Computers, Materials & Continua

Tech Science Press

https://doi.org/10.32604/cmc.2026.083713

ARTICLE

Boundary Region-Driven Feature Selection for Neighborhood Rough Sets

Wenchang Yu[1], **Xiaoqin Ma**[1,2], **Zheqing Zhang**[1] **and Kezhong Lu**[1,2,*]

[1]School of Big Data and Artificial Intelligence, Chizhou University, Chizhou, China
[2]Anhui Education Big Data Intelligent Perception and Application Engineering Research Center, Anhui Provincial Joint Construction Key Laboratory of Intelligent Education Equipment and Technology, Chizhou, China

*Corresponding Author: Kezhong Lu. Email: luke76@163.com

Received: 09 April 2026; Accepted: 01 June 2026; Published: 23 July 2026

ABSTRACT: Feature selection grounded in neighborhood rough sets has attracted sustained research attention owing to its principled treatment of classification uncertainty. However, existing forward greedy algorithms typically evaluate uncertainty over the entire object universe at each iteration, resulting in prohibitive computational complexity on large-scale datasets. To address this inefficiency, we introduce a new uncertainty index built upon Boundary Object Sets (BOS). BOS are defined as objects whose neighborhood granules intersect with multiple decision classes, thereby capturing intrinsic classification ambiguity. The proposed measure quantifies the proportion of these boundary objects relative to the total universe size. Grounded in this measure, we develop the Boundary Object Set Feature Selection (BOSFS) algorithm. BOSFS employs a double-contraction strategy that simultaneously reduces the candidate attribute pool and eliminates consistent objects from the active working set. Consequently, the algorithm restricts computationally expensive distance calculations to the monotonically shrinking BOS. Experiments on ten benchmark datasets, evaluated against six competing algorithms under three classifiers, confirm that BOSFS achieves the highest classification accuracy in 15 of 30 test cases while consuming only 25% of the runtime of the second-fastest competitor.

KEYWORDS: Neighborhood rough set; uncertainty measure; feature selection; boundary object set

1 Introduction

The proliferation of high-dimensional datasets has rendered feature selection an indispensable preprocessing step across a broad spectrum of data-driven tasks, as it alleviates the curse of dimensionality and bolsters model generalization [1,2]. Rough set theory [3,4] provides a mathematically rigorous, assumption-free framework for feature selection that has attracted sustained research attention over the past decades. Classical rough set theory characterizes uncertainty via approximations over equivalence classes, but its reliance on label-matching of attributes limits its applicability to continuous data. To bridge this functional gap, a range of generalized rough set models have been proposed. Representative advancements include hybrid fuzzy-rough frameworks [5,6], probabilistic decision-theoretic approaches [7,8], bi-variable precision mechanisms [9], preference-driven dominance relations [10], covering approximation spaces [11,12], and the recently proposed overlapping containment and cardinality rough neighborhoods [13,14]. Among these, the neighborhood rough sets [15] replace equivalence classes with neighborhood granules defined over a distance metric, enabling principled handling of numerical attributes without discretization, and have become the most widely adopted foundation for feature selection on real-valued datasets.

In rough set environments, the attribute reduction process seeks the most compact combination of variables that retains the identical distinguishing capability of the entire feature space. Existing measures for guiding this search span several paradigms, including positive-region-based dependency [3,15], conditional entropy [16,17], mutual information [18,19], measures over compromised variable-granularity neighborhoods [20], neighborhood entropy [21,22], and fuzzy scale entropy [23]. Moreover, scholars have also explored composite uncertainty measures [24], self-information measure [25], local density-weighted criteria [26], three-way decision-based methods [27,28], and incremental reduction approaches for dynamic datasets [29,30].

Building on the neighborhood rough set model specifically, a series of algorithms with increasingly sophisticated criteria have been proposed. Chen et al. [31] combined overlap degree and conditional entropy under KNN granularity. Xu et al. [32] replaced the standard Euclidean neighborhood with a shared-neighbor similarity structure to improve robustness to local density variations. Zou and Dai [24] employed a composite measure over fuzzy β-covering neighborhoods. Gao et al. [33] weighted samples by their distributional proximity to decision boundaries within a neighborhood system. Sun et al. [34] proposed a multi-label feature selection method based on fuzzy C-means clustering and weighted neighborhood mutual information, integrating whale optimization for initialization and label enhancement to improve classification performance. Liu et al. [35] proposed a reliability-enhanced partial multi-label feature selection method based on neighborhood rough set. Chang et al. [36] proposed a fuzzy neighborhood rough set-based attribute reduction method over temporal information systems for clinical efficacy evaluation.

Despite this progress, a fundamental computational bottleneck still exists in almost all forward greedy feature selection algorithms of this type. At each iteration, these methods evaluate every candidate attribute by recomputing pairwise distances and updating approximations over the entire universe. Concretely, updating the distance matrix is the dominant cost. For a decision table containing n samples and m conditional features, extracting the final reduct requires a computational cost of $O(n^2m^2)$. On large-scale or high-dimensional datasets, this design flaw leads to prohibitive runtimes, as confirmed by the empirical comparisons reported in [31–33].

To bypass this operational bottleneck, this study introduces a novel uncertainty evaluation metric driven by boundary object sets (BOS) and a corresponding accelerated feature selection algorithm. For a decision class, the BOS under attribute subset and radius δ is defined as the set of objects whose neighborhood intersects at least one other decision class. The BOS-based uncertainty measure is then defined as the ratio of the number of boundary objects to the universe size. This ratio decreases monotonically as informative attributes are added. Building on this measure, we propose the Boundary Object Set Feature Selection (BOSFS) algorithm, which embeds a double-contraction strategy. At each iteration, BOSFS shrinks both the candidate attribute pool and the active sample set. The latter is achieved by removing objects that have left the boundary and entered the positive region. Pairwise distance computation is thus confined to the monotonically shrinking BOS, reducing the effective complexity to $O(m \cdot n^2)$.

The primary innovations presented in this work include the following:

- A novel BOS-based uncertainty measure is formally introduced. Furthermore, its boundedness and monotonicity with respect to attribute subsets and neighborhood radius δ are rigorously established.
- Unlike traditional positive-region or entropy-based metrics, our proposed measure specifically targets objects exhibiting genuine classification ambiguity, offering a more precise evaluation criterion.
- The BOSFS algorithm restricts expensive $O(n^2)$ distance computations to the shrinking BOS by simultaneously contracting the candidate attribute space and the active sample set across iterations. This converts a constant-cost global search into a rapidly converging local refinement, lowering total complexity.

- Extensive experiments on 10 benchmark datasets confirm that BOSFS achieves the highest accuracy in 15 out of 30 test cases, with an average runtime of only 25% compared to the second-fastest competitor.

In summary, the proposed BOSFS method offers three main advantages: computational efficiency, focused uncertainty modeling and empirical effectiveness.

The remainder of this paper is organized as follows. Foundational principles and definitions are revisited in Section 2. Section 3 formalizes the proposed BOS-driven uncertainty criterion and details its mathematical characteristics. Section 4 outlines the operational mechanics and theoretical time constraints of the BOSFS methodology. Empirical performance validations across multiple datasets are documented in Section 5, while Section 6 provides concluding remarks and outlines potential future research directions.

2 Preliminaries

This section outlines the fundamental mathematical preliminaries related to decision frameworks and neighborhood rough set theories. Beginning with decision information systems, it then covers neighborhood approximation spaces, neighborhood rough sets, and the dependency measure.

An information system (IS) is defined as a pair (U, A), where U is a non-empty finite object set, A is an attribute set. For $x \in U$ and $a \in A$, x_a denotes the value of object x on attribute a.

A decision information system (DIS) is a particular case in which the attribute set is split into a condition part C and a decision part L satisfying $C \cap L = \varnothing$, denoted as $(U, C \cup L)$, where L typically consists of a single label-type attribute. The decision label in L induces an equivalence relation that partitions U as $U/L = \{L_1, L_2, \ldots, L_r\}$.

To quantify the dissimilarity between objects, a distance metric is indispensable, particularly for numerical data prevalent in real-world applications.

Definition 1: *Given a DIS $(U, C \cup L)$ and a conditional attribute subset $\mathcal{G} \subseteq C$. For objects $x, y \in U$, the distance between x and y under $\mathcal{G}$ is:*

$$\mathrm{dist}_{\mathcal{G}}(x, y) = \left(\sum_{a \in \mathcal{G}} d(x_a, y_a)^p \right)^{1/p} \tag{1}$$

where $p \geq 1$ (commonly $p = 2$), d is typically the absolute difference $|x_a - y_a|$.

Proposition 1: *Given a DIS $(U, C \cup L)$ and objects $x, y \in U$, if $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq C$, then $\mathrm{dist}_{\mathcal{G}_1}(x, y) \leq \mathrm{dist}_{\mathcal{G}_2}(x, y)$.*

Definition 2: *Given a DIS $(U, C \cup L)$, a conditional attribute subset $\mathcal{G} \subseteq C$ and a parameter $\delta \geq 0$. For object $x \in U$, the neighborhood of x under $\mathcal{G}$ and δ can be defined as:*

$$[x]_{\mathcal{G}}^{\delta} = \{ y \in U \mid \mathrm{dist}_{\mathcal{G}}(x, y) \leq \delta \} \tag{2}$$

The parameter δ is termed the neighborhood radius.

For $x, y \in U$, if $y \in [x]_{\mathcal{G}}^{\delta}$, then there exists a neighborhood relation from x to y, denoted as $[x]_{\mathcal{G}}^{\delta}(y)$. Let $N_{\mathcal{G}}^{\delta}$ be the family of all such relations under $\mathcal{G}$ and δ. The pair $(U, N_{\mathcal{G}}^{\delta})$ defines a neighborhood approximation space (NAS).

Proposition 2: *Given a DIS $(U, C \cup L)$, $\mathcal{G} \subseteq C$ and $\delta \geq 0$. The following conclusions hold:*

(1) *self-inclusion: $[x]_{\mathcal{G}}^{\delta}(x)$ holds for any $x \in U$;*
(2) *symmetry: if $[x]_{\mathcal{G}}^{\delta}(y) \in NAS$, then $[y]_{\mathcal{G}}^{\delta}(x) \in NAS$ for any $x, y \in U$.*

Proposition 3: *Given a DIS* $(U, C \cup L)$. *For any* $x \in U$, *these statements hold:*

(1) *if* $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq C$ *and* $\delta \geq 0$, *then* $[x]^{\delta}_{G_1} \supseteq [x]^{\delta}_{G_2}$;
(2) *if* $\mathcal{G} \subseteq C$ *and* $\delta_2 \geq \delta_1 \geq 0$, *then* $[x]^{\delta_1}_{\mathcal{G}} \subseteq [x]^{\delta_2}_{\mathcal{G}}$.

Definition 3: *Given a NAS* $(U, N^{\delta}_{\mathcal{G}})$ *and a subset* $X \subseteq U$, *the lower and upper approximations of* X *are defined as:*

$$\underline{N}^{\delta}_{\mathcal{G}}(X) = \{x \in U \mid [x]^{\delta}_{\mathcal{G}} \subseteq X\} \tag{3}$$

$$\overline{N}^{\delta}_{\mathcal{G}}(X) = \{x \in U \mid [x]^{\delta}_{\mathcal{G}} \cap X \neq \varnothing\} \tag{4}$$

If $\underline{N}^{\delta}_{\mathcal{G}}(X) = \overline{N}^{\delta}_{\mathcal{G}}(X)$, then subset X is consistent in $(U, N^{\delta}_{\mathcal{G}})$. Otherwise, it is called a neighborhood rough set.

