mathematics

MDPI

Article

A Maximal Consistent Block-Based Variable Precision Rough Set Method for Dimensional Reduction of Continuous Single-Label and Multi-Label Data

Shiqi Chen [1], Zhongying Suo [1,*] and Yuanbo Kong [2,3]

[1] Fundamentals Department, Air Force Engineering University, Xi'an 710051, China; csqcsqcsq0603@163.com
[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
* Correspondence: suozhongying@sohu.com; Tel.: +86-135-7248-6366

Abstract

To address dimensional reduction for continuous single-label and multi-label data, this paper proposes an improved variable precision rough set method based on maximal consistent blocks. We formulate dimensional reduction as an attribute reduction problem in continuous decision information systems, construct a distance-based tolerance relation, and design a maximal consistent block generation algorithm based on pivoted Bron–Kerbosch maximal clique mining for direct continuous data modeling. We establish a generalized variable precision rough set model, define β-approximation sets and distribution reduction objectives for single- and multi-label scenarios, analyze the stage-wise complexity of the procedure, separating polynomial stages from output-sensitive enumeration stages, and develop a discernibility matrix-based reduction algorithm. Experiments on fourteen public benchmark datasets against seven baselines under Equal-d (fixed feature number) and Nested-d (training-partition tuning) protocols show that the proposed method attains the lowest average rank under Equal-d, where the Friedman test indicates overall differences among methods and Holm-adjusted Wilcoxon comparisons confirm significant advantages over MCLS and the neighborhood rough-set dependency baseline; under Nested-d, the comparison with MCLS remains significant after Holm adjustment. Parameter sensitivity analysis, distance metric comparison, ablation study, and a resource audit further confirm the robustness and feasibility of the method.

Keywords: dimensional reduction; multi-label learning; decision information system; maximal consistent block; variable precision rough set; attribute reduction

MSC: 68T30; 68T37

Check for updates

Academic Editor: Jonathan Blackledge

Received: 3 July 2026
Revised: 18 August 2026
Accepted: 18 August 2026
Published: 19 August 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

In the era of data-driven intelligent systems, fields such as text classification, image annotation, and bioinformatics often deal with complex situations involving "multi-label" associations, in which a single sample is connected to multiple semantic labels. These tasks are referred to as multi-label learning [1]. Compared to single-label learning, multi-label learning more accurately reflects the intrinsic properties of the data. However, the high-dimensional feature space and intricate dependencies among labels often lead to

the "curse of dimensionality," resulting in inefficient model training and diminished generalization performance. Therefore, there is an urgent need for effective dimensionality reduction [2] methods.

Dimensional reduction encompasses two main categories of methods: feature selection and feature extraction. Feature selection is widely used in scenarios requiring explainability, such as medical diagnosis and treatment evaluation, because it preserves the original physical semantics of features [3–5]. However, when applied to continuous multi-label data, most of these methods require a discretization preprocessing step that inevitably causes information loss, and they inadequately capture the collaborative dependencies among multiple labels. Feature extraction methods [6–8] can effectively reduce dimensionality, but their high computational complexity restricts their practicality for high-dimensional continuous data.

To mitigate information loss caused by discretization in the processing of continuous multi-label data, several methods have been developed. González-López et al. [9] proposed a mutual information-based distributed continuous feature selection method. Yin et al. [10] developed a feature selection approach applicable to continuous multi-label data by constructing a robust multi-neighborhood fuzzy β-covering granular space. Ma et al. [11] introduced a discriminative feature selection method based on adaptive graph diffusion, which enhances feature discriminability by exploiting high-order structural information in continuous data. However, these methods do not sufficiently account for the correlations among multiple labels and thus fall short in meeting the requirements of multi-label collaborative classification.

From a theoretical perspective, the problem of dimensional reduction for multi-label continuous data can be naturally framed within the context of a decision information system and reformulated as an attribute reduction problem in rough set theory—that is, identifying the minimal subset of conditional attributes (features) that preserves the decision (label) information. Traditional rough set theory [12], which is based on equivalence relations, is only applicable to discrete symbolic data. To this end, scholars have extended equivalence relations by introducing neighborhood relations, dominance relations, similarity relations, and others [13–15]. It should be noted that these extended models, such as neighborhood rough sets and fuzzy rough set models, are already capable of processing continuous numerical attributes directly, and this line of research has continued to advance in recent years. Xia et al. [16] unified the Pawlak rough set and the neighborhood rough set through granular-ball computing, enabling equivalence-class-based knowledge representation on continuous data; Wang and Zhao [17] introduced a weighted k-nearest neighborhood mechanism to enhance the noise robustness of neighborhood models; Sewwandi et al. [18] performed granule-specific feature selection, which selects local feature subsets for the granules identified by neighborhood rough sets; and Zhang et al. [19] incorporated overlap functions into variable precision fuzzy rough sets, realizing fault tolerance at the fuzzy-membership level for attribute reduction. Despite their diverse mechanisms, these models share a common granulation paradigm: knowledge granules are constructed relative to individual objects, adaptive ball centers, or fuzzy similarity relations. In contrast, the tolerance maximal consistent block adopted in this study requires pairwise consistency among all its member objects together with maximality, which corresponds exactly to a maximal clique in the tolerance graph; it is therefore not centered at any specific object and does not rely on a center–radius representation. The distinction between this study and these models also lies in the reduction objective. Specifically, the reduction objectives of neighborhood rough sets, fuzzy rough sets, and classical variable precision rough sets are typically formulated at the object level, aiming to preserve the positive region or a (fuzzy) dependency degree of the decision; such union-level criteria do not constrain

the class distribution carried by each individual granule, and two attribute subsets with identical positive regions may induce granules with very different decision distributions. In contrast, the objective of this study is distribution-level: it requires the generalized decision distribution of every non-redundant knowledge granule to be preserved, with the parameter β tolerating a controlled proportion of misclassified objects within each granule, which is particularly important for continuous data susceptible to noise. Moreover, formulated via the label inclusion degree, this distribution-level objective extends naturally to multi-label scenarios by preserving the support proportion of each granule for every label, thereby retaining label-correlation information, whereas dependency-degree-based objectives have no direct multi-label counterpart.

The "maximal consistent block," defined as a set of objects satisfying both consistency and maximality, constitutes a non-redundant knowledge granule [20], which provides a suitable granular carrier for the above distribution-level objective. Subsequent studies have extended its application: Zhang et al. [21] integrated it with multi-granularity frameworks; Li et al. [22] applied it to telecommunications fraud detection; and Sun et al. [23] employed it for scale selection in incomplete multi-scale data. Sun et al. [24] proposed a variable precision rough set model based on maximal consistent blocks for incomplete information systems. In the study of maximal consistent block computation, reference [21] proposed an extended method based on the tolerance relation matrix; reference [25] derived maximal consistent blocks by constructing similarity classes and covering structures; and reference [26] introduced a novel construction method that incorporates computational complexity considerations and parallelization improvements. However, all three methods are designed for incomplete discrete data with missing values and rely on the logic of "missing value compatibility," making them unsuitable for continuous data. In the context of continuous data, the binary relation of "feature set inclusion" adopted in [24], together with its reduction objective built upon the "equality or missingness" logic, cannot quantify object differences in continuous spaces, and thus the variable precision model and the reduction objective of [24] are not directly applicable to attribute reduction for continuous data. Nevertheless, the fault tolerance of the variable precision rough set model, achieved by balancing data uncertainty and reduction accuracy through the parameter β, is particularly important for continuous data susceptible to noise interference, as it prevents overfitting caused by rigid equivalence relations. Therefore, it is necessary to improve the method proposed in [24] to enable its application to attribute reduction in continuous data settings.

In light of this, we explore the intrinsic connection between label learning and decision information systems, transforming the problem of single/multi-label dimensional reduction for continuous data into an attribute reduction problem within single/multi-label continuous decision information systems. The main contributions of this work are summarized as follows: (1) To address the limitation that classical rough-set models rely on discretization and that the approximation and reduction objectives of the existing maximal consistent block-based variable precision model [24] are built upon the discrete "equality or missingness" logic, a maximal consistent block calculation framework based on distance tolerance relations is constructed. A complete algorithm flow covering data preprocessing, tolerance matrix construction, and maximal clique mining is designed to realize direct modeling of continuous data. Combined with the fault-tolerant idea of variable precision rough sets, a generalized model with distribution-level reduction objectives suitable for continuous data is established, which improves the adaptability to noisy samples in dimensional reduction tasks. (2) A unified distribution-based reduction objective is constructed for both single-label and multi-label dimensional reduction tasks. For each scenario, β-upper and -lower approximation sets as well as distribution consistent sets are formally defined.

Following the discernibility matrix principle and Boolean function simplification, a targeted reduction solving algorithm that can remove redundant features while effectively preserving classification information and inherent label correlations is designed. (3) Empirical validation is conducted on fourteen public benchmark datasets spanning low-dimensional single-label data, dense multi-label data, sparse high-dimensional data, and continuous real-world tasks. The comparison includes seven feature-selection baselines, including a neighborhood rough-set dependency method. Equal-d and Nested-d evaluations use training-partition parameter selection, and the analysis covers primary performance, cross-dataset statistics, computational resources, distance metric comparison, parameter-selection patterns, and component ablation.

The remainder of this paper is organized as follows. Section 2 introduces basic theoretical preliminaries, including the formal definition of decision information systems, distance-based tolerance relations, and the generalized variable precision rough set model. Section 3 details the maximal consistent block construction algorithm and its stage-wise complexity analysis, and elaborates the dimensional reduction methods for single-label and multi-label continuous data, respectively. Section 4 presents the experimental setup and comprehensive result analysis. Finally, Section 5 concludes the whole work and discusses promising research directions for future improvement.

2. Preliminaries

This section systematically outlines the formal representation of decision information systems, including multi-label scenarios; introduces the concepts of tolerance relation and maximal consistent block adapted for handling continuous data; and details the construction of a generalized variable precision rough set model. This groundwork establishes the theoretical foundation for subsequent attribute reduction algorithms.

2.1. Decision Information System

Definition 1 ([27]). *Let $DS = (U, A \cup D, V, f)$ be a decision information system, where U is the universe of objects, $A \cup D$ is the full set of attributes, $A \cap D = \varnothing$, A is the set of continuous conditional attributes, and D is the set of decision attributes. In the single-label scenario, $|D| = 1$, denoted by $D = \{d_j\}$, which represents the class label of objects. Based on the equivalence relation R_D induced by the decision attributes, the universe U can be partitioned into mutually disjoint decision equivalence classes, denoted as $U/R_D = \{D_1, D_2, \ldots, D_r\}$.*

The representation is then extended to a multi-label decision table to accommodate the complex associations between objects and multiple labels.

Definition 2 ([27]). *Let $DS = (U, A \cup L, V, f)$ be a multi-label continuous decision information system, where $L = \{l_1, l_2, \ldots, l_q\}$ is a nonempty finite set of semantic labels. Each label $l \in L$ corresponds to an independent binary classification task ("has the label" or "does not have the label"), and thus the label value domain is fixed as $V_L = \{0, 1\}$. For a conditional attribute $a \in A$, $f(x, a) \in V_a$ denotes the value of object x on the continuous attribute a; for a label $l \in L$, $f(x, l) = 1$ indicates that object x has label l, while $f(x, l) = 0$ indicates that object x does not have label l.*

2.2. Tolerance Relation and Maximal Consistent Block

This section first introduces the tolerance relation based on the 2-norm [28], and then derives the related definitions of the maximal consistent block.

Definition 3. *For a continuous information system $S = (U, A, V, f)$ and any attribute subset $B \subseteq A$, the distance-based tolerance relation $T_B \subseteq U \times U$ is defined as follows:*

$$T_B = \{(x,y) \in U \times U \mid d_B(x,y) \leq \gamma\} \tag{1}$$

where $\gamma > 0$ denotes the distance threshold, and $d_B(x,y)$ represents the distance between x and y with respect to the attribute subset B:

$$d_B(x,y) = \left(\sum_{a \in B} |f(x,a) - f(y,a)|^2 \right)^{1/2} \tag{2}$$

T_B is a tolerance relation on U, and $T_B(x) = \{y \in U \mid (x,y) \in T_B\}$ denotes the tolerance class of x with respect to the attribute B.

Definition 4. *Let $S = (U, A, V, f)$ be a continuous information system, $B \subseteq A$, and $X \subseteq U$. We say X is consistent with respect to B if $(x,y) \in T_B$ for any $x, y \in X$. Furthermore, if there does not exist a subset $Y \subseteq U$ such that $X \subset Y$, and Y is consistent with respect to B, then X is called a maximal consistent block of B. We denote the set of all maximal consistent blocks determined by B as $\mathcal{C}(B)$, and the set of all maximal consistent blocks of B that contain some object $x \in U$ is denoted as $\mathcal{C}_x(B)$.*

2.3. Generalized Variable Precision Rough Set Model for Continuous Data

This section, based on the ideas presented in reference [24], extends the rough set model based on maximal consistent blocks in continuous information systems [20] to the variable precision rough set model. Before presenting the formal definitions, we fix the notation used throughout this paper: the universe of objects is always denoted by the uppercase letter U, and individual objects by lowercase letters x, y, z; approximation operators are written uniformly in lowercase "$appr$", where an underline denotes a lower approximation and an overline denotes an upper approximation; the superscript $\beta \in (0.5, 1]$ is the precision parameter, and the additional superscripts P and O distinguish the pessimistic and optimistic variants, respectively.

First, we establish the relationship between the maximal consistent blocks in a subsystem and those in the original system.

Property 1. *In a continuous information system $S = (U, A, V, f)$, where $B \subseteq A$, for any $X \in \mathcal{C}(A)$, there exists $Y \in \mathcal{C}(B)$ such that $X \subseteq Y$ holds. Moreover, for any $Y \in \mathcal{C}(B)$, there exists $Y = \bigcup\{X \mid X \in \mathcal{C}_x(A), x \in Y\}$.*

Definition 5. *In a continuous information system $S = (U, A, V, f)$, for any $B \subseteq A$, any $X \subseteq U$, and any $\beta \in (0.5, 1]$, the β-lower approximation of X with respect to B is defined as*

$$\underline{appr}_B^{\beta,P}(X) = \{x \in U \mid \forall Y \in C_x(B), D(X/Y) \geq \beta\} \tag{3}$$

and the β-upper approximation of X with respect to B is defined as

$$\overline{appr}_B^{\beta}(X) = \{x \in U \mid \exists Y \in \mathcal{C}_x(B), D(X/Y) > 1 - \beta\}. \tag{4}$$

Here, D is an inclusion degree measure on $P(U)$ and $|Y|$ represents the cardinality of the set Y, i.e.,

$$D(X/Y) = \begin{cases} \dfrac{|X \cap Y|}{|Y|} & , |Y| \neq 0, \\ 1 & , |Y| = 0. \end{cases} \tag{5}$$

This model is referred to as the pessimistic variable precision rough set model based on maximal consistent blocks.

