IFS/ACM Transactions on Machine Learning, vol. 3, no. 1, pp. 30-37, 2026

Ordinal Selection Fuzzy Rough Sets: A Dominance Relation-Driven Feature-Instance Bidirectional Selection Method

Beini Dai[a], Yan Xu[a] and Ji Feng[a,*]

[a]College of Computer and Information Science, Chongqing Normal University, Chongqing 401331, China

ARTICLE INFO

Keywords:
fuzzy rough sets
ordinal classification
feature selection
instance selection
dominance relation

ABSTRACT

Ordinal classification involves predicting labels with a natural ordering, yet the presence of feature redundancy and noisy instances can distort the ordinal relationships that fuzzy rough set models rely on. Most existing approaches tackle feature selection in isolation, overlooking how these two factors interact to degrade both the quality of selected features and the integrity of retained instances. This work introduces Ordinal Selection Fuzzy Rough Sets (OSFRS), a framework that couples feature and instance selection through dominance relations rather than treating them as separate stages. The method begins by constructing a normalized dominance-based fuzzy lower approximation to quantify feature importance, alongside an ordinal discriminability index that explicitly encodes the ordering among decision classes. Rather than evaluating all instances uniformly, the method suppresses noisy samples by measuring differences in order fuzziness – instances near decision boundaries are retained while those far from boundaries or exhibiting high ambiguity are deprioritized. A dual-criterion selection strategy then jointly optimizes ordinal discriminability and ranked feature importance, yielding a compact feature subset and a refined instance set in a single pass. Evaluation on nine UCI ordinal datasets shows that OSFRS achieves comparable classification consistency without statistically significant loss, while achieving higher reduction rates than competing methods, with no selection failures across any dataset.

1. Introduction

Ordinal classification addresses prediction problems in which the class labels possess an inherent ordering [1, 2, 3]. With the rapid growth of data in many application domains, large-scale labeled datasets with ordered class structures have become increasingly common. Representative examples include disease severity levels (e.g., mild, moderate, and severe), credit ratings (e.g., poor, fair, and excellent), and product quality grades. Unlike conventional classification, ordinal classification must exploit the ordinal relationships among class labels to reduce the cost of severe misclassification, i.e., predicting a label that is far from the true label on the ordinal scale. The main challenge is that feature redundancy and noisy instances can substantially impair the preservation of ordinal consistency. Existing fuzzy rough set methods typically evaluate feature importance using indirect dependency measures or approximation accuracy, which may fail to identify informative features, particularly in low-dimensional settings where the available discriminative information is limited. Moreover, these methods generally lack explicit indicators for quantifying deviations in ordinal relationships, and the underlying fuzzy similarity measures are sensitive to noise, leading to misidentification of representative samples during instance selection. Developing a unified dominance-relation-based bidirectional selection framework that can simultaneously handle feature redundancy and instance noise while preserving the ordinal structure remains an open problem.

This paper proposes OSFRS, a dominance relation-driven framework for feature-instance bidirectional selection. The main contributions include:

- Developing ordered evaluation measures that reflect each instance's role in maintaining the ordinal structure.
- Devising an instance selection criterion based on order fuzziness, which prioritizes boundary-adjacent samples while suppressing noisy instances.
- Establishing a feature reduction model with dual objectives – ordinal discriminability and global importance – to preserve class order under compression.

The key distinction of OSFRS from existing bidirectional selection methods lies in directly embedding dominance relations into the entire granulation process, rather than performing feature reduction and instance reduction in isolation or relying on unordered equivalence relations. The proposed framework forms a closed loop: instance filtering eliminates noisy instances to provide a denoised dataset for feature evaluation, and feature selection performs dimensionality reduction while explicitly preserving the partial order structure between classes.

2. Related Works

2.1. Fuzzy Rough Sets for Feature Selection

The study of feature selection within the fuzzy rough set framework has progressed through several lines of inquiry,

DOI: https://doi.org/10.70891/TML.2026.040030
ISSN of IFS/ACM TML: 3078-5510

*Corresponding author
jifeng@cqnu.edu.cn (J. Feng)

building on classical foundations in variable and feature selection [4]. Early work by Jensen and Shen [5] established systematic frameworks for fuzzy-rough feature selection, while Hu et al. [6] extended these methods to handle noisy data through robust fuzzy rough set models. Subsequent developments explored efficiency and weighting enhancements: Li et al. [7] introduced granular-ball divergence to construct fuzzy rough hypergraphs, and Wang et al. [8] incorporated feature weighting mechanisms to better capture attribute contributions. More recently, Zhang et al. [9] proposed a bi-selection approach that jointly handles instance and feature selection, and Li et al. [10] further examined incremental perspectives for updating selected features and instances as data evolves.

These methods all share the same problem: they rely on fuzzy equivalence or fuzzy similarity relations. Such relations do not consider directionality, so they are not suitable for ordinal problems. In ordinal problems, the dominance direction between instances carries important information. This structural mismatch limits the ability of existing granulation processes to preserve order, and this issue has been noted in the feature selection literature [11].

