The update rule for cluster centers is identical to that for the single-label scenario, which separately maintains the interval-length weighted averages of the left and right endpoints.

(2) Heterogeneous overlap removal

The heterogeneous judgment rule for multi-label scenarios is defined as follows: Let GB_i and GB_j be two granular balls. They are judged as heterogeneous if the label purity of their union fails to meet the standard, i.e.

$$P(GB_i \cup GB_j) < T. \tag{3.22}$$

Different from the single-label case where the labels are compared directly, two granular balls in multi-label tasks may share partial labels. Thus union purity judgment is adopted to tolerate partial label overlap and avoid excessive splitting. The overlap judgment formula remains consistent with that of the single-label case, which relies on the center's Hausdorff-Euclidean distance and the sum of radii for judgment:

$$\begin{aligned} d_{\text{center}}(GB_i, GB_j) &= \sqrt{\sum_{a\in B}\left[H([c^a_{GB_i}, c^a_{GB_i}], [c^a_{GB_j}, c^a_{GB_j}])\right]^2} \\ &< r_{GB_i} + r_{GB_j}. \end{aligned} \tag{3.23}$$

For any pair of heterogeneous and overlapping granular balls, split the larger one into two sub-granular-balls. This process is executed iteratively until no overlaps remain between heterogeneous granular balls, and a non-overlapping granular-ball set $\mathcal{NOLGBS}$ is finally obtained.

3.3.3. Multi-label granular-ball generation algorithm

Based on the abovementioned definition of granular balls' purity, splitting rules, and the heterogeneous overlap removal mechanism, the multi-label granular-ball generation algorithm takes MCBs as the initial granular balls. After iterative splitting and heterogeneous overlap removal, it outputs a non overlapping granular-ball set as well as the positive region.

Definition 3.6. *Let $\mathcal{NOLGBS}$ be a non-overlapping granular-ball set, let $l_p \in L$ be an arbitrary label, and let $L_p^+ = \{x \in U \mid l_p \in L_x\}$ denote the set of samples that contain the label l_p. The tolerance of the lower approximation and upper approximation of label l_p are, respectively, defined as*

$$\underline{apr}_B^{\text{tol}}(L_p^+) = \bigcup \left\{ GB \in \mathcal{NOLGBS} \,\middle|\, GB \subseteq L_p^+ \ \wedge\ P^{\text{tol}}_{\text{multi}}(GB) \geq T^{\text{tol}}_{\text{multi}} \right\}. \tag{3.24}$$

$$\overline{apr}_B^{\text{tol}}(L_p^+) = \bigcup \left\{ GB \in \mathcal{NOLGBS} \,\middle|\, GB \cap L_p^+ \neq \emptyset \ \wedge\ P^{\text{tol}}_{\text{multi}}(GB) \geq T^{\text{tol}}_{\text{multi}} \right\}. \tag{3.25}$$

The multi-label tolerance-positive region of the system is the union of the lower approximations of all label classes

$$\text{POS}_B^{\text{tol}}(L) = \bigcup_{p=1}^{q} \underline{apr}_B^{\text{tol}}(L_p^+). \tag{3.26}$$

The steps of the multi-label granular-ball generation algorithm are shown in Algorithm 4.

Algorithm 4 Multi-label granular-ball generation

Input: Multi-label interval-valued decision information system $DT = \langle U, A \cup L, V, f \rangle$, attribute subset $B \subseteq A$, multi-label matrix $Y \in \{0,1\}^{|U|\times q}$, thresholds ε, θ, multi-label purity threshold T, tolerance purity threshold $T^{\text{tol}}_{\text{multi}}$, minimum granular-ball size $LBS(B) = \max(2, 2|B|)$.

Output: Multi-label non overlapping granular-ball set $\mathcal{NOLGBS}_B$; tolerance of the lower and upper approximations for each label; multi-label positive region $\text{POS}_B^{\text{tol}}(L)$; dependency $\gamma_B^{\text{tol}}(L)$.

1: Call Algorithm 1 to obtain $\widehat{\mathcal{MCB}}(B)$; let $Blocks = \widehat{\mathcal{MCB}}(B)$.
2: **for** each $GB \in Blocks$ **do**
3: Compute multi-label Jaccard purity $P_{\text{multi}}(GB) = \frac{1}{|GB|^2}\sum_{x_i,x_j\in GB} J(L_{x_i}, L_{x_j})$, where $J(L_{x_i}, L_{x_j}) = \frac{|L_{x_i}\cap L_{x_j}|}{|L_{x_i}\cup L_{x_j}|}$.

Compute multi-label tolerance purity $P^{\text{tol}}_{\text{multi}}(GB) = \frac{1}{|GB|^2}\sum_{x_i,x_j\in GB} J(L_{x_i}, L_{x_j}) \cdot \mathbb{I}((x_i, x_j) \in T_B^{\varepsilon,\theta})$.
end for
repeat
  for each $GB \in Blocks$ **do**
    if $|GB| \le LBS(B)$ **then**
      Retain GB directly.
    else if $P_{\text{multi}}(GB) < T$ **then**
      Let k be the number of distinct label subsets in GB; split GB into k sub-granular-balls via interval k-means.
    else if $P_{\text{multi}}(GB) \ge T$ and $P^{\text{tol}}_{\text{multi}}(GB) < T^{\text{tol}}_{\text{multi}}$ **then**
      Split GB into two sub-granular-balls via interval k-means.
    else
      Retain GB.
    end if
  end for
  Deduplicate split blocks and update $Blocks$.
until no new splits occur in a round
Let $GBs_B = Blocks$.
repeat
  if two granular balls GB_i, GB_j satisfy $P_{\text{multi}}(GB_i \cup GB_j) < T$ and $d_{\text{center}}(GB_i, GB_j) < r_{GB_i} + r_{GB_j}$ **then**
    Split the granular ball with the larger sample size.
  end if
  Update GBs_B.
until no geometric overlap exists between heterogeneous granular balls
Let $\mathcal{NOLGBS}_B = GBs_B$.
for each label $l_p \in L$ **do**
  Construct set $L_p^+ = \{x \in U \mid l_p \in L_x\}$.
  Compute lower approximation $\underline{apr}^{\text{tol}}_B(L_p^+) = \bigcup\{GB \in \mathcal{NOLGBS}_B \mid GB \subseteq L_p^+ \wedge P^{\text{tol}}_{\text{multi}}(GB) \ge T^{\text{tol}}_{\text{multi}}\}$.
  Compute upper approximation $\overline{apr}^{\text{tol}}_B(L_p^+) = \bigcup\{GB \in \mathcal{NOLGBS}_B \mid GB \cap L_p^+ \ne \emptyset \wedge P^{\text{tol}}_{\text{multi}}(GB) \ge T^{\text{tol}}_{\text{multi}}\}$.
end for
Compute the positive region $\text{POS}^{\text{tol}}_B(L) = \bigcup_{p=1}^{q} \underline{apr}^{\text{tol}}_B(L_p^+)$.
Compute dependency $\gamma^{\text{tol}}_B(L) = \frac{|\text{POS}^{\text{tol}}_B(L)|}{|U|}$.
return $\mathcal{NOLGBS}_B$, approximations, $\text{POS}^{\text{tol}}_B(L)$, $\gamma^{\text{tol}}_B(L)$.

3.3.4. Multi-label attribute reduction

The algorithmic structure of multi-label attribute reduction is consistent with that of the single-label case. We only need to substitute the positive region and dependency with the multi-label tolerance-positive region and multi-label tolerance dependency.

Definition 3.7. *Let attribute subset $B \subseteq A$ and let $a \in A - B$ be a candidate attribute. The multi-label tolerance attribute significance is defined as*

$$\mathrm{SIG}_{\mathrm{tol}}(a, B, L) = \gamma^{\mathrm{tol}}_{B\cup\{a\}}(L) - \gamma^{\mathrm{tol}}_{B}(L), \tag{3.27}$$

where the multi-label tolerance dependency is defined as

$$\gamma^{\mathrm{tol}}_{B}(L) = \frac{\left|\mathrm{POS}^{\mathrm{tol}}_{B}(L)\right|}{|U|}. \tag{3.28}$$

The steps of the multi-label attribute reduction algorithm are shown in Algorithm 5.

Algorithm 5 Multi-label attribute reduction

Input: Multi-label interval-valued decision information system $DT = \langle U, A \cup L, V, f\rangle$; multi-label matrix Y; thresholds $\varepsilon, \theta, T, T^{\mathrm{tol}}_{\mathrm{multi}}$; maximum attribute count $K_{\max}$.

Output: Multi-label reduction set $R \subseteq A$.

Initialize $R = \emptyset$, *Remaining* $= A$, $\gamma_{\mathrm{old}} = 0$.
while *Remaining* $\neq \emptyset$ and $|R| < K_{\max}$ **do**
  for each candidate attribute $a \in$ *Remaining* **do**
    Let $B = R \cup \{a\}$.
    Call Algorithm 4 to generate multi-label granular balls on B and compute $\mathrm{POS}^{\mathrm{tol}}_{B}(L)$, obtaining $\gamma^{\mathrm{tol}}_{B}(L)$.
  end for
  Compute the attributes' significance $\mathrm{SIG}_{\mathrm{tol}}(a, R, L) = \gamma^{\mathrm{tol}}_{R\cup\{a\}}(L) - \gamma^{\mathrm{tol}}_{R}(L)$; select a^* with maximum significance.
  if $\gamma^{\mathrm{tol}}_{R\cup\{a^*\}}(L) > \gamma_{\mathrm{old}}$ **then**
    Update $R \leftarrow R \cup \{a^*\}$, *Remaining* $\leftarrow$ *Remaining* $\setminus \{a^*\}$, $\gamma_{\mathrm{old}} \leftarrow \gamma^{\mathrm{tol}}_{R}(L)$.
  else
    Stop forward selection.
  end if
end while
if $R = \emptyset$ **then**
  As a fallback, select the single attribute a maximizing $\gamma^{\mathrm{tol}}_{\{a\}}(L)$.
end if
return R.

Theorem 3.4 (Multi-label counterparts)**.** *(1) Since $T^{\varepsilon,\theta}_{B}$, $\mathcal{MCB}(B)$, and the generation process involve only the conditional attributes, Theorems 2.1– 3.2 apply verbatim to multi-label systems.*

(2) Under the one-hot encoding of a single-label decision with class ratios $p_d = n^{\{GB\}}_d/|GB|$,

$$P_{\mathrm{multi}}(GB) = \sum_{d} p_d^2 \le \max_{d} p_d = P_{\mathrm{single}}(GB), \tag{3.29}$$

with equality if and only if GB is decision-pure.

(3) With $T = 1$ under one-hot encoding, the union-purity heterogeneity rule $P_{\mathrm{multi}}(GB_i \cup GB_j) < T$ reduces to the single-label label-difference test.

(4) If DT is multi-label T_A-consistent, i.e., $(x, y) \in T_A^{\varepsilon,\theta} \implies L_x = L_y$, then $\mathrm{POS}_A^{\mathrm{tol}}(L) = U$; equivalently $\gamma_A^{\mathrm{tol}}(L) = 1$.

Proof. (1) Immediate from the definitions.

(2) Under one-hot encoding, $J(L_{x_i}, L_{x_j}) = 1$ if $d_{x_i} = d_{x_j}$ and 0 otherwise, so

$$P_{\mathrm{multi}}(GB) = \sum_d p_d^2 \le (\max_d p_d) \sum_d p_d = \max_d p_d; \tag{3.30}$$

the equality $\sum_d p_d(\max_d p - p_d) = 0$ forces exactly one p_d to be 1.

(3) For pure balls $GB_i \subseteq D_{c_1}$, $GB_j \subseteq D_{c_2}$, Part (2) gives $P_{\mathrm{multi}}(GB_i \cup GB_j) = 1$ if and only if $c_1 = c_2$; hence the rule with $T = 1$ fires exactly when the labels differ.

(4) Under multi-label consistency, every $X \in \mathcal{MCB}(A)$ is label-uniform (all members share an identical label subset), so all its descendant balls are label-uniform as well and satisfy $P_{\mathrm{multi}} = 1 \ge T_{\mathrm{multi}}^{\mathrm{tol}}$; fixing any label p in the common (nonempty) label subset gives $X \subseteq L_p^+$, and hence $X \subseteq \underline{apr}_A^{\mathrm{tol}}(L_p^+) \subseteq \mathrm{POS}_A^{\mathrm{tol}}(L)$, and $U = \bigcup \mathcal{MCB}(A) \subseteq \mathrm{POS}_A^{\mathrm{tol}}(L)$.

□

3.3.5. Numerical illustration of the main definitions

Consider the interval-valued decision table shown in Table 1, where D is the set of decision attributes for the single-label case and L is the set of decision attributes for the multi-label case.

Table 1. Illustrative interval-valued decision table.

Object	a_1	a_2	D	L
x_1	[1.0, 2.0]	[1.0, 2.0]	d_1	$\{l_1\}$
x_2	[1.2, 2.4]	[1.1, 2.3]	d_1	$\{l_1, l_2\}$
x_3	[1.1, 2.1]	[4.0, 5.0]	d_2	$\{l_2\}$
x_4	[1.3, 2.5]	[4.1, 5.3]	d_2	$\{l_2, l_3\}$

Set the distance threshold to 0.5, the overlap degree threshold to 0.5, the single-label tolerance purity threshold to 0.9, and both multi-label thresholds to 0.7. For objects x_1 and x_2, the Hausdorff distance and the degree of overlap aggregated over the two attributes are

$$\max_{a \in B} H(f(x_1, a), f(x_2, a)) = \max\{0.4, 0.3\} = 0.4 \le \varepsilon, \tag{3.31}$$

$$\min_{a \in B} O(f(x_1, a), f(x_2, a)) = \min\left\{\frac{0.8}{1.2}, \frac{0.9}{1.2}\right\} = \frac{2}{3} \ge \theta. \tag{3.32}$$

Thus, x_1 and x_2 satisfy the interval tolerance relation. Applying the same calculation to all object pairs yields

$$M_B = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \quad \widehat{\mathcal{MCB}}(B) = \{GB_1, GB_2\}. \tag{3.33}$$

$$GB_1 = \{x_1, x_2\}, \quad GB_2 = \{x_3, x_4\}. \tag{3.34}$$

For GB_1 on attribute a_1, the interval widths are 1 and 1.2, and the midpoints are 1.5 and 1.8. Its interval-weighted center is

$$c_{GB_1}^{a_1} = \frac{1 \times 1.5 + (1/1.2) \times 1.8}{1 + 1/1.2} = 1.6364. \tag{3.35}$$

Both granular balls are single-label pure and internally tolerant. Their single-label purity, single-label tolerance purity, approximations, positive region, and dependency are

$$P_{\text{single}}(GB_i) = P_{\text{single}}^{\text{tol}}(GB_i) = 1, \quad i = 1, 2, \tag{3.36}$$

$$\underline{\text{apr}}_B^{\text{tol}}(D_i) = \overline{\text{apr}}_B^{\text{tol}}(D_i) = GB_i, \quad i = 1, 2, \tag{3.37}$$

$$\text{POS}_B^{\text{tol}}(D) = U, \quad \gamma_B^{\text{tol}}(D) = 1. \tag{3.38}$$

For the multi-label case, the average pairwise Jaccard purity and tolerance purity of each granular ball are

$$P_{\text{multi}}(GB_i) = P_{\text{multi}}^{\text{tol}}(GB_i) = 0.75, \quad i = 1, 2. \tag{3.39}$$

The lower approximations of L_1^+ and L_2^+ contain GB_1 and GB_2, respectively. Their union covers U; therefore

$$\text{POS}_B^{\text{tol}}(L) = U, \quad \gamma_B^{\text{tol}}(L) = 1. \tag{3.40}$$

Using only attribute a_1 produces two size-limited mixed child balls in the single-label case, so its dependency equals zero. Adding a_2 increases the dependency from 0 to 1, and the significance of a_2 relative to $\{a_1\}$ is

$$\gamma_{\{a_1\}}^{\text{tol}}(D) = 0, \tag{3.41}$$

$$\text{SIG}_{\text{tol}}(a_2, \{a_1\}, D) = \gamma_{\{a_1, a_2\}}^{\text{tol}}(D) - \gamma_{\{a_1\}}^{\text{tol}}(D) = 1. \tag{3.42}$$

3.4. Enhancement strategies and fusion algorithm

Sections 3.1–3.3 have established the core mathematical contribution of this paper. The present section is of a different nature: It collects optional heuristic enhancement strategies and the fusion algorithm that integrates them into a complete implementation. These components serve only to accelerate the attribute search and to improve practical predictive performance; they neither modify any definition or theoretical result of the core model nor constitute the theoretical novelty of this paper.