Definition 6. *In a continuous information system $S = (U, A, V, f)$, for any $B \subseteq A$, any $X \subseteq U$, and any $\beta \in (0.5, 1]$, the β-lower approximation of X with respect to B is defined as*

$$\underline{appr}_B^{\beta,O}(X) = \{x \in U \mid \exists Y \in C_x(B), D(X/Y) \geq \beta\} \tag{6}$$

and the β-upper approximation of X with respect to B is defined as

$$\overline{appr}_B^{\beta}(X) = \{x \in U \mid \exists Y \in C_x(B), D(X/Y) > 1 - \beta\}. \tag{7}$$

This model is referred to as the optimistic variable precision rough set model based on maximal consistent blocks.

Remark 1 (Distinction between the pessimistic and optimistic variants)**.**

The two models differ only in the quantifier governing the lower approximation, and share an identical upper approximation:

1. *The pessimistic lower approximation adopts the universal quantifier $\forall$: an object x is accepted only if every maximal consistent block containing x is included in X with a degree of at least β. The optimistic lower approximation adopts the existential quantifier $\exists$: x is accepted as long as at least one block containing x satisfies the condition. Consequently,*

$$\underline{appr}_B^{\beta,P}(X) \subseteq \underline{appr}_B^{\beta,O}(X) \subseteq \overline{appr}_B^{\beta}(X). \tag{8}$$

2. *Intuitively, the pessimistic variant is conservative: it is robust to noisy or heterogeneous blocks, at the cost of a smaller positive region; the optimistic variant is liberal: it tolerates conflicting evidence from other blocks of the same object and yields a larger positive region.*
3. *The upper approximation uses the existential quantifier in both models because it characterizes possible membership of objects in X, which is consistent with the semantics of the upper approximation in classical variable precision rough set theory.*
4. *Both variants are monotone with respect to β: for $0.5 < \beta_1 \leq \beta_2 \leq 1$, $\underline{appr}_B^{\beta_2,*}(X) \subseteq \underline{appr}_B^{\beta_1,*}(X)$ $(* \in \{P, O\})$, while $\overline{appr}_B^{\beta_1}(X) \subseteq \overline{appr}_B^{\beta_2}(X)$.*

When $\beta = 1$, the optimistic lower approximation and the upper approximation degenerate into the lower and upper approximations of the classical maximal consistent block-based rough set model, respectively:

$$\overline{appr}_B^{1}(X) = \bigcup \{Y \in C(B) \mid Y \cap X \neq \emptyset\}, \tag{9}$$

$$\underline{appr}_B^{1,O}(X) = \bigcup \{Y \in C(B) \mid Y \subseteq X\}. \tag{10}$$

3. Attribute Reduction Algorithms on a Generalized Variable Precision Rough Set Model Based on Maximal Consistent Blocks

This section discusses the differences in attribute reduction algorithms for single-label and multi-label continuous data. Based on the generalized variable precision rough set framework presented in Section 2.3, it first outlines the detailed steps for calculating all maximal consistent blocks based on the tolerance relation, together with a stage-wise complexity analysis that separates the polynomial stages from the output-sensitive enumeration stages of the complete procedure. Then, it defines several concepts of attribute reduction algorithms and the upper and lower distribution reduction methods on this model and verifies them with examples.

3.1. Computation Method for Maximal Consistent Blocks Based on Tolerance Relation

In Section 2.2, we defined the distance-based tolerance relation T_B and the maximal consistent block $\mathcal{C}(B)$. These definitions form the foundation for constructing the subsequent generalized variable precision rough set model. The steps and algorithms for computing maximal consistent blocks are not covered in reference [24]. This section designs and elaborates on a complete algorithm (Algorithm 1) for computing maximal consistent blocks to transform the theory into computable steps and ensure that all maximal consistent blocks can be effectively and systematically extracted from continuous data. Furthermore, a stage-wise complexity analysis of the complete reduction procedure is provided.

Algorithm 1 Algorithm for Constructing Maximal Consistent Blocks.

Input: Continuous information system $S = (U, A, V, f)$, and distance threshold γ.
Output: The set of all maximal consistent blocks $\mathcal{C}(A)$.

{// Step 1: Data preprocessing}
Remove objects with missing values from U;
Let $U = \{x_1, x_2, \ldots, x_n\}$;
Construct feature matrix $M \in \mathbb{R}^{n \times |A|}$, where $M[i, j] = f(x_{i+1}, A_j)$;
{// Step 2: Tolerance relation matrix}
Initialize $R \in \{0, 1\}^{n \times n}$ with all zeros;
for $p = 1$ **to** n **do**
  for $q = p$ **to** n **do**
    Compute Euclidean distance: $d(x_p, x_q) = \sqrt{\sum_{a \in A} \left(f(x_p, a) - f(x_q, a)\right)^2}$;
    if $d(x_p, x_q) \leq \gamma$ **then**
      $R[p-1, q-1] \leftarrow 1$; $R[q-1, p-1] \leftarrow 1$;
    end if
  end for
end for
{// Step 3: Construct tolerance graph}
Construct undirected graph $G = (V, E)$ where $V \leftarrow U$ and $E \leftarrow \{(x_p, x_q) \mid R[p-1, q-1] = 1\}$;
{// Step 4: Maximal consistent block mining}
$\mathcal{C}(A) \leftarrow \emptyset$;
Apply the pivoted Bron–Kerbosch algorithm to G;
Add every returned maximal clique to $\mathcal{C}(A)$;
{// Step 5: Post-processing}
Remove duplicate blocks from $\mathcal{C}(A)$ and sort by object indices;
return $\mathcal{C}(A)$;

Let n denote the number of training objects, s the number of conditional attributes, q the number of decision labels, m the number of maximal consistent blocks, T the total object-block memberships, c the number of retained discernibility clauses, and R the number of retained hitting-set states. Table 1 separates the polynomial stages from the two output-sensitive enumeration stages. The latter stages prevent a polynomial worst-case guarantee for the complete procedure.

3.2. Attribute Reduction Method for Single-Label Continuous Data

For a single-label continuous decision information system, this section presents a specific reduction algorithm and demonstrates its application through an example of scheme evaluation.

Table 1. Stage-wise complexity of the complete maximal consistent block reduction procedure.

Stage	Time Complexity	Working Space	Quantity of Interest
Standardization	$O(ns)$	$O(ns)$	Standardization time
Pairwise distances and tolerance matrix	$O(n^2 s)$	$O(n^2)$	Tolerance-construction time and graph density
Tolerance graph	$O(n^2)$	$O(n^2)$	Number of vertices and edges
Maximal consistent blocks	Output-sensitive; $O(3^{n/3})$ in the worst case	$O(n^2 + T)$	Number and size of maximal blocks; enumeration time
Block statistics and *beta*-approximations	$O(T(s+q))$	$O(m(s+q)+nq)$	Block-statistics and approximation time
Discernibility clauses	$O(m^2 s)$	$O(cs)$	Examined block pairs and retained clauses
Clause simplification	$O(c^2 s)$ in the conservative bound	$O(cs)$	Simplification time and final clause count
Inclusion-minimal hitting sets	Output-sensitive; $O(cRs)$, $R \le 2^s$	$O(R+cs)$	Solver time and reduct-search status
Relevance and redundancy ranking	$O(ns^2)$ after relevance estimation	$O(s^2)$	Ranking time

3.2.1. β-Approximate Set for Single-Labeled Scenarios

Based on the framework outlined in Section 2.3, the "target set X" is concretized as the "decision class D_k" in a single-label decision information system, yielding the β-approximation set definition for single-label scenarios.

Definition 7. *In a single-label decision information system $DS = (U, A \cup D, V, f)$, for any $B \subseteq A$ and for each decision class $D_k \in U/R_D$, we define*

$$\underline{appr}_B^{\beta,P}(D_k) = \{x \in U \mid \forall Y \in \mathcal{C}_x(B), D(D_k/Y) \ge \beta\}, \tag{11}$$

$$\underline{appr}_B^{\beta,O}(D_k) = \{x \in U \mid \exists Y \in \mathcal{C}_x(B), D(D_k/Y) \ge \beta\}, \tag{12}$$

$$\overline{appr}_B^{\beta}(D_k) = \{x \in U \mid \exists Y \in \mathcal{C}_x(B), D(D_k/Y) > 1-\beta\}. \tag{13}$$

as the single-label pessimistic β-lower approximation, optimistic β-lower approximation, and β-upper approximation of X with respect to B, respectively.

3.2.2. Definitions Related to Single-Label Attribute Reduction

Definition 8. *Let $DS = (U, A \cup D, V, f)$ be a single-label decision information system. For any $B \subseteq A$, if $\mathcal{C}(B) = \mathcal{C}(A)$, then B is called a block consistent set. Furthermore, if for any $B' \subset B$, $\mathcal{C}(B') \neq \mathcal{C}(A)$, then B is called a block reduct.*

Definition 9. *Let $DS = (U, A \cup D, V, f)$ be a single-label decision information system, where $B \subseteq A$ and $U/R_D = \{D_1, D_2, \ldots, D_r\}$. For $\beta \in (0.5, 1]$, we denote the following:*

$$OL_B^{\beta} = \left(\underline{appr}_B^{\beta,O}(D_1), \underline{appr}_B^{\beta,O}(D_2), \ldots, \underline{appr}_B^{\beta,O}(D_r)\right), \tag{14}$$

$$PL_B^{\beta} = \left(\underline{appr}_B^{\beta,P}(D_1), \underline{appr}_B^{\beta,P}(D_2), \ldots, \underline{appr}_B^{\beta,P}(D_r)\right), \tag{15}$$

$$H_B^\beta = \left(\overline{appr}_B^\beta(D_1), \overline{appr}_B^\beta(D_2), \ldots, \overline{appr}_B^\beta(D_r)\right). \tag{16}$$

If $H_B^\beta = H_A^\beta$, then B is called a β-upper distribution consistent set of DS. Furthermore, if for any $B' \subset B$, $H_{B'}^\beta \neq H_A^\beta$, then B is called a β-upper distribution reduct of DS. (The definitions of the pessimistic β-lower distribution consistent set/reduct and optimistic β-lower distribution consistent set/reduct are analogous).

Theorem 1. *In a continuous decision information system $DS = (U, A \cup D, V, f)$, every block consistent set is a β-upper (lower) distribution consistent set.*

Proof of Theorem 1. Let B be a block consistent set, i.e., $\mathcal{C}(B) = \mathcal{C}(A)$. For any decision class $D_k \in U/R_D$: Since $\mathcal{C}(B) = \mathcal{C}(A)$, the β-upper approximations satisfy $\overline{appr}_B^\beta(D_k) = \overline{appr}_A^\beta(D_k)$. Thus, the condition for a β-upper distribution consistent set is satisfied. Similarly, the distribution vectors of the pessimistic β-lower approximations and optimistic β-lower approximations remain consistent, so B is also a pessimistic β-lower distribution consistent set and an optimistic β-lower distribution consistent set. □

Definition 10. *Let $DS = (U, A \cup D, V, f)$ be a single-label decision information system, $\beta \in (0.5, 1]$, $U/R_D = \{D_1, D_2, \ldots, D_r\}$, and $\mathcal{C}(A) = \{X_1, X_2, \ldots, X_m\}$. We denote the following:*

$$\begin{aligned} M_A^\beta(X_i) &= \left\{D_k \mid X_i \subseteq \overline{appr}_A^\beta(D_k)\right\}, \\ PG_A^\beta(X_i) &= \left\{D_k \mid X_i \subseteq \underline{appr}_A^{\beta,P}(D_k)\right\}, \\ OG_A^\beta(X_i) &= \left\{D_k \mid X_i \subseteq \underline{appr}_A^{\beta,O}(D_k)\right\}. \end{aligned} \tag{17}$$

$$D_1^\beta(X_i, X_j) = \begin{cases} \{a \in A | d_a(X_i, X_j) > \gamma_a\} & , M_A^\beta(X_i) \neq M_A^\beta(X_j), \\ A & , otherwise. \end{cases}$$

$$D_2^\beta(X_i, X_j) = \begin{cases} \{a \in A | d_a(X_i, X_j) > \gamma_a\} & , OG_A^\beta(X_i) \neq OG_A^\beta(X_j), \\ A & , otherwise. \end{cases} \tag{18}$$

$$D_3^\beta(X_i, X_j) = \begin{cases} \{a \in A | d_a(X_i, X_j) > \gamma_a\} & , PG_A^\beta(X_i) \neq PG_A^\beta(X_j), \\ A & , otherwise. \end{cases}$$

Here, the mean value of the attribute values within a maximal consistent block is used to represent the attribute features of the block. We calculate the inter-block distance as $d_a(X_i, X_j) = \left|stat_a(X_i) - stat_a(X_j)\right|$, $stat_a(X_i) = \frac{1}{|X_i|}\sum_{x \in X_i} f(x, a)$; $\gamma_a = \frac{\gamma}{\sqrt{|A|}}$ (assuming equal contribution of attributes).

We refer to $D_t^\beta(X_i, X_j)$ as the β-upper (lower) distribution discernibility attribute set for the maximal consistent blocks $X_i, X_j \in \mathcal{C}(A)$ over the attribute set A, where $t = 1, 2, 3$. We denote by $\boldsymbol{D}_t^\beta = \left(D_t^\beta(X_i, X_j), i, j \leq m\right)$ the β-upper (lower) distribution discernibility matrix of DS.

Theorem 2. *In a continuous decision information system $DS = (U, A \cup D, V, f)$, where $B \subseteq A$, the following applies:*

1. *B is a β-upper distribution consistent set if and only if for any $X_i, X_j \in \mathcal{C}(A)$, whenever $M_A^\beta(X_i) \neq M_A^\beta(X_j)$, it holds that $B \cap D_1^\beta(X_i, X_j) \neq \emptyset$.*
2. *B is an optimistic β-lower distribution consistent set if and only if for any $X_i, X_j \in \mathcal{C}(A)$, whenever $OG_A^\beta(X_i) \neq OG_A^\beta(X_j)$, it holds that $B \cap D_2^\beta(X_i, X_j) \neq \emptyset$.*
3. *B is a pessimistic β-lower distribution consistent set if and only if for any $X_i, X_j \in \mathcal{C}(A)$, whenever $PG_A^\beta(X_i) \neq PG_A^\beta(X_j)$, it holds that $B \cap D_3^\beta(X_i, X_j) \neq \emptyset$.*

Proof of Theorem 2. We take (1) as an example:

($\Rightarrow$) Let B be a β-upper distribution consistent set, i.e., $H_B^\beta = H_A^\beta$. If $M_A^\beta(X_i) \neq M_A^\beta(X_j)$, then $\{D_k \mid X_i \subseteq \overline{appr}_A^\beta(D_k)\} \neq \{D_l \mid X_j \subseteq \overline{appr}_A^\beta(D_l)\}$. Since $H_B^\beta = H_A^\beta$, we have $\overline{appr}_B^\beta(D_k) = \overline{appr}_A^\beta(D_k)$ for all $D_k \in U/R_D$. It follows that $\{D_k \mid X_i \subseteq \overline{appr}_B^\beta(D_k)\} \neq \{D_l \mid X_j \subseteq \overline{appr}_B^\beta(D_l)\}$. By Property 1, X_i and X_j are also blocks determined by the attribute set B, but not necessarily maximal consistent blocks under B. Thus, there exist maximal consistent blocks $Y_i, Y_j \in \mathcal{C}(B)$ such that $X_i \subseteq Y_i$ and $X_j \subseteq Y_j$, and $M_B^\beta(Y_i) \neq M_B^\beta(Y_j)$. Hence, there exists an attribute $a \in B$ such that $d_a(Y_i, Y_j) > \gamma_a$, which implies $d_a(X_i, X_j) > \gamma_a$. Therefore, $B \cap D_1^\beta(X_i, X_j) \neq \emptyset$.