2.2. Dominance-Based Rough Sets for Ordinal Classification

Traditional rough sets [12] rely on equivalence relations and thus cannot effectively handle partial orders, a structural constraint that has motivated several lines of theoretical extension. Dubois and Prade [13] generalized the classical framework by incorporating fuzzy set theory, yielding fuzzy rough sets that relax the rigid boundary between belonging and non-belonging. Building on this generalization, Greco et al. [14] replaced equivalence relations with dominance relations to handle multicriteria decision analysis, enabling the direct representation of preference-ordered decision classes. Subsequently, Sang et al. [15] applied fuzzy dominance rough sets to investigate attribute correlation and class separability in ordered systems, while Zadeh's theory of fuzzy information granulation [16] provided the conceptual machinery for constructing dominance-based granules.

A shortcoming shared by these methods is that they treat feature redundancy and instance noise as independent problems rather than as interacting factors. When both issues co-occur, the selected features often fail to capture the ordinal structure that the classifier depends on. Each issue may be manageable individually, but the overall performance still degrades. The following example illustrates this point.

Consider a universe $U = \{x_1, x_2, x_3, x_4\}$ with decision classes ordered as $C_1 \prec C_2 \prec C_3 \prec C_4$. Two feature subsets B_1 and B_2 yield identical dependency degrees under traditional fuzzy rough sets, yet their order preservation behaviors diverge sharply. As Table 1 shows, B_1 produces severe order inversions – predicting x_2 (true class C_2) as C_3, and x_3 (true class C_3) as C_2 – creating a reversed subsequence. By contrast, B_2 misclassifies x_2 and x_3 as well, but both predictions fall below the true labels, preserving the monotone trend. Standard dependency metrics assign B_1 and B_2 the same score, confirming that an explicit order discriminability index is needed to separate structurally distinct error patterns.

Table 1
Example of evaluation discrepancy between traditional and ordered methods.

Instance	True	B_1	B_2
x_1	C_1	C_1	C_1
x_2	C_2	C_3 (misclassified)	C_1 (misclassified)
x_3	C_3	C_2 (misclassified)	C_2 (misclassified)
x_4	C_4	C_4	C_4

From the above, three interrelated gaps in the existing literature can be identified. Traditional evaluation indicators, such as dependency degree and approximation accuracy, lack the ability to quantify violations of ordinal relationships. Fuzzy similarity measures remain vulnerable to boundary noise, which can severely distort feature evaluation when informative features are scarce. Currently, no framework coordinates instance sampling and feature selection through a shared dominance-based mechanism. OSFRS is designed precisely to embed dominance relations throughout the entire granulation process, covering both instance sampling and feature selection.

3. A Bidirectional Selection Method Based on Fuzzy Dominance Rough Sets

This section proposes a bidirectional selection method in which instance filtering and feature selection are not performed as independent stages, but are coupled through a shared dominance-based mechanism. Instances near the decision boundary carry the most information about the ordinal structure, so they are used to guide feature evaluation. At the same time, even as dimensionality decreases, the feature subset should preserve the partial order among classes. The method proceeds in two stages. An ordered fuzziness measure is defined for each instance to separate boundary anchors from noisy samples, yielding a reduced dataset that retains the ordinal skeleton of the original data. On the basis of this purified instance set, a dual-criterion evaluation model is introduced to balance importance preservation against order discriminability, thereby obtaining a compact feature subset. Since both stages share the same dominance relation throughout, noise removal and dimensionality reduction reinforce each other rather than interfere with each other.

3.1. Preliminaries and Definitions

An ordered decision system can be represented as a triple $S_{\preceq} = \langle U, A \cup D \rangle$, where U is the universe, A is the conditional feature set, and D is the decision attribute with a partial order $\preceq$. On the basis of the ordered decision system, the upper and lower ordered dominance sets can be denoted as $S_i^+ = \{x_j \mid f(x_j, d) \succeq f(x_i, d)\}$ and $S_i^- = \{x_j \mid$

$f(x_j, d) \preceq f(x_i, d)\}$, respectively, where $f(x_i, d)$ denotes the value of x_i on the decision attribute d.

Definition 3.1. *For each $x_i \in U$, the upward and downward ordered fuzzy sets of x_i induced by the partial order $\preceq$ are defined respectively as*

$$X_i^{\succeq}(x_j) = \begin{cases} 1, & \text{if } f(x_j, d) \geq f(x_i, d), \\ 0, & \text{otherwise}, \end{cases} \tag{1}$$

$$X_i^{\preceq}(x_j) = \begin{cases} 1, & \text{if } f(x_j, d) \leq f(x_i, d), \\ 0, & \text{otherwise}. \end{cases} \tag{2}$$

Definition 3.2. *Based on Definition 3.1, for a feature subset $B \subseteq A$ and $x_i \in U$, the fuzzy dominance lower and upper approximations of x_i with respect to B are defined as*

$$\underline{R_B^{\succeq}} X_i^{\succeq}(x_j) = \inf_{x_j \in U} \max\left(1 - R_B^{\succeq}(x_i, x_j), X_i^{\succeq}(x_j)\right), \tag{3}$$

$$\overline{R_B^{\succeq}} X_i^{\succeq}(x_j) = \sup_{x_j \in U} \min\left(R_B^{\succeq}(x_i, x_j), X_i^{\succeq}(x_j)\right), \tag{4}$$

where $R_B^{\succeq}(x_i, x_j) = 1 - \frac{\|x_i^B - x_j^B\|_2}{\sqrt{\sum_{c \in B}(\max V_c - \min V_c)^2}}$ *if* $f(x_i, c) \geq f(x_j, c)$ *for all* $c \in B$, *otherwise* 0.