Proposition 4: *Given a NAS* $= (U, N^{\delta}_{\mathcal{G}})$ *and a subset* $X \subseteq U$, *the following equivalences hold:*

(1) X *is consistent in* $(U, N^{\delta}_{\mathcal{G}})$;
(2) $\underline{N}^{\delta}_{\mathcal{G}}(X) = \overline{N}^{\delta}_{\mathcal{G}}(X)$;
(3) $\underline{N}^{\delta}_{\mathcal{G}}(X) = X$;
(4) $X = \overline{N}^{\delta}_{\mathcal{G}}(X)$.

Proposition 5: *Let* $(U, C \cup L)$ *be a DIS and* $X \subseteq U$. *Then the following statements hold:*

(1) *if* $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq C$ *and* $\delta \geq 0$, *then* $\underline{N}^{\delta}_{G_1}(X) \subseteq \underline{N}^{\delta}_{G_2}(X)$;
(2) *if* $\mathcal{G} \subseteq C$ *and* $\delta_2 \geq \delta_1 \geq 0$, *then* $\underline{N}^{\delta_1}_{\mathcal{G}}(X) \supseteq \underline{N}^{\delta_2}_{\mathcal{G}}(X)$;
(3) *if* $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq C$ *and* $\delta \geq 0$, *then* $\overline{N}^{\delta}_{\mathcal{G}_1}(X) \supseteq \overline{N}^{\delta}_{\mathcal{G}_2}(X)$;
(4) *if* $\mathcal{G} \subseteq C$ *and* $\delta_2 \geq \delta_1 \geq 0$, *then* $\overline{N}^{\delta_1}_{\mathcal{G}}(X) \subseteq \overline{N}^{\delta_2}_{\mathcal{G}}(X)$.

Proof:

(1) According to Proposition 3 (1), for any $x \in U$, given that $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq C$ and $\delta \geq 0$, we have $[x]^{\delta}_{\mathcal{G}_1} \supseteq [x]^{\delta}_{\mathcal{G}_2}$. Formula (3) dictates that if $x \in \underline{N}^{\delta}_{\mathcal{G}_1}(X)$, then $[x]^{\delta}_{\mathcal{G}_1} \subseteq X$. Combining these two relationships yields, $[x]^{\delta}_{\mathcal{G}_2} \subseteq [x]^{\delta}_{\mathcal{G}_1} \subseteq X$, this inclusion implies $x \in \underline{N}^{\delta}_{\mathcal{G}_2}(X)$. Consequently, it can be concluded that $\underline{N}^{\delta}_{\mathcal{G}_1}(X) \subseteq \underline{N}^{\delta}_{\mathcal{G}_2}(X)$.
(2) According to Proposition 3 (2), for any $x \in U$, given that $\mathcal{G} \subseteq C$ and $\delta_2 \geq \delta_1 \geq 0$, we have $[x]^{\delta_1}_{\mathcal{G}} \subseteq [x]^{\delta_2}_{\mathcal{G}}$. Formula (3) dictates that if $x \in \underline{N}^{\delta_2}_{\mathcal{G}}(X)$, then $[x]^{\delta_2}_{\mathcal{G}} \subseteq X$. Combining these two relationships yields, $[x]^{\delta_1}_{\mathcal{G}} \subseteq [x]^{\delta_2}_{\mathcal{G}} \subseteq X$, this inclusion implies $x \in \underline{N}^{\delta_1}_{\mathcal{G}}(X)$. Consequently, it can be concluded that $\underline{N}^{\delta_1}_{\mathcal{G}}(X) \supseteq \underline{N}^{\delta_2}_{\mathcal{G}}(X)$.

The proofs of (3) and (4) are similar to those of (1) and (2), and are therefore omitted here. □

The classic Pawlak rough set theory is shown in Fig. 1. The relationships between objects are represented by small grids; objects within the same grid are indistinguishable, while those in different grids are completely distinguishable. The irregular region X represents a target concept (decision class). The task is to determine whether each sample in U belongs to X. It is evident that objects belonging to $\underline{N}(X)$ (the green region) definitively belong to the target concept X. Therefore, $\underline{N}(X)$ is referred to as the positive region of X.

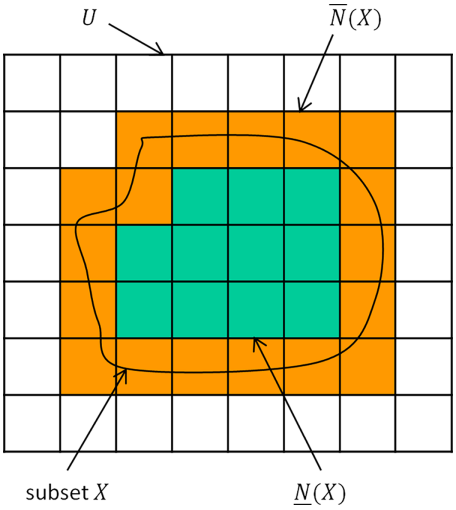

Figure 1: The $\underline{N}(X)$ and $\overline{N}(X)$ in classical rough set theory.

Consider a DIS $(U, C \cup L)$. Decision attribute in L induces U into decision partitions, denoted as $U/L = \{L_1, L_2, \ldots, L_r\}$. For any decision partition $L_i \in U/L$, its positive region under C is defined by:

$$\mathrm{POS}_C(L_i) = \underline{N}_C(L_i) \tag{5}$$

The dependency of the entire DIS is defined as:

$$\gamma_C(L) = \frac{\left|\bigcup_{L_i \in U/L} \mathrm{POS}_C(L_i)\right|}{|U|} \tag{6}$$

Here, $|\cdot|$ indicates set cardinality.

$\gamma_C(L)$ measures the uncertainty of the DIS. A larger value of $\gamma_C(L)$ indicates lower uncertainty. As illustrated in Fig. 2, $\gamma_C(L)$ calculates the proportion of objects within the positive region relative to the entire objects in U, under the conditional attribute set C. That is, the ratio of the green area to the total area in Fig. 2.

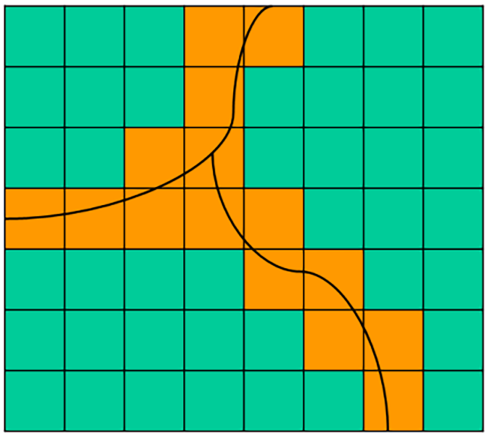

Figure 2: The positive regions of a DIS.

In forward greedy feature selection algorithms, adding a candidate attribute requires calculating the lower approximation for every decision class, resulting in $O(n^2)$ complexity. This computation must be repeated even for objects already consistent with the decision partition. Alternatively, the boundary region can also be used as a measure of uncertainty, corresponding to the yellow area in Fig. 2.

3 Boundary Region-Based Uncertainty Measure

Classical rough set theory defines the boundary region by $BN(X) = \overline{N}(X) - \underline{N}(X)$. It corresponds to the orange region in Fig. 1. According to Proposition 4, a DIS is consistent if and only if, the condition $L_i - \underline{N}(L_i) = \varnothing$ holds for each decision class $L_i \in U/L$. Conversely, the presence of any L_i where $L_i - \underline{N}(L_i) \neq \varnothing$ renders the DIS inconsistent. Motivated by this principle, our proposed uncertainty measure focuses explicitly on the boundary objects within each decision class.

Definition 4: *Given a NAS* $(U, N_{\mathcal{G}}^{\delta})$ *and a subset* $X \subseteq U$*, the boundary object set (BOS) of* X *in the NAS is defined as:*

$$\mathrm{BOS}_{\mathcal{G}}^{\delta}(X) = \{x \in X \mid \exists y \in [x]_{\mathcal{G}}^{\delta},\ y \notin X\} \tag{7}$$

Remark 1: *BOS differs from the classical rough set boundary region in two key aspects: (i) it considers only objects inside* X*, avoiding redundant counting across decision classes; (ii) it directly measures classification ambiguity at the object level. This design aligns naturally with forward greedy feature selection, where the goal is to progressively eliminate uncertainty by shrinking the BOS.*

Fig. 3 illustrates the BOS of X, where objects in the red region constitute its boundary. As Fig. 1 shows, the BOS is conceptually distinct from the classical boundary region. In Fig. 1, the grids containing objects outside the subset L_i are also treated as part of the boundary region of L_i; however, in Fig. 3, only samples in L_i are checked to determine whether they are boundary objects. This is because, in a DIS, objects outside L_i will be evaluated within other decision classes, and Definition 4 avoids repeated judgment of the same object.

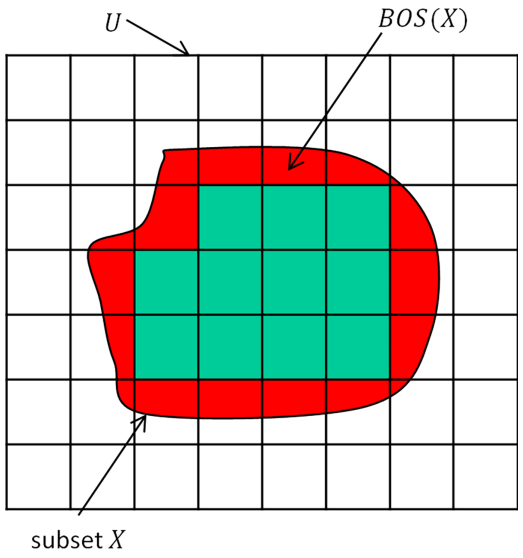

Figure 3: A schematic diagram of a boundary object.

Proposition 6: *Let* $(U, C \cup L)$ *be a DIS, the following conclusions hold, for any* $L_i \in U/L$*:*

(1) *if* $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq C$ *and* $0 \le \delta \le 1$*, then* $\mathrm{BOS}^{\delta}_{\mathcal{G}_1}(L_i) \supseteq \mathrm{BOS}^{\delta}_{\mathcal{G}_2}(L_i)$*;*
(2) *if* $0 \le \delta_1 \le \delta_2 \le 1$ *and* $\mathcal{G} \subseteq C$*, then* $\mathrm{BOS}^{\delta_1}_{\mathcal{G}}(L_i) \subseteq \mathrm{BOS}^{\delta_2}_{\mathcal{G}}(L_i)$.

Definition 5: *Let* $(U, C \cup L)$ *be a DIS,* $\mathcal{G} \subseteq C$ *and* $\delta \ge 0$*. The BOS of the entire DIS is defined as:*

$$\mathrm{BOS}^{\delta}_{\mathcal{G}}(L) = \bigcup_{L_i \in U/L} \mathrm{BOS}^{\delta}_{\mathcal{G}}(L_i) \tag{8}$$

The BOS definition of the entire DIS in Definition 5 is completely consistent with the orange region in Fig. 2.

