Information Sciences 756 (2026) 123865

Contents lists available at ScienceDirect

Information Sciences

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

Class-imbalance-aware fuzzy-rough feature selection with adaptive synergy-redundancy for interval-valued data

Weihua Xu*, Yuhang Lv

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

ARTICLE INFO

Keywords:
Class-imbalance-aware feature selection
Interval-valued information systems
Fuzzy-rough feature selection
Adaptive synergy-redundancy
Macro-F1

ABSTRACT

Feature selection for interval-valued data remains challenging when uncertainty, class imbalance, and complex feature correlations coexist. Existing fuzzy-rough methods may be biased toward majority-class discrimination and usually provide limited modeling of complementarity and redundancy. To address these issues, this paper proposes CASR-FS, a supervised class-imbalance-aware fuzzy-rough feature selection method for interval-valued data. CASR-FS integrates a macro-averaged approximation-quality dependency measure, an FJMI-based collaborative gain term, and an adaptive conditional redundancy penalty into a unified sequential selection criterion. Experiments on 14 public datasets with KNN, SVM, and DT show that CASR-FS achieves the highest average classification accuracies of 78.26%, 75.30%, and 75.41%, respectively, while selecting compact subsets on average, especially under KNN and SVM. The Macro-F1 analysis further indicates favorable class-wise performance under different imbalance ratios. These results demonstrate that CASR-FS provides an effective and practical balance among class sensitivity, discriminative ability, complementarity, redundancy control, and subset compactness.

1. Introduction

In many real-world applications, data are not only high-dimensional but also uncertain and imprecise due to measurement errors, sensor noise, and inherent variability. Feature selection has therefore become an indispensable preprocessing technique in machine learning and data mining, as it aims to retain informative and discriminative features while removing redundant ones [1]. Instance-wise joint feature selection further links feature choice with classification performance [2]. Neighborhood rough-set reduction also shows the value of selected attributes for incomplete decision systems [3]. When the value of an attribute cannot be precisely determined, representing it as an interval rather than a single point provides a natural way to preserve uncertainty information [4]. Interval-valued implication and equality structures provide another basis for reasoning with interval information [5]. Consequently, developing effective feature selection methods for interval-valued data has become an increasingly important research topic [6].

Rough set theory (RST) [7] provides a classical mathematical framework for uncertainty modeling and has been widely studied in feature selection [8]. However, conventional RST is mainly applicable to symbolic data, and the discretization of continuous attributes may lead to information loss [9]. To overcome this limitation, fuzzy rough set (FRS) theory [10] extends the rough-set framework by incorporating fuzzy similarity. Rough-set models have been used to handle fuzzy interval values in group decision-making [11]. The probabilistic treatment of fuzzy events provides an early theoretical basis for connecting fuzziness and uncertainty [12]. Recent FRS-based feature selection has also introduced robust vague quantifiers for ordinal classification [13]. From an information-theoretic

* Corresponding author.
Email addresses: chxuwh@swu.edu.cn (W. Xu), swu2024lyh@email.swu.edu.cn (Y. Lv).

https://doi.org/10.1016/j.ins.2026.123865
Received 14 April 2026; Received in revised form 26 June 2026; Accepted 2 July 2026
Available online 6 July 2026

perspective, fuzzy probabilistic approximation spaces provide information measures for fuzzy granules [14]. Fuzzy-rough minimum classification error criteria further connect fuzzy-rough evaluation with classification performance [15].

Building on this foundation, a number of studies have extended FRS-based feature selection to different interval-valued and uncertainty-aware settings. For example, Jensen et al. proposed a fuzzy set model with interval values for handling missing-value data [16]. Li et al. developed an interval-valued feature selection method based on interval dominance relations for ordered information systems [17]. Chen et al. improved attribute reduction by introducing a fuzzy discernibility matrix [18]. Dong et al. designed an incremental feature selection method with fuzzy rough sets for dynamic datasets [19]. From a structural perspective, Zhang et al. combined graph-theoretic ideas with granular-rectangular neighborhood rough sets for interval-valued feature selection [20]. These studies demonstrate the effectiveness of FRS-based models for interval-valued data, but they do not fully address the additional challenges discussed below.

However, despite its advantages in handling uncertainty, the standard FRS framework still faces an important challenge in class-imbalanced settings. Conventional FRS dependency measures are typically influenced more heavily by the majority class, which may bias feature selection toward majority-class discrimination and may overlook minority-informative features. To alleviate this issue, several studies have introduced specialized strategies. For instance, Xu et al. refined neighborhood construction by combining δ-neighborhoods with k-nearest neighbors to better handle mixed samples in imbalanced data [21], and also proposed a local composite entropy criterion for imbalanced fuzzy data [22]. Although these studies have improved class-imbalance handling to some extent, they mainly modify neighborhood structures or evaluation criteria, rather than directly addressing the bias inherent in dependency evaluation itself.

In addition to the issue of class imbalance, another important challenge lies in modeling complex feature correlations [23]. Many practical FRS-based algorithms still adopt relevance-dominant search strategies or consider only limited-order correlations, without explicitly modeling complementarity and synergy [24]. More broadly, the feature selection literature has developed a variety of strategies to balance relevance and redundancy. The minimum redundancy maximum relevance framework is a representative relevance-redundancy criterion [25]. Conditional mutual information-based selection evaluates feature usefulness under already selected variables [26]. Graph-based filtering models provide another way to rank features through structural relationships [27]. Redundancy-complementariness dispersion explicitly models redundant and complementary information in feature selection [28]. Interaction-aware feature selection further evaluates dependencies among features beyond individual relevance [29]. Feature-interaction criteria have also been applied to text categorization [30]. Fuzzy information-theoretic feature selection has further considered relevance, redundancy, and complementarity criteria [31]. FFS-MCC fuses approximation and fuzzy uncertainty measures for multi-correlation collaboration [32]. These studies provide valuable insights into correlation-aware feature selection. However, they either focus on single-valued data or do not jointly address class imbalance and feature correlation modeling within a unified FRS framework for interval-valued data.

Taken together, despite considerable progress in FRS-based feature selection, an important research gap remains: there is still a lack of a unified framework that simultaneously addresses interval-valued uncertainty, class imbalance, and complex feature correlations. Existing methods usually treat these challenges separately rather than jointly, which limits their effectiveness in real-world scenarios where these factors coexist.

The literature can be summarized from three aspects. Interval-valued fuzzy-rough methods mainly address uncertainty representation but usually do not correct majority-class bias. Imbalance-oriented rough-set methods improve class sensitivity but generally do not model higher-order complementarity and redundancy. Correlation-aware selection methods characterize relevance, redundancy, or interaction, but most are designed for single-valued data or do not incorporate class-balanced fuzzy-rough dependency evaluation. Therefore, a unified framework is still needed to connect interval-valued uncertainty modeling, class-imbalance-aware dependency evaluation, and adaptive synergy-redundancy analysis.

To address this research gap, this paper proposes the Class-Imbalance-Aware Adaptive Synergy-Redundancy Feature Selection (CASR-FS) algorithm. CASR-FS is a supervised fuzzy-rough feature ranking method for interval-valued data. It evaluates candidate features by jointly considering class sensitivity, feature collaboration, and redundancy control.

The main contributions of this paper can be summarized as follows.

- We establish a unified fuzzy-rough feature selection framework for class-imbalanced interval-valued data, in which class-balanced dependency evaluation, feature interaction modeling, and adaptive conditional redundancy control are integrated within a single selection criterion.
- We adapt macro-averaged approximation quality to the fuzzy-rough dependency evaluation of class-imbalanced interval-valued data, so that each class contributes equally to the dependency score.
- We incorporate a collaborative gain term based on fuzzy joint mutual information and an adaptive conditional redundancy penalty, enabling the selected subset to balance discriminative ability, complementarity, and redundancy control within a single evaluation criterion.
- Through extensive experiments on 14 public datasets with multiple classifiers, together with non-parametric significance tests, we show that CASR-FS achieves favorable overall performance, with consistent advantages in classification accuracy, Macro-F1 scores, and balanced performance across varying imbalance ratios.

The remainder of this paper is organized as follows. Section 2 introduces the necessary theoretical preliminaries. Section 3 details the proposed CASR-FS method. Section 4 presents and analyzes the experimental results. Finally, Section 5 concludes the paper and discusses future work.

Table 1
Summary of main notation.

Symbol	Description
$\mathcal{O} = \{o_1, \dots, o_n\}$	Object set with n samples.
$\mathcal{F} = \{f_1, \dots, f_m\}$	Conditional feature set with m features.
C, C_k, K	Decision attribute, the kth decision class, and the number of classes.
$\mathcal{P}, Q, \mathcal{T}$	Generic feature subsets used in fuzzy information measures.
S, f_q	Currently selected feature subset and a candidate feature.
$[o_i]_{\mathcal{P}}$	Fuzzy knowledge granule induced by $\mathcal{P}$ for object o_i.
$\mathcal{R}_{f_j}$, $\mathcal{R}_{\mathcal{P}}$	Fuzzy similarity relation induced by feature f_j and by subset $\mathcal{P}$.
AQ, AQ_b	Approximation quality and balanced approximation quality.
$FCMI$, $FJMI$, FII	Fuzzy conditional mutual information, fuzzy joint mutual information, and fuzzy interaction information.
$\mathcal{CREL}$, $\mathcal{COM}$, $\mathcal{ITR}$, $\mathcal{CRED}$	Conditional relevance, complementarity, interaction, and conditional redundancy.
α	Adaptive redundancy coefficient, defined as $1/\lvert S\rvert$ when $S \neq \emptyset$.
$\mathcal{R}$, S_{opt}	Generated feature ranking and final selected subset under the evaluation protocol.

2. Preliminaries

This section introduces the theoretical preliminaries used in CASR-FS. Section 2.1 defines the interval-valued information system and the corresponding fuzzy rough set model. Section 2.2 then presents the fuzzy uncertainty measures derived from fuzzy knowledge granules. For readability, Table 1 summarizes the main symbols used throughout the paper.