Based on the dominance relation, the lower approximation $\underline{R_B^{\succeq}} X_i^{\succeq}(x_j)$ measures the certainty that x_i belongs to the upward ordered fuzzy set $X_i^{\succeq}$, while the upper approximation $\overline{R_B^{\succeq}} X_i^{\succeq}(x_j)$ captures the possibility. For convenience, let $\lambda_i^+ = \underline{R_A^{\succeq}} X_i^{\succeq}(x_j)$ and $\lambda_i^- = \underline{R_A^{\preceq}} X_i^{\preceq}(x_j)$ denote the lower approximation values under the full feature set A.

3.2. Construction of Ordered Evaluation Measures

Traditional fuzzy rough sets focus on feature relevance but ignore instance-level contributions to order preservation. To address this problem, two ordered evaluation measures are constructed as follows.

Definition 3.3. *For any instance $x_i \in U$ and feature subset $B \subseteq A$, the coverage sets can be defined as*

$$Cov^+([x_i]_B^{\lambda_i^+}) = \{x_j \in S_i^+ \mid 1 - R_B(x_i, x_j) < \lambda_i^+\}, \tag{5}$$

$$Cov^-([x_i]_B^{\lambda_i^-}) = \{x_j \in S_i^- \mid 1 - R_B(x_i, x_j) < \lambda_i^-\}, \tag{6}$$

where the fuzzy similarity is defined as

$$R_B(x_i, x_j) = 1 - \frac{\|x_i^B - x_j^B\|_2}{\sqrt{\sum_{c \in B}(\max V_c - \min V_c)^2}}.$$

Here, $\lambda_i^+ = \underline{R_A^{\succeq}} X_i^{\succeq}(x_i)$ and $\lambda_i^- = \underline{R_A^{\preceq}} X_i^{\preceq}(x_i)$ denote the lower approximation values under the full feature set A. The coverage set Cov^+ contains the instances in the upper dominance set of x_i whose dissimilarities to x_i fall below the threshold λ_i^+. These instances are thus similar to x_i while respecting the dominance order. The set Cov^- is defined analogously for the lower dominance set. Together, these coverage sets describe the effective neighborhood of an instance within the ordered structure, directly supporting the quantification of instance-level importance.

Then, to quantify the contribution of each instance to preserving the ordered structure, the ordered importance index is further defined as follows.

Definition 3.4. *For any instance $x_i \in U$ and feature subset $B \subseteq A$, the ordered importance index can be defined as*

$$F_o([x_i]_B^{\lambda_i}) = \alpha \cdot \lambda_i^+ \cdot \frac{|Cov^+|}{|S_i^+|} + \beta \cdot \lambda_i^- \cdot \frac{|Cov^-|}{|S_i^-|}, \tag{7}$$

where $\alpha, \beta \in [0, 1]$ are balance weights, and $\alpha + \beta = 1$.

On the basis of above definitions, the ordered evaluation metrics that simultaneously consider the contributions of instances and features are constructed, which serve as the foundation for subsequent instance and feature selection.

3.3. Representative Instance Selection Based on Boundary Fuzziness

Evaluating all instances is too costly, but boundary-far instances contribute very little to order learning. Instance reduction techniques have been widely studied [17, 18]. To address these problems, the order fuzziness difference is defined below.

Definition 3.5. *For any instance $x_i \in U$, the order fuzziness difference is defined as*

$$\delta_\lambda(x_i) = |\lambda_i^+(x_i) - \lambda_i^-(x_i)|. \tag{8}$$

Here, $\delta_\lambda(x_i)$ represents the absolute discrepancy between the two dominance lower approximations. A smaller value indicates that the instance lies closer to the decision boundary and exhibits greater classification fuzziness.

For representative instance selection, a multi-stage forward cumulative strategy is constructed based on $\delta_\lambda(x_i)$. For the highest and lowest decision classes, instances are ranked in descending order by λ_i^+ or λ_i^-, respectively, and the top k_1 instances are selected as order boundary anchors and directly included in the retention set. For each intermediate class, a hybrid score combining normalized ordered importance and intra-class local density is computed, and the top n_{perclass} instances per class are selected as intra-class representatives and also directly included in the retention set. The strategy further supplements the retention set with all instances satisfying $\delta_\lambda(x_i) \in (t_1, t_2)$ and $F_o([x_i]_A^{\lambda_i}) > \xi$, where ξ is the importance threshold, typically set to the mean of $F_o([x_i]_A^{\lambda_i})$ over the entire dataset. This supplementation stage

does not remove any previously selected anchors or representatives, ensuring that the ordinal skeleton remains intact even if the noise thresholds fluctuate. Finally, for classes with very few samples, a fallback filling mechanism supplements neighborhood instances to prevent class disappearance, and a global minimum retention ratio guarantees a sufficient total sample size.