Proposition 7: *For a DIS* $(U, C \cup L)$*, the following conclusions hold:*

(1) *if* $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq C$ *and* $0 \le \delta \le 1$*, then* $\mathrm{BOS}^{\delta}_{\mathcal{G}_1}(L) \supseteq \mathrm{BOS}^{\delta}_{\mathcal{G}_2}(L)$*;*
(2) *if* $0 \le \delta_1 \le \delta_2 \le 1$ *and* $\mathcal{G} \subseteq C$*, then* $\mathrm{BOS}^{\delta_1}_{\mathcal{G}}(L) \subseteq \mathrm{BOS}^{\delta_2}_{\mathcal{G}}(L)$.

Since $U/L = \{L_1, L_2, \ldots, L_r\}$ is a decision partition, for any $0 \le i, j \le r$, if $i \ne j$, then $L_i \cap L_j = \varnothing$. Therefore, it follows that $|\mathrm{BOS}^{\delta}_{\mathcal{G}}(L)| = \sum_{L_i \in U/L} |\mathrm{BOS}^{\delta}_{\mathcal{G}}(L_i)|$.

Definition 6: *Let* $(U, C \cup L)$ *be a DIS,* $\mathcal{G} \subseteq C$ *and* $\delta \ge 0$*. The uncertainty of the DIS can be defined as:*

$$\rho^{\delta}_{\mathcal{G}}(L) = \frac{|\mathrm{BOS}^{\delta}_{\mathcal{G}}(L)|}{|U|} = \frac{\sum_{L_i \in U/L} |\mathrm{BOS}^{\delta}_{\mathcal{G}}(L_i)|}{|U|} \tag{9}$$

$\rho^{\delta}_{\mathcal{G}}(L)$ is essentially the proportion of boundary objects in the DIS under the attribute subset $\mathcal{G}$ and parameter δ relative to the total number of objects. A larger value of $\rho^{\delta}_{\mathcal{G}}(L)$ indicates that a greater proportion of objects fall into the boundary set relative to the total sample size, implying higher uncertainty of the DIS.

Existing measures evaluate uncertainty from the global perspective of the entire universe. The BOS-based uncertainty directly targets classification-ambiguous boundary objects, with higher pertinence for feature selection.

Proposition 8: *Let* $(U, C \cup L)$ *be a DIS,* $\mathcal{G} \subseteq C$ *and* $\delta \ge 0$*, the following properties hold:*

(1) $0 \le \rho^{\delta}_{\mathcal{G}}(L) \le 1$;
(2) $\rho^{\delta}_{\varnothing}(L) = 1$.

Proposition 9: *For a DIS* $(U, C \cup L)$*, the following conclusions hold:*

(1) *if* $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq C$ *and* $0 \le \lambda \le 1$*, then* $\rho^{\delta}_{\mathcal{G}_1}(L) \ge \rho^{\delta}_{\mathcal{G}_2}(L)$*;*
(2) *if* $0 \le \delta_1 \le \delta_2 \le 1$ *and* $\mathcal{G} \subseteq C$*, then* $\rho^{\delta_1}_{\mathcal{G}}(L) \le \rho^{\delta_2}_{\mathcal{G}}(L)$.

Proposition 10: *For a DIS* $(U, C \cup L)$*, for any* $L_i \in U/L$*,* $\mathrm{BOS}^{\delta}_{\mathcal{G}}(L_i) \subseteq \overline{N}^{\delta}_{\mathcal{G}}(L_i) - \underline{N}^{\delta}_{\mathcal{G}}(L_i)$.

Example 1: *Take a DIS* $(U, C \cup L)$*, where the universe is* $U = \{p_1, p_2, p_3, p_4, q_1, q_2, q_3\}$*. For a given* $\mathcal{G} \subseteq C$*, the distances* $\mathrm{dist}_{\mathcal{G}}$ *on* U *are reported in Table 1. The decision attribute in* L *induces a partition:* $L_1 = \{p_1, p_2, p_3, p_4\}$ *and* $L_2 = \{q_1, q_2, q_3\}$.

Representing the above DIS graphically yields Fig. 4. In Fig. 4, vertices represent objects and the radius $\delta = 0.5$*, with red square points and blue circular points representing objects in* L_1 *and* L_2*, respectively. The set of objects falling within the yellow area constitutes the BOS of this DIS.*

For L_1*,* $\mathrm{BOS}^{\delta}_{\mathcal{G}}(L_1) = \{p_4\}$ *because* $[p_4]^{\delta}_{\mathcal{G}} = \{p_2, p_3, p_4, q_1, q_2\}$*, and the objects* q_1 *and* q_2 *do not belong to* L_1*. Similarly,* $\mathrm{BOS}^{\delta}_{\mathcal{G}}(L_2) = \{q_1, q_2\}$*. Thus,* $\mathrm{BOS}^{\delta}_{\mathcal{G}}(L) = \{p_4, q_1, q_2\}$ *and* $\rho^{\delta}_{\mathcal{G}}(L) = 3/7 \approx 0.43$.

Table 1: A distance relation $\text{dist}_{\mathcal{G}}$ on U.

$\text{dist}_{\mathcal{G}}$	p_1	p_2	p_3	p_4	q_1	q_2	q_3
p_1	0.0	0.34	0.36	0.63	0.9	0.94	1.28
p_2	0.34	0.0	0.45	0.42	0.62	0.76	1.01
p_3	0.36	0.45	0.0	0.42	0.73	0.66	1.07
p_4	0.63	0.42	0.42	0.0	0.31	0.34	0.66
q_1	0.9	0.62	0.73	0.31	0.0	0.32	0.38
q_2	0.94	0.76	0.66	0.34	0.32	0.0	0.47
q_3	1.28	1.01	1.07	0.66	0.38	0.47	0.0

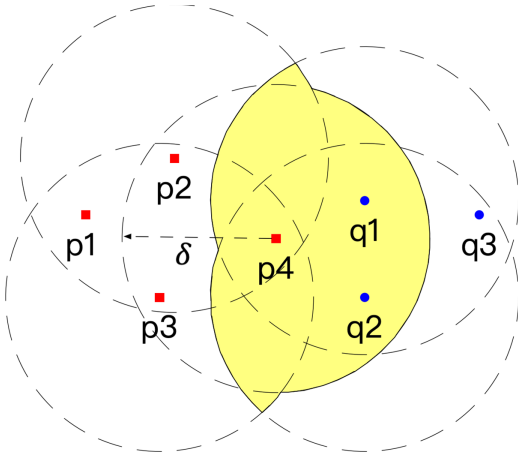

Figure 4: Visualization of the DIS: the yellow area indicates the overlap of circles with radius δ centered on objects of different classes, and the objects within this area form the BOS of the DIS.

4 BOS-Based Feature Selection Algorithm

Feature selection aims to find a compact and informative attribute subset that maintains the discriminative ability of the complete feature space. For DISs, the fundamental objective of feature selection is to quantify the impact of varying attribute combinations on object discriminability with respect to a target partition. This section introduces a novel feature selection methodology utilizing the boundary region of the DIS—as formalized in Definition 6—as the primary uncertainty measure.

4.1 The BOS-Based Feature Selection Algorithm

Definition 7: *Let* $(U, C \cup L)$ *be a DIS,* $\mathcal{G} \subseteq C$ *and* $\delta \geq 0$. *A feature subset* $\mathcal{G} \subseteq C$ *is a reduct of* C *if it meets the following conditions:*

(1) $\rho_{\mathcal{G}}^{\delta}(L) = \rho_{C}^{\delta}(L)$;
(2) *for any* $a \in G$, $\rho_{G-\{a\}}^{\delta}(L) > \rho_{\mathcal{G}}^{\delta}(L)$.

Condition (1) ensures that the reduced attribute set $\mathcal{G}$ maintains the same degree of uncertainty as the full attribute set C. Condition (2) guarantees the minimality of subset $\mathcal{G}$, indicating that no further attributes can be removed without increasing uncertainty.

The contribution of a single attribute to uncertainty reduction during forward selection is captured by the following significance measure.

Definition 8: *Let* $(U, C \cup L)$ *be a DIS,* $\mathcal{G} \subseteq C$, $\delta \geq 0$, *and* $a \in C - G$. *The contribution significance of adding attribute* a *to subset* $\mathcal{G}$ *with respect to decision attribute* L is defined as:

$$\mathrm{SIG}^{\delta}(a, G, L) = \rho^{\delta}_{\mathcal{G}}(L) - \rho^{\delta}_{G\cup\{a\}}(L) \tag{10}$$

Specifically, when the selected $\mathcal{G} = \varnothing$, $\mathrm{BOS}^{\delta}_{\mathcal{G}}(L) = U$, $\rho^{\delta}_{\mathcal{G}}(L)$ reaches its maximum value 1. Adding an attribute a to $\mathcal{G}$ reduces $\rho^{\delta}_{G\cup\{a\}}(L)$. Thus, $\mathrm{SIG}^{\delta}(a, G, L)$ quantifies the reduction in uncertainty contributed by a. Building upon Definitions 7 and 8, we propose the Boundary Object Set Feature Selection (BOSFS) algorithm.

Given a DIS with m attributes and n objects. In addition, k is a reduction factor, where $|\mathrm{BOS}^{\delta}_{\mathcal{G}\cup\{a\}}(L)| = k \cdot |\mathrm{BOS}^{\delta}_{\mathcal{G}}(L)|$; l is the final number of selected attributes in $\mathcal{G}$ and t is the iteration number. The step-by-step complexity analysis of Algorithm 1 is as follows: The outer loop (step 2): runs l times. The distance calculation (step 4): the cost per iteration decreases quadratically: $O(m \cdot (n \cdot k^{t-1})^2)$. The BOS check (step 5–7): the cost is $O(m \cdot n \cdot k^{t-1})$. The total time complexity is the sum of the geometric series across l iterations: $T(n, m, l) = \sum_{t=1}^{l} m \cdot (n \cdot k^{t-1})^2 \leq \frac{m \cdot n^2}{1-k^2}$. Because $k < 1$, the square of the reduction factor k^2 makes the series converge extremely quickly. The simplified complexity is $O(m \cdot n^2)$.

Algorithm 1: Boundary object set feature selection (BOSFS)

Input: A decision information system $(U, C \cup L)$ and a radius $\delta \geq 0$.
Output: An attribute reduct $\mathcal{G}$.

1 Initialize $\mathcal{G} = \varnothing$, $\mathrm{BOS}^{\delta}_{\mathcal{G}}(L) = U$, $\rho^{\delta}_{\mathcal{G}}(L) = 1$, $C^* = C$, $\forall x, y \in U$, $\mathrm{dist}_{\mathcal{G}}(x, y) = 0$;
2 **while** $C^* \neq \varnothing$ *and* $\mathrm{BOS}^{\delta}_{\mathcal{G}}(L) \neq \varnothing$ **do**
3 **for** $a \in C^*$ **do**
4 Compute the distance $\mathrm{dist}_{\mathcal{G}\cup\{a\}}(x, y)$, for all $x, y \in \mathrm{BOS}^{\delta}_{\mathcal{G}}(L)$;
5 **for** $L_i \in U/L$ **do**
6 Check each element in $\mathrm{BOS}^{\delta}_{\mathcal{G}}(L_i)$ to determine $\mathrm{BOS}^{\delta}_{\mathcal{G}\cup\{a\}}(L_i)$;
7 Compute the contribution value $\rho^{\delta}_{\mathcal{G}\cup\{a\}}(L)$;
8 Find the attribute a' that maximizes $\mathrm{SIG}^{\delta}(a', \mathcal{G}, L)$;
9 **if** $\mathrm{SIG}^{\delta}(a', \mathcal{G}, L) > 0$ **then**
10 $\mathcal{G} \leftarrow \mathcal{G} \cup \{a'\}$;
11 $C^* \leftarrow C^* - \{a'\}$;
12 **else**
13 **return** $\mathcal{G}$;
14 **return** the attribute reduct $\mathcal{G}$;

Example 2: *To intuitively illustrate the reduction of the BOS in Algorithm 1 as conditional attributes are added, suppose that an additional conditional attribute* a *is introduced after Example 1. According to Proposition 1, it is straightforward to conclude that the distances between objects increase. Fig. 5 visually depicts the shrinkage of the BOS as the distances among samples become larger.*

To provide a more intuitive illustration of the change in the overlapping area shown in Figs. 4 and 5, Fig. 6 directly compares the corresponding overlap regions in these two figures. It can be observed that the overlapping area decreases as more attributes are introduced.