The main frameworks introduced above (Algorithm 3 and Algorithm 5) strictly perform forward greedy selection based on the positive region dependency of rough sets, which guarantees the theoretical rigor and consistency of knowledge granulation. However, on public datasets, pure dependency-driven selection may fall into local optima or fail to fully capture discriminative information at the classifier level. For this reason, this paper further introduces the following enhancement strategies on the basis of the main framework.

(1) Granular-ball quality function

Beyond the positive region dependency, this paper constructs a granular-ball quality function $Q(B)$ to measure the granular-ball partition quality of the attribute subset B from multiple dimensions:

$$\begin{aligned} Q(B) = \omega_1 \cdot \frac{|\text{POS}_B|}{|U|} + \omega_2 \cdot \overline{P}_{\text{tol}}^{pos} + \omega_3 \cdot \overline{P}_{\text{tol}}^{all} \\ - \lambda_1 \cdot \frac{|\text{BND}_B|}{|U|} - \lambda_2 \cdot \frac{|\mathcal{NOLGBS}_B|}{|U|}, \end{aligned} \tag{3.43}$$

where $|\mathrm{POS}_B|$ and $|\mathrm{BND}_B|$ denote the number of samples in the positive region and the boundary region, respectively; $\overline{P}_{\mathrm{tol}}^{pos}$ and $\overline{P}_{\mathrm{tol}}^{all}$ stand for the average tolerance purity of granular balls in the positive region and all granular balls, respectively; and $|\mathcal{NOLGBS}_B|$ is the total number of granular balls. While retaining the dominant role of the positive region, this function takes the scale of the boundary region, the tolerance stability of granular balls, and granulation granularity into consideration simultaneously, avoiding local optima caused by a single dependency metric.

(2) Mutual information candidate pool construction

For high-dimensional data, traversing all the unselected attributes to compute granular balls in each iteration incurs a high computational cost. We first adopt the mutual information $\mathrm{MI}(a; y)$ to quantify the correlation between each conditional attribute and the decision attribute (single-label) or label set (multi-label). The top K attributes with the highest mutual information are selected to form the candidate pool $\mathcal{P}ool$. Subsequent forward selection is only carried out within the candidate pool, which significantly reduces the search cost and ensures priority evaluation of the attributes with predictive relevance. When adding a single attribute in the forward stage, the mutual information $\mathrm{MI}(a; y)$ of candidate attribute a participates in the scoring of the current round. In the post-processing refinement stage, the average mutual information of the subset is adopted

$$\overline{\mathrm{MI}}(B) = \frac{1}{|B|} \sum_{a \in B} \mathrm{MI}(a; y). \tag{3.44}$$

(3) Predictive performance verification mechanism

Traditional rough set reduction takes the positive region dependency as the sole criterion, which may produces deviation from the decision boundary of a specific classifier. This paper introduces the validation set-related predictive performance of the k-nearest neighbor (kNN) classifier as an auxiliary evaluation metric. Given an attribute subset B, we compute the average classification accuracy and macro-F1 score over R_s random training/validation splits to obtain the predictive performance score:

$$S_{\mathrm{pred}}(B) = \frac{1}{R_s} \sum_{r=1}^{R_s} \frac{\mathrm{Acc}_r(B) + \mathrm{MacroF1}_r(B)}{2}. \tag{3.45}$$

(4) Comprehensive objective function

This function integrates rough set quality, mutual information, and predictive performance, and introduces a penalty term for subset size. The comprehensive objective function is defined as:

$$J(B) = Q(B) + \alpha \cdot \overline{\mathrm{MI}}(B) + \beta \cdot S_{\mathrm{pred}}(B) - \lambda \cdot \frac{|B|}{|A|}, \tag{3.46}$$

where $\frac{|B|}{|A|}$ denotes the relative reduction size; α, β and λ are weight parameters for the mutual information term, predictive performance term, and size penalty term, respectively. This function balances the knowledge granulation ability, the predictive correlation of attributes, practical classification performance, and control of the reduction size. While maintaining consistency with rough set theory, it improves the generalization ability on real-world data.

(5) Minimum selection count and early stopping strategy

Traditional forward greedy reduction adopts a strict gain stopping criterion: The iteration terminates immediately once the positive region no longer expands. However, the positive region's dependency

may saturate rapidly after the first strong attribute, so that the algorithm degenerates into a single-feature selector. To address this issue, we introduce a minimum selection count constraint $K_{\min}$. Before the number of selected attributes reaches $K_{\min}$, the algorithm is forced to continue attribute selection even if the attribute gain of the current round is zero. After reaching $K_{\min}$, the algorithm executes the rule of stopping when the gain is insufficient. This strategy guarantees that the search process has a chance to evaluate the combinatorial effects of attributes and avoids falling into local optima prematurely.

(6) Local search and refinement strategy

Once the attribute order is determined by forward greedy selection, backtracking becomes difficult. After obtaining the forward selection sequence, this paper further performs multi-stage local refinement to alleviate the local optimum problem caused by forward selection. This stage consists of three operations: Prefix subset evaluation, attribute replacement search, and small-scale exhaustive refinement. The detailed descriptions are as follows.

1) Prefix subset evaluation: Re-evaluate the subset formed by the first k attributes of the forward sequence via $J(\cdot)$ and select the optimal prefix length to prevent over selection.
2) Attribute replacement search: Attempt to remove one attribute from the current reduction subset B and introduce a new attribute from the candidate pool. If the candidate subset B' after replacement satisfies $J(B') > J(B)$, the replacement is accepted to explore superior attribute combinations.
3) Small-scale exhaustive refinement: When the candidate pool size is small or the data dimension is low, perform an exhaustive search over attribute combinations within a limited size range using $J(\cdot)$ as the evaluation metric to finely characterize the interaction effects between attributes.

The six strategies above are independent of one another and can be combined freely. In the remainder of this section, they are integrated with the main reduction algorithms into a complete implementation workflow, and the fusion algorithms for both task types are described. Figure 2 provides an overall schematic diagram of this workflow. It should be emphasized that the fusion algorithms below are enhanced implementations of the main reduction algorithms rather than new reduction models: The single-label fusion algorithm (Algorithm 6) shares the same model definitions, the same approximation and dependency computations and the same forward greedy skeleton as Algorithm 3, and differs only in the search guidance, the candidate pool, and the stopping and refinement rules, etc.; the same relationship holds between the multi-label fusion algorithm (Algorithm 7) and the multi-label reduction algorithm (Algorithm 5).

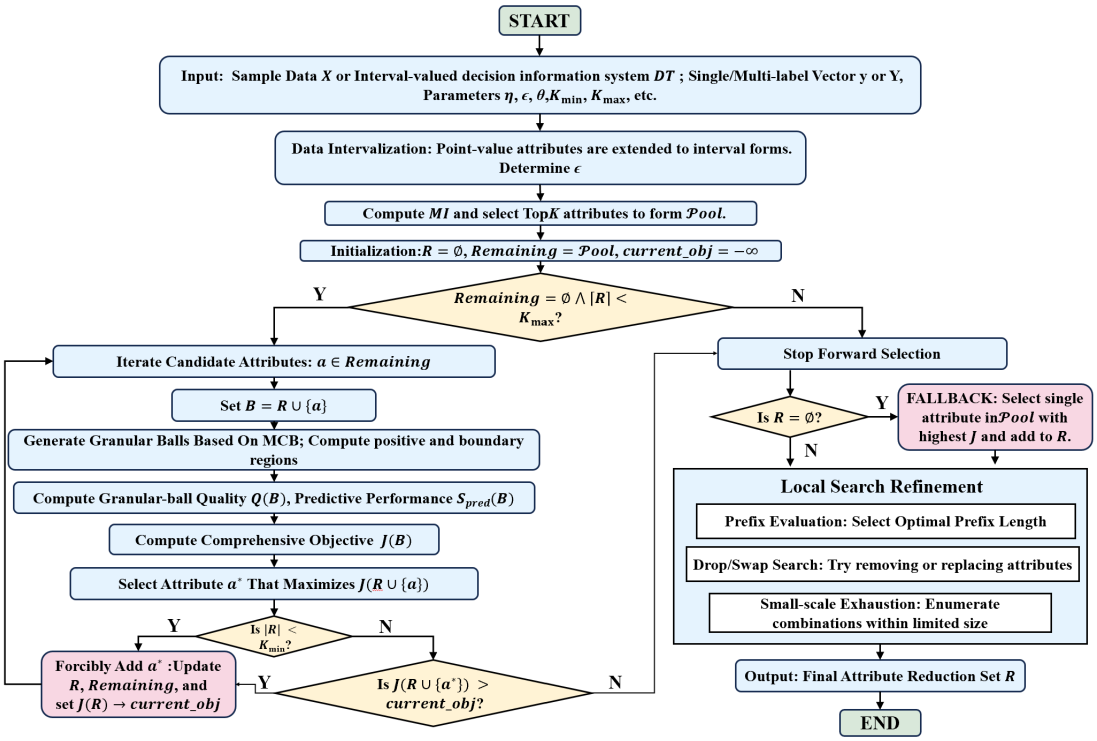

Figure 2. Simplified complete process of enhanced attribute reduction.

Algorithms 6 and 7 present the complete single-label and multi-label reduction algorithms integrated with the enhancement strategies, respectively.

Algorithm 6 Enhanced Single-label attribute reduction

Input: Sample data X or interval-valued system DT; single-label vector y; attribute types $feature\_types$; intervalization coefficient η; ε or ε-quantile(q_ε); θ; T'; $T^{\text{tol}}_{\text{single}}$; candidate pool size $PoolSize$; $K_{\min}$; $K_{\max}$; quality function weights $\omega_1, \omega_2, \omega_3, \lambda_1, \lambda_2$; comprehensive objective weights α, β, λ; local refinement parameters.

Output: Enhanced single-label reduction set R.

if input is point-valued data **then**

Perform intervalization: $I_{ia} = \begin{cases} [x_{ia} - \eta\sigma_a, x_{ia} + \eta\sigma_a], & \text{if } a \text{ is continuous;} \\ [x_{ia}, x_{ia}], & \text{if } a \text{ is discrete.} \end{cases}$ Here, the actual attribute-wise perturbation radius is $\rho_a = \eta\sigma_a$, which is an attribute-adaptive realization of the general perturbation radius ρ in Definition 2.1.

end if

if ε is not given **then**

Estimate ε by sampling sample pairs and taking the Hausdorff distance quantile.

end if

Compute the mutual information $\text{MI}(a; y)$ for each attribute $a \in A$; select the top $PoolSize$ attributes with the highest MI to form the candidate pool $\mathcal{P}ool = \text{TopK}_{a \in A}\text{MI}(a; y)$.

Initialize $R = \emptyset$, $Remaining = \mathcal{P}ool$, $current\_obj = -\infty$.

while $Remaining \neq \emptyset$ and $|R| < K_{\max}$ **do**

for each candidate attribute $a \in Remaining$ **do**
  Let $B = R \cup \{a\}$.
  Call Algorithm 2 to generate granular balls on B; enable same-label centroid initialization and boundary-region-only incremental splitting.
  Compute the boundary region $BND_B = \left(\bigcup_{i=1}^{k} \overline{apr}_B^{\mathrm{tol}}(D_i)\right) \setminus \mathrm{POS}_B^{\mathrm{tol}}(D)$.
  Compute the quality function $Q(B) = \omega_1 \frac{|\mathrm{POS}_B^{\mathrm{tol}}(D)|}{|U|} + \omega_2 \bar{P}_{\mathrm{tol}}^{\mathrm{pos}} + \omega_3 \bar{P}_{\mathrm{tol}}^{\mathrm{all}} - \lambda_1 \frac{|BND_B|}{|U|} - \lambda_2 \frac{|\mathcal{NOLGBS}_B|}{|U|}$.
  Compute the predictive performance $S_{\mathrm{pred}}(B) = \frac{1}{R_s} \sum_{r=1}^{R_s} \frac{\mathrm{Acc}_r(B)+\mathrm{MacroF1}_r(B)}{2}$.
  Compute the comprehensive objective $J(B) = Q(B)+\alpha\overline{\mathrm{MI}}(B)+\beta S_{\mathrm{pred}}(B)-\lambda\frac{|B|}{|A|}$, where $\overline{\mathrm{MI}}(B) = \frac{1}{|B|}\sum_{a\in B}\mathrm{MI}(a; y)$.
end for
Select a^* maximizing $J(R \cup \{a\})$.
if $|R| < K_{\min}$ **then**
  Forcefully add a^*: $R \leftarrow R \cup \{a^*\}$, $Remaining \leftarrow Remaining \setminus \{a^*\}$, $current\_obj \leftarrow J(R)$.
else if $J(R \cup \{a^*\}) > current\_obj$ **then**
  Update $R \leftarrow R \cup \{a^*\}$, $Remaining \leftarrow Remaining \setminus \{a^*\}$, $current\_obj \leftarrow J(R)$.
else
  Stop forward selection.
end if
end while
if $R = \emptyset$ **then**
  Fallback to select the attribute with the highest single-attribute objective value in $\mathcal{P}ool$.
end if
Perform local refinement: Evaluate prefixes of the forward sequence to select the optimal length; perform a drop/swap search, accepting improvements; if permitted, exhaustively enumerate combinations within a limited size range.
return R.

Algorithm 7 Enhanced multi-label attribute reduction

Input: Sample data X or multi-label interval-valued system DT; multi-label matrix Y; attribute types $feature\_types$; interval perturbation radius ρ; ε or ε-quantile(q_ε); θ; multi-label purity threshold T; tolerance purity threshold $T_{\mathrm{multi}}^{\mathrm{tol}}$; candidate pool size $PoolSize$; $K_{\min}$; $K_{\max}$; quality function weights $\omega_1, \omega_2, \omega_3, \lambda_1, \lambda_2$; comprehensive objective weights α, β, λ; local refinement parameters.