Through the above strategy, the influence of noisy instances is mitigated, and the ordinal skeleton is preserved. As a result, a purified space U^* is obtained, which reduces the cost of subsequent evaluation.

3.4. Order-Preserving Feature Selection under Dual-Criteria Constraints

Traditional feature selection lacks order constraints and may fail in low dimensions. Thus, an ordinal discriminability index and an importance measure are introduced to formulate a dual-criteria model. Firstly, the ordinal discriminability index is defined as follows.

Definition 3.6. *For any feature subset $B \subseteq A$, the ordinal discriminability index can be defined as*

$$OD(B) = 1 - \frac{1}{|U^*|} \sum_{x_i \in U^*} \frac{|\lambda_i^+(A) - \lambda_i^+(B)| + |\lambda_i^-(A) - \lambda_i^-(B)|}{2}. \tag{9}$$

Then, the ordered global importance sum is defined as follows, thereby ensuring global importance preservation.

Definition 3.7. *For any purified space U^*, feature subset $B \subseteq A$, the sum of ordered importance over the representative instance set can be defined as*

$$\Psi_{oU^*}(B) = \sum_{x_i \in U^*} F_o([x_i]_B^{\lambda_i}). \tag{10}$$

Finally, under the dual-criteria constraints, the feature reduction objective is defined as follows.

Definition 3.8. *For any feature subset $B \subseteq A$, the optimal reduction set is defined as*

$$\begin{aligned} B_{red} &= \arg\min |B| \\ s.t. \quad & |\Psi_{oU^*}(A) - \Psi_{oU^*}(B)| < \delta, \\ & |OD(A) - OD(B)| < \epsilon, \end{aligned} \tag{11}$$

where δ, ϵ are tolerance thresholds.

A backward deletion strategy is adopted to obtain the optimal feature subset defined in Definition 3.8. Starting from the full feature set, the strategy removes a feature when both the importance preservation constraint and the order discriminability constraint are satisfied. Compared with forward greedy selection, backward deletion avoids redundancy accumulation, and it evaluates the necessity of each feature in the context of the remaining features rather than in isolation. Combined with the ordered global importance defined in Definition 3.7, this procedure yields a compact feature subset that preserves both the ordinal structure and the classification power of the original data.

3.5. Algorithm Analysis

Algorithm 1 summarizes the complete OSFRS procedure.

Algorithm 1 OSFRS: Ordinal Selection Fuzzy Rough Sets

Require: Ordered fuzzy decision system $(U, A \cup D)$, boundary ratio k_1, intra-class ratio k_2, weights α, β, fuzziness lower bound t_1, fuzziness upper bound t_2, importance threshold ξ.

Ensure: Ordered representative instance set U^* and final feature subset A^*.

1: $U^* \leftarrow \emptyset$
2: **for** each instance $x_i \in U$ **do**
3: Compute ordered importance $F_o([x_i]_A^{\lambda_i})$ by Eq. (7).
4: Compute order fuzziness difference $\delta_\lambda(x_i)$ by Eq. (8).
5: **end for**
6: // Obtain highest/lowest decision class anchors
7: Identify D_{high} and D_{low}, select top k_1 instances ranked by λ_i^+ (for D_{high}) or λ_i^- (for D_{low}), add to U^*.
8: $n_{\text{perclass}} \leftarrow \max(1, \lfloor |U| \cdot k_2 / |D| \rfloor)$
9: **for** each decision class $d \in D$ **do**
10: Get instance indices label_{id} in class d.
11: **if** $\text{len}(\text{label}_{id}) \le n_{\text{perclass}}$ **then**
12: Add all instances in label_{id} to U^*.
13: **else**
14: Compute F_o values for instances in this class, normalize to f_{norm}.
15: Compute intra-class average similarity (centrality), normalize to c_{norm}.
16: Hybrid score $= 0.5 f_{\text{norm}} + 0.5 c_{\text{norm}}$.
17: Select top n_{perclass} instances by hybrid score, add to U^*.
18: **end if**
19: **end for**
20: // Supplement boundary-suitable instances to enrich order discrimination
21: **for** each instance $x_i \in U$ **do**
22: **if** $\delta_\lambda(x_i) \in (t_1, t_2)$ and $F_o([x_i]_A^{\lambda_i}) > \xi$ **then**
23: Add x_i to U^*.
24: **end if**
25: **end for**
26: Remove duplicates from U^*.
27: Obtain the final feature subset A^* by solving the reduction problem in Definition 3.8 (Eq. (11)).
28: **return** U^* and A^*.

The algorithm operates in two sequential stages. The first stage (Lines 1 – 26) reduces the instance space. It initializes the retention set (Line 1), then computes the ordered importance F_o and fuzziness difference δ_λ for all instances via the initial loop (Lines 2 – 5). Computing these quantities for each instance requires pairwise comparisons with all other instances, yielding an overall cost of $O(n^2)$. It then retains boundary anchors of the highest and lowest decision classes via top-k_1 ranking (Lines 6 – 7). For each intermediate class, a hybrid score that combines normalized

importance and intra-class density selects n_{perclass} representative instances (Lines 8 – 19) at an additional cost of $O(n \cdot m)$. The retention set is then supplemented with additional instances whose fuzziness difference falls within the admissible interval (t_1, t_2) and whose ordered importance exceeds the threshold ξ (Lines 20 – 25), after which duplicates are removed (Line 26), producing the purified space U^*. The second stage (Line 27) reduces dimensionality on U^* via backward deletion under the dual-criteria constraints of Definition 3.8, yielding the final feature subset A^*.