2.1. Fuzzy rough sets for interval-valued data

In an interval-valued information system (IVIS), each conditional attribute value is represented by an interval rather than a single point. Formally, an IVIS is described as a tuple $\langle \mathcal{O}, \mathcal{F} \cup C \rangle$, where $\mathcal{O} = \{o_1, o_2, \dots, o_n\}$ is a non-empty finite set of objects, $\mathcal{F} = \{f_1, f_2, \dots, f_m\}$ is a non-empty finite set of conditional features, and C denotes the decision attribute. For any object $o_i \in \mathcal{O}$ and feature $f_j \in \mathcal{F}$, the value $f_j(o_i)$ is an interval $[l_{ij}, u_{ij}]$, where l_{ij} and u_{ij} denote the lower and upper bounds, respectively. The decision attribute induces a partition $\mathcal{O}/C = \{C_1, C_2, \dots, C_K\}$ of the object set, where each decision class C_k is treated as a crisp fuzzy set and $C_k(o_i) \in \{0, 1\}$ indicates whether o_i belongs to C_k. If $o_i \in C_k$, the decision-induced granule $[o_i]_C$ is represented by the crisp fuzzy set C_k.

A relation $\mathcal{R}_{f_j}$ on the object domain $\mathcal{O}$ is called a fuzzy similarity relation if it satisfies the following properties for any $o_i, o_k \in \mathcal{O}$:

1. Reflexivity: $\mathcal{R}_{f_j}(o_i, o_i) = 1$;
2. Symmetry: $\mathcal{R}_{f_j}(o_i, o_k) = \mathcal{R}_{f_j}(o_k, o_i)$.

For interval-valued data, attribute values are first normalized to the range $[0, 1]$ using the max-min method before fuzzy similarity construction. In the experiments, real-valued features are first transformed into intervals and then normalized. For each feature, the lower and upper bounds are normalized using the same feature-level minimum and maximum parameters, so that interval widths and relative positions remain comparable after scaling. If a feature is constant and the max-min denominator is zero, all normalized lower and upper bounds are set to 0 to avoid an undefined scaling result. Then, for any $f_j \in \mathcal{F}$, the fuzzy similarity relation $\mathcal{R}_{f_j}$ between o_i and o_k on f_j is defined as

$$\mathcal{R}_{f_j}(o_i, o_k) = \frac{\max(0, \min(u_{ij}, u_{kj}) - \max(l_{ij}, l_{kj}))}{\max(u_{ij}, u_{kj}) - \min(l_{ij}, l_{kj})}. \tag{1}$$

The relation $\mathcal{R}_{f_j}$ can be represented by a fuzzy similarity matrix, i.e., $\mathbf{M}_{\mathcal{R}_{f_j}} = (\mathcal{R}_{f_j}(o_i, o_k))_{n \times n}$. In the rare case where the denominator in Eq. (1) is zero, the two intervals are identical singleton intervals; their similarity is set to 1 to preserve reflexivity. It also induces a fuzzy knowledge granule (FKG) for each object o_i with respect to feature f_j, denoted by $[o_i]_{f_j}$. Here, $[o_i]_{f_j}$ is a fuzzy set in which the membership degree of any object o_k is given by $\mathcal{R}_{f_j}(o_i, o_k)$.

For a feature subset $\mathcal{P} \subseteq \mathcal{F}$, the corresponding fuzzy similarity relation on $\mathcal{O}$ is defined as $\mathcal{R}_{\mathcal{P}}(o_i, o_k) = \min_{f_j \in \mathcal{P}}\{\mathcal{R}_{f_j}(o_i, o_k)\}$. This subset-level relation aggregates the object similarity induced by all features in $\mathcal{P}$ through the minimum operator. The minimum operator is adopted as a conjunctive aggregation of feature-wise similarities in the fuzzy rough set model. In interval-valued data, it reflects the overlap structure of the feature intervals under the selected subset and provides the basis for the induced fuzzy rough approximations.

Based on this subset-level relation, the fuzzy lower and upper approximations of a decision class C_k with respect to a feature subset $\mathcal{P}$ are defined as

$$\underline{\mathcal{R}_{\mathcal{P}}}(C_k)(o_i) = \inf_{o_j \in \mathcal{O}} \sup\{1 - \mathcal{R}_{\mathcal{P}}(o_i, o_j), C_k(o_j)\}. \tag{2}$$

$$\overline{\mathcal{R}_{\mathcal{P}}}(C_k)(o_i) = \sup_{o_j \in \mathcal{O}} \inf\{\mathcal{R}_{\mathcal{P}}(o_i, o_j), C_k(o_j)\}. \tag{3}$$

Definition 1. The dependency of the decision attribute C on a feature subset $\mathcal{P}$ is measured by the approximation quality (AQ), defined as

$$AQ_{\mathcal{P}}(C) = \frac{|\mathrm{Pos}_{\mathcal{P}}(C)|}{|\mathcal{O}|} = \frac{\sum_{o_i \in \mathcal{O}} \mathrm{Pos}_{\mathcal{P}}(C)(o_i)}{|\mathcal{O}|}, \tag{4}$$

where $\mathrm{Pos}_{\mathcal{P}}(C)$ is the fuzzy positive region of C with respect to $\mathcal{P}$, which is calculated by $\bigcup_{C_k \in \mathcal{O}/C} \underline{\mathcal{R}_{\mathcal{P}}}(C_k)$.

The approximation quality measures the proportion of objects that can be consistently classified by the feature subset $\mathcal{P}$. A higher AQ value indicates that the subset has stronger discriminative ability.

2.2. Fuzzy uncertainty measures

While approximation quality characterizes uncertainty from an algebraic perspective, information-theoretic measures provide a complementary view. In this paper, they are defined through the cardinalities of fuzzy granules rather than through statistical probabilities. They are later used to characterize conditional relevance, collaborative gain, and higher-order interaction in Section 3.3.

Definition 2. For three granule-inducing sets $\mathcal{P}$, $\mathcal{Q}$, and $\mathcal{T}$, each of which may be a feature subset or the decision partition C, the fuzzy conditional mutual information is defined as

$$FCMI(\mathcal{P};\mathcal{Q}|\mathcal{T}) = -\frac{1}{|\mathcal{O}|}\sum_{i=1}^{|\mathcal{O}|} \log_2 \frac{|[o_i]_{\mathcal{T}} \cap [o_i]_{\mathcal{P}}| \cdot |[o_i]_{\mathcal{T}} \cap [o_i]_{\mathcal{Q}}|}{|[o_i]_{\mathcal{P}} \cap [o_i]_{\mathcal{Q}} \cap [o_i]_{\mathcal{T}}| \cdot |[o_i]_{\mathcal{T}}|}, \tag{5}$$

where $\left|[o_i]_{\mathcal{P}}\right|$ denotes the cardinality of the granule induced by $\mathcal{P}$. When $\mathcal{P}$ is a feature subset, it is computed by $\sum_{j=1}^{|\mathcal{O}|} \mathcal{R}_{\mathcal{P}}(o_i, o_j)$; when $\mathcal{P} = C$, $[o_i]_C$ denotes the corresponding crisp decision-class granule defined above. For numerical stability, zero cardinalities appearing in denominators or logarithmic arguments are lower-bounded by a small constant $\varepsilon = 10^{-12}$ in implementation. Fuzzy conditional mutual information measures how much *new* information $\mathcal{P}$ provides about $\mathcal{Q}$ when the information from $\mathcal{T}$ is already known.

Definition 3. For three granule-inducing sets $\mathcal{P}$, $\mathcal{Q}$, and $\mathcal{T}$, the fuzzy joint mutual information between the joint information induced by $\mathcal{P}$ and $\mathcal{Q}$ and the target $\mathcal{T}$ is defined as

$$FJMI(\mathcal{P},\mathcal{Q};\mathcal{T}) = -\frac{1}{|\mathcal{O}|}\sum_{i=1}^{|\mathcal{O}|} \log_2 \frac{|[o_i]_{\mathcal{P}} \cap [o_i]_{\mathcal{Q}}| \cdot |[o_i]_{\mathcal{T}}|}{|\mathcal{O}| \cdot |[o_i]_{\mathcal{Q}} \cap [o_i]_{\mathcal{T}} \cap [o_i]_{\mathcal{P}}|}. \tag{6}$$

Fuzzy joint mutual information quantifies how much information the feature subsets $\mathcal{P}$ and $\mathcal{Q}$ *jointly* provide about $\mathcal{T}$.

Definition 4. For three granule-inducing sets $\mathcal{P}$, $\mathcal{Q}$, and $\mathcal{T}$, the fuzzy interaction information among them is defined as

$$FII(\mathcal{P};\mathcal{Q};\mathcal{T}) = -\frac{1}{|\mathcal{O}|}\sum_{i=1}^{|\mathcal{O}|} \log_2 \frac{|[o_i]_{\mathcal{P}} \cap [o_i]_{\mathcal{Q}} \cap [o_i]_{\mathcal{T}}| \cdot |[o_i]_{\mathcal{P}}| \cdot |[o_i]_{\mathcal{Q}}| \cdot |[o_i]_{\mathcal{T}}|}{|\mathcal{O}| \cdot |[o_i]_{\mathcal{P}} \cap [o_i]_{\mathcal{Q}}| \cdot |[o_i]_{\mathcal{P}} \cap [o_i]_{\mathcal{T}}| \cdot |[o_i]_{\mathcal{Q}} \cap [o_i]_{\mathcal{T}}|}. \tag{7}$$

Fuzzy interaction information characterizes higher-order interactions that cannot be fully captured by pairwise relations alone.

Together, these measures provide the information-theoretic foundation for the collaboration and redundancy analysis developed in Section 3.3.

3. The CASR-FS method

In this section, we present the proposed CASR-FS method. Section 3.1 introduces the overall framework. Section 3.2 presents the class-imbalance-aware dependency measure, and Section 3.3 describes the synergy and redundancy analysis. Section 3.4 then defines the unified objective function, followed by the complete CASR-FS algorithm in Section 3.5.