Output: Enhanced multi-label reduction set R.

if input is point-valued data **then**
  Perform intervalization: $I_{ia} = \begin{cases} [x_{ia} - \rho,\ x_{ia} + \rho], & \text{if } a \text{ is continuous;} \\ [x_{ia},\ x_{ia}], & \text{if } a \text{ is discrete.} \end{cases}$ When $\rho = 0$, the point-valued attributes are represented as degenerate intervals.
end if
if ε is not given **then**
  Estimate ε via the maximum Hausdorff distance quantile of random sample pairs.
end if

Compute multi-label average mutual information $\mathrm{MI}(a) = \frac{1}{q}\sum_{p=1}^{q}\mathrm{MI}(a, Y_p)$ for each attribute; select the top $PoolSize$ attributes to form $\mathcal{P}ool = \mathrm{TopK}_{a\in A}\mathrm{MI}(a)$.
Initialize $R = \emptyset$, $Remaining = \mathcal{P}ool$, $current\_obj = -\infty$.
while $Remaining \neq \emptyset$ and $|R| < K_{\max}$ **do**
  for each candidate attribute $a \in Remaining$ **do**
    Let $B = R \cup \{a\}$.
    Call Algorithm 4 to generate multi-label granular balls on B; enable label-subset-aware center initialization and boundary-region incremental splitting.
    Compute the multi-label positive region $\mathrm{POS}_B^{\mathrm{tol}}(L)$ and the boundary region $BND_B = \left(\bigcup_{p=1}^{q}\overline{apr}_B^{\mathrm{tol}}(L_p^+)\right) \setminus \mathrm{POS}_B^{\mathrm{tol}}(L)$.
    Compute the dependency $\gamma_B^{\mathrm{tol}}(L) = \frac{|\mathrm{POS}_B^{\mathrm{tol}}(L)|}{|U|}$.
    Compute the predictive score $S_{\mathrm{pred}}(B) = 0.4 \cdot \mathrm{MicroF1}(B) + 0.4 \cdot \mathrm{MacroF1}(B) + 0.2 \cdot [1 - \mathrm{HammingLoss}(B)]$.
    Compute the quality function $Q(B) = \omega_1\gamma_B^{\mathrm{tol}}(L) + \omega_2\bar{P}_{\mathrm{tol}}^{\mathrm{pos}} + \omega_3\bar{P}_{\mathrm{tol}}^{\mathrm{all}} - \lambda_1\frac{|BND_B|}{|U|} - \lambda_2\frac{|\mathcal{NOLGBS}_B|}{|U|}$.
    Compute the comprehensive objective $J(B) = Q(B) + \alpha\overline{\mathrm{MI}}(B) + \beta S_{\mathrm{pred}}(B) - \lambda\frac{|B|}{|A|}$, where $\overline{\mathrm{MI}}(B) = \frac{1}{|B|}\sum_{a\in B}\mathrm{MI}(a)$.
  end for
  Select a^* maximizing $J(R \cup \{a\})$.
  if $|R| < K_{\min}$ **then**
    Forcefully add a^*: $R \leftarrow R \cup \{a^*\}$, $Remaining \leftarrow Remaining \setminus \{a^*\}$, $current\_obj \leftarrow J(R)$.
  else if $J(R \cup \{a^*\}) > current\_obj$ **then**
    Update $R \leftarrow R \cup \{a^*\}$, $Remaining \leftarrow Remaining \setminus \{a^*\}$, $current\_obj \leftarrow J(R)$.
  else
    Stop forward selection.
  end if
end while
if $R = \emptyset$ **then**
  Fallback to select the attribute with the highest single-attribute objective value in $\mathcal{P}ool$.
end if
Perform local refinement: Select optimal prefix; execute a drop/swap search; perform small-scale exhaustive refinement if enabled.
return R and record the corresponding granular balls, approximations, and dependency history.

3.5. Computational complexity analysis

Let n be the number of objects, m be the total number of conditional attributes, $b = |B|$ be the size of the currently evaluated subset, and q be the number of labels. For subset B, e_B denotes the number of nonzero edges in the indexed tolerance graph, m_B be the total object memberships in the generated blocks or granular balls, and z_B is the aggregate work of the observed candidate-block containment checks. Here, g_s and g_m denote the accumulated object assignments during single-label and multi-label splitting, while h_s and h_m denote the candidate pairs examined during heterogeneous overlap removal; s_B is the number of distinct label signatures and j_B is the number of signature pairs whose

Jaccard similarity is evaluated.

Let $T_k(B)$ denote the cost of Algorithm k on the subset B. $\mathcal{S}_s$ and $\mathcal{S}_m$ are the collections of subsets actually visited by the strict single-label and multi-label searches; $\mathcal{S}'_s$ and $\mathcal{S}'_m$ are their candidate pool counterparts. R_s is the number of prediction-validation repetitions; $C_{\text{pred}}(B)$ and $C^{\text{ML}}_{\text{pred}}(B)$ are the classifier-dependent validation costs; and u_s, u_m, u'_s, u'_m denote the corresponding cache footprints. Table 2 gives output-sensitive bounds for the exact implementation.

Table 2. Output-sensitive time and working space complexity bounds for each algorithm.

Algorithm	Role	Output-sensitive time	Working space
Algorithm 1	MCB generation	$O(n \log n + e_B b + z_B)$	$O(nb + e_B + m_B)$
Algorithm 2	Single-label granular-ball generation	$O(T_1 + g_s b + h_s b)$	$O(nb + e_B + m_B)$
Algorithm 3	Strict single-label reduction	$O\left(\sum_{B \in \mathcal{S}_s} T_2(B)\right)$	$O(nb + e_B + u_s)$
Algorithm 4	Multi-label granular-ball generation	$O(T_1 + nq + j_B q + e_B q + g_m b + h_m b)$	$O(nb + e_B + s_B q + m_B)$
Algorithm 5	Strict multi-label reduction	$O\left(\sum_{B \in \mathcal{S}_m} T_4(B)\right)$	$O(nb + e_B + u_m)$
Algorithm 6	Enhanced single-label reduction	$O\left(nm + \sum_{B \in \mathcal{S}'_s} [T_2(B) + R_s C_{\text{pred}}(B)]\right)$	$O(nm + nb + e_B + u'_s)$
Algorithm 7	Enhanced multi-label reduction	$O\left(nmq + \sum_{B \in \mathcal{S}'_m} [T_4(B) + R_s C^{\text{ML}}_{\text{pred}}(B)]\right)$	$O(nm + nb + e_B + u'_m)$

Algorithm 1 uses endpoint indexing to construct only the tolerance edges that satisfy the original Hausdorff distance and overlap conditions. Algorithms 2 and 4 reuse this graph during granular-ball generation, and multi-label purity is aggregated over distinct label signatures and compatible edges. Algorithms 3 and 5 perform forward selection and therefore visit only the subsets in $\mathcal{S}_s$ or $\mathcal{S}_m$, rather than enumerating the power set of the attributes. Algorithms 6 and 7 add relevance screening and predictive validation; their effective cost is controlled by the sizes of $\mathcal{S}'_s$ and $\mathcal{S}'_m$. This explicit separation of the data-processing cost and the number of evaluated optimization candidates is consistent with recent optimization-based feature-weighting models [32]. The indexed implementation changes only data access and reuse: The tolerance relation, the MCB's definition, granular ball acceptance rules, and reduction objectives remain unchanged. If a tolerance graph is intrinsically dense, its output size remains an unavoidable cost; the output-sensitive form does not assume sparsity for every possible dataset.

4. Experimental results and analysis

4.1. Experimental setup

4.1.1. Datasets

To verify the effectiveness of the proposed method under diverse supervised learning scenarios, we carry out experiments on single-label classification datasets and multi-label classification datasets, respectively. The single-label experiment involves 10 public datasets, covering binary classification, multi-class classification, continuous attributes and discrete attribute settings. The multi-label

experiment adopts six public multi-label datasets. For emotions, scene, yeast, birds and medical, the official training/testing partition is utilized. CAL500 has no predefined fixed train/test split and contains a small number of samples, so three fold cross-validation is used to achieve more robust performance estimation.

The basic statistics of the two categories of datasets are summarized in Table 3, where Panel A contains single-label datasets and Panel B contains multi-label datasets.

Table 3. Summary of datasets.

Panel A: Single-label datasets.

Dataset	Samples	Features	Classes
HTRU_2	17898	8	2
Electrical	10000	12	2
Ionosphere	351	34	2
Lymphography	148	18	4
Mushroom	8124	22	2
Primary_tumor	339	17	21
Spambase	4601	57	2
Wdbc	569	30	2
Wine	178	13	3
Zoo	101	16	7

Panel B: Multi-label datasets.

Dataset	Samples	Features	Labels	Cardinality	Density	Protocol
Emotions	593	72	6	1.8685	0.3114	official_train_test
Scene	2407	294	6	1.0740	0.1790	official_train_test
Yeast	2417	103	14	4.2371	0.3026	official_train_test
Birds	645	260	19	1.0140	0.0534	official_train_test
Medical	978	1449	45	1.2454	0.0277	official_train_test
CAL500	502	68	174	26.0438	0.1497	3_fold_cv

4.1.2. Evaluation metrics and comparison methods

For single-label tasks, accuracy, macro-F1, balanced accuracy, and Matthews correlation coefficient (MCC) are adopted as primary evaluation metrics; we also record the number of selected features, reduction rate, and runtime. For multi-label tasks, micro-F1, macro-F1, Hamming loss, and average precision are taken as the core metrics. Among them, micro-F1 emphasizes the overall discriminative ability across all labels; macro-F1 is more sensitive to rare labels or minority classes; Hamming loss measures the per-label misclassification rate; average precision reflects the quality of label ranking. Since each metric focuses on distinct evaluation perspectives, we report all four metrics in multi-label experiments to comprehensively reflect the model's performance.

For single-label tasks, let n denote the total number of test samples, y_i the ground-truth class, $\hat{y}_i$ the predicted class, and C the set of all classes. Accuracy is defined as

$$\text{Accuracy} = \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}(y_i = \hat{y}_i). \tag{4.1}$$

For any class $c \in C$, let TP_c, FP_c, and FN_c be the number of true positives, false positives, and false

negatives of class c counted under the one-vs-rest strategy. Then macro-F1 is formulated as:

$$\text{Macro} - \text{F1} = \frac{1}{|C|}\sum_{c\in C}\frac{2\text{TP}_c}{2\text{TP}_c + \text{FP}_c + \text{FN}_c}. \tag{4.2}$$

Balanced accuracy computes the macro average of recall across all classes, i.e.,

$$\text{Balanced accuracy} = \frac{1}{|C|}\sum_{c\in C}\frac{\text{TP}_c}{\text{TP}_c + \text{FN}_c}. \tag{4.3}$$

MCC refers to the multi class Matthews correlation coefficient. Let M be the confusion matrix, where M_{ij} denotes the number of samples with the ground-truth class i and the predicted class j. We define $s = \sum_{ij} M_{ij}$, $c = \sum_k M_{kk}$, $t_k = \sum_j M_{kj}$, and $p_k = \sum_i M_{ik}$. The MCC formula is given by

$$\text{MCC} = \frac{cs - \sum_k p_k t_k}{\sqrt{(s^2 - \sum_k p_k^2)(s^2 - \sum_k t_k^2)}}. \tag{4.4}$$

For multi-label tasks, let q be the total number of labels and $Y, \hat{Y} \in \{0,1\}^{n\times q}$ denote the ground-truth label matrix and predicted label matrix, respectively. Micro-F1 aggregates true positives (TPs), false positives (FPs) and false negatives (FNs) over all samples and labels before calculation

$$\text{Micro} - \text{F1} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}. \tag{4.5}$$

Multi-label macro-F1 averages the F1 score calculated on each label dimension:

$$\text{Macro} - \text{F1} = \frac{1}{q}\sum_{l=1}^{q}\frac{2\text{TP}_l}{2\text{TP}_l + \text{FP}_l + \text{FN}_l}. \tag{4.6}$$

Hamming loss is defined as the average misclassification rate over all sample-label pairs:

$$\text{Hamming loss} = \frac{1}{nq}\sum_{i=1}^{n}\sum_{l=1}^{q}\mathbb{I}(Y_{il} \neq \hat{Y}_{il}). \tag{4.7}$$

Average precision is computed per label based on the prediction score matrix $S \in \mathbb{R}^{n\times q}$. For the l-th label, sort all samples in descending order of S_{il}. Let $Y_{(r)l}$ denote the ground-truth label of the r-th sample after sorting, and $P_l = \sum_i Y_{il}$. The precision at rank r is defined as $\text{Prec}_l(r) = \frac{1}{r}\sum_{u=1}^{r} Y_{(u)l}$. The average precision for label l is

$$\text{AP}_l = \frac{1}{P_l}\sum_{r=1}^{n}\text{Prec}_l(r)\, Y_{(r)l}, \tag{4.8}$$

and the overall average precision metric is

$$\text{Average precision} = \frac{1}{|\mathcal{L}^+|}\sum_{l\in\mathcal{L}^+}\text{AP}_l, \tag{4.9}$$

where $\mathcal{L}^+ = \{l \mid P_l > 0\}$ stands for the set of labels that contain at least one positive sample in the test set. All F1-type metrics are set to 0 in our experimental code when their denominators equal zero.

The method proposed in this paper, denoted as Ours, corresponds to a complete framework integrating MCB initialization, the same-label centroid initialization strategy of the GBRS′ [16], the boundary-region-only incremental splitting strategy of the GBRS$^+$ [16], and our proposed enhancement strategies. To comprehensively evaluate its effectiveness, we select a total of 12 baseline methods falling into four categories for comparison.

(1) No-reduction baseline. FullFeature directly trains the classifier using all raw features of the original dataset, which is adopted to quantify the performance degradation caused by feature reduction or feature selection algorithms.

(2) GBRS reduction methods. These methods share the granular-ball computing framework with our proposed method, and they are chosen to verify the performance improvement of our approach over existing GBRS reduction pipelines. As the core baseline of GBRS reduction, the GBRS [16] adaptively characterizes the sample neighborhoods. It can directly handle continuous attributes while retaining the interpretability of rough set equivalence classes. MaxGBRS is an enhanced comparison baseline constructed based on the GBRS. Unlike the GBRS, which initializes granular balls from the entire sample universe U, MaxGBRS takes the efficient MCB set $\widehat{\mathcal{MCB}}(B)$ generated by the interval tolerance relation as initial granular balls.

(3) Equal-budget feature selection methods. This group of methods is used for comparative experiments with a fixed feature selection budget under the single-label setting. In all experiments, the number of selected features is controlled to be identical to that of our method, so we can focus on investigating how different feature ranking and subset selection mechanisms affect the classification performance.

- The VIKOR-based Multi-target Feature Selection (VMFS) constructs a decision matrix based on the similarity between features and targets, and conducts compromise ranking via the VIKOR method [33];
- The Multi-label Feature Selection using Multi-Criteria Decision Making (MFS-MCDM) formulates feature selection as a multi-criteria decision-making problem, and fuses multi-dimensional evaluation information by combining ridge regression, the entropy weight method, and the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) [34];
- The Fast Information-theoretic Mutual-information Feature ranking (FIMF) realizes fast feature ranking based on mutual information and interaction information, which reduces the computational overhead of entropy calculation while maintaining strong feature evaluation capability [35];
- The Sparse Low-redundancy Multi-label Dynamic Structure (SLMDS) simultaneously captures dynamic local and global label structures, and selects features under sparse low-redundancy constraints [36];
- The Robust Flexible Sparse regularized multi-label Feature Selection (RFSFS) designs flexible sparse regularization to balance label-specific features and low-redundancy shared features [37];
- The Gradient-based method considering Three-Way Variable Interaction (G3WI) incorporates feature-feature-label and feature-label-label triple-wise interactions into a gradient optimization framework [38].

Comparing against these methods helps us evaluate the feature selection quality of our algorithm under the identical reduction budget from multiple perspectives, including multi criteria decision-making, information theory, sparse learning, and variable interaction modeling. For all equal-budget

feature selection baselines, single-label class labels are uniformly transformed into binary output matrices in the one-vs-rest form. In this way, multi objective/multi-label feature selection methods can be degenerated into feature rankers under the multi class output space. This set of experiments also compares the quality of feature subsets yielded by different feature evaluation criteria under the constraint of identical number of selected features.

(4) Multi-label feature selection methods. These methods explicitly model label correlations and are specially designed for multi-label tasks, covering mainstream technical routes such as dynamic graph learning, global optimization, and latent structure learning. They are selected to directly assess the overall competitiveness of our proposed method under identical task settings.