The overall time complexity is $O(n^2 + m^2 \cdot n^{*2})$ and the space complexity is $O(n^2 + n^{*2})$, where $n = |U|$ is the number of original instances, $m = |A|$ is the number of conditional features, and $n^* = |U^*|$ is the number of retained instances after instance selection. The dominant cost comes from the pairwise distance computations in both the ordered importance calculation and the backward deletion step. A notable advantage of this two-stage design is that the first-stage instance reduction lowers the data size before feature evaluation, allowing the backward deletion to operate on a cleaner input rather than directly on the original dataset.

4. Experimental Results

4.1. Experimental Setup

OSFRS is evaluated on 9 UCI ordinal datasets (Table 2) and compared against four different baselines. BSFS [9] performs bidirectional selection using ordinary fuzzy rough sets without ordinal awareness. RIS-IRFWA [8] augments BSFS with Pearson-based feature weights, thereby testing whether a weighting strategy improves ordinal performance. FRS-OD applies dominance-relation-based feature selection but performs no instance selection, while OIS-FRS applies ordered instance selection without feature reduction. The latter two serve as ablation baselines: each disables one of the two selection modules in OSFRS, allowing a more direct assessment of each module's contribution.

All methods are evaluated under 10-fold cross-validation [19] using a KNN classifier with ordinal distance weighting [20]. Prior to the experiments, all conditional attributes are first normalized to $[0, 1]$. The OSFRS parameters are configured as follows. The fuzziness bounds are set to $t_1 = 0.25$ and $t_2 = 0.75$, which define the admissible noise range by covering the middle portion of the fuzziness distribution. For the balance weights, equal values of $\alpha = \beta = 0.5$ are adopted to ensure that upper and lower dominance contributions receive identical treatment. The tolerance thresholds $\delta = 0.08$ and $\epsilon = 0.08$ govern how strictly the dual-criteria constraints are enforced during backward deletion, with both values increased to 0.12 for high-dimensional datasets. The boundary ratio k_1 and the importance screening ratio k_2 vary across datasets, with their specific values determined by the dataset size and the number of classes.

To evaluate the effectiveness of the models, three metrics are used:

Table 2
Summary of the 9 UCI datasets used in the experiments

Dataset	Instances	Features	Classes
AUTO	392	7	3
BALANCE-SCALE	625	4	3
CAR	1728	6	4
CLEVELAND	303	13	5
ERA	1000	4	9
ESL	488	4	9
LEV	1000	4	5
SWD	1000	10	4
WRED	1599	11	6

- **Ordered reduction rate:** $Red_o = 1 - (|U^*|/|U|) \times (|A^*|/|A|)$. This metric measures the combined reduction degree of data volume and dimensionality. A value closer to 1 indicates a higher degree of data reduction.
- **Ordered classification consistency:** measured by Kendall's τ^b coefficient, which accommodates ties in ordered data. Let C be the number of concordant pairs, D the number of discordant pairs, T the number of ties in the predicted sequence, and U the number of ties in the true sequence; then $\tau^b = (C - D)/\sqrt{(C + D + T)(C + D + U)}$. A value closer to 1 indicates more accurate order relation prediction.
- **Comprehensive effectiveness:** $E_{ord} = \frac{Red_o \cdot (Acc_{rank}+1)}{2}$. This metric jointly evaluates both aspects by taking the product of the reduction rate and normalized consistency. A value closer to 1 indicates that the reduction scheme preserves the core ordered structure while substantially removing redundancy.

Since Kendall's τ^b coefficient ranges from -1 to 1, it cannot be directly used in product-type composite evaluations such as E_{ord}. Therefore, a normalized consistency $Acc_{norm} = (Acc_{rank} + 1)/2$ is defined to map the raw coefficient into the $[0, 1]$ interval. The comprehensive effectiveness E_{ord} is then computed as the product of the reduction rate and this normalized consistency, providing an intuitive single quantitative reference for the reduction-performance trade-off in engineering practice.

All datasets are first normalized so that each conditional attribute lies in the $[0, 1]$ interval. For each real-valued attribute $a \in A$, the normalized value is $a(x_i) = (a(x_i) - \min_j a(x_j))/(\max_j a(x_j) - \min_j a(x_j))$. This ensures that all features contribute equally to the distance computation and that attributes with larger numerical ranges do not dominate.