3.1. The CASR-FS framework

The CASR-FS framework is illustrated in Fig. 1. Given an interval-valued information system as input, the method first constructs a fuzzy similarity relation for each feature and then induces the corresponding fuzzy knowledge granules, which provide the basis for subsequent fuzzy-rough and information-theoretic evaluation. CASR-FS is a supervised feature selection method because both the balanced approximation quality and the fuzzy information measures explicitly use the decision attribute C.

Based on this representation, CASR-FS evaluates candidate features from two complementary perspectives. From the fuzzy-rough perspective, it computes class-wise lower approximations and derives the class-imbalance-aware dependency measure AQ_b to assess class-balanced discriminative ability. From the information-theoretic perspective, it analyzes four types of feature relationships: conditional relevance, complementarity, interaction, and conditional redundancy. The first three are summarized through an FJMI-based collaborative gain term, while the last is modeled as an adaptive conditional redundancy penalty. These components are then integrated into the unified objective function introduced in Section 3.4, which is used in a sequential forward ranking procedure to generate a feature ranking. Finally, nested subsets induced by this ranking are evaluated, and the subset with the best classification performance is selected.

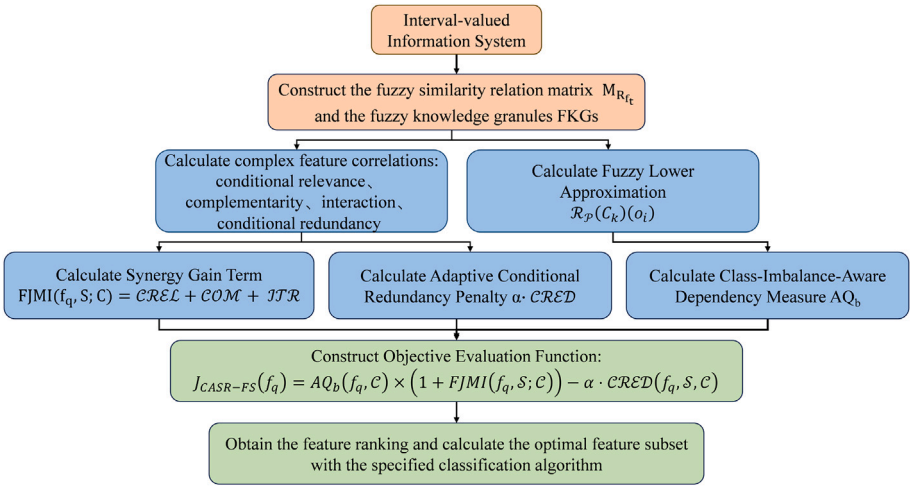


Fig. 1. The framework of the proposed CASR-FS algorithm.

3.2. Class-imbalance-aware dependency measure

Another challenge in feature selection for real-world data is class imbalance. The standard dependency measure, approximation quality, can be sensitive to this issue. Because the conventional approximation quality is calculated over the entire sample space, it functions as a micro-average, where samples from the larger majority class disproportionately dominate the final score. This bias may lead the selection process to favor features that are effective mainly for the majority class, while overlooking features that may be informative for minority-class discrimination.

To address this limitation, we define Balanced Approximation Quality (AQ_b). It replaces the conventional micro-average with a class-wise average. Instead of calculating a single dependency value over all samples, AQ_b computes approximation quality for each class and then averages the resulting scores to obtain the final dependency value, as described in Definition 5.

Definition 5. Let the set of decision classes be $\{C_1, C_2, \ldots, C_K\}$. The balanced approximation quality of a feature subset $\mathcal{P}$ is defined as

$$AQ_b(\mathcal{P}, C) = \frac{1}{K}\sum_{k=1}^{K} AQ_{\mathcal{P}}(C_k) = \frac{1}{K}\sum_{k=1}^{K} \frac{\sum_{o_i \in C_k} \underline{\mathcal{R}_{\mathcal{P}}}(C_k)(o_i)}{|C_k|}. \tag{8}$$

By design, this macro-averaging strategy ensures that every class, regardless of its size, contributes equally (with a weight of $1/K$) to the final score. As a result, the influence of the majority class is reduced, and the evaluation places greater emphasis on balanced discriminative ability across classes.

To illustrate this intuition, consider a binary imbalanced dataset with $|C_1| = 90$ and $|C_2| = 10$, where C_1 is the majority class. Suppose two candidate feature subsets, $\mathcal{P}_1$ and $\mathcal{P}_2$, produce the class-wise approximation qualities $AQ_{\mathcal{P}_1}(C_1) = 0.92$, $AQ_{\mathcal{P}_1}(C_2) = 0.30$, $AQ_{\mathcal{P}_2}(C_1) = 0.85$, and $AQ_{\mathcal{P}_2}(C_2) = 0.70$. Then

$$AQ_{\mathcal{P}_1}(C) = \frac{90 \times 0.92 + 10 \times 0.30}{100} = 0.858,$$
$$AQ_{\mathcal{P}_2}(C) = \frac{90 \times 0.85 + 10 \times 0.70}{100} = 0.835,$$
$$AQ_b(\mathcal{P}_1, C) = \frac{0.92 + 0.30}{2} = 0.61,$$
$$AQ_b(\mathcal{P}_2, C) = \frac{0.85 + 0.70}{2} = 0.775.$$

Under the conventional micro-averaged dependency measure, $\mathcal{P}_1$ would be preferred because its performance on the majority class dominates the final score. In contrast, AQ_b favors $\mathcal{P}_2$, since it provides substantially better discrimination for the minority class while maintaining good performance on the majority class. Therefore, AQ_b provides a more balanced assessment of the discriminative ability of a feature subset under class imbalance.

3.3. Synergy and redundancy analysis

Although the dependency measure in Section 3.2 evaluates the class-balanced relevance of a candidate feature, it does not explicitly characterize how that feature interacts with the subset of features already selected, denoted by S. In sequential feature selection, however, the contribution of a candidate feature depends not only on its individual relevance to the class but also on whether it provides complementary information or introduces redundant information relative to S.

To address this issue, CASR-FS analyzes the relationship among a candidate feature f_q, the selected subset S, and the decision class C through four types of correlations: conditional relevance, complementarity, interaction, and conditional redundancy. The first three characterize positive collaborative information, whereas the last captures overlapping information that should be penalized during selection.

Definition 6. Given a candidate feature f_q, the conditional relevance measures how much discriminative information the selected subset S still provides about the class C when f_q is known. It is calculated by

$$\mathcal{CREL}(f_q, S, C) = FCMI(S; C|f_q). \tag{9}$$

This term reflects the residual contribution of S to class discrimination after conditioning on f_q.

Definition 7. Conditioned on the selected subset S, complementarity measures the additional classification information provided by the new candidate feature f_q for the class C. It is calculated as

$$\mathcal{COM}(f_q, S, C) = FCMI(f_q; C|S). \tag{10}$$

This term measures how much additional discriminative information f_q contributes beyond what is already provided by S.

Definition 8. The interaction among the candidate feature f_q, the selected set S, and the class C is defined using the fuzzy interaction information:

$$\mathcal{ITR}(f_q, S, C) = FII(f_q; S; C). \tag{11}$$

This term captures the three-way interaction among f_q, S, and C, and is used to describe interaction effects that are not reflected by conditional relevance or complementarity alone.

Definition 9. Conditioned on the class C, conditional redundancy measures the information shared between the candidate feature f_q and the selected subset S. It is calculated as

$$\mathcal{CRED}(f_q, S, C) = FCMI(f_q; S|C). \tag{12}$$

This term quantifies the amount of class-conditional overlapping information between f_q and S.

Definition 10 (Collaborative Gain Analysis). For a candidate feature f_q, the selected subset S, and the class C, the sum of conditional relevance, complementarity, and interaction can be expressed as fuzzy joint mutual information. This relationship is given by

$$\begin{aligned}
&\mathcal{CREL}(f_q, S, C) + \mathcal{COM}(f_q, S, C) + \mathcal{ITR}(f_q, S, C) \\
&= FCMI(S; C|f_q) + FCMI(f_q; C|S) + FII(f_q; S; C) \\
&= FJMI(f_q, S; C)
\end{aligned}$$

Proof. For brevity, for each object $o_i \in \mathcal{O}$, let $\Gamma_i^q = |[o_i]_{f_q}|$, $\Gamma_i^S = |[o_i]_S|$, and $\Gamma_i^C = |[o_i]_C|$. In addition,

$$\Gamma_i^{qS} = |[o_i]_{f_q} \cap [o_i]_S|, \quad \Gamma_i^{qC} = |[o_i]_{f_q} \cap [o_i]_C|,$$
$$\Gamma_i^{SC} = |[o_i]_S \cap [o_i]_C|, \quad \Gamma_i^{qSC} = |[o_i]_{f_q} \cap [o_i]_S \cap [o_i]_C|.$$

Substituting these quantities into the three terms gives

$$\mathcal{CREL}(f_q, S, C) = -\frac{1}{|\mathcal{O}|} \sum_{i=1}^{|\mathcal{O}|} \log_2 \frac{\Gamma_i^{qS} \Gamma_i^{qC}}{\Gamma_i^{qSC} \Gamma_i^{q}},$$
$$\mathcal{COM}(f_q, S, C) = -\frac{1}{|\mathcal{O}|} \sum_{i=1}^{|\mathcal{O}|} \log_2 \frac{\Gamma_i^{qS} \Gamma_i^{SC}}{\Gamma_i^{qSC} \Gamma_i^{S}},$$
$$\mathcal{ITR}(f_q, S, C) = -\frac{1}{|\mathcal{O}|} \sum_{i=1}^{|\mathcal{O}|} \log_2 \frac{\Gamma_i^{qSC} \Gamma_i^{q} \Gamma_i^{S} \Gamma_i^{C}}{|\mathcal{O}| \Gamma_i^{qS} \Gamma_i^{qC} \Gamma_i^{SC}}.$$