($\Leftarrow$) Conversely, assume that for all $X_i, X_j \in \mathcal{C}(A)$, if $M_A^\beta(X_i) \neq M_A^\beta(X_j)$, then $B \cap D_1^\beta(X_i, X_j) \neq \emptyset$. Suppose, for contradiction, that $B \cap D_1^\beta(X_i, X_j) = \emptyset$. Then, for all $a \in B$, $d_a(X_i, X_j) \leq \gamma_a$. By Property 1, there exist $Y_i, Y_j \in \mathcal{C}(B)$ such that $X_i \subseteq Y_i$, $X_j \subseteq Y_j$, and $d_a(Y_i, Y_j) \leq \gamma_a$. Thus, there must exist an attribute $a \in A \setminus B$ such that $d_a(X_i, X_j) > \gamma_a$. This implies that there exists $D_k \in U/R_D$ such that $\overline{appr}_B^\beta(D_k) \neq \overline{appr}_A^\beta(D_k)$, contradicting the assumption that B is a β-upper distribution consistent set.

The proofs for (2) and (3) follow analogously. □

In a continuous decision information system $DS = (U, A \cup D, V, f)$, let $\boldsymbol{D}_t^\beta = \left(D_t^\beta(X_i, X_j), i, j \leq m\right)$ denote the β-upper (lower) distribution discernibility matrix of DS, where $A = \{a_1, a_2, \ldots, a_s\}$. We define the following:

$$M_t^\beta = \bigwedge \{ \bigvee \{a_l \mid a_l \in D^\beta(X_i, X_j), l \leq s\} \mid i, j \leq m\}, \quad t = 1, 2, 3. \tag{19}$$

M_t^β is called the β-upper (lower) distribution discernibility formula. The minimal disjunctive normal form of the discernibility formula M_t^β can be expressed as $M_t^\beta = \bigvee_{m=1}^{p} \bigwedge_{i=1}^{q} a_i$, where $q \leq s$. Let $B_m = \{a_i \mid i = 1, \ldots, q\}$; then, similar to the proof of Theorem 3.5 in [29], it can be shown that $\{B_m \mid m = 1, \ldots, p\}$ is the collection of all β-upper (lower) distribution reducts.

Let $\mu_B(Y) = \left(D(D_1/Y), D(D_2/Y), \ldots, D(D_r/Y)\right)$ denote the generalized decision distribution function of the maximal consistent block Y with respect to the attribute set B in a continuous information system.

3.2.3. Single-Label β-Upper Distribution Reduction Algorithm

Based on the discernibility matrix and Boolean function simplification, we take the β-upper distribution as an example to present the single-label β-upper distribution reduction algorithm (Algorithm 2) as follows:

Algorithm 2 Single-labe β-upper Distribution Reduction Algorithm.

Input: Decision information system $DS = (U, A \cup D, V, f)$, distance threshold γ, and tolerance parameter $\beta \in (0.5, 1]$.
Output: All β-upper distribution reducts of DS.

{// Step 1: Compute maximal consistent blocks}
Compute $\mathcal{C}(A)$ using Algorithm 1 with attribute set A;
Let $\mathcal{C}(A) = \{X_1, X_2, \ldots, X_m\}$ and $U/R_D = \{D_1, \ldots, D_r\}$;
{// Step 2: Upper approximations and block label sets}
for $k = 1$ **to** r **do**
  Compute $\overline{appr}_A^{\beta}(D_k)$;
end for
for $i = 1$ **to** m **do**
  $M_A^{\beta}(X_i) \leftarrow \{D_k \mid X_i \subseteq \overline{appr}_A^{\beta}(D_k)\}$;
  Compute $stat_a(X_i) = \frac{1}{|X_i|} \sum_{x \in X_i} f(x, a)$ for all $a \in A$;
end for
{// Step 3: Build β-upper distribution discernibility matrix}
for $i = 1$ **to** m **do**
  for $j = 1$ **to** m **do**
    if $M_A^{\beta}(X_i) \neq M_A^{\beta}(X_j)$ **then**
      $D_1^{\beta}(X_i, X_j) \leftarrow \{a \in A \mid |stat_a(X_i) - stat_a(X_j)| > \gamma_a\}$;
    else
      $D_1^{\beta}(X_i, X_j) \leftarrow A$;
    end if
  end for
end for
{// Step 4: Discernibility formula and reducts}
Construct $M_1^{\beta} = \bigwedge\{\bigvee\{a \mid a \in D_1^{\beta}(X_i, X_j)\} \mid i, j \leq m\}$;
Simplify M_1^{β} to its minimal disjunctive normal form;
return all attribute sets in the minimal form;

3.2.4. Example Calculation

Next, we use the scenario of "UAV platform performance evaluation" to verify the feasibility of the algorithm. This scenario simulates a decision-making process for selecting the optimal model from multiple UAV platforms for a specific mission (e.g., regional reconnaissance, material delivery).

Example 1. *Table 2 shows a UAV platform performance evaluation table* $(U, A \cup D, V, f)$*, where* $U = \{x_1, x_2, \ldots, x_{20}\}$ *is the set of UAV platforms to be evaluated, and* $A = \{a_1, a_2, \ldots, a_5\}$ *is the set of performance indicators. The meanings of* a_1 *to* a_5 *are endurance time, control radius, maximum load, wind resistance rating, and communication rate.* $V \in [0, 3]$ *represents the normalized score for each indicator, where a higher score indicates better performance. We set* $\gamma = 2.5$*.*

In this example, the comprehensive rating rules are as follows:

- *A platform with 5 evaluation scores* ≥ 2 *is rated as "Excellent"* (d_3)*.*
- *A platform with 3–4 evaluation scores* ≥ 2 *is rated as "Good"* (d_2)*.*
- *A platform with 2 or fewer evaluation scores* ≥ 2 *is rated as "Average"* (d_1)*.*

Taking $\beta = 0.7$ *as an example, we calculate its upper distribution reduct.*

Table 2. Evaluation form of projects.

U	a_1	a_2	a_3	a_4	a_5	D
x_1	1.1	1.3	2.2	1.5	2.1	d_1
x_2	3.0	2.9	2.1	2.8	2.5	d_3
x_3	2.3	1.8	0.7	1.5	2.2	d_1
x_4	2.1	1.0	0.6	1.4	1.9	d_1
x_5	2.7	2.0	2.2	2.9	3.0	d_3
x_6	3.0	3.0	2.1	2.5	2.6	d_3
x_7	1.0	2.1	2.9	2.0	2.1	d_2
x_8	1.2	0.9	2.1	2.2	2.1	d_2
x_9	1.0	2.4	2.2	1.1	0.9	d_1
x_{10}	2.9	3.0	1.5	3.0	2.5	d_2
x_{11}	0.7	0.8	1.3	1.4	2.0	d_1
x_{12}	3.0	2.9	1.1	2.1	3.0	d_2
x_{13}	2.2	2.1	2.0	0.6	2.1	d_2
x_{14}	2.0	2.1	2.1	0.9	1.1	d_2
x_{15}	1.1	1.1	3.0	0.6	1.2	d_1
x_{16}	3.0	2.2	2.9	2.9	2.0	d_3
x_{17}	2.0	1.2	3.0	2.2	2.1	d_2
x_{18}	2.2	2.5	1.1	1.2	0.8	d_1
x_{19}	2.1	2.0	1.2	1.4	1.7	d_1
x_{20}	2.8	3.0	2.5	2.1	2.1	d_3

The set of maximal consistent blocks determined by the conditional attribute set A is as follows:

$$\begin{aligned} C(A) = \{&Y_1 = \{x_1, x_3, x_4, x_8, x_{11}, x_{13}, x_{14}, x_{19}\}, \\ &Y_2 = \{x_1, x_3, x_4, x_{13}, x_{14}, x_{18}, x_{19}\}, \\ &Y_3 = \{x_1, x_3, x_5, x_8, x_{17}, x_{19}\}, \ldots, \\ &Y_{17} = \{x_3, x_6, x_{13}, x_{17}, x_{19}, x_{20}\}, \\ &Y_{18} = \{x_5, x_7, x_8, x_{16}, x_{17}, x_{19}\}, \\ &Y_{19} = \{x_6, x_7, x_{13}, x_{17}, x_{19}, x_{20}\}\} \end{aligned} \tag{20}$$

The partition of U induced by the decision attribute D is as follows:

$$\begin{aligned} U/R_D = \{&d_1 = \{x_1, x_3, x_4, x_9, x_{11}, x_{15}, x_{18}, x_{19}\}, \\ &d_2 = \{x_7, x_8, x_{10}, x_{12}, x_{13}, x_{14}, x_{17}\}, \\ &d_3 = \{x_2, x_5, x_6, x_{16}, x_{20}\}\} \end{aligned} \tag{21}$$

The generalized decision distribution functions of the maximal consistent blocks are as follows:

$$\begin{aligned} &\mu_A(Y_1) = (0.6250, 0.3750, 0.0000), \quad \mu_A(Y_2) = (0.7143, 0.2857, 0.0000), \\ &\mu_A(Y_3) = (0.5000, 0.3333, 0.1667), \quad \ldots, \quad \mu_A(Y_{17}) = (0.3333, 0.3333, 0.3333), \\ &\mu_A(Y_{18}) = (0.1667, 0.5000, 0.3333), \quad \mu_A(Y_{19}) = (0.1667, 0.5000, 0.3333) \end{aligned} \tag{22}$$

The 0.7-upper approximations of the decision equivalence classes are as follows:

$$\begin{aligned} &\overline{appr}_A^{0.7}(d_1) = \{x_1, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{17}, x_{18}, x_{19}, x_{20}\}, \\ &\overline{appr}_A^{0.7}(d_2) = \{x_1, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{17}, x_{19}, x_{20}\}, \\ &\overline{appr}_A^{0.7}(d_3) = \{x_1, x_2, x_3, x_5, x_6, x_7, x_8, x_{10}, x_{12}, x_{13}, x_{16}, x_{17}, x_{19}, x_{20}\} \end{aligned} \tag{23}$$

Thus, we have the following:

$$\begin{aligned} &M_A^{0.7}(Y_1) = \{d_1, d_2\}, \quad M_A^{0.7}(Y_2) = \{d_1\}, \\ &M_A^{0.7}(Y_3) = \{d_1, d_2, d_3\}, \quad \ldots, \quad M_A^{0.7}(Y_{17}) = \{d_1, d_2, d_3\}, \\ &M_A^{0.7}(Y_{18}) = \{d_2, d_3\}, \quad M_A^{0.7}(Y_{19}) = \{d_1, d_2, d_3\} \end{aligned} \tag{24}$$

The discernibility attribute sets are as follows:

$$\begin{aligned} D_1^{\beta}(Y_1, Y_{15}) &= \{a_2\}, \\ &\vdots \\ D_1^{\beta}(Y_2, Y_{15}) &= \{a_4\}, \\ D_1^{\beta}(Y_5, Y_{15}) &= \{a_1\}, \\ &\vdots \\ D_1^{\beta}(Y_{10}, Y_{15}) &= \{a_1\}, \\ D_1^{\beta}(Y_{10}, Y_{16}) &= \{a_1\} \end{aligned} \tag{25}$$

Using the formula $M_t^{\beta} = \bigwedge\{\bigvee\{a_l \mid a_l \in D^{\beta}(X_i, X_j), l \leq s\} \mid i, j \leq m\}$, *we perform Boolean function simplification to obtain the* 0.7*-upper distribution discernibility formula for this decision information system:*

$$M^{0.7} = a_1 \wedge a_2 \wedge a_4 \tag{26}$$

Thus, $\{a_1, a_2, a_4\}$ *is the* β*-upper distribution reduct of the UAV platform performance evaluation table.*

Compared with the traditional discernibility matrix based on objects, the proposed method based on maximal consistent blocks compresses the size of the discernibility matrix, further simplifying the calculation process and saving storage space.

3.3. Attribute Reduction Method for Multi-Label Continuous Data

For multi-label continuous decision information systems, this section presents a specific reduction algorithm.

3.3.1. β-Approximate Set for Multi-Labeled Scenarios

Traditional rough set methods were developed primarily for single-label scenarios and are not applicable to continuous data settings with multiple labels per object. In this subsection, based on the general framework introduced in Section 2.3, we concretize the "target set X" as the set of objects that "satisfy the semantics of label l" in the multi-label decision table, i.e., $\{x \in U \mid f(x, l) = 1\}$. This leads to the definition of the β-approximation set in the context of multi-label scenarios.

Definition 11. *In a multi-label decision information system* $DS = (U, A \cup L, V, f)$, *let* $B \subseteq A$ *and* $L = \{l_1, l_2, \ldots, l_q\}$ *be the set of labels. We define*

$$\underline{appr}_B^{\beta,P}(l) = \{x \in U \mid \forall Y \in \mathcal{C}_x(B), D_l(Y) \geq \beta\}, \tag{27}$$

$$\underline{appr}_B^{\beta,O}(l) = \{x \in U \mid \exists Y \in \mathcal{C}_x(B), D_l(Y) \geq \beta\}, \tag{28}$$

$$\overline{appr}_B^{\beta}(l) = \{x \in U \mid \exists Y \in \mathcal{C}_x(B), D_l(Y) > 1 - \beta\}. \tag{29}$$

as the multi-label pessimistic β-lower approximation, optimistic β-lower approximation, and β-upper approximation of X with respect to B, respectively. Here, $D_l(Y)$ is the label inclusion degree, which measures the support proportion of the set Y for label l. It is defined as

$$D_l(Y) = \begin{cases} \dfrac{|\{x \in Y \mid f(x,l) = 1\}|}{|Y|} & , |Y| \neq 0, \\ 1 & , |Y| = 0. \end{cases} \tag{30}$$

3.3.2. Definitions Related to Multi-Label Attribute Reduction

Definition 12. *Let $DS = (U, A \cup L, V, f)$ be a multi-label decision information system. For any $B \subseteq A$, if $\mathcal{C}(B) = \mathcal{C}(A)$, then B is called a block consistent set. Furthermore, if for any $B' \subset B$, $\mathcal{C}(B') \neq \mathcal{C}(A)$, then B is called a block reduct.*

Definition 13. *Let $DS = (U, A \cup L, V, f)$ be a multi-label decision information system, where $B \subseteq A$ and $L = \{l_1, l_2, \ldots, l_q\}$. For $\beta \in (0.5, 1]$, we denote the following:*

$$OL_B^\beta = \left(\underline{appr}_B^{\beta,O}(l_1), \underline{appr}_B^{\beta,O}(l_2), \ldots, \underline{appr}_B^{\beta,O}(l_q)\right), \tag{31}$$

$$PL_B^\beta = \left(\underline{appr}_B^{\beta,P}(l_1), \underline{appr}_B^{\beta,P}(l_2), \ldots, \underline{appr}_B^{\beta,P}(l_q)\right), \tag{32}$$

$$H_B^\beta = \left(\overline{appr}_B^{\beta}(l_1), \overline{appr}_B^{\beta}(l_2), \ldots, \overline{appr}_B^{\beta}(l_q)\right). \tag{33}$$

If $H_B^\beta = H_A^\beta$, then B is called a β-upper distribution consistent set of DS. Furthermore, if for any $B' \subset B$, $H_{B'}^\beta \neq H_A^\beta$, then B is called a β-upper distribution reduct of DS. (The definitions of the pessimistic β-lower distribution consistent set/reduct and optimistic β-lower distribution consistent set/reduct are analogous).