4.2. Effectiveness Analysis

Table 3 presents the average performance across all datasets. OSFRS achieves the highest average reduction rate (0.679) while matching BSFS in consistency (0.499). BSFS and RIS-IRFWA show similar consistency levels but slightly

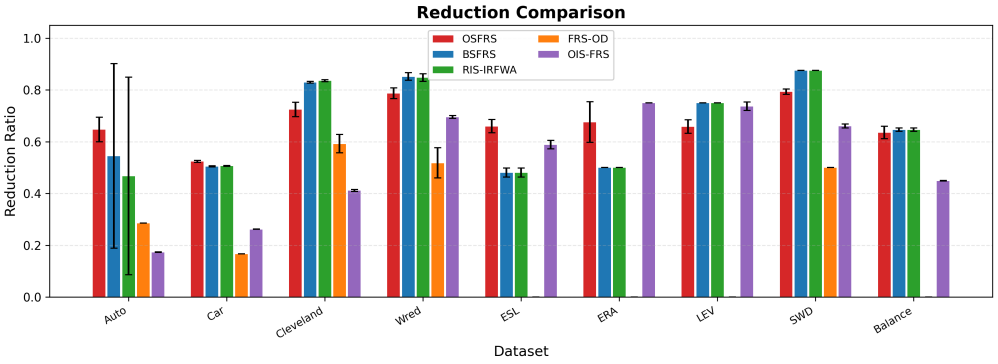

Figure 1: Comparison of reduction rate.

Table 3
Average performance summary

Method	Red_o	Acc_{rank}	E_{ord}
BSFS	0.665	0.499	0.493
FRS-OD	0.229	0.579	0.171
OIS-FRS	0.526	0.476	0.384
RIS-IRFWA	0.657	0.485	0.482
OSFRS	**0.679**	**0.499**	**0.504**

lower reduction rates (0.665 and 0.657, respectively). OIS-FRS trades reduction for consistency, while FRS-OD is dominated across both metrics.

The performance of OSFRS on datasets with different characteristics is discussed next. On the CAR dataset, which contains six features and four ordered classes, OSFRS attains a reduction rate of 0.524 and consistency of 0.720, outperforming BSFS (0.392) and RIS-IRFWA (0.351) in reduction. On ERA and ESL – both four-feature, nine-class datasets where FRS-OD fails completely – OSFRS still achieves reductions of 0.676 and 0.660, respectively. On the higher-dimensional SWD and WRED datasets, the reduction rates reach 0.794 and 0.787. This pattern indicates that OSFRS does not merely pursue the highest reduction rate on favorable datasets. It also maintains stable performance when the feature space is small relative to the number of classes.

Figure 1 compares the reduction rates of all methods across the nine datasets, highlighting the stable advantage of OSFRS in low-dimensional and moderate-dimensional scenarios.

Kendall's τ^b provides a stricter consistency measure than simple accuracy, as it penalizes order inversions rather than merely counting misclassifications. Under this metric, OSFRS attains an average consistency of 0.499, matching BSFS at 0.499 and exceeding RIS-IRFWA at 0.485 and OIS-FRS at 0.476. FRS-OD records the highest average at 0.579, but this value deserves closer examination. On the BALANCE-SCALE, ERA, ESL, and LEV datasets, FRS-OD performs no reduction at all, retaining the full feature and instance space and therefore preserving the raw order structure by default. A consistency score computed without any compression conflates preservation of the original signal with algorithmic merit. In contrast, OSFRS achieves the same consistency level as BSFS while producing a substantially higher reduction rate (0.679 versus 0.665), indicating that dominance-based selection preserves ordinal information more effectively per unit of compression.

Table 4
Detailed comparison of reduction rate Red_o on each dataset.

Dataset	BSFS	FRS-OD	OIS-FRS	RIS-IRFWA	OSFRS
AUTO	0.545	0.286	0.174	0.468	0.647
BALANCE-SCALE	0.646	0.000	0.450	0.646	0.636
CAR	0.505	0.167	0.262	0.507	0.524
CLEVELAND	0.830	0.592	0.412	0.836	0.725
ERA	0.500	0.000	0.750	0.500	0.676
ESL	0.481	0.000	0.589	0.481	0.660
LEV	0.750	0.000	0.737	0.750	0.659
SWD	0.875	0.500	0.661	0.875	0.794
WRED	0.852	0.518	0.696	0.848	0.787

Table 5
Detailed comparison of classification consistency Acc_{rank} on each dataset.

Dataset	BSFS	FRS-OD	OIS-FRS	RIS-IRFWA	OSFRS
AUTO	0.460	0.651	0.552	0.401	0.594
BALANCE-SCALE	0.672	0.553	0.386	0.672	0.510
CAR	0.392	0.911	0.536	0.351	0.720
CLEVELAND	0.352	0.315	0.412	0.346	0.484
ERA	0.366	0.342	0.401	0.366	0.351
ESL	0.847	0.852	0.796	0.847	0.710
LEV	0.544	0.609	0.489	0.544	0.418
SWD	0.449	0.460	0.430	0.449	0.364
WRED	0.409	0.516	0.278	0.390	0.343

Tables 4 – 6 report the per-dataset performance of all algorithms in terms of reduction rate Red_o, classification consistency Acc_{rank}, and comprehensive effectiveness E_{ord}, respectively. These results reveal dataset-specific variations that the averages in Table 3 cannot capture.