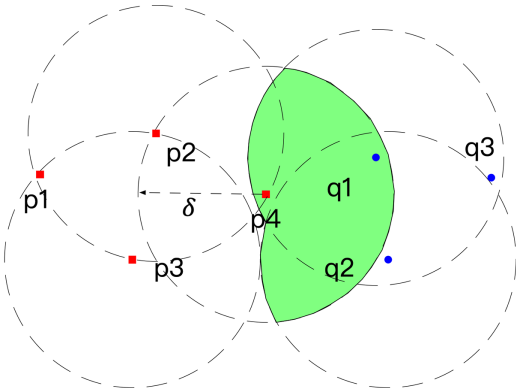

Figure 5: The green region represents the overlapping area, under radius δ, among samples from different classes after selecting attribute a.

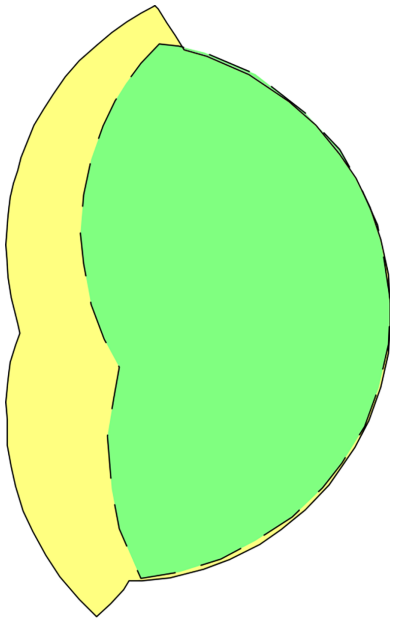

Figure 6: Comparison of the overlapping regions in Figs. 4 and 5, showing that the BOS shrinks after attribute a is added.

It should be noted that Algorithm 1 measures the size of the BOS by the number of objects contained in the overlapping region, rather than by the area of the region itself. In this example, after adding attribute a, the distance between q_2 and p_4 becomes greater than δ. Consequently, the BOS changes from $\{p_4, q_1, q_2\}$ to $\{p_4, q_1\}$, indicating that the number of objects in the set is also reduced.

4.2 Computational Efficiency of BOSFS

Traditional forward greedy selection algorithms reliant on the positive region suffer from computational bottlenecks due to repeated scanning of the entire universe U. In contrast, the BOSFS algorithm achieves superior efficiency via a "double-contraction" strategy targeting both attributes and samples.

The key efficiency gain stems from replacing global universe scans with targeted BOS evaluation. Traditional methods recompute distances and update approximations over all n objects at every iteration, including those that have been correctly classified, which is redundant and increases with dataset size. In BOSFS, once an object exits the boundary and enters the positive region, it will no longer be involved in subsequent iterations, so the effective working set shrinks monotonically.

A second source of speedup is the quadratic contraction of distance computation costs. Conventional algorithms incur $O(n^2)$ distance calculations per candidate attribute per iteration, yielding a per-iteration cost of approximately $O(m \cdot n^2)$. Since BOSFS restricts these computations to the shrinking BOS, if the boundary set contracts by a factor $k < 1$ per iteration, the distance cost at iteration t is $O(m \cdot (n \cdot k^{(t-1)})^2)$, which converges rapidly as a geometric series. Highly discriminative attributes speed up the reduction of BOS, further reducing the effective computation load.

The above algorithm analysis is based on the assumption that $k < 1$. Proposition 7 (1) theoretically guarantees that $k \leq 1$. In the worst case, $k = 1$, which indicates that the number of samples in the BOS does not decrease after adding any unselected attribute. In this situation, SIG = 0, and Algorithm 1 terminates the attribute selection process at Line 13. Therefore, whenever Algorithm 1 performs more than one iteration of attribute selection, the average value of k is strictly less than 1.

Compared with existing acceleration methods [37,38], the main differences of BOSFS lie in both the reduction procedure and the reduction objects.

Zhang et al. [37] achieves acceleration through object selection, which differs from BOSFS in two aspects. First, in [37], object selection and attribute selection are conducted separately: object selection is performed first, followed by attribute selection on the selected object subset. In contrast, BOSFS alternates between boundary object selection and attribute selection, during which the number of samples in the boundary object set continuously decreases. Second, the object selection strategy in [37] mainly removes noisy and atypical objects, whereas BOSFS removes objects that do not contribute to uncertainty. Consequently, BOSFS eliminates more objects, leading to a more significant acceleration effect.

The difference between Yu et al. [38] and BOSFS mainly lies in the research perspective. The acceleration strategy in [38] is achieved by controlling the number of pairwise similarity relations, whereas BOSFS accelerates the process by reducing the number of involved objects.

By dynamically filtering out "solved" data and focusing exclusively on the shrinking boundary, BOSFS effectively converts a high-cost global search into a low-cost local refinement. The inclusion of highly informative attributes accelerates the depletion of the BOS, further expediting algorithm execution.

5 Experiments

This section compares BOSFS with six state-of-the-art feature selection methods. The compared algorithms are briefly described below.

- IOFS [37] (2022): An attribute reduction algorithm based on the sum of instance importance degrees.
- RCE [31] (2024): Utilizes KNN granularity, applying overlap degree for pre-sorting attributes and conditional entropy for uncertainty measurement.

- SNRS [32] (2024): Employs shared-neighbor similarity for neighborhood calculation, using positive-domain dependence degree as the uncertainty metric.
- ARBCM [24] (2025): Calculates neighborhoods based on a multi-β covering dataset, using a composite measure driven by lower-to-upper approximation ratios.
- AMG [33] (2025): Computes sample weights based on distribution, utilizing the average margin of weighted samples as the uncertainty criterion.
- CSFS [38] (2025): An attribute selection algorithm based on the sum of fuzzy similarities between objects from different decision classes.

Algorithm performance is evaluated along three dimensions:

- The cardinality of the chosen attribute subset.
- The time cost incurred during the reduction process.
- The classification efficacy of the resulting subset.

All experiments were carried out using Python 3.8 with the scikit-learn library. The hardware environment consisted of an Intel Xeon Gold 5318Y processor, 128 GB of RAM.

5.1 Experiment Setup

In our experimental study, ten publicly accessible datasets were adopted as benchmarks for algorithm evaluation. Among them, six datasets came from the UCI [39], while the remaining were obtained from the KRBD [40]. Table 2 provides detailed descriptions of these datasets.

Table 2: Data description.

No.	Dataset	Samples	Attributes	Classes
D1	wine	178	13	3
D2	glass	214	9	3
D3	ionosphere	351	34	2
D4	gallstone	319	38	2
D5	german	1000	20	2
D6	liverdisorder	345	6	2
D7	breastCancer	97	24,481	2
D8	DLBCLOutcome	58	7129	2
D9	DLBCLNIH	240	7399	2
D10	AMLALL	72	7129	2

To compare and assess the quality of the reducts, we selected three classic classifiers that utilize the selected subsets as their feature set for training and testing the data. The performance of the feature subsets was evaluated using KNN, SVM, and DT (Decision Tree) classifiers. All classifier training and testing were conducted using ten-fold cross-validation. Statistical performance was quantified using the mean accuracy alongside its corresponding standard deviation.

It is necessary to specify the parameter settings for the models. Except for AMG, the other comparative algorithms require the setting of certain hyperparameters. To ensure a fair comparison in the experiments, considering efficiency and the number of parameters, Table 3 provides the detailed hyperparameter configuration scheme.

Table 3: Hyperparameter configuration scheme.

No	Algorithm	Parameter
1	BOSFS	δ from 0.02 to 0.2 in increments of 0.02
2	IOFS	δ from 0.0001 to 0.01 in increments of 0.001
3	RCE	k' from 0.05 to 0.5 in increments of 0.05
4	SNRS	δ from 0.02 to 0.2 in increments of 0.02
5	ARBCM	β from 0.5 to 1 in increments of 0.05
6	CSFS	λ from 0.1 to 1 in increments of 0.1

5.2 Experimental Analysis

Using the three previously defined metrics, we compare our algorithm with the benchmark methods in this subsection. These results confirm that BOSFS reliably identifies compact, high-quality feature subsets with substantially lower computational cost.

The reduct sizes of the seven algorithms across the ten datasets are presented in Table 4. Cross-referencing Table 4 with the original dimensions in Table 2, all seven algorithms achieve substantial dimensionality reduction: the mean selected subset sizes range from 2.2 to 48.6, corresponding to a dimensionality reduction rate of about 98.7%–99.9%.

The peak classification performance attained by the feature subsets extracted via the seven competitive algorithms is summarized in Tables 5–7. These results span multiple benchmark datasets and are validated across three classifiers. The bold entries indicate the highest classification accuracy reached by the features selected by the corresponding algorithm. Across all three tables, BOSFS consistently achieves competitive or leading classification accuracy. Among the 30 experimental settings (10 datasets × 3 classifiers), BOSFS attains the highest classification accuracy in 15 instances. SNRS leads by securing the top accuracy in 8 experimental scenarios, closely followed by AMG and CSFS with 7 instances of optimal results. IOFS, RCE and ARBCM achieve the highest accuracy in only 5, 3 and 1 case(s), respectively. Overall, compared to the other six feature selection algorithms, BOSFS shows clearly superior performance across the three classifiers.

The temporal costs associated with each selection method are documented in Table 8. These durations correspond to the specific parameter configurations that yielded maximum precision for each classifier. Figs. 7–9 provide a more intuitive comparison of the performance of various feature selection algorithms. Note that the vertical axis in Figs. 7–9 uses a logarithmic scale to accommodate the wide range of runtimes. Among all compared methods, BOSFS is the fastest by a substantial margin. CSFS ranks second, yet BOSFS's average runtime is only 25% of CSFS's—a gap that persists consistently across all 10 datasets, as detailed in Table 8 and illustrated in Figs. 7–9.

Table 4: Size of the reduct under optimal parameters.