Theorem 3. *In a continuous decision information system $DS = (U, A \cup L, V, f)$, every block consistent set is a β-upper (lower) distribution consistent set.*

Proof of Theorem 3. Similar to the proof of Theorem 1. □

Definition 14. *Let $DS = (U, A \cup L, V, f)$ be a multi-label decision information system, $\beta \in (0.5, 1]$, and $\mathcal{C}(A) = \{X_1, X_2, \ldots, X_m\}$. We denote the following:*

$$\begin{aligned} M_A^\beta(X_i) &= \left\{l \in L \mid X_i \subseteq \overline{appr}_A^\beta(l)\right\}, \\ PG_A^\beta(X_i) &= \left\{l \in L \mid X_i \subseteq \underline{appr}_A^{\beta,P}(l)\right\}, \\ OG_A^\beta(X_i) &= \left\{l \in L \mid X_i \subseteq \underline{appr}_A^{\beta,O}(l)\right\}. \end{aligned} \tag{34}$$

$$D_1^\beta(X_i, X_j) = \begin{cases} \{a \in A | d_a(X_i, X_j) > \gamma_a\} & , M_A^\beta(X_i) \neq M_A^\beta(X_j), \\ A & , otherwise. \end{cases}$$

$$D_2^\beta(X_i, X_j) = \begin{cases} \{a \in A | d_a(X_i, X_j) > \gamma_a\} & , OG_A^\beta(X_i) \neq OG_A^\beta(X_j), \\ A & , otherwise. \end{cases} \tag{35}$$

$$D_3^\beta(X_i, X_j) = \begin{cases} \{a \in A | d_a(X_i, X_j) > \gamma_a\} & , PG_A^\beta(X_i) \neq PG_A^\beta(X_j), \\ A & , otherwise. \end{cases}$$

Here, the definitions of $d_a(X_i, X_j)$, $stat_a(X_i)$, and γ_a are the same as in the single-label case. We refer to $D_t^\beta(X_i, X_j)$ as the β-upper (lower) distribution discernibility attribute set for the maximal consistent blocks $X_i, X_j \in \mathcal{C}(A)$ over the attribute set A, where $t = 1, 2, 3$. We denote by $\boldsymbol{D}_t^\beta = \left(D_t^\beta(X_i, X_j), i, j \leq m\right)$ the β-upper (lower) distribution discernibility matrix of DS.

Theorem 4. *In a continuous decision information system $DS = (U, A \cup L, V, f)$, where $B \subseteq A$, the following applies:*

1. *B is a β-upper distribution consistent set if and only if for any $X_i, X_j \in \mathcal{C}(A)$, whenever $M_A^\beta(X_i) \neq M_A^\beta(X_j)$, it holds that $B \cap D_1^\beta(X_i, X_j) \neq \emptyset$.*
2. *B is an optimistic β-lower distribution consistent set if and only if for any $X_i, X_j \in \mathcal{C}(A)$, whenever $OG_A^\beta(X_i) \neq OG_A^\beta(X_j)$, it holds that $B \cap D_2^\beta(X_i, X_j) \neq \emptyset$.*
3. *B is a pessimistic β-lower distribution consistent set if and only if for any $X_i, X_j \in \mathcal{C}(A)$, whenever $PG_A^\beta(X_i) \neq PG_A^\beta(X_j)$, it holds that $B \cap D_3^\beta(X_i, X_j) \neq \emptyset$.*

Proof of Theorem 4. Similar to the proof of Theorem 2. □

In a multi-label continuous decision information system $DS = (U, A \cup L, V, f)$, let $\boldsymbol{D}_t^\beta = \left(D_t^\beta(X_i, X_j), i, j \leq m\right)$ denote the β-upper (lower) distribution discernibility matrix of DS, where $A = \{a_1, a_2, \ldots, a_s\}$. The definition of M_t^β is consistent with that in the single-label case.

3.3.3. Multi-Label β-Upper Distribution Reduction Algorithm

Based on the discernibility matrix and Boolean function simplification, we take the β-upper distribution as an example to present the multi-label β-upper distribution reduction algorithm (Algorithm 3). The steps are symmetrical to those of the single-label algorithm, with only the "β-class-level approximate calculation" phase replaced by the label-level scenario to accommodate multi-label classification objectives. The steps are as follows:

Remark 2. *The generalized variable precision rough set model and the reduction algorithms presented in this section follow the conceptual skeleton of β-approximations and distribution reduction established in [24], while several essential adaptations are introduced to accommodate continuous data.*

First, the tolerance relation is induced by a metric distance threshold instead of the "equality or missingness" logic, so that consistency among objects is measured directly in the continuous feature space; accordingly, Property 1 is re-established under the metric tolerance relation, and the computation of maximal consistent blocks is transformed into maximal clique mining on the distance-induced tolerance graph (Algorithm 1).

Second, the discernibility between maximal consistent blocks is redefined in a metric form: each block is characterized by the mean values of its member objects on each attribute, and two blocks are regarded as discernible on an attribute only when the difference in their means exceeds the corresponding threshold. This redesign is indispensable for continuous data—since continuous attribute values rarely coincide, the Boolean discernibility logic of [24] would mark almost every attribute as discernible for every pair of blocks, and the resulting discernibility matrix would degenerate and yield only the trivial reduct A.

Third, the framework is extended from single-decision systems to multi-label decision information systems through the label inclusion degree, so that the support proportion of each granule for every label is explicitly preserved during reduction (Definitions 11–14 and Theorem 4), a scenario that the framework of [24] does not cover.

Algorithm 3 Multi-label β-upper Distribution Reduction Algorithm.

Input: Multi-label continuous system $DS = (U, A \cup L, V, f)$, distance threshold γ, parameter $\beta \in (0.5, 1]$.
Output: All β-upper distribution reducts of DS.

{// Step 1: Compute maximal consistent blocks}
Compute $\mathcal{C}(A)$ using Algorithm 1 with attribute set A;
Let $\mathcal{C}(A) = \{X_1, X_2, \ldots, X_m\}$ and $L = \{l_1, \ldots, l_q\}$;
{// Step 2: Label-level upper approximations}
for $j = 1$ **to** q **do**
  Compute $\overline{appr}_A^{\beta}(l_j)$;
end for
for $i = 1$ **to** m **do**
  $M_A^{\beta}(X_i) \leftarrow \{l_j \mid X_i \subseteq \overline{appr}_A^{\beta}(l_j)\}$;
  Compute $stat_a(X_i) = \frac{1}{|X_i|} \sum_{x \in X_i} f(x, a)$ for all $a \in A$;
end for
{// Step 3: Build discernibility matrix}
for $i = 1$ **to** m **do**
  for $j = 1$ **to** m **do**
    if $M_A^{\beta}(X_i) \neq M_A^{\beta}(X_j)$ **then**
      $D_1^{\beta}(X_i, X_j) \leftarrow \{a \in A \mid |stat_a(X_i) - stat_a(X_j)| > \gamma_a\}$;
    else
      $D_1^{\beta}(X_i, X_j) \leftarrow A$;
    end if
  end for
end for
{// Step 4: Discernibility formula and reducts}
Construct $M_1^{\beta} = \bigwedge\{\bigvee\{a \mid a \in D_1^{\beta}(X_i, X_j)\} \mid i, j \leq m\}$;
Simplify M_1^{β} to its minimal disjunctive normal form;
return all attribute sets in the minimal form;

4. Experiments and Analysis

This section evaluates the proposed method on fourteen continuous single-label and multi-label datasets. The experiments cover the data and protocol settings, method comparisons under Equal-d and Nested-d, cross-dataset statistical tests, running time and scalability, distance-metric settings and parameter selection analysis, a comparison of candidate distance metrics, and component ablation.

4.1. Experimental Setup

4.1.1. Datasets

The evaluation corpus comprises five single-label and nine multi-label benchmark datasets. Dry Bean [30], Glass Identification [31], Iris [32], Leaf [33], and Wine [34] are obtained from the UCI Machine Learning Repository; Bibtex [35], Corel5k [36], Delicious [37], Emotions [38], Enron [39], Mediamill [40], Scene [41], TMC2007 [42], and Yeast [43] are obtained from the MULAN benchmark collection. Dataset provenance is documented through the original publication associated with each benchmark. Glass Identification, for which the repository provides no introductory publication, is cited as an electronic dataset resource. Mediamill-10k contains 10,000 instances. Table 3 summarizes the dataset characteristics and primary score. Macro-F1 is used for single-label datasets and Micro-F1 for multi-label datasets. Each dataset uses five outer splits, and parameter and feature count selection is performed by three-fold cross-validation within the corresponding outer training partition.

Table 3. Dataset characteristics.

Dataset	Task	Public Objects	Features	Classes/Labels	Primary Score
Dry Bean	Single-label	13,611	16	7	Macro-F1
Glass	Single-label	214	9	6	Macro-F1
Iris	Single-label	150	4	3	Macro-F1
Leaf	Single-label	340	14	30	Macro-F1
Wine	Single-label	178	13	3	Macro-F1
Bibtex	Multi-label	7395	1836	159	Micro-F1
Corel5k	Multi-label	5000	499	374	Micro-F1
Delicious	Multi-label	16,105	500	983	Micro-F1
Emotions	Multi-label	593	72	6	Micro-F1
Enron	Multi-label	1702	1001	53	Micro-F1
Mediamill-10k	Multi-label	10,000	120	101	Micro-F1
Scene	Multi-label	2407	294	6	Micro-F1
TMC2007-500	Multi-label	28,596	500	22	Micro-F1
Yeast	Multi-label	2417	103	14	Micro-F1

4.1.2. Comparison Methods and Evaluation Metrics

The proposed method in this paper is denoted as Ours, and it is compared with seven baseline methods. These methods encompass mainstream technical approaches such as information theory, multi-criteria decision-making, and manifold learning:

1. The Fast Information-theoretic Mutual-information Feature ranking (FIMF) [44]: An information-theoretic method that optimizes feature scoring by limiting label combination cardinality and reusing entropy calculations for efficient large-scale feature selection.
2. The VIKOR-based Multi-target Feature Selection (VMFS) [45]: A multi-criteria decision-making approach that constructs a decision matrix using cosine similarity and ranks features via the VIKOR method.
3. The Multi-label Feature Selection using Multi-Criteria Decision Making (MFS-MCDM) [46]: A method that employs ridge regression to build feature-label decision matrices and combines entropy weighting with TOPSIS for feature ranking.
4. The manifold-based constraint Laplacian score (MCLS) [47]: A manifold learning-based algorithm that transforms logical labels into numerical ones and constructs scale-invariant similarity matrices to integrate local structural features.
5. The manifold regularization based multi-label feature selection (MDFS) [48]: A manifold regularization approach that constructs low-dimensional embeddings to capture label correlations and uses $l_{2,1}$-norm regularization for sparse feature selection.
6. The Shared latent Structure feature Selection (SSFS) [49]: A graph regularization method based on latent structure sharing that simultaneously models feature spaces and label structures for high-dimensional data.
7. The neighborhood rough-set dependency ranking (NRS-Dep) [28,50]: NRS-Dep denotes the neighborhood rough-set dependency ranking used in this comparison. After min–max scaling within the training partition, each feature is scored by the relative size of its neighborhood positive region. For multi-label data, the positive region uses a label-wise lower-approximation construction. The neighborhood radius is selected from $\{0.10, 0.20, 0.30\}$ within each outer training partition, and dependency ties are resolved by neighborhood discernibility.

All methods are implemented under a unified data splitting, classifier, and evaluation pipeline.

The experiments adopt 5-fold cross-validation, with the k-nearest neighbors as the classifier where $k = 5$. For the single-label datasets, Accuracy and Macro-F1 are reported; for the multi-label datasets, Accuracy, Macro-F1, Micro-F1 and Hamming loss are reported.

The mathematical definitions of all evaluation metrics on the r-th test fold are given as follows. Let I_r denote the set of test samples in the r-th fold, with a total number of n_r samples. For single-label tasks, y_i and $\hat{y}_i$ represent the true label and predicted label of sample i, respectively. For multi-label tasks, $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iL})$ and $\hat{\boldsymbol{y}}_i = (\hat{y}_{i1}, \ldots, \hat{y}_{iL})$ denote the true label vector and predicted label vector of sample i, respectively, where L is the total number of labels. $\mathbb{I}(\cdot)$ is the indicator function, which takes value 1 when the condition inside holds, and 0 otherwise.

The evaluation metrics for single-label tasks are as follows:

1. **Accuracy**: Accuracy represents the proportion of samples with completely correct category prediction:
$$\text{Accuracy}^{(r)} = \frac{1}{n_r} \sum_{i \in I_r} \mathbb{I}(y_i = \hat{y}_i). \tag{36}$$

2. **Macro-F1**: Let $\mathcal{C}$ be the set of all categories. For each category $c \in \mathcal{C}$, the true positive (TP_c), false positive (FP_c) and false negative (FN_c) are defined as
$$TP_c = \sum_{i \in I_r} \mathbb{I}(y_i = c, \hat{y}_i = c), \tag{37}$$
$$FP_c = \sum_{i \in I_r} \mathbb{I}(y_i \neq c, \hat{y}_i = c), \tag{38}$$
$$FN_c = \sum_{i \in I_r} \mathbb{I}(y_i = c, \hat{y}_i \neq c). \tag{39}$$
The F1 score of category c is calculated as
$$F1_c = \begin{cases} \dfrac{2TP_c}{2TP_c + FP_c + FN_c}, & 2TP_c + FP_c + FN_c > 0, \\ 0, & 2TP_c + FP_c + FN_c = 0. \end{cases} \tag{40}$$
Macro-F1 is the unweighted average of F1 scores across all categories:
$$\text{Macro-F1}^{(r)} = \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} F1_c. \tag{41}$$

The evaluation metrics for multi-label tasks are as follows:

1. **Accuracy (Subset Accuracy)**: For multi-label tasks, Accuracy adopts subset accuracy, which means a sample is counted as correct only when all its labels are predicted correctly:
$$\text{Accuracy}^{(r)} = \frac{1}{n_r} \sum_{i \in I_r} \mathbb{I}(\boldsymbol{y}_i = \hat{\boldsymbol{y}}_i) = \frac{1}{n_r} \sum_{i \in I_r} \mathbb{I}(\forall \ell \in \{1, \ldots, L\}, y_{i\ell} = \hat{y}_{i\ell}). \tag{42}$$

2. **Macro-F1**: For each label $\ell \in \{1, \ldots, L\}$, the true positive ($TP_\ell$), false positive ($FP_\ell$) and false negative ($FN_\ell$) are defined as
$$TP_\ell = \sum_{i \in I_r} \mathbb{I}(y_{i\ell} = 1, \hat{y}_{i\ell} = 1), \tag{43}$$
$$FP_\ell = \sum_{i \in I_r} \mathbb{I}(y_{i\ell} = 0, \hat{y}_{i\ell} = 1), \tag{44}$$

$$FN_\ell = \sum_{i \in I_r} \mathbb{I}(y_{i\ell} = 1, \hat{y}_{i\ell} = 0). \tag{45}$$

The F1 score of label ℓ is

$$F1_\ell = \begin{cases} \dfrac{2TP_\ell}{2TP_\ell + FP_\ell + FN_\ell}, & 2TP_\ell + FP_\ell + FN_\ell > 0, \\ 0, & 2TP_\ell + FP_\ell + FN_\ell = 0. \end{cases} \tag{46}$$

Macro-F1 is the unweighted average of F1 scores across all labels:

$$\text{Macro-F1}^{(r)} = \frac{1}{L} \sum_{\ell=1}^{L} F1_\ell. \tag{47}$$

3. **Micro-F1**: Micro-F1 first accumulates TP, FP and FN over all labels, then calculates the overall F1 score:

$$\text{Micro-F1}^{(r)} = \frac{2\sum_{\ell=1}^{L} TP_\ell}{2\sum_{\ell=1}^{L} TP_\ell + \sum_{\ell=1}^{L} FP_\ell + \sum_{\ell=1}^{L} FN_\ell}. \tag{48}$$

If the denominator is 0, the corresponding Micro-F1 value is set to 0.