OSFRS achieves the highest reduction rate on AUTO, CAR, ERA, ESL, and WRED, while its consistency remains competitive on the remaining datasets. As discussed earlier, FRS-OD yields zero reduction on several low-dimensional datasets, which explains both its inflated average consistency

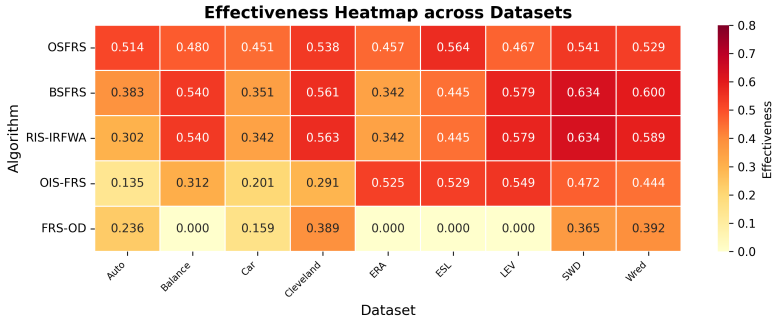

Figure 2: Heatmap of comprehensive effectiveness E_{ord} for all algorithms across nine datasets.

Table 6
Detailed comparison of comprehensive effectiveness E_{ord} on each dataset.

Dataset	BSFS	FRS-OD	OIS-FRS	RIS-IRFWA	OSFRS
AUTO	0.383	0.236	0.135	0.302	0.514
BALANCE-SCALE	0.540	0.000	0.312	0.540	0.480
CAR	0.351	0.159	0.201	0.342	0.451
CLEVELAND	0.561	0.389	0.291	0.563	0.538
ERA	0.342	0.000	0.525	0.342	0.457
ESL	0.445	0.000	0.529	0.445	0.564
LEV	0.579	0.000	0.549	0.579	0.467
SWD	0.634	0.365	0.472	0.634	0.541
WRED	0.600	0.392	0.444	0.589	0.529

and its poor comprehensive effectiveness. BSFS and RIS-IRFWA produce nearly identical values on all three metrics, indicating that the Pearson-based weighting in RIS-IRFWA does not alter selection behavior when the underlying fuzzy rough set model lacks ordinal information. In terms of E_{ord}, OSFRS ranks first on AUTO, CAR, and ESL, and second on CLEVELAND and SWD.

4.3. Statistical Significance Analysis

The experiments employed the Friedman and Wilcoxon signed-rank tests [21]. For Red_o, the Friedman test detected highly significant differences, with $\chi^2 = 19.634$ and $p < 0.001$; the Wilcoxon test confirmed that OSFRS significantly outperformed FRS-OD, with $W = 0$ and $p = 0.0039$. For Acc_{rank}, the Friedman test yielded $p = 0.2685$, and the pairwise test between OSFRS and BSFS showed no significant difference, with $W = 20$ and $p = 0.8203$. Therefore, this result only indicates that the two methods do not exhibit a significant difference on this metric, rather than proving that their performance is identical.

OSFRS's main contribution is not a significant improvement in ordinal prediction accuracy. It delivers more efficient dimensionality reduction while keeping classification performance statistically equivalent to that of the baselines, and provides a robust selection framework.

4.4. Heatmap Analysis of E_{ord} Across Individual Datasets

Figure 2 visualizes E_{ord} values as a heatmap, with darker cells indicating higher effectiveness. The heatmap reveals a clear structural difference among the methods.

OSFRS performs well on all nine datasets, never falling into the worst category. But this stability comes at a cost: on CLEVELAND, LEV, SWD, and WRED, BSFS outperforms OSFRS because these datasets have larger feature spaces and the ordinal structure is easier to exploit. However, BSFS and FRS-OD degrade noticeably on the other datasets. FRS-OD produces no results at all on BALANCE-SCALE, ERA, LEV, and ESL, indicating that it basically fails on low-dimensional problems. BSFS and RIS-IRFWA also perform poorly on AUTO, CAR, and ERA, because the non-ordinal fuzzy rough set model they use struggles to preserve the ordinal structure. OIS-FRS achieves medium-dark cells only on ERA and ESL, confirming that instance selection alone without feature reduction is insufficient for ordinal classification.

The heatmap reflects a trade-off between stability and peak performance. BSFS and FRS-OD perform well on favorable datasets but suffer sharp declines on others. Although OSFRS sacrifices some peak performance, it behaves more consistently across different datasets. RIS-IRFWA performs almost identically to BSFS, which again indicates that if the underlying model lacks ordinal awareness, simply adding Pearson-based weighting does not substantially change the selection behavior.

5. Conclusion

This paper proposes OSFRS, a bidirectional feature-instance selection method for ordinal classification. It combines instance filtering based on dominance relations with dual-criteria feature reduction, rather than treating feature and instance selection as separate stages or relying on unordered equivalence relations as existing methods do. OSFRS carries dominance relations through the entire granulation process, so that the ordinal structure among decision classes is preserved during both noise removal and dimensionality reduction. Across nine ordinal datasets, OSFRS consistently achieves high reduction rates and never suffers from selection failure, a result that none of the compared methods can match. Statistical tests show that OSFRS does not cause a significant drop in classification consistency relative to the baselines. The heatmap also indicates that OSFRS

maintains stable effectiveness across datasets, whether low-dimensional with many classes or high-dimensional with a moderate number of classes.