Adding the above three equalities and using $\log_2 a + \log_2\ b + \log_2 c = \log_2(abc)$, we obtain

$$\begin{aligned}
&\mathcal{CREL}(f_q, S, C) + \mathcal{COM}(f_q, S, C) + \mathcal{ITR}(f_q, S, C) \\
&= -\frac{1}{|\mathcal{O}|} \sum_{i=1}^{|\mathcal{O}|} \log_2 \left(\frac{\Gamma_i^{qS} \Gamma_i^{qC}}{\Gamma_i^{qSC} \Gamma_i^{q}} \cdot \frac{\Gamma_i^{qS} \Gamma_i^{SC}}{\Gamma_i^{qSC} \Gamma_i^{S}} \cdot \frac{\Gamma_i^{qSC} \Gamma_i^{q} \Gamma_i^{S} \Gamma_i^{C}}{|\mathcal{O}| \Gamma_i^{qS} \Gamma_i^{qC} \Gamma_i^{SC}} \right) \\
&= -\frac{1}{|\mathcal{O}|} \sum_{i=1}^{|\mathcal{O}|} \log_2 \frac{\Gamma_i^{qS} \Gamma_i^{C}}{|\mathcal{O}| \Gamma_i^{qSC}}
\end{aligned}$$

$$= -\frac{1}{|\mathcal{O}|}\sum_{i=1}^{|\mathcal{O}|}\log_2\frac{|[o_i]_{f_q}\cap[o_i]_S|\cdot|[o_i]_C|}{|\mathcal{O}|\cdot|[o_i]_{f_q}\cap[o_i]_S\cap[o_i]_C|}$$

$$= FJMI(f_q, S; C).$$

□

According to the above analysis, conditional relevance, complementarity, and interaction can be treated jointly as positive collaborative information and summarized by the FJMI term, whereas conditional redundancy acts as a penalty term. This decomposition also makes the different roles of the four relationships explicit and provides the basis for the unified objective function in Section 3.4, where collaborative contribution and redundancy control are incorporated simultaneously during feature evaluation. Here, FJMI is interpreted as a fuzzy-granule-based collaborative information score. Under the adopted definitions and the ε-smoothed implementation, the non-degenerate values used in our experiments are non-negative. Nevertheless, because fuzzy-granule information measures are not probability measures in the strict statistical sense, pathological numerical cases may yield very small or slightly negative values.

3.4. Objective function

To evaluate each remaining candidate feature during sequential selection, CASR-FS uses a unified objective function that jointly considers class-balanced dependency, collaborative gain, and conditional redundancy.

Definition 11 (Objective Function). For a candidate feature $f_q \in \mathcal{F} \setminus S$ with $|S| \geq 1$, its objective score is defined as

$$J_{CASR-FS}(f_q) = AQ_b(\{f_q\}, C) \times \big(1 + FJMI(f_q, S; C)\big) - \alpha \cdot \mathcal{CRED}(f_q, S, C), \tag{13}$$

where $AQ_b(\{f_q\}, C)$ denotes the balanced approximation quality of the singleton feature subset $\{f_q\}$, and the adaptive coefficient α is defined as $1/|S|$.

When $S = \emptyset$, the first feature is initialized by maximizing $AQ_b(\{f_q\}, C)$. In this case, the collaborative gain term and the redundancy penalty are not involved.

The objective function combines the following three considerations.

- The term $AQ_b(\{f_q\}, C)$ measures the class-balanced discriminative ability of the candidate feature f_q. By using the macro-averaged dependency evaluation in Section 3.2, this term reduces the bias of conventional dependency measures toward majority classes.
- The factor $\big(1 + FJMI(f_q, S; C)\big)$ serves as a collaborative gain term. According to Definition 10, it reflects the joint contribution of the candidate feature and the selected subset through conditional relevance, complementarity, and interaction. A larger $FJMI$ value leads to a larger contribution from the dependency term. If $FJMI$ is very small or slightly negative in a pathological numerical case, this factor weakens the singleton dependency contribution; when the candidate is synergistic with the selected subset, it strengthens the contribution.
- The term $\alpha \cdot \mathcal{CRED}(f_q, S, C)$ acts as an adaptive conditional redundancy penalty. It suppresses candidate features that share excessive information with the current subset. Here, "adaptive" specifically refers to the subset-size-dependent coefficient $\alpha = 1/|S|$. Since $\alpha = 1/|S|$, the redundancy penalty is relatively stronger when the selected subset is small and becomes gradually moderated as the subset grows, which helps avoid excessive penalization in later stages.

Therefore, the proposed objective function provides a unified criterion for balancing class-sensitive discrimination, collaborative information, and redundancy control during sequential feature selection. The singleton dependency term $AQ_b(\{f_q\}, C)$ is used instead of $AQ_b(S \cup \{f_q\}, C)$ to keep the first factor candidate-specific and stable during each iteration, while the influence of the current subset is introduced through $FJMI$ and $\mathcal{CRED}$. Thus, $J_{CASR-FS}$ is used as a heuristic sequential ranking criterion rather than a monotonic set-function guarantee; the score of a candidate may change with S because collaborative and redundant relations are subset-dependent.

3.5. The CASR-FS algorithm

Algorithm 1 summarizes the complete procedure of CASR-FS. The method first computes the singleton balanced approximation quality $AQ_b(\{f_t\}, C)$ for each feature $f_t \in \mathcal{F}$. The feature with the largest singleton dependency score is then selected to initialize the selected subset S and the ranking sequence $\mathcal{R}$. After initialization, CASR-FS performs sequential forward ranking. For each remaining candidate feature $f_q \in \mathcal{F} \setminus S$, the algorithm evaluates $J_{CASR-FS}(f_q)$ using the corresponding $FJMI$ and $\mathcal{CRED}$ terms, and appends the best-scoring feature to S and $\mathcal{R}$. The ranking generation is the algorithmic core of CASR-FS. After the ranking is obtained, the nested subsets induced by $\mathcal{R}$ are evaluated by a specified classifier, and the subset with the highest classification accuracy is returned as S_{opt}. In the experiments of this paper, this subset-evaluation step is implemented using five-fold cross-validation. Ties are broken by the current candidate order and then by the smallest subset size.

Let n, m, and K denote the numbers of objects, features, and classes, respectively. The singleton precomputation stage requires $O(mKn^2)$, since each feature needs an $n \times n$ similarity matrix and class-wise lower approximations. After initialization, the ranking stage evaluates the remaining candidates iteratively, and each $FJMI/\mathcal{CRED}$ computation costs $O(n^2)$ when the singleton similarity

Algorithm 1: The CASR-FS Algorithm

Input: IVIS = $\langle \mathcal{O}, \mathcal{F} \cup C \rangle$ and a classifier for subset evaluation.
Output: $\mathcal{S}_{opt}$.

$\mathcal{R} \leftarrow \emptyset$, $\mathcal{S} \leftarrow \emptyset$
for $t = 1$ **to** m **do**
  Calculate the fuzzy similarity relation on f_t using Eq. (1)
  Obtain the fuzzy similarity matrix and the set of fuzzy knowledge granules induced by f_t
  Calculate the fuzzy lower approximations using Eq. (2)
  Calculate $AQ_b(\{f_t\}, C)$ using Eq. (8)
end
$f_{best} \leftarrow \arg\max_{f_t \in \mathcal{F}} AQ_b(\{f_t\}, C)$ $\mathcal{S} \leftarrow \{f_{best}\}$
$\mathcal{R} \leftarrow \mathcal{R} \oplus f_{best}$
while $|\mathcal{S}| < |\mathcal{F}|$ **do**
  foreach *candidate feature* $f_q \in \mathcal{F} \setminus \mathcal{S}$ **do**
    Calculate $FJMI(f_q, \mathcal{S}; C)$ according to Definition 3
    Calculate $C\mathcal{RED}(f_q, \mathcal{S}, C)$ according to Eq. (12)
    Calculate $J_{CASR-FS}(f_q)$ using Eq. (13)
  end
  $f_{best} \leftarrow \arg\max_{f_q \in \mathcal{F} \setminus \mathcal{S}} J_{CASR-FS}(f_q)$ $\mathcal{S} \leftarrow \mathcal{S} \cup \{f_{best}\}$
  $\mathcal{R} \leftarrow \mathcal{R} \oplus f_{best}$
end
Let $\{\mathcal{R}_k\}_{k=1}^{|\mathcal{R}|}$ be the set of nested subsets induced by $\mathcal{R}$, where $\mathcal{R}_k = \{\mathcal{R}[1], \ldots, \mathcal{R}[k]\}$
for $k = 1$ **to** $|\mathcal{R}|$ **do**
  Calculate the average classification accuracy $ACC(\mathcal{R}_k)$ using the specified classifier
end
$k_{opt} \leftarrow \arg\max_{k \in \{1,\ldots,|\mathcal{R}|\}} ACC(\mathcal{R}_k)$ $\mathcal{S}_{opt} \leftarrow \mathcal{R}_{k_{opt}}$
return $\mathcal{S}_{opt}$

matrices and the current subset-level relation are cached. In implementation, after a feature is appended to $\mathcal{S}$, the subset relation is incrementally updated by an elementwise minimum operation between the previous subset matrix and the newly selected feature matrix, avoiding reconstruction from all selected features. Therefore, the classifier-independent ranking stage has complexity $O\left(n^2(mK + m^2)\right)$. The final subset-evaluation stage examines m nested subsets, contributing $O\left(\sum_{r=1}^{m} T_{\text{eval}}(n, r)\right)$, where $T_{\text{eval}}(n, r)$ denotes the cost of evaluating an r-feature subset under the specified classifier and validation protocol. Hence, the overall complexity of CASR-FS is $O\left(n^2(mK + m^2) + \sum_{r=1}^{m} T_{\text{eval}}(n, r)\right)$.

4. Experiments and analysis

This section presents the experimental setup and results used to evaluate CASR-FS. The proposed method is compared with eight representative feature selection methods on fourteen public datasets.

All experiments were performed within the PyCharm integrated development environment using Python. The hardware platform was a personal computer equipped with an AMD Ryzen 7 5800H processor and 16 GB of RAM, running a 64-bit Windows 10 operating system.