4. **Hamming Loss**: Hamming loss represents the average prediction error rate per sample and per label:

$$\text{HL}^{(r)} = \frac{1}{n_r L} \sum_{i \in I_r} \sum_{\ell=1}^{L} \mathbb{I}(y_{i\ell} \neq \hat{y}_{i\ell}). \tag{49}$$

For Hamming loss, a lower value indicates better performance; for all other metrics above, a higher value indicates better performance.

To avoid confusion of evaluation metrics across different tasks, the main result ranking and statistical tests adopt primary evaluation metrics: Macro-F1 for single-label tasks and Micro-F1 for multi-label tasks. The complete experimental results of all evaluation metrics are presented in Appendix A.

4.1.3. Parameter and Feature Number Protocol

The density-adaptive ranking profile is used throughout the evaluation. For each outer training partition, β is selected from $\{0.60, 0.70, 0.85\}$ and γ_{quantile} from $\{0.20, 0.40\}$ by three-fold inner cross-validation. Under Equal-d, the feature count is fixed at $\min(\text{kref}, 20)$. Under Nested-d, the candidate set is $\{5, 10, 20, \min(\text{kref}, 20)\}$ after duplicate removal. The selected configuration maximizes the inner lower confidence bound, defined as the mean primary score minus its standard error. Model selection is completed before evaluation of the corresponding outer test partition. Here, γ_{quantile} is the gamma-quantile level used to derive the tolerance threshold γ from the empirical distribution of pairwise Euclidean distances in the current outer training partition.

Table 4 summarizes the modal parameter choices across the five outer training partitions for each dataset under both protocols, along with the full range of observed k values. Figure 1 aggregates the selected β–γ_{quantile} combinations across Equal-d and Nested-d and displays their frequencies for each dataset.

Table 4. Parameter selections within the outer training partitions; each mode is calculated across five outer folds.

Dataset	Protocol	β Mode	γ_q (γ_{quantile}) Mode	k Mode	Observed k Values
Bibtex	Equal-d	$\beta = 0.6$ (5/5)	$\gamma_q = 0.2$ (3/5)	k = 20 (5/5)	20
	Nested-d	$\beta = 0.6$ (4/5)	$\gamma_q = 0.2$ (3/5)	k = 10/20 (2/5)	5; 10; 20
Corel5k	Equal-d	$\beta = 0.6/0.7$ (2/5)	$\gamma_q = 0.2$ (3/5)	k = 20 (5/5)	20
	Nested-d	$\beta = 0.7/0.85$ (2/5)	$\gamma_q = 0.2$ (4/5)	k = 20 (3/5)	10; 20
Delicious	Equal-d	$\beta = 0.6$ (5/5)	$\gamma_q = 0.2$ (3/5)	k = 20 (5/5)	20
	Nested-d	$\beta = 0.6$ (4/5)	$\gamma_q = 0.2$ (4/5)	k = 10 (4/5)	10; 20
Dry Bean	Equal-d	$\beta = 0.85$ (3/5)	$\gamma_q = 0.2$ (3/5)	k = 8 (5/5)	8
	Nested-d	$\beta = 0.6$ (5/5)	$\gamma_q = 0.2$ (4/5)	k = 5 (3/5)	5; 16
Emotions	Equal-d	$\beta = 0.85$ (3/5)	$\gamma_q = 0.2$ (3/5)	k = 4 (5/5)	4
	Nested-d	$\beta = 0.6$ (4/5)	$\gamma_q = 0.2$ (4/5)	k = 20 (2/5)	4; 5; 10; 20
Enron	Equal-d	$\beta = 0.7$ (3/5)	$\gamma_q = 0.2$ (3/5)	k = 20 (5/5)	20
	Nested-d	$\beta = 0.6/0.85$ (2/5)	$\gamma_q = 0.2$ (3/5)	k = 5 (3/5)	5; 10; 20
Glass	Equal-d	$\beta = 0.7$ (3/5)	$\gamma_q = 0.2$ (3/5)	k = 5 (5/5)	5
	Nested-d	$\beta = 0.6$ (3/5)	$\gamma_q = 0.2$ (4/5)	k = 5 (3/5)	5; 9
Iris	Equal-d	$\beta = 0.6$ (5/5)	$\gamma_q = 0.2$ (5/5)	k = 3 (5/5)	3
	Nested-d	$\beta = 0.6$ (5/5)	$\gamma_q = 0.2$ (5/5)	k = 3 (5/5)	3
Leaf	Equal-d	$\beta = 0.6/0.85$ (2/5)	$\gamma_q = 0.2$ (4/5)	k = 8 (5/5)	8
	Nested-d	$\beta = 0.6$ (3/5)	$\gamma_q = 0.2$ (3/5)	k = 8/10 (2/5)	8; 10; 14
Mediamill-10k	Equal-d	$\beta = 0.6/0.85$ (2/5)	$\gamma_q = 0.2$ (3/5)	k = 20 (5/5)	20
	Nested-d	$\beta = 0.7$ (3/5)	$\gamma_q = 0.2$ (4/5)	k = 20 (3/5)	10; 20
Scene	Equal-d	$\beta = 0.85$ (3/5)	$\gamma_q = 0.2$ (3/5)	k = 20 (5/5)	20
	Nested-d	$\beta = 0.7/0.85$ (2/5)	$\gamma_q = 0.4$ (3/5)	k = 20 (3/5)	10; 20
TMC2007-500	Equal-d	$\beta = 0.6/0.85$ (2/5)	$\gamma_q = 0.2$ (3/5)	k = 20 (5/5)	20
	Nested-d	$\beta = 0.6$ (3/5)	$\gamma_q = 0.4$ (3/5)	k = 20 (3/5)	10; 20
Wine	Equal-d	$\beta = 0.6$ (3/5)	$\gamma_q = 0.4$ (3/5)	k = 7 (5/5)	7
	Nested-d	$\beta = 0.6$ (4/5)	$\gamma_q = 0.2$ (4/5)	k = 10 (4/5)	10; 13
Yeast	Equal-d	$\beta = 0.6$ (3/5)	$\gamma_q = 0.2$ (4/5)	k = 20 (5/5)	20
	Nested-d	$\beta = 0.6$ (3/5)	$\gamma_q = 0.2$ (4/5)	k = 20 (4/5)	5; 20

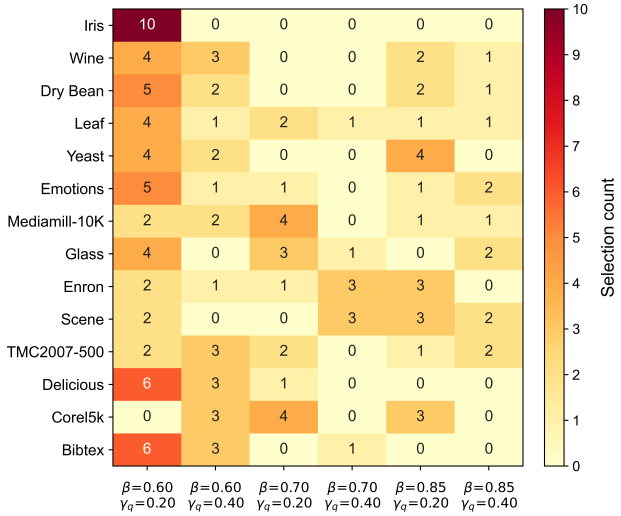


Figure 1. Frequency of β and γ_{quantile} selections within the outer training partitions.

4.2. Main Experimental Results

Tables 5 and 6 report Macro-F1 on the five single-label datasets under Equal-d and Nested-d, respectively; Tables 7 and 8 report Micro-F1 on the nine multi-label datasets. Bold underlining marks the largest score in each dataset row, and italic underlining marks the second-largest score. All scores are means over five outer-test partitions, with parameters and feature counts selected only within the corresponding outer training partition. Figure 2 displays the primary F1 scores across the fourteen evaluation datasets.

Table 5. Equal-d results on the five single-label datasets (Macro-F1). Bold underlining = best; italic underlining = second best.

Dataset	Ours	FIMF	VMFS	MFS-MCDM	MCLS	MDFS	SSFS	NRS-Dep
Dry Bean	*0.8801*	0.8683	0.8745	**0.8865**	0.8760	0.8490	0.8683	0.8567
Glass	**0.4484**	0.3721	0.4357	0.3976	0.3971	*0.4455*	0.3829	0.4348
Iris	**0.9831**	0.8612	**0.9831**	**0.9831**	0.9475	*0.9661*	0.9323	**0.9831**
Leaf	**0.5644**	0.5205	0.5316	0.5406	0.5376	0.4198	0.5174	*0.5435*
Wine	0.9683	0.9683	0.9571	**0.9881**	0.8937	0.9474	0.9461	*0.9779*

Table 6. Nested-d results on the five single-label datasets (Macro-F1). Bold underlining = best; italic underlining = second best.

Dataset	Ours	FIMF	VMFS	MFS-MCDM	MCLS	MDFS	SSFS	NRS-Dep
Dry Bean	0.8447	**0.8687**	0.8468	0.8561	*0.8668*	0.8174	**0.8687**	0.8461
Glass	0.4234	0.4314	0.4126	**0.4455**	0.3980	*0.4428*	0.4156	0.4420
Iris	**0.9831**	0.8967	**0.9831**	**0.9831**	*0.9475*	*0.9475*	0.9323	**0.9831**
Leaf	*0.5672*	0.5409	0.5061	**0.5706**	0.5462	0.5462	0.5473	0.5393
Wine	*0.9683*	0.9460	*0.9683*	*0.9683*	0.9460	0.9571	0.9548	**0.9881**

Table 7. Equal-d results on the nine multi-label datasets (Micro-F1). Bold underlining = best; italic underlining = second best.

Dataset	Ours	FIMF	VMFS	MFS-MCDM	MCLS	MDFS	SSFS	NRS-Dep
Bibtex	0.0299	0.0538	**0.1208**	0.0350	0.0000	0.0438	*0.0971*	0.0059
Corel5k	*0.0332*	0.0202	0.0219	0.0268	0.0172	0.0312	**0.0452**	0.0049
Delicious	0.1035	0.0733	0.0665	*0.1058*	0.0505	0.0898	**0.1153**	0.0346
Emotions	0.4934	0.4481	*0.5022*	**0.5050**	0.4595	0.4490	0.4280	0.4934
Enron	*0.4396*	**0.4444**	0.4095	0.4077	0.4292	0.3690	0.4290	0.4264
Mediamill-10k	*0.4845*	**0.4943**	0.4586	0.4639	0.4519	0.4741	0.4766	0.4784
Scene	*0.3791*	0.3619	0.3780	0.3281	0.3486	**0.4885**	0.3222	0.3216
TMC2007-500	0.3579	0.3786	0.3331	*0.4211*	0.3013	**0.4327**	0.3575	0.2678
Yeast	*0.5383*	0.5071	**0.5602**	0.5364	0.5290	0.4923	0.5042	0.5239

Table 8. Nested-d results on the nine multi-label datasets (Micro-F1). Bold underlining = best; italic underlining = second best.

Dataset	Ours	FIMF	VMFS	MFS-MCDM	MCLS	MDFS	SSFS	NRS-Dep
Bibtex	0.0454	0.0552	**0.1261**	0.0454	0.0000	0.0438	*0.0971*	0.0059
Corel5k	*0.0384*	0.0124	0.0045	0.0173	0.0172	**0.0473**	0.0320	0.0319
Delicious	**0.1291**	0.0610	0.0658	0.1121	0.0505	0.0927	*0.1189*	0.0174
Emotions	0.5588	0.4867	**0.5965**	*0.5652*	0.5323	0.5210	0.4701	0.5583
Enron	0.4386	**0.4415**	0.4079	0.3934	*0.4408*	0.3841	0.4251	0.4282
Mediamill-10k	*0.4758*	**0.4934**	0.4559	0.4677	0.4462	0.4681	0.4737	0.4516
Scene	0.3442	*0.3744*	0.3553	0.2681	0.3152	**0.4952**	0.3222	0.3622
TMC2007-500	0.3360	*0.3910*	0.3416	0.3713	0.3013	**0.3930**	0.3758	0.2704
Yeast	*0.5309*	0.5187	**0.5515**	0.5170	0.5179	0.5074	0.5067	0.5239

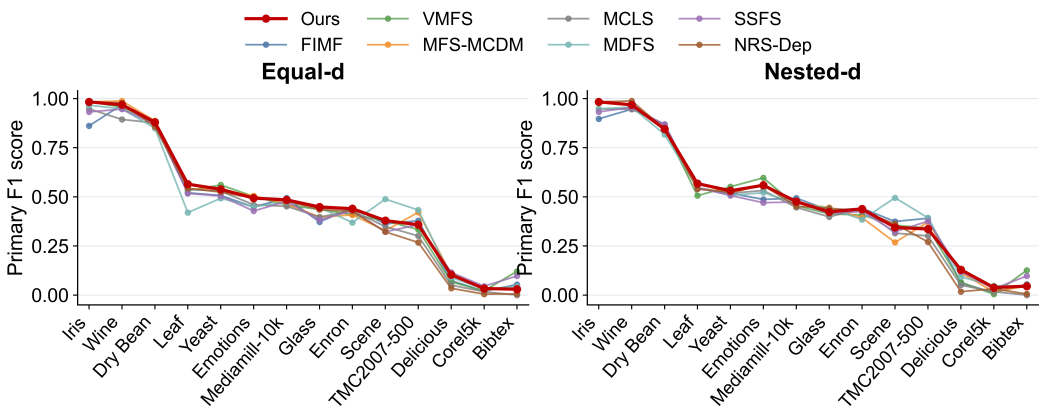


Figure 2. Primary F1 scores across the fourteen evaluation datasets. Macro-F1 is shown for the five single-label datasets and Micro-F1 for the nine multi-label datasets.