	SVM						DT							KNN							
	RCE	**SNRS**	**ARBCM**	**AMG**	**IOFS**	**CSFS**	**BOSFS**	**RCE**	**SNRS**	**ARBCM**	**AMG**	**IOFS**	**CSFS**	**BOSFS**	**RCE**	**SNRS**	**ARBCM**	**AMG**	**IOFS**	**CSFS**	**BOSFS**
D1	13	6	2	11	13	8	7	5	6	2	11	6	4	4	10	5	2	11	13	5	7
D2	9	7	3	9	6	9	8	8	6	3	9	7	8	6	5	6	3	9	6	9	6
D3	29	11	4	32	1	25	9	16	11	4	32	1	25	12	16	6	4	32	1	4	7
D4	7	10	2	13	10	10	11	4	15	2	13	2	3	12	4	5	2	13	2	9	11
D5	14	9	2	17	20	19	10	13	12	2	17	20	13	10	13	10	2	17	20	19	10
D6	6	6	3	5	6	5	5	6	6	3	5	6	5	5	6	6	3	5	6	5	5
D7	5	4	2	117	1	5	5	6	5	2	117	1	3	3	7	4	3	117	1	6	5
D8	7	4	1	96	45	5	4	8	4	1	96	52	6	4	5	2	1	96	47	5	4
D9	9	6	2	53	176	6	7	11	5	2	53	107	7	3	11	6	2	53	164	6	7
D10	8	2	1	133	45	4	2	5	2	1	133	42	3	4	4	2	1	133	41	4	2
Ave.	10.7	6.5	2.2	48.6	32.3	9.6	6.8	8.2	7.2	2.2	48.6	24.4	7.7	6.3	8.1	5.2	2.3	48.6	30.1	7.2	6.4

Table 5: Accuracy (%) of attribute subsets with KNN classifier.

Dataset	RCE	SNRS	ARBCM	AMG	IOFS	CSFS	BOSFS
D1	96.7 ± 4.44	**98.3 ± 3.56**	90 ± 7.35	**98.3 ± 2.6**	96.6 ± 3.71	**98.3 ± 3.56**	97.8 ± 3.69
D2	**68.7 ± 11.7**	**68.7 ± 11.8**	65 ± 11.3	67.3 ± 8.98	64.9 ± 9.78	67.3 ± 8.98	**68.7 ± 11.8**
D3	85.2 ± 7.55	89.7 ± 5.31	91.4 ± 4.44	84 ± 5.29	35.9 ± 1.35	86.6 ± 6.43	**92.3 ± 4.56**
D4	70.8 ± 9.85	67.1 ± 7.24	55.8 ± 9.53	71.8 ± 7.62	**73.9 ± 9.45**	68 ± 9.28	67.4 ± 6.16
D5	73.2 ± 6.48	**73.6 ± 4.86**	62.7 ± 2.41	71.8 ± 7.6	73.2 ± 6.48	72.7 ± 6.5	**73.6 ± 4.86**
D6	63.2 ± 9.54	64 ± 9.39	57.4 ± 5.87	63.2 ± 5.91	63.2 ± 9.54	**66.1 ± 5.22**	**66.1 ± 5.22**
D7	76.1 ± 13.4	77.4 ± 10.1	72.4 ± 15.9	**83.8 ± 10.1**	47.3 ± 3.49	76.3 ± 14.9	77.7 ± 12.4
D8	72.3 ± 13.6	76 ± 7.72	69 ± 18	74.3 ± 15.1	72.7 ± 17.4	81 ± 13.8	**83.3 ± 14.9**
D9	59.2 ± 7.64	61.7 ± 8.29	56.2 ± 10.4	66.2 ± 9.76	**70 ± 10.5**	63.3 ± 10.2	61.7 ± 8.7
D10	90.4 ± 10.6	95.7 ± 6.55	66.6 ± 17.7	95.7 ± 6.55	87.7 ± 11.4	**100 ± 0**	97.1 ± 5.71

Note: Bold values indicate the highest classification accuracy obtained under the corresponding dataset.

Table 6: Accuracy (%) of attribute subsets with SVM classifier.

Dataset	RCE	SNRS	ARBCM	AMG	IOFS	CSFS	BOSFS
D1	**98.3 ± 2.55**	**98.3 ± 2.55**	92.2 ± 7.92	**98.3 ± 2.55**	**98.3 ± 2.55**	**98.3 ± 2.55**	**98.3 ± 2.55**
D2	65 ± 9.73	65 ± 9.73	**65.5 ± 9.03**	65 ± 9.73	64.5 ± 9.48	65 ± 9.73	**65.5 ± 9.49**
D3	94 ± 4.51	94 ± 5.02	91.5 ± 4.59	93.4 ± 5.12	64.1 ± 1.35	94 ± 4.69	**94.3 ± 5.57**
D4	71.5 ± 8.05	71.5 ± 8.67	59.5 ± 13.4	71.8 ± 8.36	72.4 ± 7.32	67.4 ± 5.34	**73.6 ± 5.72**
D5	**74.7 ± 8.28**	74.5 ± 5.16	70.3 ± 1.9	73.7 ± 5.93	73.1 ± 7.71	74 ± 7.56	74 ± 6.02
D6	67.3 ± 4.66	67.3 ± 4.66	57.1 ± 5.08	**69.6 ± 8.82**	67.3 ± 4.66	68.1 ± 3.64	68.1 ± 3.64
D7	77.2 ± 12.3	77.6 ± 12.5	75.3 ± 17.4	**87.8 ± 11.6**	52.7 ± 3.49	78.6 ± 12.2	80.6 ± 12.1
D8	73 ± 13.1	71.3 ± 14.8	55.7 ± 14.4	83 ± 14.9	73 ± 19.9	78.7 ± 14.9	**86.7 ± 16.3**
D9	61.7 ± 7.64	61.3 ± 6.19	59.6 ± 5.29	**72.1 ± 5.91**	65.8 ± 6.67	66.7 ± 7.68	67.1 ± 7.32
D10	94.6 ± 8.64	94.5 ± 9.33	62.5 ± 8.89	**100 ± 0**	**100 ± 0**	98.6 ± 4.29	97.1 ± 5.71

Note: Bold values indicate the highest classification accuracy obtained under the corresponding dataset.

Table 7: Accuracy (%) of attribute subsets with DT classifier.

Dataset	RCE	SNRS	ARBCM	AMG	IOFS	CSFS	BOSFS
D1	89.9 ± 4.09	**95 ± 6.78**	88.9 ± 7.83	86.5 ± 7.67	91 ± 7.37	**95 ± 5.24**	94.4 ± 6.09
D2	62.1 ± 15.2	**65.9 ± 17.1**	60.3 ± 10.1	59.8 ± 16	56.6 ± 13.7	62.1 ± 15.2	**65.9 ± 17.1**
D3	93.2 ± 5.4	92.3 ± 3.85	87.5 ± 4.73	88.3 ± 4.51	64.1 ± 1.35	90 ± 4.28	**93.5 ± 3.83**
D4	68.6 ± 15.3	69.8 ± 11.1	58 ± 7.22	65.8 ± 10.2	**73.9 ± 9.75**	62.4 ± 9.37	72.7 ± 8.49
D5	68.7 ± 3.23	**71 ± 4.4**	56.5 ± 3.56	66.7 ± 3.93	67.8 ± 2.75	68.1 ± 4.37	70.6 ± 4.48
D6	62.9 ± 6.49	62.9 ± 6.49	54.8 ± 8.95	62.6 ± 7.03	62.9 ± 6.49	**64.1 ± 9.45**	**64.1 ± 9.45**
D7	72.4 ± 13.8	76.4 ± 11.7	66.8 ± 11.7	72 ± 9.7	52.7 ± 3.49	68.4 ± 17.5	**85.8 ± 7.9**
D8	75.7 ± 15.7	79.6 ± 13.52	63.7 ± 14.8	62 ± 15.3	65.3 ± 16.9	72 ± 18.4	**80.7 ± 10.2**
D9	57.9 ± 8.63	62.5 ± 12.2	60 ± 9.9	59.2 ± 10.5	61.7 ± 8.08	**63.7 ± 11.9**	62.5 ± 9.32
D10	94.5 ± 6.8	**98.6 ± 4.29**	69.3 ± 10.5	90.5 ± 10.5	90.7 ± 10	97.1 ± 5.71	97.3 ± 5.37

Note: Bold values indicate the highest classification accuracy obtained under the corresponding dataset.

Table 8: Time consumption (seconds) under optimal parameters.

	KNN						SVM							DT							
	RCE	**SNRS**	**ARBCM**	**AMG**	**IOFS**	**CSFS**	**BOSFS**	**RCE**	**SNRS**	**ARBCM**	**AMG**	**IOFS**	**CSFS**	**BOSFS**	**RCE**	**SNRS**	**ARBCM**	**AMG**	**IOFS**	**CSFS**	**BOSFS**
D1	7.13	0.585	8.31	3.98	3.57	0.457	0.096	5.84	0.597	8.63	3.98	3.57	0.537	0.066	6.99	0.597	8.63	3.98	2.9	0.422	0.09
D2	11.5	0.33	15	5.17	5.32	0.533	0.077	4.42	0.379	15	5.17	5.32	0.533	0.067	4.42	0.821	15	5.17	4.67	0.309	0.067
D3	130	13.5	251	24.7	8.24	3.6	1.07	130	6.98	251	24.7	8.24	8	1.27	61.4	4.39	132	24.7	8.24	7.37	0.788
D4	172	4.62	90.3	16.7	8.45	4.67	1.17	255	2.69	85	16.7	8.45	5.06	1.26	84.8	2.69	85	16.7	8.45	3.06	0.606
D5	901	39.1	540	87.9	127	24	4.08	948	34.4	540	87.9	127	24	4.08	820	53.5	540	87.9	127	23.4	4.08
D6	6	0.509	27.8	2.62	2.63	0.735	0.098	6	0.509	27.8	2.62	2.63	0.735	0.098	4.4	0.509	27.8	2.62	2.63	0.735	0.098
D7	1490	353	3340	12800	397	216	71.6	1900	460	3300	12800	397	191	47.9	1490	353	6110	12800	397	149	71.6
D8	116	63.6	214	1470	713	27.3	8.6	112	70.1	214	1470	713	27.3	8.6	49.3	20.4	214	1470	777	27.5	8.6
D9	4430	642	7490	7240	36447	392	102	2380	830	6140	7240	39000	392	46.5	4430	830	5980	7240	23400	414	102
D10	74.7	95.3	235	2600	882	29.8	6.71	85.3	57	235	2600	882	29.2	11.6	144	45.5	235	2600	936	26.5	6.71
Ave.	734	121	1220	2430	3850	69.9	19.6	583	146	1080	2430	4130	67.8	12.1	710	131	1330	2430	2570	65.2	19.5

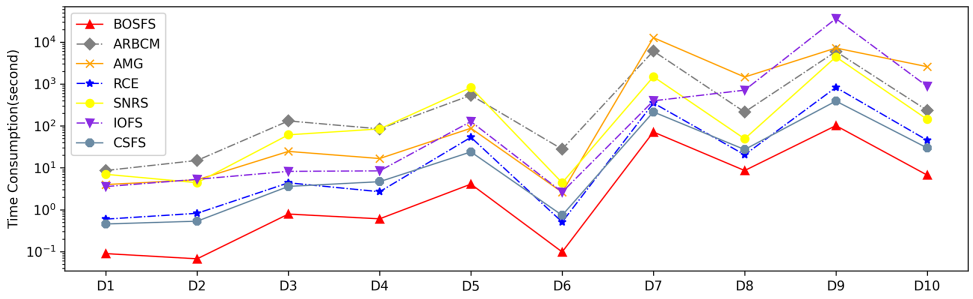

Figure 7: The runtime using optimal parameters for KNN.