- The multi-label feature selection via Latent Representation learning and Dynamic Graph constraints (LRDG) exploits supervised information hidden in instance correlations via latent representation learning and dynamic graph constraints [39];
- The Global Relevance and Redundancy Optimization (GRRO) from the perspective of global optimization, simultaneously takes feature correlation, label correlation, and feature redundancy into consideration [40];
- The Shared latent Structure feature Selection (SSFS) characterizes the correlation between feature structures and label structures through latent structure sharing terms [41].

Unless otherwise specified, bold and underlined values in all performance comparison tables represent the optimal results within the current comparison group, while italicized and underlined values correspond to the suboptimal results. Smaller values indicate better performance for Hamming loss and average rank; all other evaluation metrics are positively correlated with the model's performance.

4.2. Single-label classification experiments

4.2.1. Comparison with GBRS reduction methods

Table 4 presents the average experimental results of the proposed method compared with FullFeature, the GBRS, and the MaxGBRS. Our method achieves an average accuracy of 0.8624 and an average macro-F1 of 0.8000 over the 10 single-label datasets, and both metrics are superior to the GBRS and MaxGBRS.

In terms of runtime, our method consumes 128.94 seconds on average, which lies between the GBRS (30.76 s) and the MaxGBRS (584.60 s). Despite incorporating MCB initialization and enhanced search strategies, the computational overhead is controlled within a reasonable range.

Compared with FullFeature, under an average reduction rate of 68.30% (i.e., an average retained feature proportion of 31.70%), our method yields slightly higher mean values for all four classification metrics. This validates that the selected attribute subset maintains high information retention efficiency.

The results reveal that even though the GBRS and the MaxGBRS attain higher compression rates, overly aggressive granularity compression easily causes a loss of discriminative information. Benefiting from MCB initialization, tolerance purity constraints, mutual information candidate pool construction, and in-training validation search, our method achieves a favorable trade-off between classification performance and feature compression at a more moderate reduction granularity.

Table 4. Comparison results with GBRS-type methods on single-label tasks.

Method	Accuracy	Macro-F1	Balanced acc.	MCC	Features	Reduction	Runtime(s)
FullFeature	*0.8617*	*0.7898*	*0.7889*	*0.7694*	22.7000	0.0000	0.0448
GBRS	0.6844	0.5720	0.6022	0.4513	1.3667	0.9236	30.7564
MaxGBRS	0.7841	0.7009	0.7137	0.6297	2.3000	0.8673	584.6022
Ours	0.8624	0.8000	0.8020	0.7735	5.5000	0.6830	128.9394

4.2.2. Comparison with equal-budget feature selection methods

Table 5 summarizes the average experimental results of the proposed method against six equal-budget feature selection baselines. Since all methods are constrained to select the identical number of features, this group of comparisons mainly quantifies the quality gap among distinct feature ranking and subset selection strategies.

The proposed method achieves an average macro-F1 of 0.8000, which is 0.0496 higher than the suboptimal baseline VMFS; its average accuracy reaches 0.8624, outperforming VMFS by 0.0352. Meanwhile, our method attains the highest balanced accuracy and MCC values across all competitors.

These results demonstrate that, under the constraint of identical reduction scale, the feature subset yielded by our proposed method possesses stronger discriminative capability.

Table 5. Comparison results with feature selection baselines on single-label tasks.

Method	Accuracy	Macro-F1	Balanced Acc.	MCC	Features	Reduction	Runtime(s)
Ours	0.8624	0.8000	0.8020	0.7735	5.5000	0.6830	128.9394
VMFS	*0.8272*	*0.7504*	*0.7526*	*0.7124*	5.5000	0.6830	0.0174
MFS-MCDM	0.8197	0.7394	0.7511	0.6980	5.5000	0.6830	0.0191
FIMF	0.8059	0.7247	0.7306	0.6643	5.5000	0.6830	0.1378
SLMDS	0.7670	0.6873	0.7013	0.6119	5.5000	0.6830	10.8633
RFSFS	0.7621	0.6854	0.6978	0.6097	5.5000	0.6830	0.0274
G3WI	0.7518	0.6738	0.6937	0.5742	5.5000	0.6830	0.4343

From the perspective of datasets, our method achieves the optimal macro-F1 on 8 out of the 10 datasets, only failing to attain the top rank on HTRU_2 and WDBC. Figure 3 provides the dataset-wise comparison of macro-F1. The HTRU_2 dataset has low raw feature dimensionality and a distinct class structure, so several lightweight ranking methods can already deliver near-saturated performance. Likewise, WDBC is a low-dimensional high-quality medical diagnosis dataset, and excessive feature compression may trigger a slight loss of discriminative information. In contrast, our method shows more remarkable superiority on datasets such as the ionosphere, spambase, zoo, and mushroom datasets. This verifies that the MCB granular-ball structure and in-training validation search strategy help identify stable discriminative attribute combinations in high-dimensional scenarios or cases with complex decision boundaries.

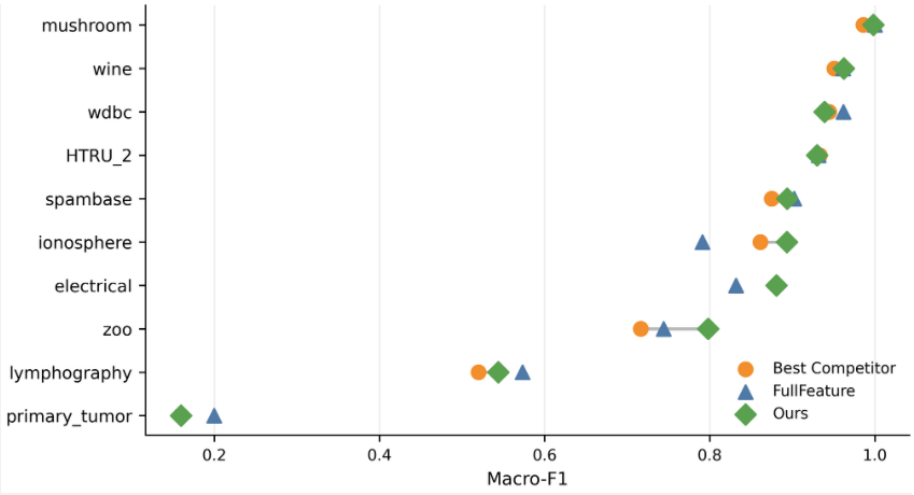

Figure 3. Comparison of macro-F1 across all datasets for single-label tasks.

Figure 4 further reveals the relationship between our method's performance and the reduction rate. Most datasets retain a high macro-F1 even when the reduction rate exceeds 50%. This demonstrates that our method does not simply cut down the number of features; instead, it preserves critical information related to classification boundaries under strict reduction constraints. The low absolute macro-F1 of primary_tumor directly stems from its 21 classes and severely imbalanced class distribution. All baseline methods suffer universal performance degradation on such challenging multi-class imbalanced datasets, and this result can act as a reference for analyzing models' robustness under multi-class imbalance scenarios.

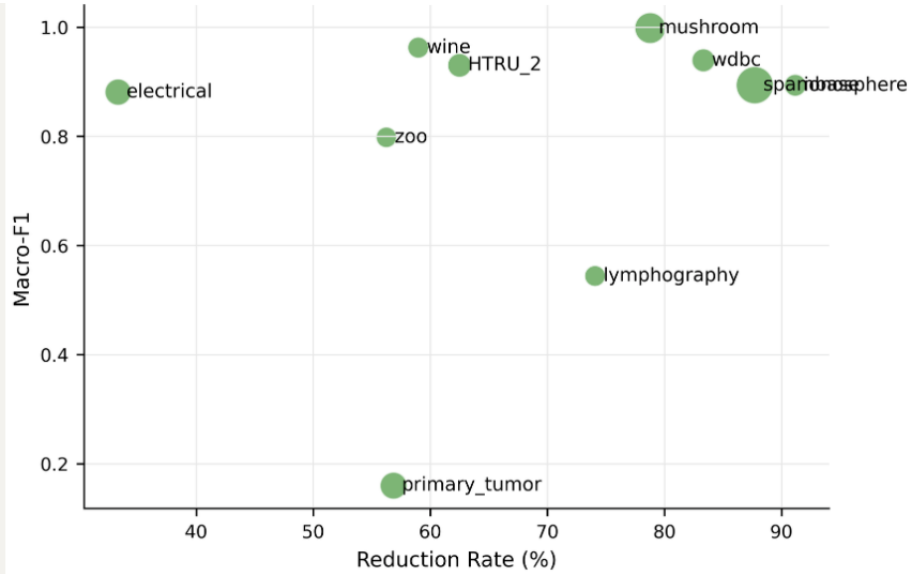

Figure 4. Relationship between performance and reduction rate on single-label tasks.

4.2.3. Statistical tests

To verify the statistical significance of the aforementioned performance gaps, we conduct the Friedman test and the Wilcoxon signed-rank paired test with Holm correction for the two groups of

comparative experiments, respectively. The statistical test results for equal-budget feature selection comparisons are listed in Table 6, while those for GBRS methods are summarized in Table 7.

For the equal-budget feature selection comparisons, the Friedman test yields statistically significant results across all four evaluation metrics. The corresponding p-values for accuracy, macro-F1, balanced accuracy, and MCC are 0.0001, 0.0002, 0.0007, and 0.0001, respectively, demonstrating that there are statistically distinguishable performance differences among all competing methods. Our method achieves average ranks of 1.7000, 1.7000, 1.6000, and 1.7000 on the four metrics, which are the lowest (optimal) ranks among all baselines. Further pairwise Wilcoxon signed-rank tests with Holm correction reveal that our method significantly outperforms all six equal-budget baselines on every metric, with corrected p-values ranging from 0.0059 to 0.0098.

The closest competing baseline is VMFS. The win/tie/loss (W/T/L) count of our method against VMFS is 7/1/2. All corresponding Holm-corrected p-values are less than 0.05, indicating that the performance superiority remains statistically significant.

In the comparisons against GBRS methods, the Friedman test also returns significant results for all four metrics, which confirms an overall performance discrepancy across FullFeature, GBRS, MaxGBRS, and our proposed method. Subsequent pairwise tests further indicate that our method significantly surpasses the GBRS and MaxGBRS on all four metrics, with identical Holm-corrected p-values of 0.0029. However, the performance difference between our method and FullFeature does not reach the level of statistical significance.

Accordingly, the statistical conclusion of this group of experiments is summarized as follows: Our method significantly outperforms GBRS-type reduction methods, and it retains classification performance comparable with the full-feature baseline while achieving substantial feature compression.

Table 6. Statistical test results for equal-budget feature selection on single-label tasks.

Metric	F-χ^2	F-p	CD	Ours Rank	Sig. Wins	Holm p	N-Baseline	W/T/L
Accuracy	27.7148	0.0001	2.8483	1.7000	6/6	0.0059–0.0098	VMFS	7/1/2
Macro-F1	26.6919	0.0002	2.8483	1.7000	6/6	0.0059–0.0098	VMFS	7/1/2
Balanced accuracy	23.1459	0.0007	2.8483	1.6000	6/6	0.0059–0.0098	VMFS	7/1/2
MCC	27.8811	0.0001	2.8483	1.7000	6/6	0.0059–0.0098	VMFS	7/1/2

Note: F-χ^2: Friedman χ^2; F-p: Friedman p; Holm p: Holm p range; N-baseline: Nearest baseline.

Table 7. Statistical test results for GBRS methods on single-label tasks.

Metric	F-χ^2	F-p	Ours Rank	vs. FullFeature	vs. GBRS	vs. MaxGBRS
Accuracy	22.4400	0.000053	1.6000	4/0/6; p=0.6523	10/0/0; p=0.0029	10/0/0; p=0.0029
Macro-F1	22.4400	0.000053	1.6000	4/0/6; p=0.5000	10/0/0; p=0.0029	10/0/0; p=0.0029
Balanced accuracy	22.5600	0.000050	1.5000	5/0/5; p=0.4229	10/0/0; p=0.0029	10/0/0; p=0.0029
MCC	22.4400	0.000053	1.6000	4/0/6; p=0.6875	10/0/0; p=0.0029	10/0/0; p=0.0029

Note: F-χ^2: Friedman χ^2; F-p: Friedman p. The last three columns report the multi-comparison results in the format of "win/tie/loss; Holm p".

4.3. Multi-label classification experiments

4.3.1. Comparison with multi-label feature selection methods

Table 8 summarizes the average performance, average number of selected features, and average reduction rate of all competing methods across the six multi-label datasets. Our proposed method achieves competitive performance in terms of average micro-F1 and Hamming loss. Its superiority mainly lies in the overall label discriminative ability and per-label error control.

Table 8. Average performance and reduction results on multi-label tasks.

Method	Micro-F1	Macro-F1	Hamming loss	Average precision	Features	Reduction
Ours	0.5969	0.3899	0.1375	0.4334	34.6667	0.7809
FullFeature	0.5777	0.3957	0.1388	0.4382	376.1667	0.0000
LRDG	0.5701	0.3803	0.1443	0.4220	34.6667	0.7809
GRRO	0.5490	0.3584	0.1485	0.3882	34.6667	0.7809
SSFS	0.5342	0.3397	0.1614	0.3737	34.6667	0.7809

As illustrated in Figure 5, our method exhibits powerful dimensional reduction capacity in high-dimensional multi-label scenarios. For the medical dataset, only 24 out of the original 1449 features are retained, corresponding to a reduction rate of 98.34%; for the birds dataset, 16 out of 260 features are selected, with a reduction rate of 93.85%. The yeast, CAL500, and scene datasets also have extremely high compression rates. These observations verify that our method can generate compact feature subsets for high-dimensional multi-label data.

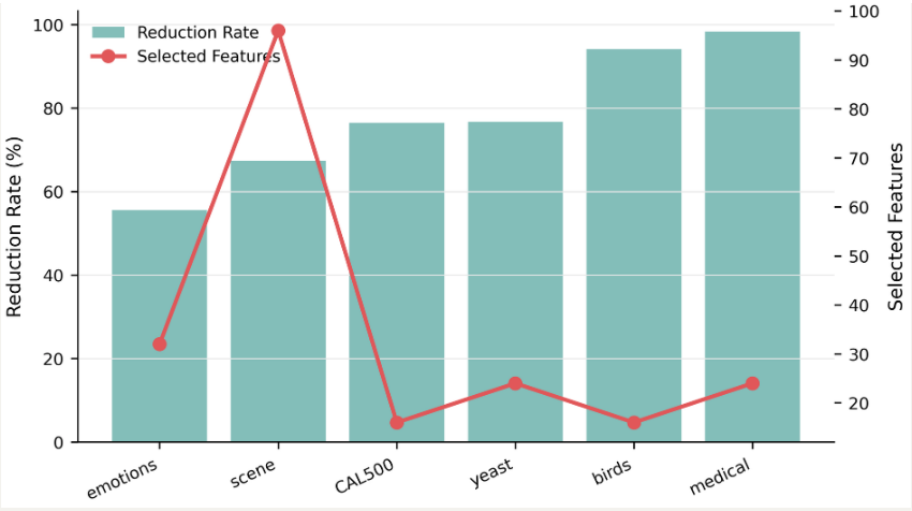

Figure 5. Number of selected features and reduction rate on multi-label tasks.

4.3.2. Statistical tests

The statistical test results for multi-label experiments are summarized in Table 9. For the full comparison group (including the FullFeature baseline), the Friedman test detects statistically significant inter method differences for micro-F1, macro-F1, and average precision, with corresponding p-values of 0.0130, 0.0092, and 0.0041, respectively. For the subgroup consisting solely of feature

selection approaches, the Friedman test returns significant results on all four evaluation metrics, verifying that statistically distinguishable performance discrepancies exist among multi-label feature selection methods.