4.1. Datasets and data preprocessing

To evaluate CASR-FS on public benchmarks, we selected 14 publicly available benchmark datasets (available from the UCI Machine Learning Repository: https://archive.ics.uci.edu/). Detailed statistical information is summarized in Table 2.

An additional column "IR" is introduced to indicate the imbalance ratio, calculated as the ratio of samples in the majority class to those in the minority class. For multi-class datasets, IR is computed as the ratio between the largest and smallest class sizes.

Because CASR-FS is designed for interval-valued data, the real-valued benchmark datasets were converted into interval-valued form before feature selection. This preprocessing step simulates uncertainty around real-valued observations, so the experimental results should be interpreted under this intervalization rule. Given the real-valued information system $\langle \mathcal{O}, \mathcal{F} \cup C \rangle$, for each feature $f_j \in \mathcal{F}$ and each object $o_i \in \mathcal{O}$, we set $l_{ij} = f_j(o_i) - 2\text{std}$ and $u_{ij} = f_j(o_i) + 2\text{std}$ [33], where std denotes the standard deviation of the feature values within the class to which o_i belongs under feature f_j.

4.2. Performance evaluation and classifiers

To provide a comprehensive and unbiased assessment, we employed three distinct classifiers to gauge the quality of the selected feature subsets: K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Decision Tree (DT). The classifiers were implemented

Table 2
The basic statistical information of experimental datasets.

No.	Datasets	Abbreviation	Samples	Features	Classes	IR
1	GLIOMA	GLIOMA	50	4434	4	2.14
2	colon	colon	62	2000	2	1.82
3	LSVT Voice Rehabilitation	LSVT	126	313	2	2.0
4	Toxicity	Toxicity	171	1203	2	2.05
5	Darwin	Darwin	174	450	2	1.1
6	Parkinsons	Parkinsons	195	23	2	3.1
7	Connectionist Bench (Sonar, Mines vs. Rocks)	Sonar	208	60	2	1.1
8	Dermatology	Dermatology	366	34	6	5.6
9	Breast Cancer Wisconsin (Diagnostic)	BCWD	569	30	2	1.7
10	Hill-Valley	Hill-Valley	606	100	2	1.0
11	HCV data	HCV	615	12	4	25.7
12	South German credit	SGC	1000	20	2	2.3
13	Image segmentation	Image	2310	19	7	1.0
14	Abalone	Abalone	4177	8	28	689.0

with fixed parameters to avoid dataset-specific tuning: KNN used $k = 5$, uniform neighbor weights, and Euclidean distance; SVM used an RBF kernel with $C = 1$ and $\gamma = \text{scale}$; and DT used the Gini criterion with the best-split strategy.

The evaluation focused on four aspects. Classification accuracy served as the primary indicator of subset effectiveness, while Macro-F1 was additionally reported to assess balanced class-wise performance under class imbalance. The final subset was selected according to cross-validated accuracy because it directly reflects the downstream classification performance of the selected feature subset. For a K-class problem, Macro-F1 is formally defined as

$$\text{Macro-F1} = \frac{1}{K}\sum_{k=1}^{K} F1_k. \tag{14}$$

$$F1_k = \frac{2 \cdot \text{Precision}_k \cdot \text{Recall}_k}{\text{Precision}_k + \text{Recall}_k}. \tag{15}$$

Each class therefore contributes equally to the final Macro-F1 score, regardless of its sample size. The number of selected features and the algorithm's running time were further recorded to evaluate subset compactness and computational efficiency. The performance of each feature selection algorithm was evaluated using a five-fold cross-validation strategy to ensure the reliability and stability of the results, and all reported results are averages over the five folds. The folds were generated by stratified five-fold partitioning so that the class distribution of each fold was as close as possible to that of the full dataset. In addition, Friedman and Wilcoxon signed-rank tests were used in Section 4.4.5 to examine the statistical significance of the observed performance differences.

4.3. Compared algorithms

To evaluate CASR-FS, we compared it with eight representative feature selection methods and a baseline using the full original feature set. A brief description of each comparison method is given below.

(1) Original dataset: Classification is performed using all conditional attributes of the complete dataset.
(2) KND-UFS [34]: An unsupervised method that integrates fuzzy c-means clustering and k-nearest neighbors rough sets.
(3) W-MGMN [35]: An incremental matrix-based method for evaluating feature importance.
(4) LFSACIE [22]: A fuzzy feature selection method for unbalanced data that uses a local composite entropy to evaluate features.
(5) GLSFS [20]: A graph-driven feature selection method for interval-valued data that combines graph theory with granular-rectangular neighborhood rough sets.
(6) FFS-MCC [32]: This method combines approximation and fuzzy uncertainty measures to evaluate features based on multi-correlation collaboration.
(7) INF-FS [27]: A graph-based filter framework in which features are treated as nodes and the edges between them represent correlation and redundancy.
(8) KNCMI [21]: This method employs an iterative strategy that integrates the δ-neighborhood with k-nearest neighbors and uses information entropy to evaluate feature importance and perform feature selection.
(9) CMIM [26]: A filter method that sequentially selects features by maximizing the conditional mutual information with the class label, given the already selected features.

For comparison methods originally designed for single-valued information systems, had each interval-valued feature converted into a single-point representation by midpoint extraction before applying the corresponding algorithm.

4.4. Evaluation results

Fig. 2 presents the evaluation framework used in the experiments. The compared methods differ in their output forms: KND-UFS, W-MGMN, and LFSACIE directly output feature subsets, whereas GLSFS, FFS-MCC, INF-FS, KNCMI, CMIM, and CASR-FS produce

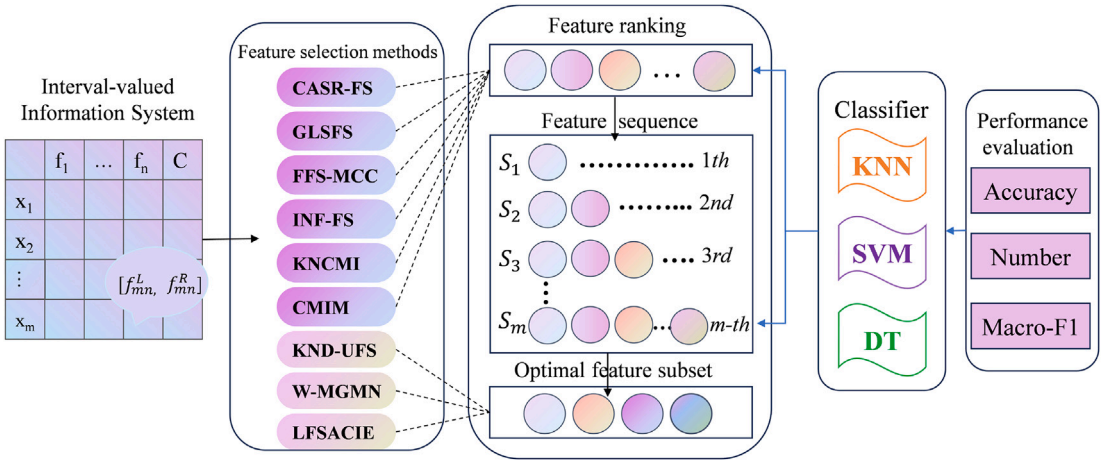


Fig. 2. The evaluation framework for feature selection methods.

feature rankings. For ranking-based methods, a common post-processing procedure was adopted: nested subsets induced by the ranking were evaluated with the designated classifier, and the subset with the highest classification accuracy was selected. Therefore, this classifier-based subset evaluation should be regarded as a unified comparison protocol rather than as part of the intrinsic ranking mechanism of the original methods. This protocol gives ranking-based methods an additional subset-selection step compared with methods that directly output a subset. We retain it because otherwise ranking methods would not yield a unique subset size, but this difference should be considered when interpreting subset compactness and accuracy comparisons.

4.4.1. Classification accuracy

Tables 3–5 report the classification accuracy of all algorithms under the KNN, SVM, and DT classifiers. The "Average" row gives the overall mean accuracy of each algorithm on the 14 datasets, and the bold numbers identify the highest accuracy on each dataset. The following observations are obtained from Tables 3–5.

- CASR-FS achieved the highest average classification accuracy across the three classifiers: 78.26% with KNN, 75.30% with SVM, and 75.41% with DT. Although CASR-FS was not the top-performing method on every individual dataset, these averages still indicate favorable overall performance across different classifiers.
- CASR-FS also performed well on high-dimensional data. On BCWD, GLIOMA, and colon, it achieved the highest accuracy under multiple classifiers, and it remained competitive on Darwin and LSVT. Even on the highly imbalanced Abalone dataset, where all methods experienced substantial performance degradation, CASR-FS still produced the best accuracy among the compared methods. This pattern is consistent with the unified objective, which combines class-balanced dependency evaluation, collaborative feature effects, and redundancy control.
- Finally, although the absolute performance of a selected subset depends on the downstream classifier, the subsets produced by CASR-FS showed relatively stable behavior across different classifiers on several datasets. For example, on BCWD, the subset selected by CASR-FS yielded accuracies of 97.36% with KNN, 95.96% with SVM, and 95.26% with DT. This result suggests that the selected features retain discriminative information under different classifier biases, rather than being effective only for a single learning algorithm.

Table 3
The comparison of classification accuracy (mean ± std. dev. %) on KNN.