4.3. Statistical Significance Test

Tables 9–11 summarize the cross-dataset rank and statistical analyses. Lower average ranks indicate better performance. Figure 3 visually compares the average rank of all methods under both evaluation protocols. Under Equal-d, Ours achieves the lowest average rank (2.643), and the Friedman test indicates an overall difference among methods ($\chi^2 = 17.861$, $p = 0.01261$). Holm-adjusted Wilcoxon comparisons are significant for Ours–MCLS ($p = 0.000854$; 14/0/0 win/tie/loss) and Ours–NRS-Dep ($p = 0.0232$; 11/1/2). Under Nested-d, Ours has an average rank of 3.429, and the Ours–MCLS comparison is significant after Holm adjustment ($p = 0.0120$; 12/0/2).

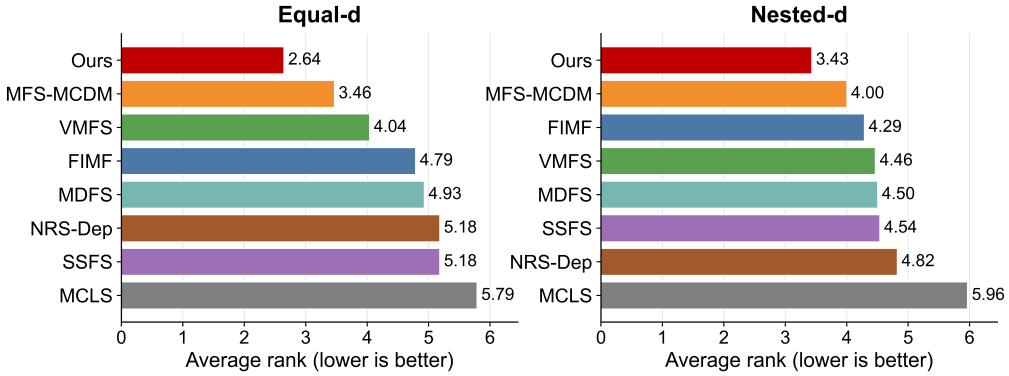


Figure 3. Average ranks across the fourteen evaluation datasets under the Equal-d and Nested-d protocols.

Table 9. Average ranks under the Equal-d and Nested-d protocols. Values in the table that are bold and underlined indicate the best results; values in italics and underlined indicate the second-best results.

Method	Equal-d Average Rank	Nested-d Average Rank
Ours	**2.643**	**3.429**
FIMF	4.786	4.286
VMFS	4.036	4.464
MFS-MCDM	*3.464*	*4.000*
MCLS	5.786	5.964
MDFS	4.929	4.500
SSFS	5.179	4.536
NRS-Dep	5.179	4.821

Table 10. Friedman test results.

Protocol	Friedman χ^2	df	p
Equal-d	17.861	7	0.01261
Nested-d	8.761	7	0.2703

Table 11. Planned two-sided Wilcoxon signed-rank comparisons with Holm adjustment; Δ = Ours − comparator.

Protocol	Comparator	Mean Δ	W/T/L	Raw p	Holm p	Holm < 0.05
Equal-d	FIMF	0.0237	9/1/4	0.04445	0.2222	No
Equal-d	VMFS	0.0051	10/1/3	0.1025	0.314	No
Equal-d	MFS-MCDM	0.0056	7/1/6	0.4896	0.4896	No
Equal-d	MCLS	0.0332	14/0/0	0.0001221	0.0008545	Yes
Equal-d	MDFS	0.0147	11/0/3	0.1726	0.3452	No
Equal-d	SSFS	0.0201	11/0/3	0.07849	0.314	No
Equal-d	NRS-Dep	0.0251	11/1/2	0.003863	0.02318	Yes
Nested-d	FIMF	0.0118	7/0/7	0.4263	1	No
Nested-d	VMFS	0.0044	6/2/6	0.753	1	No
Nested-d	MFS-MCDM	0.0073	6/3/5	0.5069	1	No
Nested-d	MCLS	0.0256	12/0/2	0.001709	0.01196	Yes
Nested-d	MDFS	0.0015	10/0/4	0.2958	1	No
Nested-d	SSFS	0.0103	11/0/3	0.2412	1	No
Nested-d	NRS-Dep	0.0168	9/1/4	0.1164	0.6982	No

4.4. Running Time and Scalability

Table 12 summarizes the observed computational range over five outer training partitions.

Table 12. Observed resource audit for the full ranking profile; entries are minima–maxima over five outer training partitions.

Dataset	MCBs	Clique Time (s)	Fit Time (s)	Peak Memory (MiB)
Bibtex	55–75	0.00–0.08	0.51–1.18	316–323
Corel5k	52–80	0.00–0.01	0.07–0.18	225–226
Delicious	46–62	0.00–0.01	0.15–0.32	467–469
Dry Bean	112–191	0.00–0.01	0.09–0.40	193–196
Emotions	228–808	0.00–0.02	0.62–3.33	219–238
Enron	50–71	0.00–0.01	0.51–1.05	240–292
Glass	79–174	0.01	0.14–0.18	186–188
Iris	33–40	0.00	0.03	184–185
Leaf	834–2411	0.00–0.14	2.66–22.12	232–383
Mediamill-10k	179–242	0.00–0.01	0.63–7.52	296–322
Scene	254–462	0.00–0.02	1.21–3.76	253–289
TMC2007-500	50–89	0.00–0.01	0.13–0.24	277–280
Wine	91–183	0.00–0.01	0.06–0.20	183–190
Yeast	423–3000	0.01–0.03	3.59–5.39	248–421

Across the fourteen datasets, the observed number of maximal consistent blocks ranged from 33–40 on Iris to 834–2411 on Leaf; Yeast spanned 423–3000. Leaf and Yeast showed the widest fold-to-fold ranges and also required longer fitting times than most datasets.

Clique enumeration remained short, with a maximum observed time of 0.14 s, whereas total fitting time ranged from 0.03 s on Iris to 22.12 s on Leaf. The gap between clique time and total fitting time indicates that the subsequent construction of block statistics,

discernibility clauses, and inclusion-minimal hitting sets accounts for a substantial part of the computational cost. Peak memory ranged from 183 to 469 MiB, and the largest memory values did not occur on the datasets with the largest numbers of maximal consistent blocks, indicating that memory use depends on the complete fitting pipeline rather than block enumeration alone.

4.5. Metric Settings and Parameter Sensitivity Analysis

All numerical features are standardized using statistics computed from the corresponding training partition, and Euclidean distance is used to construct the tolerance relation for every dataset and outer split. The same metric is retained throughout inner selection and outer evaluation, so β, γ_{quantile}, and the feature count are evaluated under a common geometry.

For each outer training partition, the tolerance threshold γ is set to the empirical γ_{quantile}-quantile of its pairwise Euclidean distances. The quantile rule adapts the distance scale to the current training data while preserving the same prespecified construction across datasets. A lower γ_{quantile} produces a more local tolerance relation because it selects a smaller quantile of the training-partition distance distribution.

Statistics and labels from the outer test partition are excluded from standardization, threshold construction, parameter selection, and feature count selection. Fixing the distance metric and restricting all data-dependent quantities to the training partition isolates the effects of β, γ_{quantile}, and k and keeps the comparison between Equal-d and Nested-d consistent.

Under this configuration, we further examine how the three key parameters are actually selected by the inner-validation procedure across datasets and protocols.

β and γ are critical parameters that affect the approximation relations of the model, the structure of the maximal consistent blocks, and the final reduction results.

The selected values in Table 4 and Figure 1 are obtained independently within each outer training partition. Across the twenty-eight dataset–protocol settings, $\gamma_{\text{quantile}} = 0.20$ is the modal choice in twenty-five settings, while the selected β values span the complete candidate set. Under Nested-d, the selected feature count varies across outer partitions on thirteen datasets and remains $k = 3$ on every Iris partition. The observed frequencies characterize dataset-specific and partition-specific selection under the prespecified inner-validation rule. Because $\gamma_{\text{quantile}} = 0.20$ is the lower candidate quantile, its predominance corresponds to a more local tolerance relation than $\gamma_{\text{quantile}} = 0.40$ within the same training partition. These frequencies report inner-validation choices and contain no information from the corresponding outer test partitions.

4.6. Distance Metric Comparison

To verify the rationality of the selected distance metric, this section compares three distance metrics, namely, Euclidean, Manhattan, and cosine, under the same reduction framework.

All three distance metrics are evaluated using $\beta = 0.70$ and the quantile parameter $\gamma_{\text{quantile}} = 0.20$, the same Equal-d feature number, identical five-fold partitions, and the same evaluation population. The distance definition is the sole varying factor; the outer test folds are excluded from configuration decisions. Table 13 reports Macro-F1 for the single-label datasets and Micro-F1 for the multi-label datasets.

Table 13. Comparison of different distance metrics.

Dataset	Euclidean	Manhattan	Cosine
Iris	**0.9831**	**0.9831**	*0.9475*
Wine	**0.9888**	0.9562	*0.9563*
Glass	**0.4164**	*0.4162*	0.4036
Leaf	**0.5818**	*0.5669*	0.5535
Dry Bean	**0.8904**	*0.8756*	0.8482
Emotions	0.5361	*0.5437*	**0.5542**
Scene	0.4346	**0.4822**	*0.4673*
Yeast	**0.5519**	*0.4994*	0.4973
Enron	**0.4473**	0.3415	*0.4023*
Bibtex	0.0299	**0.0407**	*0.0393*
Corel5k	**0.0283**	0.0273	*0.0275*
Delicious	*0.1028*	**0.1120**	0.0996
TMC2007-500	**0.3725**	0.3426	*0.3520*
Mediamill-10k	**0.4914**	*0.4840*	0.4749

Note: Values are means over five outer test folds. Higher values indicate better performance. Bold underlined and italic underlined values denote the best and second-best results, respectively.

The results demonstrate that no single distance metric achieves optimal performance on all datasets. Euclidean distance obtained the lowest average rank of 1.5357 and achieved the highest or tied-highest mean score on 10 of the 14 datasets. Manhattan and cosine distances obtained average ranks of 2.0357 and 2.4286, with the highest or tied-highest scores on four datasets and one dataset, respectively. Euclidean distance outperformed Manhattan on nine datasets, tied on one, and underperformed on four. It outperformed cosine on 11 datasets and underperformed on three. The lowest average rank and the broadest coverage of best-performing datasets show that Euclidean distance provides the most stable aggregate performance across the evaluated single-label and multi-label datasets. Euclidean distance is therefore adopted as the default setting in the main experiments. This comparison indicates that the distance metric affects the tolerance relation and the structure of the maximal consistent blocks, and the distance form can be further adjusted according to data types in practical applications.

4.7. Ablation Study

The ablation study examines the structural and relevance terms in the ranking profile used by the proposed method. The Full profile combines maximal-block structural evidence, label relevance, the soft-reduct membership contribution, and redundancy control. The structure-removed profile suppresses the maximal-block structural terms and the soft-reduct membership contribution while retaining label relevance and redundancy control. The relevance-removed profile sets the label-relevance contribution to zero while retaining the structural terms, the soft-reduct membership contribution, and redundancy control. The three profiles use identical outer folds, candidate budgets, classifiers, and evaluation metrics. Parameters and feature counts are selected within the corresponding outer training partition for each profile. Macro-F1 is used for the five single-label datasets, and Micro-F1 is used for the nine multi-label datasets.

Table 14 reports the primary F1 scores under Equal-d and Nested-d, Table 15 summarizes the paired dataset-level contrasts, and Figure 4 displays the score differences between the Full profile and each ablated profile.

Table 14. Component ablation on the evaluation datasets. Within each protocol triplet, bold underlining = best; italic underlining = second best.

Dataset	Eq Full	Eq −Struct.	Eq −Rel.	Nest Full	Nest −Struct.	Nest −Rel.
Bibtex	0.0299	*0.0407*	**0.0756**	0.0454	**0.0507**	*0.0468*
Corel5k	**0.0332**	*0.0275*	0.0152	**0.0384**	*0.0235*	0.0122
Delicious	*0.1035*	**0.1111**	0.0997	**0.1291**	*0.1290*	0.1023
Dry Bean	**0.8801**	*0.8588*	0.7991	0.8447	**0.8757**	*0.8517*
Emotions	**0.4934**	*0.4852*	0.4825	0.5588	**0.6002**	*0.5764*
Enron	**0.4396**	0.2787	*0.4146*	**0.4386**	0.3247	*0.4146*
Glass	*0.4484*	**0.4754**	0.4079	*0.4234*	**0.4660**	0.4195
Iris	**0.9831**	**0.9831**	*0.8764*	**0.9831**	**0.9831**	*0.9120*
Leaf	**0.5644**	0.5164	*0.5177*	0.5672	**0.5739**	*0.5677*
Mediamill-10k	*0.4845*	0.4763	**0.5041**	*0.4758*	0.4652	**0.4958**
Scene	0.3791	**0.4826**	*0.4090*	0.3442	**0.4436**	*0.4090*
TMC2007-500	0.3579	**0.3881**	*0.3714*	0.3360	**0.4099**	*0.3371*
Wine	**0.9683**	*0.9460*	0.9342	**0.9683**	0.9468	*0.9667*
Yeast	**0.5383**	*0.5357*	0.5009	*0.5309*	**0.5487**	0.5252

Table 15. Paired component contrasts; Δ = Full − ablated profile.

Protocol	Contrast	Mean Δ	W/T/L	Two-Sided p
Equal-d	Full vs. structure-removed	0.0070	8/1/5	0.6602
Equal-d	Full vs. relevance-removed	0.0211	10/0/4	0.104
Nested-d	Full vs. structure-removed	−0.0112	5/1/8	0.2856
Nested-d	Full vs. relevance-removed	0.0033	7/0/7	0.583

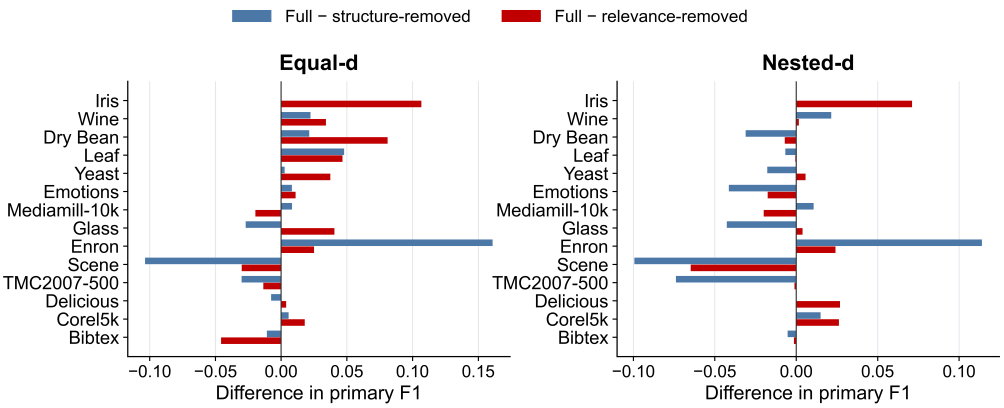


Figure 4. Component-ablation score differences across the evaluated datasets.