Pairwise tests further quantify the statistically significant performance gaps between all competing algorithms under multi-label settings. When compared against the Multi-Label k-Nearest Neighbor (MLKNN) baseline method FullFeature-MLKNN, our method achieves five wins and one loss on the micro-F1 metric. The only inferior case lies on the scene dataset: Our method achieves a micro-F1 of 0.7133, whereas FullFeature-MLKNN reaches 0.7284. The scene dataset originally has 294 features and 6 labels, and the raw full-feature representation already delivers strong discriminative capability. During feature reduction, redundant attributes are filtered out, yet some weak discriminative features that differentiate highly similar scene categories may also be discarded, leading to a minor performance decline. In pairwise comparisons with other feature selection baselines, our method significantly outperforms SSFS on micro-F1, macro-F1 and average precision, with a consistent Holm-corrected p-value of 0.0469 for all three metrics.

Table 9. Statistical test results for multi-label tasks.

Metric	All χ^2	All F-p	FS χ^2	FS F-p	vs. Full	FF Holm p	Sig. FS Pairwise
Micro-F1	12.6667	0.0130	9.8000	0.0203	5/0/1	0.1094	vs. SSFS (p=0.0469)
Macro-F1	13.4667	0.0092	9.8000	0.0203	3/0/3	0.6562	vs. SSFS (p=0.0469)
Hamming loss	7.8667	0.0966	8.6000	0.0351	4/0/2	0.4219	None
Average precision	15.3333	0.0041	11.0000	0.0117	2/0/4	0.8438	vs. SSFS (p=0.0469)

Note: All F-p: All Friedman p; FS F-p: FS Friedman p; vs full: vs Full W/T/L; FF Holm p: FullFeature Holm p.

4.4. Ablation study

We evaluated the method through an algorithm-level comparison and a separate leave-one-component-out analysis. The algorithm-level comparison reuses the full-sample results in Table 4. the GBRS, MaxGBRS, and ours obtain mean macro-F1 values of 0.5720, 0.7009, and 0.8000 over the 10 single-label datasets, respectively; accuracy, balanced accuracy, and MCC exhibit the same ordering. In the existing four-method comparison family, which retains FullFeature as the non-reduction reference, the Friedman test yields $p \leq 5.3 \times 10^{-5}$ for each metric. Ours wins on all 10 datasets against both the GBRS and MaxGBRS, with Holm-adjusted Wilcoxon p-values of 0.0029. These comparisons concern complete configured algorithms. In particular, the GBRS and MaxGBRS differ in MCB initialization, tolerance scaling, quality and relevance weights, predictive weighting, subset-size penalties, candidate-pool settings, minimum selected-feature counts, and validation settings. Their performance difference therefore quantifies the algorithm-level progression between the two configurations and does not isolate the effect of MCB initialization.

The leave-one-component-out analysis used a separate controlled protocol. Nine single-label datasets used 120 stratified observations in two-fold cross-validation, with 60 training and 60 test observations per fold; zoo retained all 101 observations, with 50/51 training-test fold sizes. Each multi-label dataset used 120 training observations and retained its official test set. All variants within this protocol used identical partitions, classifiers, feature budgets, and random seeds. The complete method achieves mean macro-F1 values of 0.6686 and 0.3503 in the single-label and multi-label panels, respectively. These protocol-specific values are not directly compared with the full-sample

mean of 0.8000, and the two analyses are not interpreted as successive stages of a numerical effect decomposition. Table 10 defines each change as the ablated result minus the complete result within the controlled protocol; a negative value therefore indicates a positive contribution from the component being removed.

Table 10. Performance change in leave-one-component-out ablation study.

Removed component	Single-label: Change in macro-F1 / selected features	Multi-label: Change in macro-F1/selected features
MCB initialization	−0.0284/ − 0.05	N/A
Interval overlap constraint	+0.0020/ + 0.35	N/A
Mutual information guidance	−0.0050/0.00	−0.0229/ − 4.40
Predictive validation	−0.0408/ − 0.20	−0.0128/ − 13.60
Local refinement	−0.0238/ + 0.75	N/A
Granular ball quality; GBNRS$'^{+}$ splitting; subset-size regularization	0.0000/0.00 for each	0.0000/0.70 for quality and size; splitting N/A

Within the controlled protocol, MCB initialization, predictive validation, mutual-information guidance, and local refinement provide positive mean contributions to single-label macro-F1, with predictive validation producing the largest decrease when removed. In the multi-label panel, mutual-information guidance and predictive validation improve the mean macro-F1 by 0.0229 and 0.0128, respectively. The interval-overlap constraint trades a negligible macro-F1 difference of 0.0020 for a more compact subset containing 0.35 fewer attributes on average. The granular-ball quality term, GBNRS$'^{+}$ splitting, and subset-size regularization return the same final subsets under the controlled settings.

4.5. Parameter sensitivity analysis

4.5.1. Parameter roles and default settings

All sensitivity experiments used the settings of the main experiments as the reference configuration. We divided the parameters into model-defining and search-control groups. Model-defining parameters determine the interval construction, tolerance neighborhoods, granular-ball acceptance, and subset quality. Search-control parameters restrict the candidate space and penalize unstable validation scores; they affect the computational cost and generalization but do not alter the mathematical definition of the tolerance relation.

The positive region dependency, tolerance purity terms, and boundary ratio are bounded within $[0, 1]$, while the granular ball count is normalized by the sample size to measure partition fragmentation. We therefore fixed the positive region coefficient at one and assigned smaller coefficients to the tolerance purity rewards and the boundary and fragmentation penalties. This preserves positive region dependency as the primary criterion. In the comprehensive objective, mutual information and predictive performance refine the ranking of subsets with similar rough-set quality, while the weak relative-size penalty prevents unnecessary attributes from being retained. Table 11 reports the shared defaults and their roles.

Table 11. Default parameter settings and corresponding selection rationales.

Group	Parameters	Default	Selection rationale
Interval construction	η, ρ	0.10, 0	η scales the attribute's standard deviation in single-label intervalization; $\rho = 0$ avoids artificial widening of deterministic multi-label benchmarks.
Tolerance relation	$q_\varepsilon(\boldsymbol{\varepsilon}\text{-quantile}), \theta$	0.20, 0.40	Robust lower-tail distance scale with a moderate overlap requirement.
Purity acceptance	T', $T^{\text{tol}}_{\text{single}}$	1.00, 0.90	Class purity remains strict, with limited interval inconsistency allowed.
Multi-label acceptance	T, $T^{\text{tol}}_{\text{multi}}$	0.70, 0.70	Balances label-set consistency and tolerance consistency.
Quality rewards	$\omega_1, \omega_2, \omega_3$	1.00, 0.12, 0.08	Sets positive-region dependency as the primary term and treats purity as auxiliary evidence.
Quality penalties	λ_1, λ_2	0.16, 0.004	Discourages boundary-heavy and excessively fragmented partitions.
Comprehensive objective	α, β, λ	0.30, 0.62, 0.003	Balances relevance, predictive evidence, and weak subset-size regularization.
Single-label search	$PoolSize$, $K_{\min}$, $K_{\max}$	10, 1, 8	Retains a compact candidate set while avoiding an unnecessarily restrictive feature budget.
Stable validation	R_{cv}, ξ, ζ	2, 0.018, 0.035	Uses repeated validation with lower-order stability reward and dispersion penalty.

For multi-label tasks, the distance quantile was fixed at $q_\varepsilon = 0.15$. Because the original multi-label attributes are deterministic point values, the default interval-width coefficient was $\rho = 0$, and the corresponding overlap threshold was set to zero. The maximum feature count and candidate-pool size were determined from data dimensionality and label count. For emotions, scene, yeast, birds, medical, and CAL500, the maximum feature counts were 32, 96, 48, 64, 24, and 16, and the candidate-pool sizes were 64, 96, 96, 96, 48, and 32, respectively. These values denote search-budget upper bounds

rather than the number of features that must be selected.

The paired parameters have complementary roles. Increasing q_ε relaxes the distance condition, whereas increasing θ strengthens the overlap condition. Increasing β gives predictive validation greater influence, whereas increasing λ favors smaller subsets. Their joint effects were therefore evaluated by two 3×3 factorial grids rather than inferred from independent one-factor curves.

4.5.2. Sensitivity and interaction results

Figure 6(a) summarizes the single-label one-factor analysis over all 10 datasets. The default configuration achieved a mean macro-F1 of 0.8000, a mean accuracy of 0.8624, and a mean selected-feature count of 5.50. Increasing $K_{\max}$ from 8 to 10 changed macro-F1 by only +0.00006, reduced accuracy by 0.0010, and increased the selected-feature count to 5.77. In contrast, setting $K_{\max}$ to 6 or 4 reduced macro-F1 by 0.0252 and 0.0502. Reducing the candidate pool to 8 shortened the average runtime from 130.89 s to 59.11 s but decreased macro-F1 by 0.0112. Removing the stability prior changed macro-F1 by −0.0004. These results locate the default feature budget and candidate-pool size near a performance-compression plateau.

Figure 6(b) reports the one-factor analysis on the four multi-label sensitivity datasets. Perturbations of T, $T^{\mathrm{tol}}_{\mathrm{multi}}$, ρ, and candidate-pool size changed mean micro-F1 by −0.0025 to +0.0132 relative to the default. The changes were substantially smaller than those caused by setting the maximum feature count to 10, 6, or 4, which reduced micro-F1 by 0.0416, 0.0862, and 0.1057. The multi-label model is therefore most sensitive to an excessively restrictive feature budget, whereas the purity and interval-width parameters exhibit moderate local sensitivity.

Figures 6(c) and 6(d) show two 3×3 interaction grids evaluated by three fold cross-validation on the complete wine, ionosphere, and zoo datasets. Across the (q_ε, θ) grid, mean macro-F1 ranged from 0.8076 to 0.8129, and the default pair $(0.20, 0.40)$ attained the maximum. Across the (β, λ) grid, the range was 0.8023–0.8129; the default pair $(0.62, 0.003)$ also lie on the maximum performance plateau. The mean selected-feature count remained between 3.00 and 3.11 throughout the objective grid and was 3.11 throughout the tolerance grid, indicating that the observed stability was not produced by abrupt changes in reduction size.

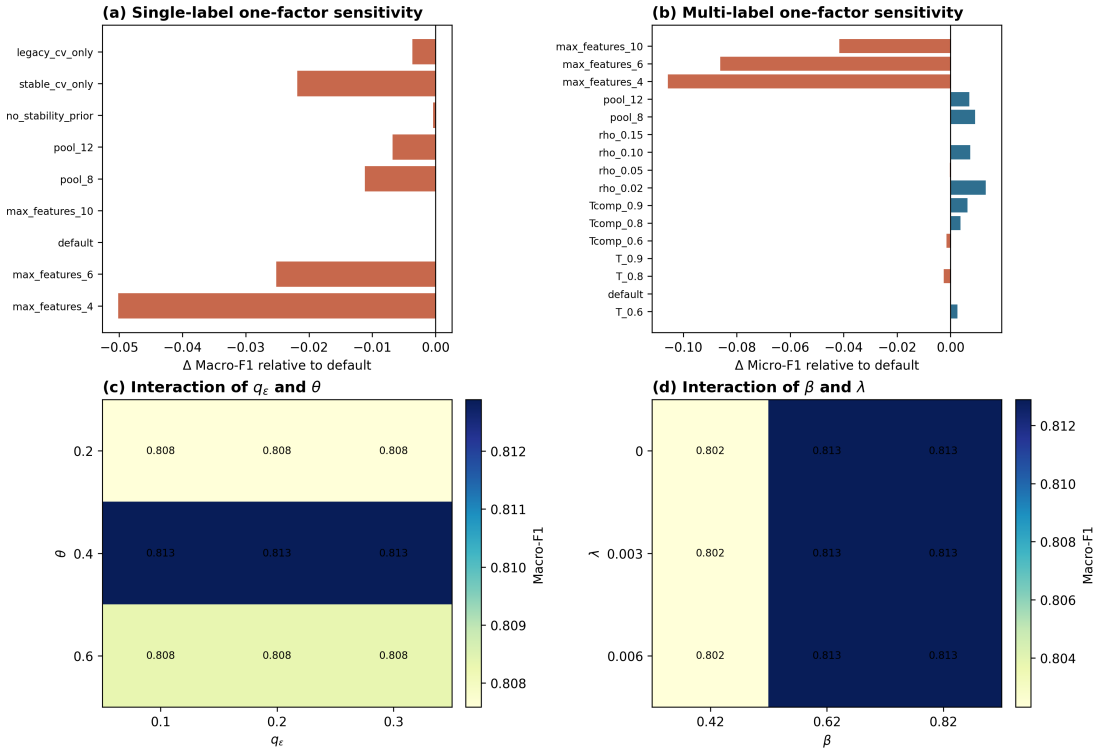

Figure 6. One-factor sensitivity and pairwise interaction analysis. (a) Single-label macro-F1 changes relative to the default configuration; (b) multi-label micro-F1 changes relative to the default configuration; (c) the interaction between the distance quantile q_ε and the overlap threshold θ; (d) the interaction between the predictive-performance weight β and the subset size penalty λ. Values in (c) and (d) denote the mean macro-F1.

4.6. Computational efficiency and scalability

After excluding one warm-up run, we measured the direct execution time of Algorithms 1– 7, and the results are shown in Table 12. Each reported value is the mean and standard deviation over three repeated measurements. The single-label profiling configuration was $(n, m, q) = (178, 13, 1)$, and the multi-label profiling configuration contained $(n, m, q) = (200, 72, 6)$. These measurements isolate attribute reduction from downstream classifier fitting and outer performance evaluation.

Table 12. Measured execution time of Algorithms 1–7.

Algorithm	Task	Runtime (s)
Algorithm 1	Single-label MCB generation	0.0053 ± 0.0005
Algorithm 2	Single-label granular-ball generation	0.0058 ± 0.0005
Algorithm 3	Strict single-label reduction	0.7348 ± 0.0028
Algorithm 4	Multi-label granular-ball generation	0.4766 ± 0.0009
Algorithm 5	Strict multi-label reduction	17.1696 ± 0.0765
Algorithm 6	Enhanced single-label reduction	9.6968 ± 0.0534
Algorithm 7	Enhanced multi-label reduction	5.9204 ± 0.0491

The generation algorithms completed in less than 0.5 s in both task settings.

We examined sample-size scalability for the enhanced single-label and multi-label algorithms while fixing the feature dimensionality, parameter values, and search budgets. The number of objects was varied over 100, 200, 400, and 800, and every setting was repeated twice. Figure 7(a) decomposes the end-to-end reduction time. For the single-label algorithm, search control and local refinement accounted for 56.03% of the runtime, followed by granular-ball processing (21.25%) and predictive validation (17.23%). For the multi-label algorithm, granular-ball processing was the dominant component (87.42%), while predictive validation, candidate scoring, and MCB generation accounted for 4.31%, 3.42%, and 2.93%, respectively.

Figure 7(b) reports the sample size results. The mean single-label runtime increased from 4.08 s at 100 objects to 99.46 s at 800 objects, whereas the mean multi-label runtime increased from 1.87 s to 32.41 s. A log–log regression of median runtime on sample size yielded empirical scaling exponents of 1.59 and 1.38, respectively. The largest individual run completed within 105.64 s, and the largest mean incremental memory use across the tested settings was 123.54 MB. The indexed tolerance graph avoids materializing absent object pairs, while the mutual information candidate pool and cached subset evaluation restrict repeated granular reconstruction. Consequently, the implementation remains practically tractable over the examined range. The empirical exponents characterize these data and fixed search budgets; they are not presented as universal worst-case bounds.