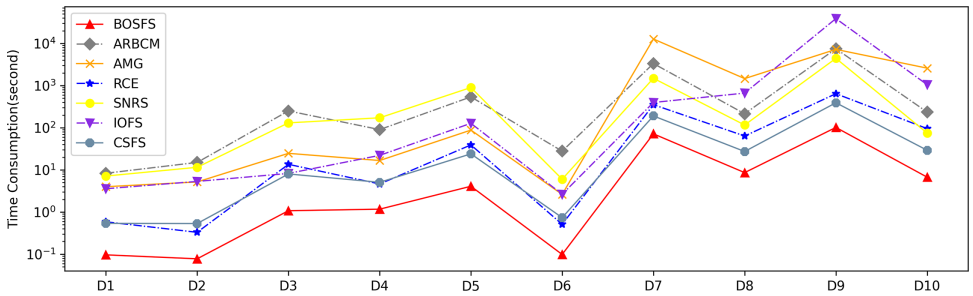

Figure 8: The runtime using optimal parameters for SVM.

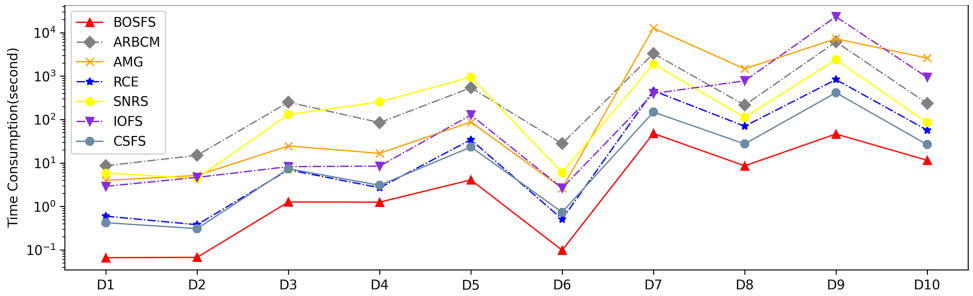

Figure 9: The runtime using optimal parameters for DT.

Among all compared algorithms, BOSFS exhibits the most stable performance in terms of both classification accuracy and runtime. Its standard deviations are consistently low across datasets, and its runtime remains nearly constant regardless of dataset dimensionality, whereas other methods show high variability.

5.3 Statistical Test

To ensure the reliability of our findings, we conduct a statistical analysis to comprehensively evaluate the superiority of algorithm BOSFS in classification accuracy and runtime. During the upcoming statistical tests, we employ three widely used statistical testing methods: the Friedman test, the Bonferroni-Dunn test and the pairwise Wilcoxon signed-rank test [41], to thoroughly assess the significance of our findings.

Initially, drawing from Tables 5–7, we present the rankings of all algorithms across each dataset as outlined in Table 9.

Table 9: The rankings of all algorithms across each dataset.

	KNN							SVM							DT						
	RCE	**SNRS**	**ARBCM**	**AMG**	**IOFS**	**CSFS**	**BOSFS**	**RCE**	**SNRS**	**ARBCM**	**AMG**	**IOFS**	**CSFS**	**BOSFS**	**RCE**	**SNRS**	**ARBCM**	**AMG**	**IOFS**	**CSFS**	**BOSFS**
D1	5	2	7	2	6	2	4	3.5	3.5	7	3.5	3.5	3.5	3.5	5	1.5	6	7	4	1.5	3
D2	2	2	6	4.5	7	4.5	2	4.5	4.5	1.5	4.5	7	4.5	1.5	3.5	1.5	5	6	7	3.5	1.5
D3	5	3	2	6	7	4	1	3	3	6	5	7	3	1	2	3	6	5	7	4	1
D4	3	6	7	2	1	4	5	4.5	4.5	7	3	2	6	1	4	3	7	5	1	6	2
D5	3.5	1.5	7	6	3.5	5	1.5	1	2	7	5	6	3.5	3.5	3	1	7	6	5	4	2
D6	5	3	7	5	5	1.5	1.5	5	5	7	1	5	2.5	2.5	4	4	7	6	4	1.5	1.5
D7	5	3	6	1	7	4	2	5	4	6	1	7	3	2	3	2	6	4	7	5	1
D8	6	3	7	4	5	2	1	4.5	6	7	2	4.5	3	1	3	2	6	7	5	4	1
D9	6	4.5	7	2	1	3	4.5	5	6	7	1	4	3	2	7	2.5	5	6	4	1	2.5
D10	5	3.5	7	3.5	6	1	2	5	6	7	1.5	1.5	3	4	4	1	7	6	5	3	2
Ave.	4.55	3.15	6.3	3.6	4.85	3.1	2.45	4.1	4.45	6.25	2.75	4.75	3.5	2.2	3.85	2.15	6.2	5.8	4.9	3.35	1.75

We set the significance level α to 0.05 for rigorous statistical testing. With 7 algorithms and 10 datasets, the threshold for Friedman test is $F_{0.05}(6, 54) = 2.266$. By consulting Table 9, we are able to determine the values of the Friedman test statistics for Classifiers KNN, SVM, and DT, which are 5.31, 5.73, and 15.75, respectively. Clearly, these statistical values are all greater than the threshold, indicating significant differences in performance among these algorithms.

To demonstrate the statistical superiority of Algorithm BOSFS over other algorithms, we further conduct a Bonferroni-Dunn test. Given the presence of 7 algorithms and 10 datasets, we can determine the critical value for the Bonferroni-Dunn test to be $CD_\alpha = 2.489$. Fig. 10 presents the CD diagram for the three classifiers, with the red line indicating the value of CD.

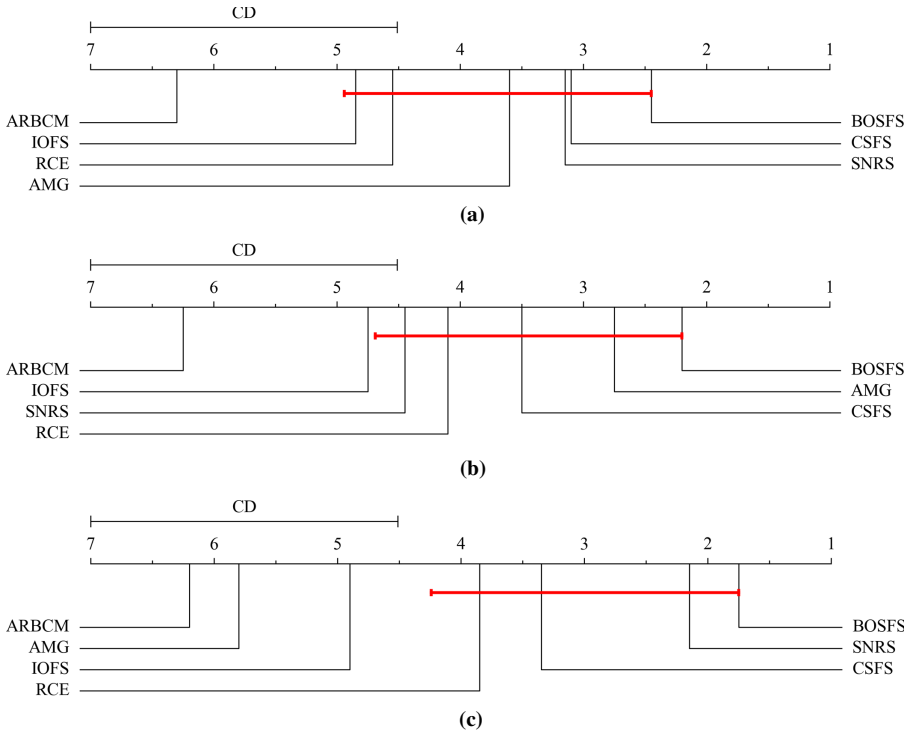

Figure 10: CD diagram of seven algorithms. (**a**) KNN; (**b**) SVM; (**c**) DT.

The differences between BOSFS and ARBCM, IOFS, and AMG exceed the CD threshold on at least one classifier, confirming that BOSFS significantly outperforms ARBCM on all three classifiers, significantly outperforms IOFS on the SVM and DT classifiers, and significantly outperforms AMG on the DT classifier. Although there is no significant statistical difference between BOSFS and SNRS, CSFS and RCE in terms of feature selection quality, BOSFS still holds a clear advantage over them when considering the feature selection required time.

To further validate the superiority of BOSFS at the dataset level, we conduct pairwise one-sided Wilcoxon signed-rank tests between BOSFS and each of the six competitors. To control the family-wise error rate across six simultaneous comparisons per classifier, Holm–Bonferroni correction is applied. The effect size is quantified by Cliff's d, with thresholds $|d| < 0.147$ (negligible), $0.147 \leq |d| < 0.330$ (small), $0.330 \leq |d| < 0.474$ (medium), and $|d| \geq 0.474$ (large). The results of the pairwise one-sided Wilcoxon signed-rank tests conducted on classification accuracy and running time are shown in Figs. 11 and 12, respectively.

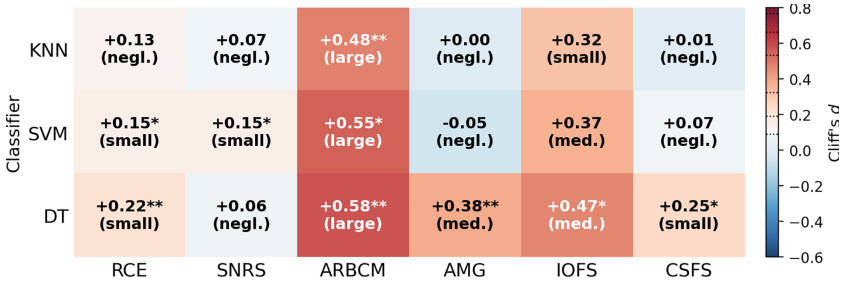

Figure 11: Accuracy: Cliff's d (BOSFS vs. competitor). **Holm $p < 0.01$, *Holm $p < 0.05$.

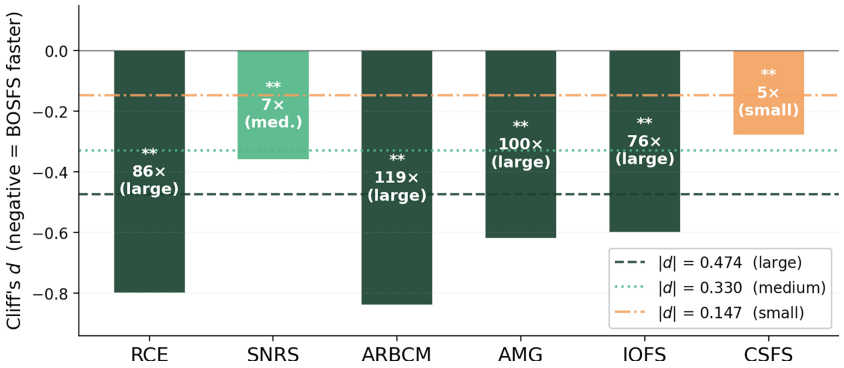

Figure 12: Runtime: Cliff's d, BOSFS vs. competitor. **Holm $p < 0.01$.

Accuracy results show that BOSFS significantly outperforms ARBCM across all three classifiers. Under the DT classifier, BOSFS is significantly superior to five of the six competitors. Under SVM, BOSFS significantly outperforms RCE, SNRS, IOFS, and ARBCM. Although no significant difference is found between BOSFS and AMG/CSFS under KNN and SVM, all Cliff's d values are non-negative, indicating that BOSFS never performs worse in expectation.