Datasets	Original	KND-UFS	W-MGMN	LFSACIE	GLSFS	FFS-MCC	INF-FS	KNCMI	CMIM	CASR-FS
GLIOMA	80.00 ± 6.83	72.05 ± 5.91	65.28 ± 6.43	66.36 ± 7.26	78.67 ± 9.08	82.64 ± 5.99	78.08 ± 7.65	83.32 ± 5.26	83.76 ± 6.22	**84.00 ± 6.19**
colon	72.18 ± 6.47	65.71 ± 5.44	59.52 ± 8.03	60.61 ± 6.67	71.76 ± 10.12	74.99 ± 5.25	71.48 ± 8.95	78.08 ± 6.82	75.74 ± 7.82	**82.05 ± 6.93**
LSVT	77.78 ± 4.02	62.68 ± 4.22	59.48 ± 4.28	65.08 ± 2.97	70.58 ± 8.74	77.02 ± 2.66	62.65 ± 8.42	83.26 ± 6.96	76.12 ± 8.88	**89.63 ± 5.46**
Toxicity	62.61 ± 5.98	57.94 ± 5.00	48.53 ± 7.58	51.06 ± 6.16	62.33 ± 8.31	62.49 ± 4.54	61.88 ± 6.46	62.73 ± 4.63	62.97 ± 7.28	**63.47 ± 5.64**
Darwin	64.29 ± 9.04	76.94 ± 6.87	44.15 ± 10.21	55.11 ± 7.14	65.93 ± 14.82	76.99 ± 7.51	74.57 ± 13.23	79.00 ± 5.06	81.55 ± 7.60	**81.61 ± 7.57**
Parkinsons	71.28 ± 12.29	78.46 ± 6.61	86.67 ± 6.36	70.26 ± 21.85	75.90 ± 7.71	80.51 ± 8.97	76.41 ± 4.41	85.64 ± 6.80	80.51 ± 8.97	**87.18 ± 8.43**
Sonar	54.25 ± 12.94	55.33 ± 9.20	71.72 ± 11.49	52.92 ± 6.98	65.96 ± 12.47	**74.08 ± 10.76**	60.60 ± 7.93	59.55 ± 8.07	60.52 ± 15.65	72.61 ± 15.70
Dermatology	96.99 ± 3.40	76.78 ± 3.41	47.57 ± 5.63	24.06 ± 4.01	85.80 ± 2.50	96.45 ± 1.40	86.89 ± 2.53	83.61 ± 3.78	95.36 ± 1.37	**97.27 ± 1.50**
BCWD	96.66 ± 1.40	89.46 ± 1.73	65.73 ± 3.32	89.46 ± 3.96	92.79 ± 2.18	93.32 ± 1.72	92.79 ± 2.18	96.84 ± 1.31	92.97 ± 2.21	**97.36 ± 0.96**
Hill-Valley	54.46 ± 2.57	52.65 ± 5.90	49.17 ± 3.20	53.47 ± 3.22	55.61 ± 3.75	57.26 ± 3.84	55.61 ± 2.37	53.31 ± 4.23	57.26 ± 3.00	**57.43 ± 4.77**
HCV	88.94 ± 2.75	87.15 ± 2.02	86.99 ± 0.00	88.62 ± 0.89	91.06 ± 1.36	91.06 ± 0.89	90.73 ± 1.10	89.76 ± 1.90	**92.03 ± 1.95**	90.24 ± 1.15
SGC	70.70 ± 4.99	69.60 ± 1.46	70.00 ± 0.00	70.00 ± 0.00	65.10 ± 2.71	65.80 ± 2.77	65.80 ± 2.34	72.50 ± 4.34	70.00 ± 0.00	**75.20 ± 3.03**
Image	93.33 ± 2.48	90.65 ± 2.12	59.09 ± 3.33	63.81 ± 5.37	90.22 ± 3.42	93.68 ± 1.62	90.22 ± 3.42	**94.24 ± 2.57**	90.22 ± 3.42	93.33 ± 2.56
Abalone	23.73 ± 4.47	15.73 ± 3.21	10.00 ± 4.69	10.00 ± 4.32	21.73 ± 4.42	23.42 ± 2.77	21.25 ± 3.73	23.95 ± 3.38	23.70 ± 3.92	**24.23 ± 3.71**
Average	71.94 ± 5.69	67.94 ± 4.51	58.85 ± 5.33	58.63 ± 5.77	70.96 ± 6.54	74.98 ± 4.33	70.64 ± 5.34	74.70 ± 4.65	74.48 ± 5.59	**78.26 ± 5.26**

Table 4
The comparison of classification accuracy (mean ± std. dev. %) on SVM.

Datasets	Original	KND-UFS	W-MGMN	LFSACIE	GLSFS	FFS-MCC	INF-FS	KNCMI	CMIM	CASR-FS
GLIOMA	76.00 ± 6.27	64.51 ± 3.45	62.27 ± 4.48	68.34 ± 4.89	73.51 ± 3.60	73.78 ± 2.90	69.58 ± 6.74	76.09 ± 5.04	75.03 ± 4.63	**78.00 ± 7.27**
colon	77.31 ± 8.63	68.29 ± 3.11	66.54 ± 6.40	70.62 ± 4.06	72.65 ± 3.64	75.52 ± 2.72	71.24 ± 7.09	78.49 ± 4.43	75.83 ± 5.53	**79.10 ± 6.62**
LSVT	66.68 ± 1.68	66.68 ± 1.68	66.68 ± 1.68	66.68 ± 1.68	66.68 ± 1.68	**80.18 ± 3.39**	66.68 ± 1.68	69.08 ± 4.41	72.22 ± 5.67	73.78 ± 3.36
Toxicity	67.26 ± 8.89	56.35 ± 2.65	56.19 ± 5.20	59.08 ± 3.25	62.46 ± 3.05	63.49 ± 2.46	62.80 ± 8.88	68.06 ± 3.31	67.13 ± 4.58	**68.67 ± 6.73**
Darwin	72.76 ± 21.38	51.14 ± 0.57	42.42 ± 11.11	51.14 ± 0.57	51.14 ± 0.57	51.14 ± 0.57	65.14 ± 21.26	79.40 ± 0.19	61.46 ± 4.19	**80.89 ± 14.50**
Parkinsons	75.38 ± 1.26	75.38 ± 1.26	75.38 ± 1.26	78.46 ± 12.31	78.46 ± 12.31	80.00 ± 3.40	75.38 ± 2.05	80.51 ± 4.76	80.00 ± 4.97	**82.05 ± 4.59**
Sonar	62.92 ± 14.06	56.27 ± 11.66	**74.63 ± 12.38**	62.43 ± 15.20	67.29 ± 6.39	64.40 ± 9.22	62.01 ± 8.53	62.92 ± 15.43	62.97 ± 6.17	72.62 ± 10.56
Dermatology	96.72 ± 1.10	75.14 ± 4.66	50.27 ± 1.03	35.25 ± 1.59	92.89 ± 2.81	**97.81 ± 1.40**	92.89 ± 2.81	82.79 ± 2.35	95.91 ± 2.84	96.72 ± 1.10
BCWD	95.08 ± 1.52	84.19 ± 4.59	62.74 ± 0.39	89.46 ± 2.53	86.12 ± 1.50	89.63 ± 2.80	68.18 ± 3.42	95.08 ± 1.52	91.38 ± 1.99	**95.96 ± 1.05**
Hill-Valley	50.33 ± 0.17	50.66 ± 2.95	51.49 ± 2.82	51.66 ± 3.09	51.32 ± 3.39	50.99 ± 0.62	51.49 ± 2.82	50.17 ± 0.33	**52.31 ± 1.03**	50.33 ± 0.17
HCV	87.80 ± 0.00	86.99 ± 2.91	87.80 ± 0.00	87.48 ± 0.40	87.80 ± 0.00	88.94 ± 0.40	87.80 ± 0.00	87.80 ± 0.00	88.29 ± 0.65	**89.43 ± 0.51**
SGC	70.50 ± 4.72	69.60 ± 1.16	70.00 ± 0.00	67.30 ± 4.26	70.00 ± 1.30	70.00 ± 0.32	70.20 ± 0.81	72.60 ± 5.75	70.50 ± 6.06	**73.50 ± 5.80**
Image	85.06 ± 4.39	49.65 ± 7.15	52.73 ± 1.21	57.75 ± 4.40	78.10 ± 2.92	82.51 ± 3.97	54.37 ± 6.88	86.41 ± 5.09	61.21 ± 3.38	**86.45 ± 4.90**
Abalone	26.36 ± 3.30	13.68 ± 3.57	11.66 ± 2.86	16.21 ± 3.57	25.67 ± 2.83	23.20 ± 2.11	18.46 ± 3.54	26.17 ± 3.06	23.87 ± 2.75	**26.70 ± 3.91**
Average	72.15 ± 5.53	62.04 ± 3.67	59.34 ± 3.63	61.56 ± 4.41	68.86 ± 3.29	70.83 ± 2.59	65.44 ± 5.47	72.54 ± 3.98	69.86 ± 3.89	**75.30 ± 5.08**

Table 5
The comparison of classification accuracy (mean ± std. dev. %) on DT.