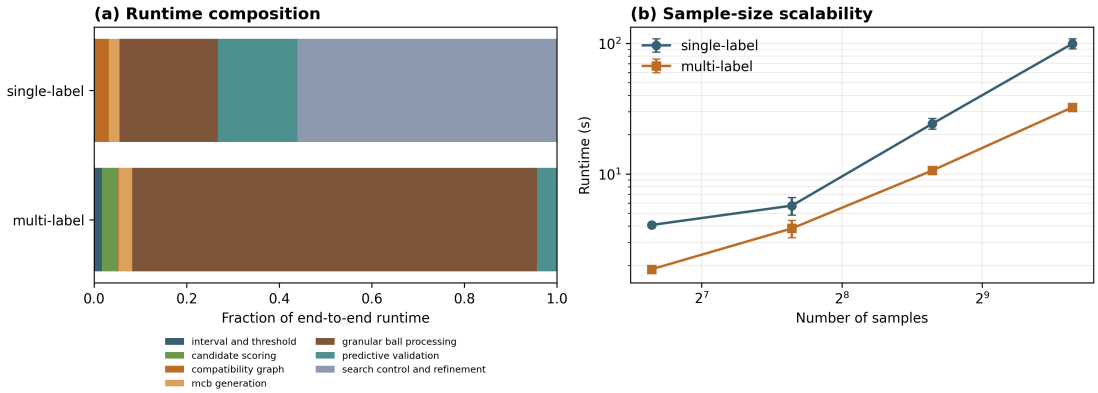

Figure 7. Computational efficiency and scalability. (a) Stage-wise fractions of end-to-end runtime for the enhanced single-label and multi-label algorithms. (b) Runtime as the number of objects increases from 100 to 800; markers and error bars denote the mean and standard deviation over two repeated measurements.

4.7. Noise robustness analysis

We evaluated robustness to training set corruption on three single-label datasets (wine, ionosphere, and zoo) and two genuine multi-label datasets (emotions and birds). The original binary label matrices of emotions and birds were used directly. The single-label experiment adopted stratified two-fold cross-validation, whereas the predefined training/test partitions were used for the multi-label datasets. Noise was added exclusively to the training portion of each split, and both the test attributes and test labels remained unchanged. All methods used the same corrupted training instance and a fixed random seed of 17.

Two noise mechanisms were examined. For attribute noise, an exact proportion of the training cells was selected. A selected continuous value was perturbed as

$$x'_{ia} = x_{ia} + v_{ia}, \quad v_{ia} \sim \mathcal{N}\left(0, (0.75\sigma_a)^2\right), \tag{4.10}$$

where σ_a denotes the standard deviation of attribute a estimated from the training split. A selected discrete value was replaced by a different category observed in the same training split. For decision noise, a selected single-label decision was replaced by another class. In a multi-label task, a selected entry of the binary label matrix was flipped according to

$$Y'_{ip} = 1 - Y_{ip}. \tag{4.11}$$

The common noise levels were 0%, 10%, and 20%; an additional 5% level was included for the single-label curves. The GBRS and MaxGBRS were used as the single-label comparators, while LRDG and SSFS were used in the multi-label experiment. The downstream classifiers and their settings were identical to those in the main experiments.

Let $S(r)$ denote macro-F1 for a single-label task or micro-F1 for a multi-label task at the noise level r. We measured performance retention by the clean-score-normalized area under the noise-performance curve

$$R_{\mathrm{AUC}} = \frac{1}{S(0)(r_{\max} - r_{\min})} \int_{r_{\min}}^{r_{\max}} S(r)\, dr, \tag{4.12}$$

where the integral was evaluated by the trapezoidal rule. Values close to 1 indicate that performance is retained across the examined range; values above 1 may occur when moderate random perturbation produces a regularization effect.

Table 13 summarizes the predictive performance under attribute noise and decision noise. When averaged over the selected datasets, the proposed method retained the highest absolute predictive score at the maximum noise level in all four task-noise settings. Its mean normalized area exceeded MaxGBRS under both single-label noise mechanisms and exceeded LRDG and SSFS under both multi-label mechanisms. The GBRS had a marginally larger mean normalized area under single-label attribute noise (1.0232 versus 1.0218), but its mean macro-F1 at 20% noise was 0.5014, compared with 0.7978 for the proposed method. These results support robustness in terms of maintained absolute performance and competitive relative retention over the examined noise range.

Table 13. Prediction performance under attribute noise and decision noise.

Task	Method	**Clean score**	Score (20% attr. noise)	Score (20% dec. noise)	R_{AUC} (attr.)	R_{AUC} (dec.)
Single-label (macro-F1)	Ours	0.8228	0.7978	0.7659	1.0218	1.0022
Single-label (macro-F1)	MaxGBRS	0.6956	0.6403	0.6253	0.9920	0.9786
Single-label (macro-F1)	GBRS	0.5376	0.5014	0.5001	1.0232	0.9834
Multi-label (micro-F1)	Ours	0.5075	0.5324	0.3141	1.0359	0.8546
Multi-label (micro-F1)	LRDG	0.4937	0.5122	0.3105	1.0185	0.7831
Multi-label (micro-F1)	SSFS	0.4646	0.4335	0.3055	1.0070	0.7768

Note: Clean score denotes the mean clean score; score (20% attr. noise) denotes the mean score at 20% attribute noise; score (20% dec. noise) denotes the mean score at 20% decision noise; R_{AUC} (attr.) denotes the mean attribute R_{AUC}; R_{AUC} (dec.) denotes the mean decision R_{AUC}.

4.8. Real-world maritime surveillance application

We further examined the proposed method in a dual-use coastal maritime surveillance application based on passive acoustic target classification. We used 300 quality-controlled passages,

comprising 100 motorboats, 100 sailboats, and 100 yachts, drawn from a publicly available database of underwater radiated noise from small vessels in coastal areas [42]. This case was analyzed separately from the public benchmark comparison because its purpose was to evaluate the complete interval-valued workflow in a defined application context.

Each passage was described by 10 interval-valued condition attributes. Speed and length were represented by uncertainty envelopes corresponding to the approximately 7% estimation error reported in the data descriptor [42]. The acoustic spectrum was aggregated into eight bands from 31.5 Hz to 20 kHz; the lower and upper bounds of each acoustic attribute were the minimum and maximum observed one-third-octave levels within the corresponding band. The 10 paired lower and upper bounds were supplied directly to the interval tolerance relation without point-to-interval conversion. Anonymous sample and source-record identifiers were excluded from the condition attributes.

We applied five-fold stratified cross-validation with three repetitions. Attribute reduction and every parameter-dependent decision were performed within the training fold. For downstream k-nearest neighbors (kNN) classification, each retained interval was represented by its midpoint and half-width, preserving both the location and width information. FullFeature, GBRS, MaxGBRS, and Ours used identical outer partitions and classification settings. Table 14 reports the mean and standard deviation over all 15 test folds.

Table 14. Prediction results on the coastal maritime surveillance dataset.

Method	Macro-F1 ↑	Accuracy ↑	Balanced Acc. ↑	MCC ↑	Features ↓	Reduction ↑
FullFeature	0.8261 ± 0.0460	0.8267 ± 0.0463	0.8267 ± 0.0463	0.7431 ± 0.0703	10.00 ± 0.00	0.0000 ± 0.0000
GBRS	0.7033 ± 0.1559	0.7078 ± 0.1479	0.7078 ± 0.1479	0.5683 ± 0.2185	1.00 ± 0.00	0.9000 ± 0.0000
MaxGBRS	0.8444 ± 0.0216	0.8444 ± 0.0217	0.8444 ± 0.0217	0.7691 ± 0.0336	2.00 ± 0.00	0.8000 ± 0.0000
Ours	0.8563 ± 0.0387	0.8567 ± 0.0389	0.8567 ± 0.0389	0.7873 ± 0.0585	3.20 ± 0.75	0.6800 ± 0.0748

Note: ↑ indicates higher values corresponding to better performance; ↓ indicates lower values corresponding to better performance.

Ours improved macro-F1 by 0.0302 over FullFeature, by 0.1529 over GBRS, and by 0.0119 over MaxGBRS. Ours retained 3.20 attributes on average and reduced the original attribute set by 68.0%. Vessel length was selected in all 15 folds, while the 31.5–63 Hz and 500–1000 Hz acoustic bands were each selected in 11 folds (73.3%). These results show that vessel length and the two acoustic bands were selected consistently across folds; they indicate selection stability only and do not directly establish the predictive contribution, discriminative strength, or a causal acoustic mechanism. The complete method required 44.145 s per fold for offline reduction, compared with 2.518 s for MaxGBRS, reflecting the cost of training-only predictive validation and subset refinement. The result supports the applicability of the proposed reduction framework to public, non-sensitive field observations.

5. Conclusions

This paper addresses the dimensional reduction problem for interval-valued data in single-label and multi-label scenarios. Within the framework of interval-valued decision information systems, an interval tolerance relation is designed by fusing the Hausdorff distance and the degree of interval overlap, and its basic properties are established. On this basis, MCBs are introduced to replace the global dataset as the initial granular-ball set, and GBRS attribute reduction models are established for

both single-label and multi-label scenarios, together with their theoretical properties. To make the reduction procedure more effective in practice, several algorithmic enhancement strategies, including a granular-ball quality function, a mutual information candidate pool, predictive performance validation, and local refinement, are further integrated into the abovementioned models as a complete algorithmic workflow. In the experimental section, the proposed method was compared with GBRS, maximal GBRS, and several representative feature selection methods on multiple public datasets covering single-label and multi-label tasks. The results show that the proposed method has clear advantages in terms of the classification accuracy and reduction rate, and exhibits good stability under parameter perturbations and attribute/decision noise, while the additional computational cost remains tractable over the examined data scales. Its practical value was further verified in a real-world maritime surveillance application.

Despite these advantages, the proposed method still has two limitations. First, the interval tolerance relation treats all conditional attributes equally through max–min aggregation, without accounting for the varying importance of different attributes; meanwhile, in multi-label scenarios, label correlations are captured only through the pairwise Jaccard similarity of label subsets, and higher-order label dependencies are not explicitly modeled. Second, the computational cost depends on the density of the induced tolerance graph: When the graph is intrinsically dense, the output size becomes an unavoidable cost. In addition, the current framework is restricted to interval-valued numerical data, and point-valued data must be converted by an intervalization step whose quality affects the subsequent reduction results.

Future research can proceed in the following directions. On the theoretical side, we plan to introduce attribute-weighted tolerance relations that reflect the varying importance of conditional attributes, and to incorporate richer label correlation modeling, such as label embeddings, into the granular-ball construction process for multi-label tasks; meanwhile, the proposed models will be extended to more general data representations, such as fuzzy or intuitionistic fuzzy interval-valued data, so as to relax the dependence on the intervalization step. On the practical side, incremental MCB updating and granular-ball reconstruction mechanisms will be investigated for large-scale dynamic data, and parallel strategies for granular-ball generation and attribute evaluation will be designed to further improve the scalability on ultra large-scale datasets. We will also apply the proposed method to a broader range of real-world decision problems, such as medical diagnosis and industrial fault diagnosis, to further examine its practical value.

Author contributions

Shiqi Chen: Conceptualization, methodology, writing – original draft, formal analysis; Zhongying Suo: Resources, methodology, funding acquisition, writing – review and editing; Yuanbo Kong: Validation, visualization, data curation; Songlei Xue: Data curation, investigation; Zhuoluo Wang: Writing – review and editing, supervision. All authors have read and approved the final version of the manuscript for publication.

Use of Generative AI tools declaration

The authors used an AI-assisted language tool for language polishing and grammar checking only. All scientific content was reviewed and approved by the authors.

Acknowledgments

This work was supported by the Shaanxi Fundamental Science Research Project for Mathematics and Physics (No. 25JSQ047).

Conflict of interest

The authors declare that they have no conflicts of interest.

References

1. M. Chavent, Y. Lechevallier, Dynamical clustering of interval data: Optimization of an adequacy criterion based on Hausdorff distance, In: *Classification, clustering, and data analysis: Recent advances and applications*, Berlin, Heidelberg: Springer, 2002, 53–60. https://doi.org/10.1007/978-3-642-56181-8_5
2. A. Sengupta, T. K. Pal, On comparing interval numbers, *Eur. J. Oper. Res.*, **127** (2000), 28–43. https://doi.org/10.1016/S0377-2217(99)00319-7
3. M. L. Zhang, Z. H. Zhou, A review on multi-label learning algorithms, *IEEE Trans. Knowl. Data Eng.*, **26** (2013), 1819–1837. https://doi.org/10.1109/TKDE.2013.39
4. A. N. Gorban, I. Y. Tyukin, Blessing of dimensionality: Mathematical foundations of the statistical physics of data, *Philos. Trans. R. Soc. A*, **376** (2018), 20170237. https://doi.org/10.1098/rsta.2017.0237
5. Y. Yu, *Multi-label learning: Theory, method and application*, (In Chinese), Beijing: Science Press, 2020.
6. Y. Omae, Y. Sakai, H. Takahashi, Features gradient-based signals selection algorithm of linear complexity for convolutional neural networks, *AIMS Mathematics*, **9** (2024), 792–817. https://doi.org/10.3934/math.2024041
7. M. Pang, Z. Li, A novel profit-based validity index approach for feature selection in credit risk prediction, *AIMS Mathematics*, **9** (2024), 974–997. https://doi.org/10.3934/math.2024049
8. Y. Liu, S. Guo, Weighted linear discriminant analysis: An effective feature extraction method for multi-class imbalanced datasets, *Symmetry*, **16** (2024), 1656. https://doi.org/10.3390/sym16121656
9. D. Yu, M. Zhang, X. Liu, F. Yao, Fault feature extraction and fusion method for AUV with weak thruster fault based on variational mode decomposition and D-S evidence theory, *Math. Biosci. Eng.*, **19** (2022), 9335–9356. https://doi.org/10.3934/mbe.2022434
10. Z. Pawlak, Rough sets, *Int. J. Comput. Inf. Sci.*, **11** (1982), 341–356. https://doi.org/10.1007/BF01001956
11. Q. Hu, D. Yu, J. Liu, C. Wu, Neighborhood rough set based heterogeneous feature subset selection, *Inf. Sci.*, **178** (2008), 3577–3594. https://doi.org/10.1016/j.ins.2008.05.024

12. H. Fang, J. Li, W. Song, Failure mode and effects analysis: An integrated approach based on rough set theory and prospect theory, *Soft Comput.*, **24** (2020), 6673–6685. https://doi.org/10.1007/s00500-019-04305-8

13. S. Senthil Kumar, H. Hannah Inbarani, Cardiac arrhythmia classification using multi-granulation rough set approaches, *Int. J. Mach. Learn. Cyber.*, **9** (2018), 651–666. https://doi.org/10.1007/s13042-016-0594-z

14. X. Chu, J. Xu, X. Chu, B. Sun, Y. Zhang, K. Bao, et al., A nonadditive rough set model for long-term clinical efficacy evaluation of chronic diseases in real-world settings, *Artif. Intell. Rev.*, **57** (2024), 33. https://doi.org/10.1007/s10462-023-10672-4

15. T. Fujita, The hyperfuzzy VIKOR and hyperfuzzy DEMATEL methods for multi-criteria decision-making, *Spectr. Decis. Mak. Appl.*, **3** (2026), 292–315. https://doi.org/10.31181/sdmap31202654

16. S. Xia, C. Wang, G. Wang, X. Gao, W. Ding, J. Yu, et al., GBRS: A unified granular-ball learning model of Pawlak rough set and neighborhood rough set, *IEEE Trans. Neural Networks Learn. Syst.*, **36** (2023), 1719–1733. https://doi.org/10.1109/TNNLS.2023.3325199

17. X. Lian, S. Xia, B. Sang, G. Wang, X. Gao, GBFRS: Robust fuzzy rough sets via granular ball computing, *IEEE Trans. Neural Networks Learn. Syst.*, 2026, in press. https://doi.org/10.1109/TNNLS.2026.3681370

18. W. Ye, W. Xu, Innovative multi-granularity granular-balls rough set for feature selection: driving generalized multi-granularity rough set evolution with Zentropy integration, *Inf. Sci.*, **718** (2025), 122411. https://doi.org/10.1016/j.ins.2025.122411

19. X. Peng, Y. Gong, X. Hou, Z. Tang, Y. Shao, A novel covering rough set model based on granular-ball computing for data with label noise, *Int. J. Approx. Reason.*, **182** (2025), 109420. https://doi.org/10.1016/j.ijar.2025.109420

20. W. Xu, *Ordered information system and rough set*, (In Chinese), Beijing: Science Press, 2013.

21. W. Ali, T. Shaheen, I. U. Haq, H. G. Toor, T. Alballa, H. A. E.-W. Khalifa, A novel interval-valued decision theoretic rough set model with intuitionistic fuzzy numbers based on power aggregation operators and their application in medical diagnosis, *Mathematics*, **11** (2023), 4153. https://doi.org/10.3390/math11194153

22. W. Li, H. Zhou, W. Xu, X. Z. Wang, P. Pedrycz, Interval dominance-based feature selection for interval-valued ordered data, *IEEE Trans. Neural Networks Learn. Syst.*, **34** (2022), 6898–6912. https://doi.org/10.1109/TNNLS.2022.3184120

23. B. Sang, H. Chen, Y. Yang, T. Li, W. Xu, C. Luo, Feature selection for dynamic interval-valued ordered data based on fuzzy dominance neighborhood rough set, *Knowl. Based Syst.*, **227** (2021), 107223. https://doi.org/10.1016/j.knosys.2021.107223

24. K. Atanassov, G. Gargov, Interval valued intuitionistic fuzzy sets, *Fuzzy Sets Syst.*, **31** (1989), 343–349. https://doi.org/10.1016/0165-0114(89)90205-4

25. Z. Xu, Methods for aggregating interval-valued intuitionistic fuzzy information and their application to decision making, *Control Decis.*, **22** (2007), 215–219.