The component effects vary across datasets. Under Equal-d, the full profile records 8/1/5 wins/ties/losses against the structure-removed profile and 10/0/4 against the relevance-removed profile. Under Nested-d, the corresponding counts are 5/1/8 and 7/0/7. The positive mean differences for both Equal-d contrasts and for the Nested-d relevance contrast indicate an average benefit from the corresponding complete profile terms within those comparisons. The Nested-d structure contrast varies in the opposite direction on more datasets. These results characterize dataset-dependent contributions of structural evidence and label relevance without implying a universal component effect.

Across the fourteen datasets, Ours obtains the lowest average rank under Equal-d. The Equal-d Friedman test detects an overall method difference, and Holm-adjusted comparisons are significant against MCLS and NRS-Dep. Under Nested-d, the Holm-adjusted comparison with MCLS is significant. The parameter selections in Table 4 are confined to

the outer training partitions, and the resource audit reports the observed computational range of maximal-block and inclusion-minimal hitting-set enumeration.

5. Conclusions

Dimensional reduction for continuous single-label and multi-label data is a critical yet challenging task in pattern recognition and machine learning. Targeting the shortcomings of existing rough set-based reduction methods in continuous data processing, this paper proposes a novel dimensional reduction method based on maximal consistent blocks and variable precision rough sets. The proposed method constructs maximal consistent blocks directly from continuous data via distance tolerance relations, avoiding information loss caused by discretization preprocessing. By introducing the variable precision approximation mechanism, the model achieves favorable fault tolerance for noisy samples. Furthermore, unified distribution-based reduction objectives and corresponding discernibility matrix solving algorithms are established for both single-label and multi-label scenarios. Across the fourteen evaluated datasets, Ours attained the lowest average rank under Equal-d, where the Friedman test indicated an overall difference among methods and the Holm-adjusted comparisons with MCLS and NRS-Dep were significant. Under Nested-d, the planned comparison with MCLS remained significant after Holm adjustment. Maximal-block enumeration and inclusion-minimal hitting-set search constitute the output-sensitive stages of the procedure; the resource audit in Section 4.4 reports their observed computational range.

For future work, we will carry out research from four aspects. First, we will extend the current framework to handle continuous multi-label data with missing labels to improve adaptability to real-world incomplete scenarios. Second, we will explore lightweight optimization strategies for the proposed algorithm to further improve computational efficiency for large-scale continuous dimensional reduction tasks. Specifically, since maximal consistent block enumeration is exponential in the worst case, we plan to control its practical cost from three directions: (i) regulating the density of the tolerance graph through the quantile-based distance threshold γ so that degenerate dense graphs—which are the main source of the exponential blow-up—are avoided in practice; (ii) adopting pivot-optimized Bron–Kerbosch enumeration with degeneracy ordering, together with pruning and parallelization, so that the enumeration cost remains output-sensitive, i.e., proportional to the number of blocks actually generated rather than to the worst-case bound; and (iii) developing approximate or sampling-based block generation schemes that trade a bounded loss of blocks for substantial speed-ups on massive data. Third, we will extend the model to dynamic incremental continuous multi-label data and develop an incremental dimensional reduction algorithm that avoids full recalculation when samples or features are updated and improves the applicability of the method in dynamic data stream scenarios. In this setting, maximal consistent blocks will be maintained incrementally—only the blocks affected by the updated objects are recomputed and merged—which directly alleviates the enumeration overhead discussed above and provides a natural implementation path for the lightweight strategies in the second aspect. Fourth, we will explore the application of the proposed framework to data preprocessing and feature screening in intelligent systems, such as heterogeneous vehicle platooning and multi-agent collaborative control [51], where high-dimensional continuous monitoring data must be compressed online under strict time and resource budgets. Combining the granularity reduction ability of the proposed rough set method with predefined-time performance constraints and adaptive event-triggered scheduling mechanisms in such systems may further improve the efficiency of downstream decision and control modules, which is worthy of in-depth exploration and verification in follow-up studies.

Author Contributions: Methodology, S.C. and Z.S.; writing—original draft preparation, S.C.; validation, S.C. and Y.K.; writing—review and editing, Z.S.; data curation, Y.K.; supervision, Z.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are openly available in Github at https://github.com/1933211129/LDR (accessed on 17 August 2026).

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Complete Experimental Results of All Evaluation Metrics

The following tables report complete outcomes for the eight-method comparison. Ours denotes the full ranking profile, and the values follow the Equal-d and Nested-d protocols used in the main comparison.

Table A1. Complete all-metric results for single-label datasets.

Dataset	Protocol	Metric	Ours	FIMF	VMFS	MFS-MCDM	MCLS	MDFS	SSFS	NRS-Dep
Dry Bean	Equal-d	Accuracy	0.8667	*0.8750*	0.8667	**0.8833**	0.8667	0.8583	*0.8750*	0.8583
		Macro-F1	*0.8801*	0.8683	0.8745	**0.8865**	0.8760	0.8490	0.8683	0.8567
	Nested-d	Accuracy	0.8500	**0.8750**	0.8417	*0.8667*	0.8583	0.8250	**0.8750**	0.8417
		Macro-F1	0.8447	**0.8687**	0.8468	0.8561	*0.8668*	0.8174	**0.8687**	0.8461
Glass	Equal-d	Accuracy	**0.6583**	0.6333	*0.6500*	0.6417	**0.6583**	0.6333	0.6000	*0.6500*
		Macro-F1	**0.4484**	0.3721	0.4357	0.3976	0.3971	*0.4455*	0.3829	0.4348
	Nested-d	Accuracy	0.6417	*0.6833*	0.6500	0.6667	0.6333	0.6667	0.6583	**0.7000**
		Macro-F1	0.4234	0.4314	0.4126	**0.4455**	0.3980	*0.4428*	0.4156	0.4420
Iris	Equal-d	Accuracy	**0.9833**	0.8667	**0.9833**	**0.9833**	0.9500	*0.9667*	0.9333	**0.9833**
		Macro-F1	**0.9831**	0.8612	**0.9831**	**0.9831**	0.9475	*0.9661*	0.9323	**0.9831**
	Nested-d	Accuracy	**0.9833**	0.9000	**0.9833**	**0.9833**	*0.9500*	*0.9500*	0.9333	**0.9833**
		Macro-F1	**0.9831**	0.8967	**0.9831**	**0.9831**	*0.9475*	*0.9475*	0.9323	**0.9831**
Leaf	Equal-d	Accuracy	**0.6000**	0.5720	0.5760	0.5880	0.5800	0.4640	0.5640	*0.5920*
		Macro-F1	**0.5644**	0.5205	0.5316	0.5406	0.5376	0.4198	0.5174	*0.5435*
	Nested-d	Accuracy	*0.6000*	0.5840	0.5440	**0.6080**	0.5840	0.5840	0.5840	0.5840
		Macro-F1	*0.5672*	0.5409	0.5061	**0.5706**	0.5462	0.5462	0.5473	0.5393
Wine	Equal-d	Accuracy	0.9660	0.9660	0.9549	**0.9889**	0.8863	0.9431	0.9444	*0.9778*
		Macro-F1	0.9683	0.9683	0.9571	**0.9881**	0.8937	0.9474	0.9461	*0.9779*
	Nested-d	Accuracy	*0.9660*	0.9438	*0.9660*	*0.9660*	0.9438	0.9549	0.9549	**0.9889**
		Macro-F1	*0.9683*	0.9460	*0.9683*	*0.9683*	0.9460	0.9571	0.9548	**0.9881**

Note: Values are means across the five outer test partitions. Higher values are preferable for Accuracy, Macro-F1, and Micro-F1; lower values are preferable for Hamming loss. Bold underlining marks the best value and italic underlining marks the second-best value within each dataset–protocol–metric row; ties at the reported precision receive the same mark.

Table A2. Complete all-metric results for multi-label datasets (Part I: Bibtex, Corel5k, and Delicious).

Dataset	Protocol	Metric	Ours	FIMF	VMFS	MFS-MCDM	MCLS	MDFS	SSFS	NRS-Dep
Bibtex	Equal-d	Accuracy	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	**0.0277**	*0.0000*	*0.0000*
		Macro-F1	0.0028	0.0033	**0.0095**	0.0029	0.0000	0.0044	*0.0060*	0.0008
		Micro-F1	0.0299	0.0538	**0.1208**	0.0350	0.0000	0.0438	*0.0971*	0.0059
		Hamming loss	0.0177	0.0179	**0.0170**	0.0182	0.0176	0.0175	*0.0171*	0.0178
	Nested-d	Accuracy	0.0000	0.0000	*0.0077*	0.0000	0.0000	**0.0277**	0.0000	0.0000
		Macro-F1	0.0057	0.0031	**0.0096**	0.0057	0.0000	0.0044	*0.0060*	0.0008
		Micro-F1	0.0454	0.0552	**0.1261**	0.0454	0.0000	0.0438	*0.0971*	0.0059
		Hamming loss	0.0176	0.0180	**0.0168**	0.0176	0.0176	0.0175	*0.0171*	0.0176
Corel5k	Equal-d	Accuracy	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**
		Macro-F1	0.0010	0.0005	0.0007	**0.0013**	0.0002	*0.0012*	*0.0012*	0.0003
		Micro-F1	*0.0332*	0.0202	0.0219	0.0268	0.0172	0.0312	**0.0452**	0.0049
		Hamming loss	0.0118	0.0103	0.0099	*0.0097*	0.0104	0.0108	0.0108	**0.0096**
	Nested-d	Accuracy	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**
		Macro-F1	*0.0012*	0.0004	0.0002	0.0007	0.0002	**0.0019**	0.0008	0.0004
		Micro-F1	*0.0384*	0.0124	0.0045	0.0173	0.0172	**0.0473**	0.0320	0.0319
		Hamming loss	0.0126	*0.0099*	0.0101	**0.0098**	0.0104	0.0115	0.0106	0.0111

Table A2. *Cont.*

Dataset	Protocol	Metric	Ours	FIMF	VMFS	MFS-MCDM	MCLS	MDFS	SSFS	NRS-Dep
Delicious	Equal-d	Accuracy	**0.0083**	*0.0000*	*0.0000*	**0.0083**	*0.0000*	*0.0000*	*0.0000*	*0.0000*
		Macro-F1	**0.0069**	0.0028	0.0034	**0.0069**	0.0014	0.0046	*0.0068*	0.0008
		Micro-F1	0.1035	0.0733	0.0665	*0.1058*	0.0505	0.0898	**0.1153**	0.0346
		Hamming loss	0.0199	0.0200	0.0200	**0.0196**	0.0198	0.0202	*0.0197*	0.0199
	Nested-d	Accuracy	*0.0000*	**0.0083**	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	**0.0083**
		Macro-F1	**0.0081**	0.0031	0.0036	0.0059	0.0014	0.0052	*0.0071*	0.0003
		Micro-F1	**0.1291**	0.0610	0.0658	0.1121	0.0505	0.0927	*0.1189*	0.0174
		Hamming loss	*0.0197*	*0.0197*	*0.0197*	0.0200	0.0198	0.0200	0.0199	**0.0195**

Note: Values are means across the five outer test partitions. Higher values are preferable for Accuracy, Macro-F1, and Micro-F1; lower values are preferable for Hamming loss. Bold underlining marks the best value and italic underlining marks the second-best value within each dataset–protocol–metric row; ties at the reported precision receive the same mark.

Table A3. Complete all-metric results for multi-label datasets (Part II: Emotions, Enron, and Mediamill-10k).

Dataset	Protocol	Metric	Ours	FIMF	VMFS	MFS-MCDM	MCLS	MDFS	SSFS	NRS-Dep
Emotions	Equal-d	Accuracy	0.1511	0.1101	*0.1923*	**0.2025**	0.1395	0.0986	0.0996	0.1874
		Macro-F1	0.4571	0.3842	**0.4708**	*0.4649*	0.4402	0.4113	0.3567	0.4326
		Micro-F1	0.4934	0.4481	*0.5022*	**0.5050**	0.4595	0.4490	0.4280	0.4934
		Hamming loss	0.2873	0.3038	*0.2749*	0.2769	0.2943	0.3029	0.2948	**0.2739**
	Nested-d	Accuracy	0.1669	0.1512	**0.2509**	*0.2274*	0.1879	0.1838	0.1322	0.2084
		Macro-F1	0.5230	0.4210	**0.5405**	0.5218	0.5015	0.4885	0.4019	*0.5243*
		Micro-F1	0.5588	0.4867	**0.5965**	*0.5652*	0.5323	0.5210	0.4701	0.5583
		Hamming loss	0.2526	0.2788	**0.2285**	0.2454	0.2586	0.2588	0.2656	*0.2435*
Enron	Equal-d	Accuracy	0.0071	0.0291	0.0220	**0.0875**	0.0000	0.0071	*0.0314*	0.0214
		Macro-F1	0.0492	**0.0615**	0.0518	0.0552	0.0398	0.0374	*0.0567*	0.0547
		Micro-F1	*0.4396*	**0.4444**	0.4095	0.4077	0.4292	0.3690	0.4290	0.4264
		Hamming loss	0.0611	**0.0591**	0.0622	0.0621	0.0653	0.0671	0.0636	*0.0609*
	Nested-d	Accuracy	0.0162	0.0291	0.0200	**0.0363**	0.0000	0.0071	0.0100	*0.0305*
		Macro-F1	0.0496	**0.0589**	0.0497	0.0539	0.0407	0.0394	0.0490	*0.0559*
		Micro-F1	0.4386	**0.4415**	0.4079	0.3934	*0.4408*	0.3841	0.4251	0.4282
		Hamming loss	0.0624	**0.0598**	0.0687	0.0632	0.0643	0.0669	0.0644	*0.0600*
Mediamill-10k	Equal-d	Accuracy	**0.0497**	0.0000	0.0163	0.0262	0.0163	0.0167	*0.0413*	0.0250
		Macro-F1	*0.0417*	**0.0430**	0.0345	0.0377	0.0335	0.0395	0.0405	0.0407
		Micro-F1	*0.4845*	**0.4943**	0.4586	0.4639	0.4519	0.4741	0.4766	0.4784
		Hamming loss	0.0359	**0.0350**	0.0370	0.0367	0.0375	0.0366	*0.0358*	0.0361
	Nested-d	Accuracy	**0.0574**	0.0247	0.0247	*0.0342*	0.0330	0.0080	0.0330	0.0247
		Macro-F1	*0.0417*	**0.0427**	0.0364	0.0394	0.0330	0.0382	0.0401	0.0327
		Micro-F1	*0.4758*	**0.4934**	0.4559	0.4677	0.4462	0.4681	0.4737	0.4516
		Hamming loss	0.0361	**0.0356**	0.0376	0.0367	0.0376	0.0363	*0.0358*	0.0382

Note: Values are means across the five outer test partitions. Higher values are preferable for Accuracy, Macro-F1, and Micro-F1; lower values are preferable for Hamming loss. Bold underlining marks the best value and italic underlining marks the second-best value within each dataset–protocol–metric row; ties at the reported precision receive the same mark.