Runtime results are even more decisive. The one-sided Wilcoxon test rejects H_0 at $p < 0.01$ for every competitor. The effect sizes are large against RCE, ARBCM, AMG, and IOFS, with mean speedup ratios of 86×, 119×, 100×, and 76×, respectively. Even against the second-fastest competitor CSFS, the test is significant with a mean speedup of 4.7×. These results provide statistically rigorous confirmation that the computational advantage of BOSFS is not a dataset-specific artifact but a consistent, significant property.

5.4 Analysis of the Size of the BOS

In Section 4.1, an attribute reduction factor k was introduced, and the time complexity of the algorithm was analyzed based on this factor. In this section, we further conduct a statistical analysis of the empirical values of k on real-world datasets. Since the parameter δ influences the value of k, we uniformly adopt the value of δ that yields the highest classification accuracy under the SVM classifier for the estimation of k.

Fig. 13 illustrates the relationship between the iteration rounds of feature selection and the ρ value for each dataset. As observed from Fig. 13, the size of the BOS exhibits an exponential decreasing trend with respect to the feature selection iterations. Consequently, as the number of iterations increases, the number of objects involved in the computation of ρ rapidly declines, leading to a significant reduction in computational cost.

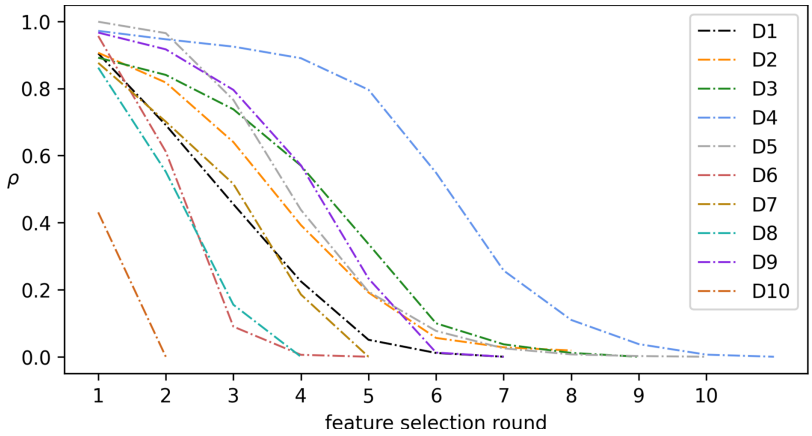

Figure 13: The relationship between the ρ value and the feature selection rounds.

Table 10 reports the value of k at each iteration, along with the average k value for each dataset. As observed from Table 10, the average values of k across different datasets range from 0.21 to 0.64, with an overall mean of 0.495 over the ten datasets. According to the analysis in Section 4.1, the size of the BOS exhibits an exponential relationship with respect to k as the iterations proceed. Specifically, a smaller value of k leads to a faster reduction in the size of the BOS.

Table 10: The relationship between k value and the feature rounds.

Round	1	2	3	4	5	6	7	8	9	10	11	Ave.
D1	0.9	0.76	0.66	0.49	0.23	0.22	0.0	–	–	–	–	0.47
D2	0.91	0.9	0.78	0.61	0.49	0.29	0.5	0.67	–	–	–	0.64
D3	0.89	0.94	0.88	0.77	0.59	0.3	0.37	0.31	0.0	–	–	0.56
D4	0.97	0.97	0.98	0.96	0.89	0.69	0.47	0.43	0.34	0.17	0.0	0.62
D5	0.99	0.97	0.79	0.57	0.44	0.4	0.32	0.28	0.29	0.0	–	0.51
D6	0.96	0.64	0.15	0.06	0.0	–	–	–	–	–	–	0.36
D7	0.88	0.8	0.74	0.36	0.0	–	–	–	–	–	–	0.56
D8	0.86	0.64	0.28	0.0	–	–	–	–	–	–	–	0.45
D9	0.97	0.95	0.87	0.72	0.41	0.05	0.0	–	–	–	–	0.57
D10	0.43	0.0	–	–	–	–	–	–	–	–	–	0.21

The above two observations, derived from empirical data statistics, validate the theoretical complexity analysis presented in Section 4.1 as well as the efficiency analysis of the BOSFS algorithm in Section 4.2.

5.5 Analysis of the Hyperparameter δ

The radius parameter δ, which controls neighborhood granularity, significantly affects both the reduct size and the final classification accuracy. To explore how δ affects selection results across datasets with different dimensionalities, we tested δ values from 0.005 to 0.4 in increments of 0.005 for each dataset. Based on dimensionality, the datasets were divided into two groups for analysis: low-dimensional datasets (dimensionality < 1000: D1, D2, . . ., D6) and high-dimensional datasets (dimensionality ≥ 1000: D7, D8, D9, and D10). We analyzed both the cardinality of the attribute subsets obtained under different δ values and the attribute subsets' performance in three classifiers.

5.5.1 Influence on Reduct Cardinality

To analyze the relationship linking δ to the cardinality of the reduct, we employed the selection ratio, defined as the ratio of the selected attribute subset to the complete attribute set. Fig. 14 illustrates the relationship between δ and the selection ratios for low-dimensional and high-dimensional datasets. Fig. 14 reveals several consistent patterns:

(1) The neighborhood radius δ exhibits an overall positive correlation with the selection ratio, which is consistent with theoretical expectations. As δ increases, the neighborhood sample sets of individual objects expand, leading to an increase in the BOS of the target partition, which correspondingly increases the DIS uncertainty. Consequently, more condition attributes are required to reduce the DIS uncertainty.

(2) The relationship between δ and the selection ratio does not uniformly exhibit the expected positive correlation. In certain limiting cases—for example, D5 yields an empty attribute subset when δ is in the interval $[0.2, 0.4]$—this occurs because in Algorithm 1, when δ takes an excessively large value, $\mathrm{SIG}^{\delta}(a, G, L)$ equals zero for any newly added condition attribute, as selecting any condition attribute fails to decrease the value of $\rho_{G}^{\delta}(L)$. Consequently, the attribute selection algorithm terminates.

(3) Compared with low-dimensional datasets, the BOSFS algorithm exhibits more effective attribute selection for high-dimensional datasets. Even when $\delta = 0.4$, the selection ratio remains within the range of 0.05%–2%.

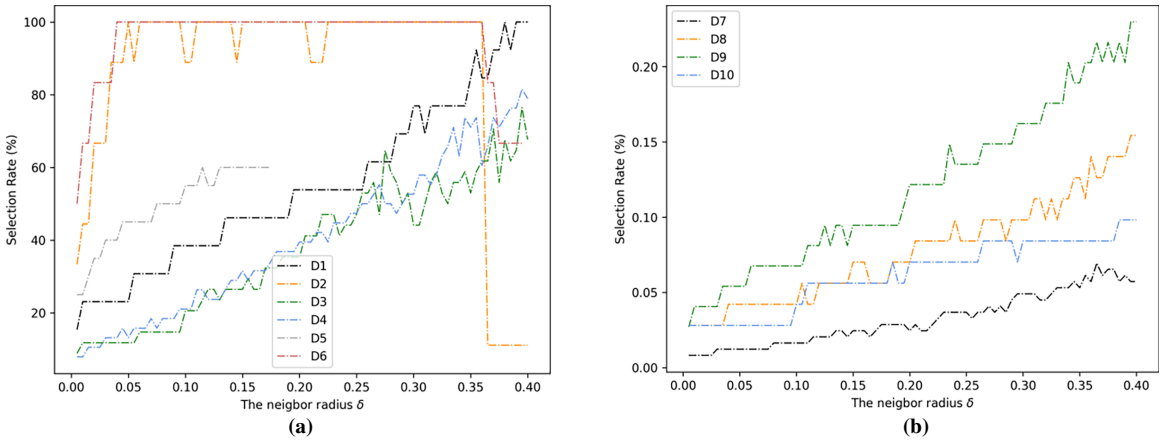

Figure 14: Attribute selection rate vs. δ on (**a**) low-dimensional and (**b**) high-dimensional datasets.

5.5.2 Influence on Classification Accuracy

A comparative analysis of classification accuracies relative to δ is presented in Fig. 15. Unlike the monotonic trend seen for selection ratio, classifier accuracy shows no consistent relationship with δ for either low- or high-dimensional datasets.

Based on the above analysis, the influence of δ on the BOSFS algorithm can be summarized as follows:

(1) A larger value of δ generally leads to a higher feature selection rate, i.e., a larger number of selected features.

(2) The value of δ shows no clear positive or negative correlation with the classification performance of the selected feature subset.

(3) An excessively large δ may cause the algorithm to terminate prematurely.

(4) For the same δ interval, the feature selection rate on high-dimensional datasets is significantly lower than that on low-dimensional datasets, indicating that the feature selection benefit is more pronounced for high-dimensional data.

Based on these observations, the practical guidelines for choosing δ are summarized as follows:

Tip 1: The dataset should first be normalized to the range $[0,1]$ using Min-Max normalization to eliminate the influence of different attribute scales.

Tip 2: Setting δ within the interval $[0.01, 0.1]$ is generally appropriate, as it prevents the selected feature subset from becoming excessively large while maintaining stable performance in downstream models.

Tip 3: If the feature selection rate is of particular concern, we recommend $\delta = 0.05$ as a suitable choice.

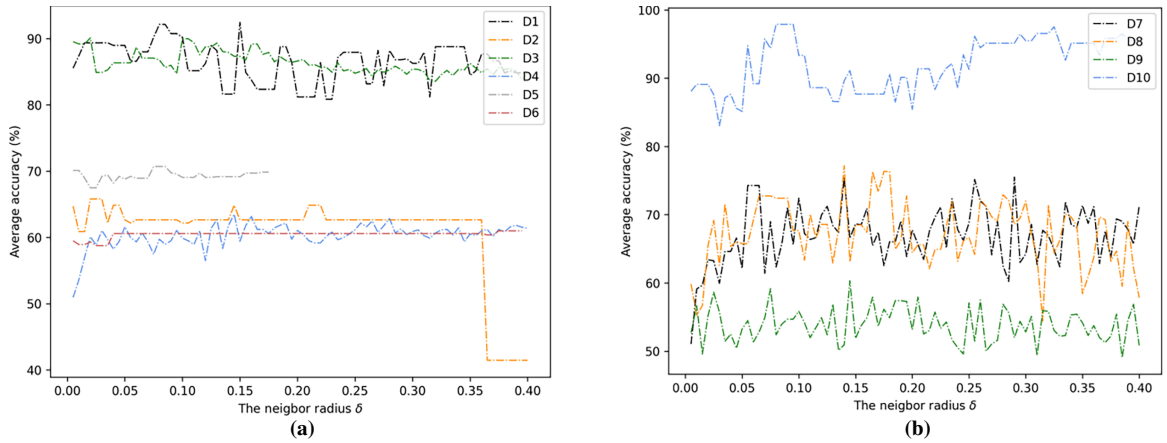

Figure 15: Classification accuracy vs. δ on (**a**) low-dimensional and (**b**) high-dimensional datasets.