26. S. P. Wan, J. Y. Dong, S. M. Chen, Y. Gao, Group decision making based on novel interval-valued intuitionistic fuzzy best-worst method with multiplicative consistency, *Eng. Appl. Artif. Intell.*, **167** (2026), 113713. https://doi.org/10.1016/j.engappai.2025.113713
27. Y. Leung, D. Li, Maximal consistent block technique for rule acquisition in incomplete information systems, *Inf. Sci.*, **153** (2003), 85–106. https://doi.org/10.1016/S0020-0255(03)00061-6
28. R. Li, H. Chen, S. Liu, K. Wang, B. Wang, X. Hu, TFD-IIS-CRMCB: Telecom fraud detection for incomplete information systems based on correlated relation and maximal consistent block, *Entropy*, **25** (2023), 112. https://doi.org/10.3390/e25010112
29. Y. Sun, W. Z. Wu, X. Wang, Maximal consistent block based optimal scale selection for incomplete multi-scale information systems, *Int. J. Mach. Learn. Cybern.*, **14** (2023), 1797–1809. https://doi.org/10.1007/s13042-022-01728-y
30. Y. Sun, J. Mi, T. Feng, L. Li, M. Liang, Maximum consistent block based variable precision rough set model and attribute reduction, *J. Front. Comput. Sci. Technol.*, **14** (2020), 892–900. https://doi.org/10.3778/j.issn.1673-9418.1905011
31. M. Kryszkiewicz, Rough set approach to incomplete information systems, *Inf. Sci.*, **112** (1998), 39–49. https://doi.org/10.1016/S0020-0255(98)10019-1
32. Y. Chen, X. Liu, M. Deveci, Z. Wang, S. Zhang, Preference disaggregation-based multiclass Mahalanobis-Taguchi system applied to medical insurance fraud, *Eng. Appl. Artif. Intell.*, **167** (2026), 113934. https://doi.org/10.1016/j.engappai.2026.113934
33. A. Hashemi, M. B. Dowlatshahi, H. Nezamabadi-pour, VMFS: A VIKOR-based multi-target feature selection, *Expert Syst. Appl.*, **182** (2021), 115224. https://doi.org/10.1016/j.eswa.2021.115224
34. A. Hashemi, M. B. Dowlatshahi, H. Nezamabadi-Pour, MFS-MCDM: Multi-label feature selection using multi-criteria decision making, *Knowl. Based Syst.*, **206** (2020), 106365. https://doi.org/10.1016/j.knosys.2020.106365
35. J. Lee, D. W. Kim, Fast multi-label feature selection based on information-theoretic feature ranking, *Pattern Recognit.*, **48** (2015), 2761–2771. https://doi.org/10.1016/j.patcog.2015.04.009
36. Y. Li, L. Hu, W. Gao, Robust sparse and low-redundancy multi-label feature selection with dynamic local and global structure preservation, *Pattern Recognit.*, **134** (2023), 109120. https://doi.org/10.1016/j.patcog.2022.109120
37. Y. Li, L. Hu, W. Gao, Multi-label feature selection via robust flexible sparse regularization, *Pattern Recognit.*, **134** (2023), 109074. https://doi.org/10.1016/j.patcog.2022.109074
38. Y. Zou, X. Hu, P. Li, Gradient-based multi-label feature selection considering three-way variable interaction, *Pattern Recognit.*, **145** (2024), 109900. https://doi.org/10.1016/j.patcog.2023.109900
39. Y. Zhang, W. Huo, J. Tang, Multi-label feature selection via latent representation learning and dynamic graph constraints, *Pattern Recognit.*, **151** (2024), 110411. https://doi.org/10.1016/j.patcog.2024.110411
40. J. Zhang, Y. Lin, M. Jiang, S. Li, Y. Tang, J. Long, et al., Fast multilabel feature selection via global relevance and redundancy optimization, *IEEE Trans. Neural Networks Learn. Syst.*, **35** (2024), 5721–5734. https://doi.org/10.1109/TNNLS.2022.3208956

41. W. Gao, Y. Li, L. Hu, Multilabel feature selection with constrained latent structure shared term, *IEEE Trans. Neural Networks Learn. Syst.*, **34** (2023), 1253–1262. https://doi.org/10.1109/TNNLS.2021.3105142

42. M. Shipton, J. Obradović, F. Ferreira, N. Mišković, T. Bulat, N. Cukrov, et al., A database of underwater radiated noise from small vessels in the coastal area, *Sci. Data*, **12** (2025), 289. https://doi.org/10.1038/s41597-025-04584-x

Appendix

Full experimental results

This appendix presents the complete dataset-wise, method-wise, and metric-wise experimental results that are not elaborated in the main text. The main text only retains the average performance, statistical tests, and key visualizations for concise argumentation, while this appendix provides detailed data for result verification and experimental reproduction.

Unless otherwise specified, values formatted with bold font and underline represent the optimal result for the corresponding dataset and metric; values formatted with italic font and underline represent the suboptimal result. For evaluation metrics, a smaller Hamming loss indicates better performance, while all other metrics are optimized with larger values.

Results on single-label tasks

Table 15. Full comparison results across all datasets and metrics for single-label tasks.

Dataset	Method	Accuracy	Macro-F1	Balanced Acc.	MCC	Features	Reduction	Runtime(s)
HTRU_2	Ours	0.9777	0.9300	0.9117	0.8613	3.0000	0.6250	90.9581
HTRU_2	FullFeature	*0.9784*	*0.9320*	0.9116	*0.8656*	8.0000	0.0000	0.0734
HTRU_2	GBRS	0.9759	0.9234	0.9009	0.8492	1.0000	0.8750	163.2082
HTRU_2	MaxGBRS	0.9769	0.9272	0.9080	0.8558	2.0000	0.7500	5370.4401
HTRU_2	VMFS	0.9788	0.9334	0.9156	0.8681	3.0000	0.6250	0.0254
HTRU_2	MFS-MCDM	0.9788	0.9334	0.9156	0.8681	3.0000	0.6250	0.0255
HTRU_2	FIMF	0.9766	0.9262	0.9067	0.8541	3.0000	0.6250	0.0475
HTRU_2	SLMDS	0.9779	0.9310	0.9154	0.8629	3.0000	0.6250	68.7875
HTRU_2	RFSFS	0.9779	0.9312	0.9160	0.8635	3.0000	0.6250	0.0366
HTRU_2	G3WI	0.9780	0.9307	0.9106	0.8631	3.0000	0.6250	0.0703
Electrical	Ours	0.8927	0.8809	0.8724	0.7650	8.0000	0.3333	140.2517
Electrical	FullFeature	0.8515	0.8320	0.8199	0.6728	12.0000	0.0000	0.2741
Electrical	GBRS	0.6231	0.5641	0.5646	0.1398	1.0000	0.9167	64.3250
Electrical	MaxGBRS	0.6713	0.6276	0.6243	0.2619	2.0000	0.8333	283.9872
Electrical	VMFS	0.8927	0.8809	0.8724	0.7650	8.0000	0.3333	0.0888
Electrical	MFS-MCDM	0.8927	0.8809	0.8724	0.7650	8.0000	0.3333	0.0889
Electrical	FIMF	0.7390	0.7047	0.6985	0.4167	8.0000	0.3333	0.0990
Electrical	SLMDS	0.7873	0.7643	0.7589	0.5312	8.0000	0.3333	19.9441
Electrical	RFSFS	*0.8525*	*0.8362*	*0.8285*	*0.6756*	8.0000	0.3333	0.1012
Electrical	G3WI	0.7284	0.6911	0.6847	0.3910	8.0000	0.3333	0.1508
Ionosphere	Ours	0.9060	0.8934	0.8795	0.7951	3.0000	0.9118	58.8044
Ionosphere	FullFeature	0.8291	0.7911	0.7706	0.6304	34.0000	0.0000	0.0068
Ionosphere	GBRS	0.8462	0.8252	0.8154	0.6594	1.0000	0.9706	2.0199
Ionosphere	MaxGBRS	0.8575	0.8451	0.8452	0.6973	2.0000	0.9412	2.7442
Ionosphere	VMFS	0.8689	0.8481	0.8314	0.7142	3.0000	0.9118	0.0021
Ionosphere	MFS-MCDM	0.8376	0.8051	0.7860	0.6453	3.0000	0.9118	0.0024

Dataset	Method	Accuracy	Macro-F1	Balanced Acc.	MCC	Features	Reduction	Runtime(s)
Ionosphere	FIMF	0.8091	0.7909	0.7883	0.5847	3.0000	0.9118	0.0075
Ionosphere	SLMDS	*0.8803*	*0.8613*	*0.8473*	*0.7386*	3.0000	0.9118	0.0210
Ionosphere	RFSFS	0.8519	0.8283	0.8129	0.6724	3.0000	0.9118	0.0062
Ionosphere	G3WI	0.7863	0.7064	0.7094	0.5386	3.0000	0.9118	0.2002
Lymphography	Ours	*0.7707*	*0.5440*	*0.5466*	*0.5565*	4.6667	0.7407	40.2857
Lymphography	FullFeature	0.8248	0.5734	0.5778	0.6591	18.0000	0.0000	0.0066
Lymphography	GBRS	0.5743	0.2264	0.2929	0.0806	1.3333	0.9259	1.9498
Lymphography	MaxGBRS	0.6820	0.4338	0.4519	0.3949	2.3333	0.8704	2.4914
Lymphography	VMFS	0.7227	0.4060	0.4122	0.4586	4.6667	0.7407	0.0019
Lymphography	MFS-MCDM	0.6890	0.3896	0.3965	0.3873	4.6667	0.7407	0.0023
Lymphography	FIMF	0.7158	0.4032	0.4094	0.4413	4.6667	0.7407	0.0288
Lymphography	SLMDS	0.7705	0.4333	0.4408	0.5504	4.6667	0.7407	0.0181
Lymphography	RFSFS	0.7498	0.5200	0.5261	0.5040	4.6667	0.7407	0.0048
Lymphography	G3WI	0.6279	0.3737	0.4286	0.2650	4.6667	0.7407	0.2302
Mushroom	Ours	*0.9983*	*0.9983*	*0.9983*	*0.9966*	4.6667	0.7879	240.4924
Mushroom	FullFeature	1.0000	1.0000	1.0000	1.0000	22.0000	0.0000	0.0397
Mushroom	GBRS	0.9921	0.9921	0.9918	0.9843	4.0000	0.8182	37.1836
Mushroom	MaxGBRS	0.9935	0.9935	0.9936	0.9870	4.0000	0.8182	50.3826
Mushroom	VMFS	0.9579	0.9576	0.9566	0.9180	4.6667	0.7879	0.0243
Mushroom	MFS-MCDM	0.9861	0.9861	0.9857	0.9723	4.6667	0.7879	0.0394
Mushroom	FIMF	0.9389	0.9387	0.9379	0.8793	4.6667	0.7879	0.0447
Mushroom	SLMDS	0.9233	0.9189	0.9206	0.8625	4.6667	0.7879	14.1068
Mushroom	RFSFS	0.7954	0.7803	0.7881	0.6381	4.6667	0.7879	0.0490
Mushroom	G3WI	0.9449	0.9440	0.9436	0.8934	4.6667	0.7879	0.1919
Primary_tumor	Ours	*0.3540*	*0.1597*	*0.1904*	*0.2515*	7.3333	0.5686	153.3674
Primary_tumor	FullFeature	0.3894	0.1999	0.2245	0.3080	17.0000	0.0000	0.0069
Primary_tumor	GBRS	0.2478	0.0229	0.0577	0.0000	1.0000	0.9412	3.6047
Primary_tumor	MaxGBRS	0.2625	0.0674	0.1067	0.1243	2.0000	0.8824	6.5037
Primary_tumor	VMFS	0.3068	0.1171	0.1499	0.1819	7.3333	0.5686	0.0024
Primary_tumor	MFS-MCDM	0.3510	0.1363	0.1763	0.2405	7.3333	0.5686	0.0031
Primary_tumor	FIMF	0.3097	0.1300	0.1500	0.1911	7.3333	0.5686	0.9946
Primary_tumor	SLMDS	0.2507	0.0865	0.1140	0.1003	7.3333	0.5686	0.1251
Primary_tumor	RFSFS	0.3068	0.1234	0.1507	0.2091	7.3333	0.5686	0.0050
Primary_tumor	G3WI	0.3127	0.1588	0.1677	0.2012	7.3333	0.5686	1.9718
Spambase	Ours	*0.8991*	*0.8936*	*0.8910*	*0.7878*	7.0000	0.8772	404.3434
Spambase	FullFeature	0.9072	0.9023	0.9002	0.8051	57.0000	0.0000	0.0244
Spambase	GBRS	0.4047	0.3399	0.4932	-0.0246	1.0000	0.9825	31.7583
Spambase	MaxGBRS	0.8192	0.8091	0.8066	0.6189	2.0000	0.9649	125.8939
Spambase	VMFS	0.7602	0.7560	0.7724	0.5480	7.0000	0.8772	0.0224
Spambase	MFS-MCDM	0.7009	0.6968	0.7296	0.4739	7.0000	0.8772	0.0221
Spambase	FIMF	0.8820	0.8751	0.8714	0.7513	7.0000	0.8772	0.0570
Spambase	SLMDS	0.5579	0.5448	0.6137	0.2467	7.0000	0.8772	5.5417
Spambase	RFSFS	0.4797	0.4536	0.5502	0.1280	7.0000	0.8772	0.0528
Spambase	G3WI	0.5414	0.5357	0.5906	0.1951	7.0000	0.8772	0.9072
WDBC	Ours	0.9438	0.9392	0.9381	0.8803	5.0000	0.8333	79.3454
WDBC	FullFeature	0.9649	0.9619	0.9567	0.9252	30.0000	0.0000	0.0076
WDBC	GBRS	0.8841	0.8736	0.8733	0.7506	1.0000	0.9667	2.1337
WDBC	MaxGBRS	0.9121	0.9059	0.9061	0.8118	2.0000	0.9333	1.9797
WDBC	VMFS	*0.9491*	*0.9446*	*0.9413*	*0.8913*	5.0000	0.8333	0.0025
WDBC	MFS-MCDM	0.9297	0.9237	0.9192	0.8492	5.0000	0.8333	0.0028
WDBC	FIMF	0.9139	0.9075	0.9075	0.8181	5.0000	0.8333	0.0081
WDBC	SLMDS	0.8347	0.8215	0.8174	0.6446	5.0000	0.8333	0.0450
WDBC	RFSFS	0.9174	0.9105	0.9103	0.8247	5.0000	0.8333	0.0090
WDBC	G3WI	0.9122	0.9045	0.8995	0.8115	5.0000	0.8333	0.1645
Wine	Ours	0.9608	0.9625	0.9663	0.9417	5.3333	0.5897	44.7750
Wine	FullFeature	*0.9606*	*0.9618*	*0.9649*	*0.9416*	13.0000	0.0000	0.0019