Datasets	Original	KND-UFS	W-MGMN	LFSACIE	GLSFS	FFS-MCC	INF-FS	KNCMI	CMIM	CASR-FS
GLIOMA	66.00 ± 9.06	57.30 ± 8.21	54.44 ± 7.15	56.26 ± 4.64	67.06 ± 7.67	67.21 ± 5.88	66.73 ± 5.91	67.49 ± 5.73	67.47 ± 6.04	**68.00 ± 5.55**
colon	71.03 ± 10.03	64.04 ± 8.26	63.05 ± 7.22	63.21 ± 5.38	76.15 ± 8.32	75.74 ± 5.41	74.80 ± 7.34	75.41 ± 5.29	76.75 ± 7.18	**77.31 ± 5.67**
LSVT	83.38 ± 9.12	61.88 ± 6.62	81.69 ± 6.09	66.68 ± 1.68	84.86 ± 10.89	85.66 ± 6.56	**87.29 ± 5.32**	86.49 ± 4.12	86.49 ± 7.44	86.52 ± 6.94
Toxicity	52.59 ± 8.64	44.27 ± 7.45	45.37 ± 6.61	44.49 ± 4.99	56.50 ± 7.29	58.73 ± 5.65	56.75 ± 6.60	57.53 ± 4.66	58.58 ± 6.31	**60.22 ± 5.75**
Darwin	69.43 ± 13.51	67.65 ± 15.50	56.92 ± 10.27	67.82 ± 7.95	77.53 ± 7.77	81.03 ± 5.91	78.64 ± 11.54	77.89 ± 3.10	82.08 ± 9.48	**83.31 ± 7.15**
Parkinsons	69.74 ± 14.45	73.33 ± 13.03	66.15 ± 7.50	83.59 ± 8.21	80.00 ± 6.76	83.59 ± 8.21	81.54 ± 5.23	81.54 ± 6.36	**86.67 ± 10.05**	85.13 ± 10.05
Sonar	60.22 ± 13.73	61.53 ± 3.41	62.11 ± 9.54	64.45 ± 8.10	69.33 ± 8.62	68.79 ± 11.61	68.40 ± 11.20	65.99 ± 13.13	69.28 ± 4.83	**69.76 ± 6.33**
Dermatology	93.18 ± 3.10	73.77 ± 6.44	35.80 ± 1.18	35.25 ± 1.59	**93.46 ± 3.65**	93.18 ± 2.57	92.35 ± 4.11	85.52 ± 2.92	91.53 ± 2.20	93.45 ± 3.03
BCWD	91.73 ± 2.42	92.27 ± 1.71	84.70 ± 3.65	92.97 ± 2.28	94.73 ± 0.96	**95.26 ± 2.04**	94.03 ± 0.65	94.38 ± 1.72	94.03 ± 1.28	**95.26 ± 1.31**
Hill-Valley	57.75 ± 3.44	50.82 ± 4.54	53.46 ± 2.97	52.15 ± 5.24	60.89 ± 3.03	59.90 ± 1.52	57.76 ± 2.01	56.61 ± 2.67	59.57 ± 2.44	**62.22 ± 4.99**
HCV	90.41 ± 2.74	62.44 ± 17.20	86.02 ± 2.93	87.80 ± 1.85	90.57 ± 2.44	90.08 ± 1.30	89.92 ± 2.33	89.27 ± 3.14	90.24 ± 2.24	**90.73 ± 2.16**
SGC	66.70 ± 4.88	63.70 ± 5.94	67.30 ± 3.26	**70.00 ± 0.00**	66.10 ± 3.68	67.70 ± 2.62	67.70 ± 4.03	69.10 ± 3.93	**70.00 ± 0.00**	68.20 ± 4.32
Image	95.28 ± 1.32	94.59 ± 1.58	50.91 ± 3.54	20.48 ± 3.34	95.15 ± 1.51	95.02 ± 1.42	95.15 ± 2.06	94.81 ± 1.74	95.45 ± 1.46	**95.58 ± 1.53**
Abalone	19.75 ± 4.68	11.02 ± 5.73	10.00 ± 4.04	10.00 ± 3.33	18.79 ± 3.14	19.31 ± 3.52	18.54 ± 3.38	19.76 ± 3.53	19.59 ± 2.75	**20.09 ± 3.17**
Average	70.51 ± 7.22	62.76 ± 7.54	58.42 ± 5.43	58.23 ± 4.18	73.65 ± 5.41	74.37 ± 4.59	73.54 ± 5.12	72.98 ± 4.43	74.84 ± 4.55	**75.41 ± 4.85**

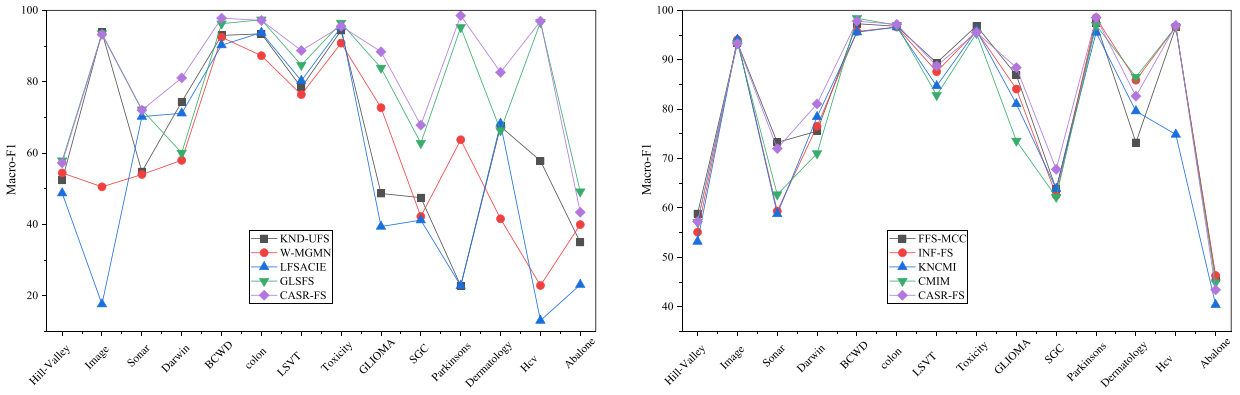
4.4.2. Macro-F1 performance analysis

To further assess the effectiveness of CASR-FS in class-imbalanced settings, we analyzed the experimental results using Macro-F1 from two complementary perspectives: average performance across all datasets and trend analysis under different imbalance ratios. Compared with overall accuracy, Macro-F1 places greater emphasis on balanced class-wise performance and is therefore particularly suitable for evaluating feature selection methods under class imbalance.

Fig. 6 presents the average Macro-F1 scores of the nine feature selection methods on the 14 datasets under the three classifiers. CASR-FS achieves the highest average Macro-F1 score in all three cases. This observation suggests that the subsets selected by CASR-FS provide more balanced class-wise discrimination across datasets, which is consistent with the class-balanced dependency design in the unified objective function. The similar tendency across KNN, SVM, and DT also suggests that this behavior is not limited to a single downstream classifier.

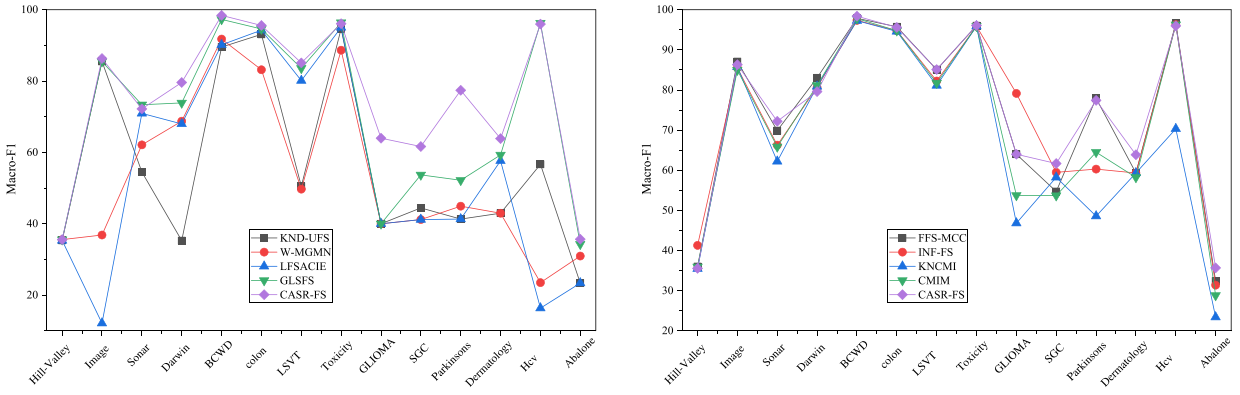
To further investigate how CASR-FS performs under varying degrees of class imbalance, Figs. 3–5 present the Macro-F1 scores on all 14 datasets ordered by imbalance ratio (IR) under the three classifiers. For visual clarity, the eight comparative methods are divided into two groups in each figure: panel (a) includes KND-UFS, W-MGMN, LFSACIE, and GLSFS, whereas panel (b) includes FFS-MCC, INF-FS, KNCMI, and CMIM.

Several observations can be drawn from this trend analysis. On balanced datasets such as Hill-Valley (IR = 1.0) and Image (IR = 1.0), CASR-FS remains competitive, indicating that its class-imbalance-aware design does not noticeably reduce effectiveness on balanced data. On moderately imbalanced datasets such as colon (IR = 1.82), LSVT (IR = 2.0), Toxicity (IR = 2.05), and GLIOMA (IR = 2.14), CASR-FS shows a more noticeable advantage over several compared methods. On more imbalanced datasets such as Dermatology (IR = 5.6), HCV (IR = 25.7), and especially Abalone (IR = 689), the Macro-F1 values of all methods decline, but CASR-FS still maintains relatively strong and stable behavior across different imbalance levels. Beyond these Macro-F1 trends, the overall accuracy and classifier-specific results also include cases where CASR-FS is not the best method. For example, CMIM performs better on HCV under KNN, FFS-MCC is stronger on LSVT and Dermatology under SVM, and CMIM or LFSACIE can be competitive on some DT settings. These cases suggest that when the class structure is already well captured by a conditional mutual-information criterion or when a simple tree classifier benefits from a different partitioning bias, the collaborative redundancy modeling in CASR-FS may not always yield the top score. Thus, the results should be interpreted as overall favorable rather than uniformly dominant.



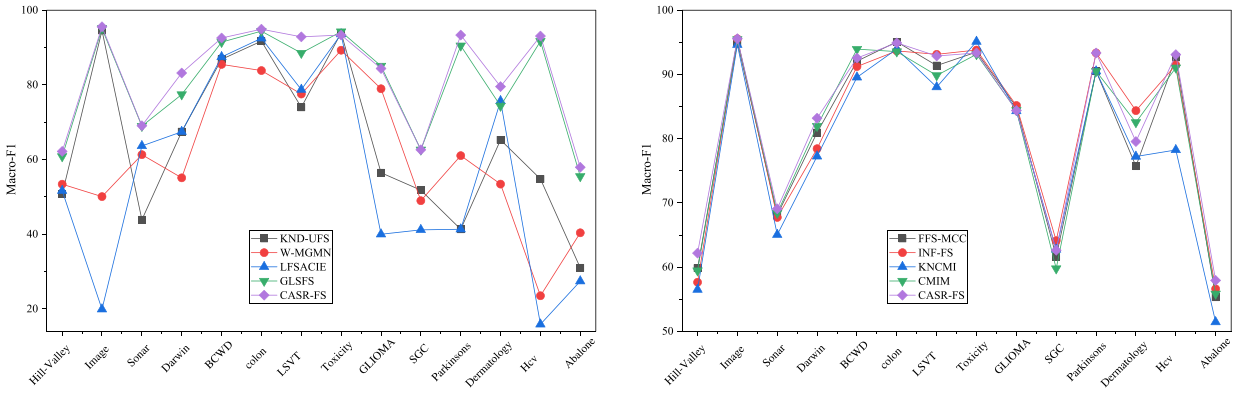
(a) (b)

Fig. 3. Macro-F1 scores for the KNN classifier across all datasets sorted in ascending order of imbalance ratio.