6 Conclusion

In this paper, we addressed the computational inefficiency of neighborhood rough set-based feature selection caused by repeatedly scanning the entire object universe. A novel BOS-based uncertainty measure is proposed, together with a corresponding forward greedy feature selection algorithm called BOSFS. The BOS-based uncertainty measure $\rho_{\mathcal{G}}^{\delta}(L)$ quantifies the proportion of boundary objects relative to the total sample size. The core innovation of BOSFS lies in its "double-contraction" strategy, which restricts computation exclusively to the shrinking boundary object sets across iterations, reducing the effective computational cost from $O(m^2 \cdot n^2)$ down to $O(m \cdot n^2)$. Experiments on ten benchmark datasets, conducted against six state-of-the-art algorithms under three classifiers, show that BOSFS achieves the highest classification accuracy in 15 of 30 cases while running in only 25% of the time required by the second-fastest competitor. Nevertheless, several limitations remain. The performance of BOSFS is sensitive to the neighborhood radius δ, and an adaptive selection strategy for this parameter warrants further investigation. When δ is excessively large, the algorithm may terminate prematurely with a suboptimal reduct, requiring a more robust mechanism. Furthermore, the current framework relies on a fixed Minkowski distance metric, and extending it to fuzzy relations or adaptive distance measures for heterogeneous data is a valuable future direction. Integrating BOSFS with wrapper-based strategies and evaluating its scalability on large-scale datasets are also open issues for subsequent research.

Acknowledgement: None.

Funding Statement: This work was supported by the Anhui Provincial Department of Education University Research Project (2024AH051375, 2024AH051368) and industry-sponsored research project from Nanjing Wenstone Information Technology Co., Ltd. (2025HX0249).

Author Contributions: The authors confirm their contribution to the paper as follows: study conception and design: Wenchang Yu; analysis and interpretation of results: Xiaoqin Ma, Zheqing Zhang; draft manuscript preparation: Kezhong Lu. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: Not applicable.

Ethics Approval: Not applicable.

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

References

1. Wang C, Wang C, Qian Y, Leng Q. Feature selection based on weighted fuzzy rough sets. IEEE Trans Fuzzy Syst. 2024;32(7):4027–37. doi:10.1109/TFUZZ.2024.3387571.
2. Xia S, Bai X, Wang G, Cheng Y, Meng D, Gao X, et al. An efficient and accurate rough set for feature selection, classification, and knowledge representation. IEEE Trans Knowl Data Eng. 2022;35(8):7724–35. doi:10.1109/TKDE.2022.3220200.
3. Pawlak Z. Rough sets. Int J Comput Inf Sci. 1982;11(5):341–56. doi:10.1007/BF01001956.
4. Pawlak Z. Rough sets: theoretical aspects of reasoning about data. Dordrecht, The Netherlands: Kluwer Academic Publishers; 1991. doi:10.1007/978-94-011-3534-4.
5. Fatima S, Akram M, Zafar F. A hybrid decision-making technique based on extended entropy and trapezoidal fuzzy rough number. J Appl Math Comput. 2024;70(5):4755–92. doi:10.1007/s12190-024-02150-z.
6. Kumar A, Prasad PS. Scalable feature subset selection with fuzzy rough sets and fuzzy min-max neural network in hybrid decision system. IEEE Trans Fuzzy Syst. 2024;33(2):669–79. doi:10.1109/TFUZZ.2024.3488074.
7. Chen L, Deng Y. GDTRSET: a generalized decision-theoretic rough sets based on evidence theory. Artif Intell Rev. 2023;56(3):3341–62. doi:10.1007/s10462-023-10605-1.
8. Ali W, Shaheen T, Toor HG, Alballa T, Alburaikan A, Khalifa HA. An improved intuitionistic fuzzy decision-theoretic rough set model and its application. Axioms. 2023;12(11):1003. doi:10.3390/axioms12111003.
9. Yu B, Hu Y, Dai J. A bi-variable precision rough set model and its application to attribute reduction. Inf Sci. 2023;645(1):119368. doi:10.1016/j.ins.2023.119368.
10. Sang B, Chen H, Yang L, Wan J, Li T, Xu W. Feature selection considering multiple correlations based on soft fuzzy dominance rough sets for monotonic classification. IEEE Trans Fuzzy Syst. 2022;30(12):5181–95. doi:10.1109/TFUZZ.2022.3169625.
11. Dai J, Zou X, Qian Y, Wang X. Multi-fuzzy β-covering approximation spaces and their information measures. IEEE Trans Fuzzy Syst. 2022;31(3):955–69. doi:10.1109/TFUZZ.2022.3193448.
12. Zhang K, Dai J. Redefined fuzzy rough set models in fuzzy β-covering group approximation spaces. Fuzzy Sets Syst. 2022;442(8):109–54. doi:10.1016/j.fss.2021.10.012.
13. Al-shami TM, Mhemdi A. Overlapping containment rough neighborhoods and their generalized approximation spaces with applications. J Appl Math Comput. 2025;71(1):869–900. doi:10.1007/s12190-024-02261-7.
14. Al-shami TM, Hosny RA, Mhemdi A, Hosny M. Cardinality rough neighborhoods with applications. AIMS Math. 2024;9(11):31366–92. doi:10.3934/math.20241511.
15. Hu Q, Yu D, Xie Z. Neighborhood classifiers. Expert Syst Appl. 2008;34(2):866–76. doi:10.1016/j.eswa.2006.10.043.
16. Yang L, Sang B, Xu W, Chen H, Yuan Z, Qin K. QFIG: a novel attribute reduction method using conditional entropy in quantified fuzzy approximation space. Fuzzy Sets Syst. 2025;516(2):109443. doi:10.1016/j.fss.2025.109443.

17. Hu Q, Yu D, Xie Z, Liu J. Fuzzy probabilistic approximation spaces and their information measures. IEEE Trans Fuzzy Syst. 2006;14(2):191–201. doi:10.1109/TFUZZ.2005.864086.
18. Xu J, Meng X, Qu K, Sun Y, Hou Q. Feature selection using relative dependency complement mutual information in fitting fuzzy rough set model. Appl Intell. 2023;53(15):18239–62. doi:10.1007/s10489-022-04445-9.
19. Chen Y, Zhang Z, Zheng J, Ma Y, Xue Y. Gene selection for tumor classification using neighborhood rough sets and entropy measures. J Biomed Inform. 2017;67:59–68. doi:10.1016/j.jbi.2017.02.007.
20. Song S, Zhang X. Feature selections using systematic and optimal algebra-information fusion measures based on compromised variable-granularity neighborhoods. Expert Syst Appl. 2026;304(1):131973. doi:10.1016/j.eswa.2026.131973.
21. Hu Q, Zhang L, Chen D, Pedrycz W, Yu D. Gaussian kernel based fuzzy rough sets: model, uncertainty measures and applications. Int J Approx Reason. 2010;51(4):453–71. doi:10.1016/j.ijar.2010.01.004.
22. Wang X, Tsang ECC, Zhao S, Chen D, Yeung DS. Learning fuzzy rules from fuzzy samples based on rough set technique. Inf Sci. 2007;177(20):4493–514. doi:10.1016/j.ins.2007.04.010.
23. Wang J, Huang Z, Weng Z, Li J. Feature subset selection using fuzzy scale entropy-based uncertainty measures for multi-scale fuzzy relation decision systems. Fuzzy Sets Syst. 2026;526:109665. doi:10.1016/j.fss.2025.109665.
24. Zou X, Dai J. A fuzzy β-covering rough set model for attribute reduction by composite measure. Fuzzy Sets Syst. 2025;507:109314. doi:10.1016/j.fss.2025.109314.
25. Wang C, Huang Y, Shao M, Hu Q, Chen D. Feature selection based on neighborhood self-information. IEEE Trans Cybern. 2020;50(9):4031–42. doi:10.1109/TCYB.2019.2923430.
26. Ma Q, Zhu X, Zhao X, Zhao B, Fu G, Zhang R. An equidistance index intuitionistic fuzzy c-means clustering algorithm based on local density and membership degree boundary. Appl Intell. 2024;54(4):3205–21. doi:10.1007/s10489-024-05297-1.
27. Gao C, Hu J, Zhou J, Ding W, Pedrycz W. Average weight margin-based feature selection with three-way decision. Pattern Recogn. 2025;174(6):113009. doi:10.1016/j.patcog.2025.113009.
28. Chen Z, Chen G, Gao C, Zhou J, Wen J. Robust weighted fuzzy margin-based feature selection with three-way decision. Int J Approx Reason. 2024;173(9):109253. doi:10.1016/j.ijar.2024.109253.
29. Li Y, Wu X, Wang X. Incremental reduction methods based on granular ball neighborhood rough sets and attribute grouping. Int J Approx Reason. 2023;160(8):108974. doi:10.1016/j.ijar.2023.108974.
30. Pan Y, Xu W, Ran Q. An incremental approach to feature selection using the weighted dominance-based neighborhood rough sets. Int J Mach Learn Cybern. 2023;14(4):1217–33. doi:10.1007/s13042-022-01695-4.
31. Chen J, Zhang X, Yuan Z. Feature selection based on overlap degrees and granulation measures for k-NN rough sets. Pattern Recogn. 2024;156(18):110837. doi:10.1016/j.patcog.2024.110837.
32. Xu F, Cai M, Li Q, Wang H, Fujita H. Shared neighbors rough set model and neighborhood classifiers. Expert Syst Appl. 2024;244(1):122965. doi:10.1016/j.eswa.2023.122965.
33. Gao C, Zhou J, Wang X, Pedrycz W. Granule margin-based feature selection in weighted neighborhood systems. IEEE Trans Cybern. 2025;55(4):2151–64. doi:10.1109/TCYB.2025.3544693.
34. Sun L, Guo J, Wu X, Xu J. Fuzzy C-means clustering-based multi-label feature selection via weighted neighborhood mutual information. Inf Sci. 2025;718:122389. doi:10.1016/j.ins.2025.122389.
35. Liu J, Huang S, Fan Y, Zhang H, Gou J, Lin Y. Reliability-enhanced partial multi-label feature selection based on neighborhood rough set. Pattern Recogn. 2026;179:113762. doi:10.1016/j.patcog.2026.113762.
36. Chang X, Sun B, Liu X, Ye J, Chu X. Fuzzy neighborhood rough set-based attribute reduction over temporal information systems with application to clinical efficacy evaluation. Inf Process Manag. 2026;63(6):104730. doi:10.1016/j.ipm.2026.104730.
37. Zhang X, Mei C, Li J, Yang Y, Qian T. Instance and feature selection using fuzzy rough sets: a bi-selection approach for data reduction. IEEE Trans Fuzzy Syst. 2023;31(6):1981–94. doi:10.1109/TFUZZ.2022.3216990.
38. Yu W, Ma X, Zhang Z, Zhang Q. A method for fast feature selection utilizing cross-similarity within the context of fuzzy relations. Comput Mater Contin. 2025;83(1):1195–218. doi:10.32604/cmc.2025.060833.
39. Dua D, Graff C. UCI machine learning repository [Internet]. Irvine, CA, USA: University of California, School of Information and Computer Science; 2019 [cited 2026 Jan 13]. Available from: https://archive.ics.uci.edu/.

40. Cano A, Masegosa A, Moral S. Kent ridge biomedical data set [Internet]; 2005 [cited 2026 Jan 13]. Available from: http://leo.ugr.es/elvira/DBCRepository/.
41. Demšar J. Statistical comparisons of classifiers over multiple data sets. J Mach Learn Res. 2006;7:1–30.