Dataset	Method	Accuracy	Macro-F1	Balanced Acc.	MCC	Features	Reduction	Runtime(s)
Wine	GBRS	0.6920	0.6949	0.7019	0.5398	1.0000	0.9231	0.6584
Wine	MaxGBRS	0.8539	0.8542	0.8594	0.7765	2.0000	0.8462	0.6098
Wine	VMFS	0.9439	0.9442	0.9478	0.9197	5.3333	0.5897	0.0020
Wine	MFS-MCDM	0.9494	0.9510	0.9545	0.9252	5.3333	0.5897	0.0023
Wine	FIMF	0.9327	0.9354	0.9401	0.9022	5.3333	0.5897	0.0126
Wine	SLMDS	0.8260	0.8365	0.8483	0.7545	5.3333	0.5897	0.0172
Wine	RFSFS	0.8879	0.8949	0.8994	0.8336	5.3333	0.5897	0.0049
Wine	G3WI	0.8648	0.8634	0.8701	0.8024	5.3333	0.5897	0.0927
Zoo	Ours	0.9207	0.7980	0.8254	0.8993	7.0000	0.5625	36.7701
Zoo	FullFeature	*0.9112*	*0.7443*	0.7630	*0.8863*	16.0000	0.0000	0.0065
Zoo	GBRS	0.6037	0.2573	0.3299	0.5338	1.3333	0.9167	0.7229
Zoo	MaxGBRS	0.8122	0.5418	0.6349	0.7656	2.6667	0.8333	0.9897
Zoo	VMFS	0.8910	0.7164	0.7268	0.8619	7.0000	0.5625	0.0022
Zoo	MFS-MCDM	0.8815	0.6907	*0.7749*	0.8530	7.0000	0.5625	0.0023
Zoo	FIMF	0.8417	0.6352	0.6967	0.8039	7.0000	0.5625	0.0780
Zoo	SLMDS	0.8619	0.6749	0.7370	0.8273	7.0000	0.5625	0.0266
Zoo	RFSFS	0.8018	0.5751	0.5964	0.7478	7.0000	0.5625	0.0050
Zoo	G3WI	0.8217	0.6301	0.7324	0.7809	7.0000	0.5625	0.3635

Results on multi-label tasks

Table 16. Full comparison results across all datasets and metrics for multi-label tasks.

Dataset	Method	Micro-F1	Macro-F1	Hamming loss	Average precision	Features	Reduction
CAL500	Ours	0.4719	0.1685	0.2204	0.1814	16	0.7647
CAL500	FullFeature	*0.4706*	0.1693	*0.2224*	*0.1883*	68	0.0000
CAL500	GRRO	0.4680	*0.1686*	0.2245	0.1868	16	0.7647
CAL500	LRDG	0.4685	0.1672	0.2228	0.1927	16	0.7647
CAL500	SSFS	0.4654	0.1648	0.2249	0.1797	16	0.7647
Birds	Ours	0.3698	0.2024	0.0666	0.2479	16	0.9410
Birds	FullFeature	0.2976	*0.1963*	0.0715	*0.2286*	271	0.0000
Birds	GRRO	*0.3306*	0.1425	0.0660	0.1802	16	0.9410
Birds	LRDG	0.2760	0.1615	0.0624	0.1929	16	0.9410
Birds	SSFS	0.2800	0.1487	*0.0645*	0.1809	16	0.9410
Emotions	Ours	0.7197	0.7044	0.2005	*0.7018*	32	0.5556
Emotions	FullFeature	*0.7092*	*0.6989*	0.2145	0.6986	72	0.0000
Emotions	GRRO	0.6850	0.6710	*0.2079*	0.6866	32	0.5556
Emotions	LRDG	0.6936	0.6897	0.2376	0.7151	32	0.5556
Emotions	SSFS	0.6478	0.6360	0.2665	0.6715	32	0.5556
Medical	Ours	0.6507	*0.1555*	0.0181	*0.2317*	24	0.9834
Medical	FullFeature	*0.6103*	0.2006	0.0219	0.2651	1449	0.0000
Medical	GRRO	0.4951	0.0805	0.0232	0.1260	24	0.9834
Medical	LRDG	0.6045	0.1180	*0.0196*	0.1904	24	0.9834
Medical	SSFS	0.5646	0.1064	0.0213	0.1769	24	0.9834
Scene	Ours	0.7133	*0.7209*	0.1095	*0.7702*	96	0.6735
Scene	FullFeature	0.7284	0.7342	*0.1017*	0.7751	294	0.0000
Scene	GRRO	0.6566	0.6709	0.1394	0.6774	96	0.6735
Scene	LRDG	*0.7143*	0.7160	0.0992	0.7620	96	0.6735
Scene	SSFS	0.6219	0.6224	0.1427	0.6311	96	0.6735
Yeast	Ours	0.6557	0.3877	*0.2097*	0.4672	24	0.7670
Yeast	FullFeature	0.6503	0.3751	0.2004	*0.4735*	103	0.0000
Yeast	GRRO	*0.6587*	*0.4169*	0.2300	0.4720	24	0.7670
Yeast	LRDG	0.6635	0.4296	0.2242	0.4789	24	0.7670
Yeast	SSFS	0.6257	0.3600	0.2484	0.4022	24	0.7670

Parameter sensitivity analysis

Analysis on single-label tasks

Table 17. Parameter sensitivity analysis results for single-label tasks.

Variant	Macro-F1	Accuracy	Features	Runtime(s)	Delta Macro-F1
max_features_10	0.8000	*0.8614*	5.7667	131.8761	0.0001
default	0.8000	0.8624	5.5000	130.8898	*0.0000*
no_stability_prior	*0.7996*	0.8609	5.5667	130.5861	-0.0004
legacy_cv_only	0.7963	0.8581	5.6000	120.7487	-0.0036
pool_12	0.7932	0.8591	5.9667	125.1795	-0.0068
pool_8	0.7888	0.8552	4.7000	59.1108	-0.0112
stable_cv_only	0.7781	0.8554	5.3667	132.1565	-0.0219
max_features_6	0.7748	0.8428	4.8333	126.3888	-0.0252
max_features_4	0.7498	0.8348	3.5667	84.6583	-0.0502

Analysis on multi-label tasks

Table 18. Parameter sensitivity analysis results for multi-label tasks.

Variant	Delta Micro-F1	Delta Macro-F1	Delta Hamming loss	Delta AP	Features
T_0.6	0.0026	-0.0027	0.0100	-0.0018	44.0000
T_0.8	-0.0025	-0.0124	0.0025	-0.0093	42.0000
T_0.9	-0.0001	-0.0006	0.0109	*0.0011*	46.0000
Tcomp_0.6	-0.0015	0.0027	0.0046	-0.0058	40.0000
Tcomp_0.8	0.0037	0.0022	*-0.0041*	0.0008	44.0000
Tcomp_0.9	0.0063	0.0044	0.0012	-0.0058	42.0000
default	0.0000	0.0000	0.0000	0.0000	46.0000
max_features_10	-0.0416	-0.0495	0.0305	-0.0723	10.0000
max_features_4	-0.1057	-0.0988	0.0637	-0.1362	4.0000
max_features_6	-0.0862	-0.0996	0.0524	-0.1216	6.0000
pool_12	0.0070	0.0055	0.0009	0.0023	48.0000
pool_8	*0.0092*	0.0112	-0.0046	0.0007	46.0000
rho_0.02	0.0132	*0.0083*	-0.0015	-0.0010	46.0000
rho_0.05	-0.0003	-0.0019	-0.0016	-0.0039	44.0000
rho_0.10	0.0073	0.0010	0.0022	-0.0072	46.0000
rho_0.15	-0.0002	0.0025	0.0099	-0.0063	46.0000

Summary of main notation

Table 19. Summary of main notations.

Notation	Description
(1) Interval-valued decision information system	
$DT = \langle U, AT, V, f \rangle$	Interval-valued decision information system (Definition 1)
U	Universe, i.e., the set of all objects (samples)
n	Number of objects, i.e., $n = \lvert U \rvert$
x, x_i	Objects (samples) in U
A	Set of conditional attributes; $a \in A$ denotes one conditional attribute
m	Number of conditional attributes, i.e., $m = \lvert A \rvert$

Table 19. Summary of main notations (continued).

Notation	Description
D	Set of decision attributes ($D = \{d\}$ in the single-label case)
$AT = A \cup D$	Full attribute set
V_a	Value domain of attribute a; $V = \bigcup_{a \in AT} V_a$
f	Information function; $f(x, a)$ is the value of object x on attribute a
$z_{x,a}$	Point-valued attribute value before intervalization (Definition 1)
D_i	The i-th decision class (single-label)
L	Label set in multi-label systems; q denotes the number of labels
L_x	Label subset associated with object x (multi-label)
L_p^+	The set of objects having label p
σ_a	Standard deviation of attribute a (Algorithms 6, Eq (4.10))
ρ_a	Attribute-wise perturbation radius, $\rho_a = \eta\sigma_a$ (Algorithms 6)
(2) Interval measures	
$I = [l, r]$	A closed interval with lower endpoint l and upper endpoint r
$l_{x,a}, r_{x,a}$	Lower and upper endpoints of $f(x, a)$
$\|I\|$	Length of interval I, i.e., $\|I\| = r - l$
$m_{x,a}$	Midpoint of $f(x, a)$, i.e., $(l_{x,a} + r_{x,a})/2$
$H(I_1, I_2)$	Hausdorff distance between two intervals (Definition 2)
$O(I_1, I_2)$	Normalized interval overlap degree (Definition 2)
ε_l	Small positive guarding constant in O (Definition 2)
ρ	Perturbation radius for converting a point value into an interval (Definition 1)
(3) Interval tolerance relation and MCBs	
ε	Distance threshold in the interval tolerance relation (Definition 3)
θ	Overlap threshold in the interval tolerance relation (Definition 3)
$T_B^{\varepsilon,\theta}$	Interval tolerance relation under attribute subset $B \subseteq A$ (Definition 3)
$N_B(x)$	Tolerance class of object x under B (Eq (3.1))
$\mathcal{MCB}(B)$	Collection of all MCBs under B (Definition 4)
$\widehat{\mathcal{MCB}}(B)$	Approximate MCB collection returned by Algorithm 1
X	A (maximal) consistent block, i.e., $X \in \mathcal{MCB}(B)$
M_{ij}	Tolerance indicator, $M_{ij} = \mathbb{I}((x_i, x_j) \in T_B^{\varepsilon,\theta})$ (Eq (3.2))
$\deg_X(x_i)$	Tolerance degree of x_i within candidate block X (Eq (3.3))
$Cand, MaxCand$	Candidate block set and its maximality-screened subset (Algorithm 1)
C	Repaired clique produced by GreedyClique (Algorithm 1)
(4) GBRS model	
GB, GB_i	Granular balls
c_{GB}^a	Center of granular ball GB on attribute a (Definition 7)
r_{GB}	Radius of granular ball GB (Definition 7)
$\omega_{x,a}$	Interval length weight for center computation (Definition 7)
$\omega'_{x,a}$	Interval length weight in interval k-means (Section 3.2.2)
$P_{single}(GB)$	Single-label purity of GB (Definition 8)

Table 19. Summary of main notations (continued).

Notation	Description
$P^{tol}_{single}(GB)$	Single-label tolerance purity of GB (Definition 8)
T'	Single-label purity threshold
T^{tol}_{single}	Single-label tolerance purity threshold
$LBS(B)$	Lower bound of splitting size, $LBS(B) = \max\{2, 2\|B\|\}$ (Section 3.2.2)
$NOLGBs_B$	Non-overlapping granular-ball set generated under B
$P_{multi}(GB)$	Multi-label purity of GB (Definition 11)
$P^{tol}_{multi}(GB)$	Multi-label tolerance purity of GB (Definition 11)
T	Multi-label purity threshold
T^{tol}_{multi}	Multi-label tolerance purity threshold
$J(L_{x_i}, L_{x_j})$	Jaccard similarity between two label subsets (Definition 11)
(5) Approximations and attribute reduction	
$\underline{apr}^{tol}_B(\cdot)$	Tolerance of lower approximation (Definitions 9 and 12)
$\overline{apr}^{tol}_B(\cdot)$	Tolerance of upper approximation (Definitions 9 and 12)
$POS^{tol}_B(D)$, $POS^{tol}_B(L)$	Positive regions for single-label and multi-label systems
$\gamma^{tol}_B(D)$, $\gamma^{tol}_B(L)$	Dependency functions for single-label and multi-label systems
$SIG^{tol}(a, B, \cdot)$	Significance of attribute a with respect to B (Definitions 10 and 13)
BND_B	Boundary region under attribute subset B
(6) Enhancement strategies	
$Q(B)$	Granular-ball quality function under attribute subset B (Section 3.4)
$MI(a; y)$	Mutual information between attribute a and the decision/labels (Section 3.4)
$S_{pred}(B)$	Predictive performance score of attribute subset B (Section 3.4)
$J(B)$	Comprehensive objective function for the attribute search (Section 3.4)
$K_{\max}, K_{\min}$	Upper and lower bounds on the number of selected attributes
$PoolSize$	Size of the mutual information candidate pool (Algorithms 6 and 7)
q_ε	Distance quantile used to determine ε when it is not given (Algorithms 6 and 7)
R_s	Number of random training/validation splits in predictive performance verification (Section 3.4)
R_{cv}	Number of repetitions in stable validation (Table 11)
ξ, ζ	Stability reward and dispersion penalty coefficients in stable validation (Table 11)
η	Intervalization coefficient in the enhanced workflow (Algorithms 6)
α, β, λ	Weighting coefficients in the comprehensive objective function
$\omega_1, \omega_2, \omega_3, \lambda_1, \lambda_2$	Weighting coefficients in the granular-ball quality function
(7) Complexity analysis	
b	Size of the currently evaluated subset, $b = \|B\|$ (Section 3.5)
e_B	Number of nonzero edges in the indexed tolerance graph under B

Table 19. Summary of main notations (continued).

Notation	Description
m_B	Total object memberships in the generated blocks or granular balls under B
z_B	Aggregate work of the observed candidate-block containment checks
g_s, g_m	Accumulated object assignments during single-label and multi-label splitting
h_s, h_m	Candidate pairs examined during heterogeneous overlap removal
s_B	Number of distinct label signatures under B
j_B	Number of signature pairs whose Jaccard similarity is evaluated
$T_k(B)$	Cost of Algorithm k on subset B
$\mathcal{S}_s, \mathcal{S}_m$	Collections of subsets visited by the strict single-label and multi-label searches
$\mathcal{S}'_s, \mathcal{S}'_m$	Candidate-pool counterparts of $\mathcal{S}_s$ and $\mathcal{S}_m$
$C_{\text{pred}}(B), C^{\text{ML}}_{\text{pred}}(B)$	Classifier-dependent validation costs for single-label and multi-label tasks
u_s, u_m, u'_s, u'_m	Cache footprints of the corresponding searches

AIMS Press