The current framework still has limitations. Both the instance selection and feature selection stages require pairwise distance computations, which impose a computational burden that may become prohibitive on very large-scale datasets. The method also depends on several parameters, including the boundary ratio k_1, the fuzziness bounds t_1 and t_2, and the tolerance thresholds δ and ϵ. These parameters may require careful tuning to achieve optimal performance on new datasets. Moreover, the current formulation assumes that all conditional attributes are real-valued and normalized, while its applicability to mixed-type or categorical ordinal data has not been fully explored.

Future work will address these issues along several directions. The framework will be extended to more complex data scenarios, such as multi-label ordinal data and incomplete ordinal data, where the dominance relation can be generalized to handle partial orders among label vectors. With sampling strategies and parallel computing, OSFRS can be further accelerated to better handle large-scale datasets. On the theoretical side, the mathematical properties of the proposed ordered importance measure and order discriminability index, such as their convergence behavior and sensitivity to parameter perturbations, can be analyzed further. Incremental learning is another promising direction, allowing the bidirectional selection framework to adapt to dynamically growing ordinal datasets without requiring full recomputation each time.

Acknowledgement

The authors declare that there is no funding and no conflict of interest.

References

[1] P. A. Gutiérrez, M. Pérez-Ortiz, J. Sánchez-Monedero, F. Fernández-Navarro, C. Hervás-Martínez, Ordinal regression methods: Survey and experimental study, IEEE Transactions on Knowledge and Data Engineering, 2016, vol. 28, no. 1, pp. 127–146.

[2] E. Frank, M. Hall, A simple approach to ordinal classification, Proceedings of the 12th European Conference on Machine Learning, Springer, Freiburg, Germany, 2001, pp. 145–156.

[3] P. McCullagh, Regression models for ordinal data, Journal of the Royal Statistical Society: Series B (Methodological), 1980, vol. 42, no. 2, pp. 109–127.

[4] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, Journal of Machine Learning Research, 2003, vol. 3, pp. 1157–1182.

[5] R. Jensen, Q. Shen, New approaches to fuzzy-rough feature selection, IEEE Transactions on Fuzzy Systems, 2009, vol. 17, no. 4, pp. 824–838.

[6] Q. Hu, L. Zhang, S. An, D. Zhang, D. Yu, On robust fuzzy rough set models, IEEE Transactions on Fuzzy Systems, 2012, vol. 20, no. 4, pp. 636–651.

[7] Y. Li, L. Yang, B. Sang, G. Wang, S. Xia, W. Xu, J. Yu, An efficient feature selection method using the granular-ball divergence-based fuzzy rough hypergraph, IEEE Transactions on Knowledge and Data Engineering, 2026, pp. 1–14.

[8] C. Wang, C. Wang, Y. Qian, Q. Leng, Feature selection based on weighted fuzzy rough sets, IEEE Transactions on Fuzzy Systems, 2024, vol. 32, no. 7, pp. 4027–4037.

[9] X. Zhang, C. Mei, J. Li, Y. Yang, T. Qian, Instance and feature selection using fuzzy rough sets: A bi-selection approach, IEEE Transactions on Fuzzy Systems, 2023, vol. 31, no. 6, pp. 1981–1994.

[10] J. Li, Y. She, X. He, T. Qian, W. Zheng, Incremental perspective for bi-selection of instances and features by employing fuzzy rough set technique, Fuzzy Sets and Systems, 2026, vol. 528, pp. 109705:1–109705:18.

[11] H. Liu, L. Yu, Toward integrating feature selection algorithms for classification and clustering, IEEE Transactions on Knowledge and Data Engineering, 2005, vol. 17, no. 4, pp. 491–502.

[12] Z. Pawlak, Rough sets, International Journal of Computer and Information Sciences, 1982, vol. 11, no. 5, pp. 341–356.

[13] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, International Journal of General Systems, 1990, vol. 17, no. 2-3, pp. 191–209.

[14] S. Greco, B. Matarazzo, R. Slowinski, Rough sets theory for multicriteria decision analysis, European Journal of Operational Research, 2001, vol. 129, no. 1, pp. 1–47.

[15] B. Sang, L. Yang, H. Chen, T. Li, W. Xu, Robust attribute reduction exploring class-separability and attribute-correlation for ordered decision systems, IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2025, vol. 55, no. 6, pp. 3941–3953.

[16] L. A. Zadeh, Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, Fuzzy Sets and Systems, 1997, vol. 90, no. 2, pp. 111–127.

[17] D. R. Wilson, T. R. Martinez, Reduction techniques for instance-based learning algorithms, Machine Learning, 2000, vol. 38, no. 3, pp. 257–286.

[18] J. A. Olvera-López, J. A. Carrasco-Ochoa, J. F. Martínez-Trinidad, J. Kittler, A review of instance selection methods, Artificial Intelligence Review, 2010, vol. 34, no. 2, pp. 133–143.

[19] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, Proceedings of the 14th International Joint Conference on Artificial Intelligence, Morgan Kaufmann, Montreal, Quebec, Canada, 1995, pp. 1137–1143.

[20] L. E. Peterson, K-nearest neighbor, Scholarpedia, 2009, vol. 4, no. 2, p. 1883.

[21] J. Demšar, Statistical comparisons of classifiers over multiple data sets, Journal of Machine Learning Research, 2006, vol. 7, pp. 1–30.