Table A4. Complete all-metric results for multi-label datasets (Part III: Scene, TMC2007-500, and Yeast).

Dataset	Protocol	Metric	Ours	FIMF	VMFS	MFS-MCDM	MCLS	MDFS	SSFS	NRS-Dep
Scene	Equal-d	Accuracy	*0.3044*	0.2620	0.2843	0.2571	0.2820	**0.4006**	0.2212	0.2734
		Macro-F1	*0.3061*	0.2682	0.2896	0.2737	0.2823	**0.4228**	0.2501	0.2637
		Micro-F1	*0.3791*	0.3619	0.3780	0.3281	0.3486	**0.4885**	0.3222	0.3216
		Hamming loss	0.1913	*0.1815*	0.1868	0.1869	0.1902	**0.1566**	0.1855	0.1930
	Nested-d	Accuracy	0.2730	0.2696	0.2419	0.1926	0.2559	**0.4086**	0.2212	*0.2900*
		Macro-F1	0.2805	0.2752	0.2629	0.2108	0.2542	**0.4224**	0.2501	*0.3079*
		Micro-F1	0.3442	*0.3744*	0.3553	0.2681	0.3152	**0.4952**	0.3222	0.3622
		Hamming loss	0.2033	0.1844	0.1871	0.1922	0.1989	**0.1526**	0.1855	*0.1784*

Table A4. *Cont.*

Dataset	Protocol	Metric	Ours	FIMF	VMFS	MFS-MCDM	MCLS	MDFS	SSFS	NRS-Dep
TMC2007-500	Equal-d	Accuracy	0.0484	0.0501	0.0542	**0.1115**	*0.0676*	0.0600	0.0336	0.0521
		Macro-F1	0.1117	0.0701	0.0741	**0.1380**	0.0343	*0.1220*	0.0628	0.0515
		Micro-F1	0.3579	0.3786	0.3331	*0.4211*	0.3013	**0.4327**	0.3575	0.2678
		Hamming loss	0.1079	0.1022	0.1088	*0.0965*	0.1151	**0.0920**	0.1074	0.0993
	Nested-d	Accuracy	0.0407	0.0666	0.0437	**0.0759**	*0.0676*	0.0604	0.0336	0.0420
		Macro-F1	**0.1101**	0.0749	0.0936	0.0897	0.0343	*0.1011*	0.0654	0.0390
		Micro-F1	0.3360	*0.3910*	0.3416	0.3713	0.3013	**0.3930**	0.3758	0.2704
		Hamming loss	0.1084	0.1012	0.1113	0.0998	0.1151	**0.0952**	*0.0994*	0.1122
Yeast	Equal-d	Accuracy	**0.0981**	0.0491	0.0404	0.0824	0.0698	0.0698	*0.0841*	0.0307
		Macro-F1	0.2959	**0.3221**	*0.3119*	0.2830	0.2746	0.2628	0.2769	0.2970
		Micro-F1	*0.5383*	0.5071	**0.5602**	0.5364	0.5290	0.4923	0.5042	0.5239
		Hamming loss	0.2586	0.2706	*0.2562*	**0.2526**	0.2593	0.2626	0.2711	0.2598
	Nested-d	Accuracy	*0.0807*	0.0391	0.0470	0.0657	0.0544	0.0544	**0.1014**	0.0307
		Macro-F1	0.3025	**0.3264**	*0.3058*	0.2754	0.2714	0.2752	0.2694	0.2970
		Micro-F1	*0.5309*	0.5187	**0.5515**	0.5170	0.5179	0.5074	0.5067	0.5239
		Hamming loss	0.2605	0.2689	*0.2577*	**0.2564**	0.2649	0.2632	0.2636	0.2598

Note: Values are means across the five outer test partitions. Higher values are preferable for Accuracy, Macro-F1, and Micro-F1; lower values are preferable for Hamming loss. Bold underlining marks the best value and italic underlining marks the second-best value within each dataset–protocol–metric row; ties at the reported precision receive the same mark.

References

1. Zhou, Z.H.; Zhang, M.L., Multi-label learning. In *Encyclopedia of Machine Learning and Data Mining*; Springer: New York, NY, USA, 2017; pp. 875–881.
2. Liu, Z.; Tang, C.; Abhadiomhen, S.E.; Shen, X.J.; Li, Y. Robust Label and Feature Space Co-Learning for Multi-Label Classification. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 11846–11859. [CrossRef]
3. Ruan, J.; Wang, M.; Liu, D.; Chen, M.; Gao, X. Multi-Label Feature Selection with Feature–Label Subgraph Association and Graph Representation Learning. *Entropy* **2024**, *26*, 992. [CrossRef] [PubMed]
4. Zhang, J.; Wu, H.; Jiang, M.; Liu, J.; Li, S.; Tang, Y.; Long, J. Group-preserving label-specific feature selection for multi-label learning. *Expert Syst. Appl.* **2023**, *213*, 118861. [CrossRef]
5. He, Z.; Lin, Y.; Lin, Z.; Wang, C. Multi-label feature selection via similarity constraints with non-negative matrix factorization. *Knowl.-Based Syst.* **2024**, *297*, 111948. [CrossRef]
6. Li, H.; Fang, M.; Wang, P. Dual dimensionality reduction on instance-level and feature-level for multi-label data. *Neural Comput. Appl.* **2023**, *35*, 24773–24782. [CrossRef]
7. Hao, P.; Liu, K.; Gao, W. Uncertainty-Aware Global-View Reconstruction for Multi-View Multi-Label Feature Selection. In *Proceedings of the AAAI Conference on Artificial Intelligence*; AAAI: Palo Alto, CA, USA, 2025; Volume 39, pp. 17068–17076.
8. Wan, M.; Cai, M.; Yang, G. Robust exponential graph regularization non-negative matrix factorization technology for feature extraction. *Mathematics* **2023**, *11*, 1716. [CrossRef]
9. González-López, J.; Ventura, S.; Cano, A. Distributed selection of continuous features in multilabel classification using mutual information. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *31*, 2280–2293. [CrossRef] [PubMed]
10. Yin, T.; Chen, H.; Wan, J.; Zhang, P.; Horng, S.J.; Li, T. Exploiting feature multi-correlations for multilabel feature selection in robust multi-neighborhood fuzzy β covering space. *Inf. Fusion* **2024**, *104*, 102150. [CrossRef]
11. Ma, J.; Xu, F.; Rong, X. Discriminative multi-label feature selection with adaptive graph diffusion. *Pattern Recognit.* **2024**, *148*, 110154. [CrossRef]
12. Pawlak, Z. *Rough Sets: Theoretical Aspects of Reasoning About Data*, 1st ed.; Kluwer Academic Publishers: London, UK, 1991; pp. 1–55.
13. Qian, D.; Liu, K.; Wang, J.; Zhang, S.; Yang, X. Attribute reduction based on directional semi-neighborhood rough set. *Int. J. Mach. Learn. Cybern.* **2025**, *16*, 2523–2535. [CrossRef]
14. Greco, S.; Matarazzo, B.; Slowinski, R. Rough sets theory for multicriteria decision analysis. *Eur. J. Oper. Res.* **2001**, *129*, 1–47. [CrossRef]
15. Dai, J.; Gao, S.; Zheng, G. Generalized rough set models determined by multiple neighborhoods generated from a similarity relation. *Soft Comput.* **2018**, *22*, 2081–2094. [CrossRef]
16. Xia, S.; Wang, C.; Wang, G.; Gao, X.; Ding, W.; Yu, J.; Zhai, Y.; Chen, Z. GBRS: A unified granular-ball learning model of pawlak rough set and neighborhood rough set. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *36*, 1719–1733. [CrossRef] [PubMed]
17. Wang, N.; Zhao, E. A new method for feature selection based on weighted k-nearest neighborhood rough set. *Expert Syst. Appl.* **2024**, *238*, 122324. [CrossRef]

18. Sewwandi, M.A.N.D.; Li, Y.; Zhang, J. Granule-specific feature selection for continuous data classification using neighborhood rough sets. *Expert Syst. Appl.* **2024**, *238*, 121765. [CrossRef]
19. Zhang, X.; Ou, Q.; Wang, J. Variable precision fuzzy rough sets based on overlap functions with application to tumor classification. *Inf. Sci.* **2024**, *666*, 120451. [CrossRef]
20. Leung, Y.; Li, D. Maximal consistent block technique for rule acquisition in incomplete information systems. *Inf. Sci.* **2003**, *153*, 85–106. [CrossRef]
21. Zhang, Z.; Yang, X. Tolerance-based multigranulation rough sets in incomplete systems. *Front. Comput. Sci.* **2014**, *8*, 753–762. [CrossRef]
22. Li, R.; Chen, H.; Liu, S.; Wang, K.; Wang, B.; Hu, X. TFD-IIS-CRMCB: Telecom fraud detection for incomplete information systems based on correlated relation and maximal consistent block. *Entropy* **2023**, *25*, 112. [CrossRef] [PubMed]
23. Sun, Y.; Wu, W.Z.; Wang, X. Maximal consistent block based optimal scale selection for incomplete multi-scale information systems. *Int. J. Mach. Learn. Cybern.* **2023**, *14*, 1797–1809. [CrossRef]
24. Yan, S.; Jusheng, M.; Tao, F.; Leijun, L.; Meishe, L. Maximum consistent block based variable precision rough set model and attribute reduction. *J. Front. Comput. Sci. Technol.* **2020**, *14*, 892.
25. Leung, Y.; Wu, W.Z.; Zhang, W.X. Knowledge acquisition in incomplete information systems: A rough set approach. *Eur. J. Oper. Res.* **2006**, *168*, 164–180. [CrossRef]
26. Mroczek, T.; Zheng, R. A new approach to constructing maximal consistent blocks for mining incomplete data. *Procedia Comput. Sci.* **2022**, *207*, 1047–1056. [CrossRef]
27. Zhang, W.; Qiu, G. *Uncertain Decision Making Based on Rough Sets*, 1st ed.; Tsinghua University Press: Beijing, China, 2005; pp. 92–93.
28. Hu, Q.; Yu, D.; Liu, J.; Wu, C. Neighborhood rough set based heterogeneous feature subset selection. *Inf. Sci.* **2008**, *178*, 3577–3594. [CrossRef]
29. Mi, J.; Wu, W.Z.; Zhang, W.X. Knowledge reducts based on variable precision rough set theory. *Syst. Eng.-Theory Pract.* **2004**, *24*, 116–120.
30. Koklu, M.; Ozkan, I.A. Multiclass classification of dry beans using computer vision and machine learning techniques. *Comput. Electron. Agric.* **2020**, *174*, 105507. [CrossRef]
31. German, B. Glass Identification. 1987. Available online: https://archive.ics.uci.edu/dataset/42/glass+identification (accessed on 20 June 2026).
32. Fisher, R.A. The use of multiple measurements in taxonomic problems. *Ann. Eugen.* **1936**, *7*, 179–188. [CrossRef]
33. Silva, P.F.B. Development of a System for Automatic Plant Species Recognition. Master's Thesis, Universidade do Porto (Portugal), Porto, Portugal, 2013.
34. Aeberhard, S.; Coomans, D.; De Vel, O. Comparative analysis of statistical pattern recognition methods in high dimensional settings. *Pattern Recognit.* **1994**, *27*, 1065–1077. [CrossRef]
35. Katakis, I.; Tsoumakas, G.; Vlahavas, I. Multilabel text classification for automated tag suggestion. *ECML PKDD Discov. Chall.* **2008**, *75*, 2008.
36. Duygulu, P.; Barnard, K.; de Freitas, J.F.; Forsyth, D.A. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In *Proceedings of the European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2002; pp. 97–112.
37. Tsoumakas, G.; Katakis, I.; Vlahavas, I. Effective and efficient multilabel classification in domains with large number of labels. In *Proceedings of the ECML/PKDD 2008 Workshop on Mining Multidimensional Data (MMD'08)*; Springer: Berlin/Heidelberg, Germany, 2008; Volume 21, pp. 53–59.
38. Trohidis, K.; Tsoumakas, G.; Kalliris, G.; Vlahavas, I. Multi-label classification of music by emotion. *EURASIP J. Audio Speech Music Process.* **2011**, *2011*, 4. [CrossRef]
39. Klimt, B.; Yang, Y. The enron corpus: A new dataset for email classification research. In *Proceedings of the European Conference on Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 217–226.
40. Snoek, C.G.; Worring, M.; Van Gemert, J.C.; Geusebroek, J.M.; Smeulders, A.W. The challenge problem for automated detection of 101 semantic concepts in multimedia. In *Proceedings of the 14th ACM International Conference on Multimedia*; ACM: San Jose, CA, USA, 2006; pp. 421–430.
41. Boutell, M.R.; Luo, J.; Shen, X.; Brown, C.M. Learning multi-label scene classification. *Pattern Recognit.* **2004**, *37*, 1757–1771. [CrossRef]
42. Srivastava, A.N.; Zane-Ulman, B. Discovering recurring anomalies in text reports regarding complex space systems. In *Proceedings of the 2005 IEEE Aerospace Conference*; IEEE: New York, NY, USA, 2005; pp. 3853–3862.
43. Elisseeff, A.; Weston, J. A kernel method for multi-labelled classification. *Adv. Neural Inf. Process. Syst.* **2001**, *14*, 681–687. [CrossRef]

44. Lee, J.; Kim, D.W. Fast multi-label feature selection based on information-theoretic feature ranking. *Pattern Recognit.* **2015**, *48*, 2761–2771. [CrossRef]
45. Hashemi, A.; Dowlatshahi, M.B.; Nezamabadi-pour, H. VMFS: A VIKOR-based multi-target feature selection. *Expert Syst. Appl.* **2021**, *182*, 115224. [CrossRef]
46. Hashemi, A.; Dowlatshahi, M.B.; Nezamabadi-pour, H. MFS-MCDM: Multi-label feature selection using multi-criteria decision making. *Knowl.-Based Syst.* **2020**, *206*, 106365. [CrossRef]
47. Huang, R.; Jiang, W.; Sun, G. Manifold-based constraint Laplacian score for multi-label feature selection. *Pattern Recognit. Lett.* **2018**, *112*, 346–352. [CrossRef]
48. Zhang, J.; Luo, Z.; Li, C.; Zhou, C.; Li, S. Manifold regularized discriminative feature selection for multi-label learning. *Pattern Recognit.* **2019**, *95*, 136–150. [CrossRef]
49. Gao, W.; Li, Y.; Hu, L. Multilabel Feature Selection With Constrained Latent Structure Shared Term. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *34*, 1253–1262. [CrossRef] [PubMed]
50. Liu, J.; Lin, Y.; Li, Y.; Weng, W.; Wu, S. Online multi-label streaming feature selection based on neighborhood rough set. *Pattern Recognit.* **2018**, *84*, 273–287. [CrossRef]
51. Zhao, Y.; Liu, Y.; Zhang, H.; Zhao, N.; Shi, Y. Funnel-based predefined-time output feedback fuzzy adaptive event-triggered platoon control for heterogeneous vehicles. *IEEE Internet Things J.* **2026**, *early access*. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.