(a) (b)

Fig. 4. Macro-F1 scores of the SVM classifier across all datasets sorted in ascending order of imbalance ratio.



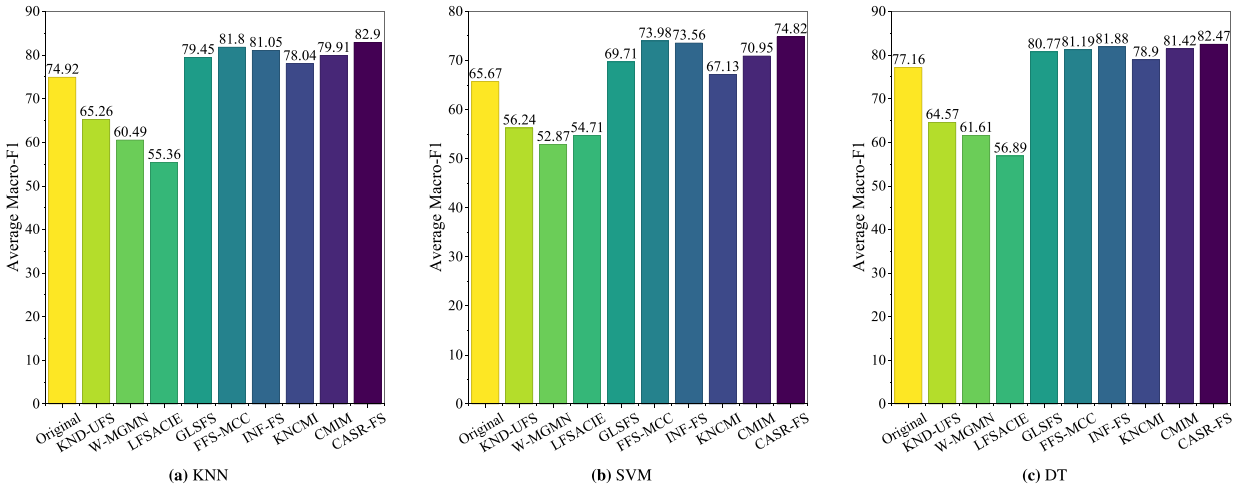
(a) (b)

Fig. 5. Macro-F1 scores of the DT classifier across all datasets sorted in ascending order of imbalance ratio.

4.4.3. Number of selected features

This section analyzes the dimensionality reduction capabilities of the compared algorithms, with detailed results on the number of selected features presented in Tables 6–8. The "Average" column reports the mean number of selected features for each algorithm across all datasets.

Although all methods achieve a substantial reduction in feature space, a crucial challenge in feature selection lies in striking an optimal balance between reducing dimensionality and preserving classification performance. An analysis of the results reveals a clear



(a) KNN **(b)** SVM **(c)** DT

Fig. 6. The average Macro-F1 scores of nine FS methods using three classification algorithms across all datasets.

Table 6
Number of selected features on KNN.

Datasets	Original	KND-UFS	W-MGMN	LFSACIE	GLSFS	FFS-MCC	INF-FS	KNCMI	CMIM	CASR-FS
GLIOMA	4434	209	235	348	2647	925	1113	1024	1068	92
colon	2000	97	105	137	1043	446	481	399	402	39
LSVT	313	8	176	22	111	145	26	118	61	32
Toxicity	1203	68	72	85	541	250	161	189	228	67
Darwin	450	12	18	10	334	18	19	42	22	4
Parkinsons	23	6	3	4	7	5	2	4	5	4
Sonar	60	4	34	4	24	28	3	5	10	39
Dermatology	34	6	5	3	29	27	33	14	28	31
BCWD	30	4	5	4	28	28	25	27	26	29
Hill-Valley	100	7	15	10	72	14	21	8	11	9
HCV	12	4	1	4	8	10	12	3	2	4
SGC	20	6	4	2	12	19	13	8	11	8
Image	19	12	2	2	14	16	13	12	17	11
Abalone	8	3	1	2	6	6	6	4	7	5
Average	621.86	31.86	48.29	45.50	348.29	138.36	137.71	132.64	135.57	26.71

Table 7
Number of selected features on SVM.

Datasets	Original	KND-UFS	W-MGMN	LFSACIE	GLSFS	FFS-MCC	INF-FS	KNCMI	CMIM	CASR-FS
GLIOMA	4434	209	235	348	1011	683	1171	1024	534	78
colon	2000	97	105	137	552	266	410	363	227	46
LSVT	313	8	176	22	96	84	16	6	48	24
Toxicity	1203	68	72	85	403	170	218	294	138	91
Darwin	450	12	18	10	408	76	6	24	27	80
Parkinsons	23	6	3	4	7	10	11	12	13	3
Sonar	60	4	34	4	16	15	17	58	9	7
Dermatology	34	6	5	3	32	33	33	14	33	32
BCWD	30	4	5	4	24	11	13	29	19	8
Hill-Valley	100	7	15	10	2	4	19	2	7	8
HCV	12	4	1	4	8	4	6	1	1	7
SGC	20	6	4	2	12	19	14	12	11	14
Image	19	12	2	2	14	17	12	10	17	11
Abalone	8	3	1	2	7	7	6	5	7	6
Average	621.86	31.86	48.29	45.50	185.14	99.93	139.43	132.43	77.93	29.64

trade-off between the number of selected features and classification performance. Methods such as KND-UFS and LFSACIE generally select the smallest feature subsets, or among the smallest, on average. However, as detailed in Section 4.4.1, this aggressive reduction strategy often comes at the cost of reduced classification performance. This suggests that purely minimizing feature count may discard features carrying important discriminative information.

Table 8
Number of selected features on DT.

Datasets	Original	KND-UFS	W-MGMN	LFSACIE	GLSFS	FFS-MCC	INF-FS	KNCMI	CMIM	CASR-FS
GLIOMA	4434	209	235	348	3469	2086	2146	2454	2278	303
colon	2000	97	105	137	1487	782	975	1180	881	153
LSVT	313	8	176	22	205	194	120	208	305	185
Toxicity	1203	68	72	85	917	507	637	542	477	302
Darwin	450	12	18	10	434	48	428	28	88	95
Parkinsons	23	6	3	4	15	3	8	6	2	12
Sonar	60	4	34	4	19	31	22	55	15	49
Dermatology	34	6	5	3	30	26	32	14	27	33
BCWD	30	4	5	4	21	14	13	16	19	5
Hill-Valley	100	7	15	10	31	74	24	10	12	15
HCV	12	4	1	4	10	7	7	12	10	11
SGC	20	6	4	2	13	19	8	9	7	18
Image	19	12	2	2	18	17	11	18	14	12
Abalone	8	3	1	2	7	5	5	5	6	6
Average	621.86	31.86	48.29	45.50	476.86	272.36	316.86	325.50	295.79	85.64

Table 9
The comparison of the running time (in seconds) of nine FS methods.

Datasets	KND-UFS	W-MGMN	LFSACIE	GLSFS	FFS-MCC	INF-FS	KNCMI	CMIM	CASR-FS
GLIOMA	6.00	0.01	22.48	17.27	2500.00	3.08	440.36	1155.68	53.46
colon	1.76	0.01	24.48	9.23	1422.47	1.09	257.34	303.50	17.65
LSVT	1.07	0.02	14.62	1.09	64.85	0.27	899.45	10.66	4.53
Toxicity	4.63	0.03	26.48	9.59	639.36	0.72	436.37	256.40	50.29
Darwin	1.79	0.09	44.28	2.55	132.50	0.48	16.18	50.16	12.92
Parkinsons	0.20	0.04	17.88	0.10	0.38	0.01	7.31	0.09	0.12
Sonar	0.35	0.03	29.94	0.40	2.44	0.01	29.04	0.66	0.48
Dermatology	1.17	0.24	12.24	0.01	1.00	0.01	4.65	0.35	1.23
BCWD	0.80	0.22	31.52	0.38	0.87	0.01	19.47	0.41	1.95
Hill-Valley	3.69	0.16	15.52	1.18	6.65	0.04	386.10	4.66	20.18
HCV	1.83	0.19	12.02	0.07	0.22	0.01	10.62	0.07	0.51
SGC	1.20	0.56	13.22	0.04	0.48	0.01	30.33	0.32	3.41
Image	12.24	1.93	16.36	0.36	1.28	0.03	152.59	0.65	18.92
Abalone	151.63	3.11	3.24	0.01	0.30	0.01	8.92	0.17	13.57
Average	13.45	0.47	20.31	3.02	340.91	0.41	192.77	127.41	14.23

In contrast, CASR-FS shows a favorable balance between these two competing objectives. Although it is not designed to produce the smallest possible subset, CASR-FS selects relatively compact subsets under KNN and SVM, with average sizes of 26.71 and 29.64, respectively, compared with 621.86 original features on average. Under DT, the average subset size is larger (85.64), indicating that the selected subset size is classifier-dependent under the adopted evaluation protocol. Combined with the accuracy results reported earlier, this suggests that CASR-FS can retain competitive classification performance while often maintaining relatively compact subsets. This behavior is consistent with its objective function, which emphasizes class-balanced discriminative ability, collaborative feature effects, and redundancy control. Overall, the results suggest that CASR-FS provides a favorable trade-off between subset compactness and classification performance.

4.4.4. Computational efficiency

The running time of all feature selection methods, recorded in seconds, is presented in Table 9. The results indicate that computational efficiency varies significantly depending on the underlying algorithmic approach. Simpler subset-based methods, such as W-MGMN and INF-FS, consistently exhibit the lowest run-times due to their more direct feature evaluation mechanisms.

In contrast, CASR-FS incurs a higher computational cost than simpler methods, mainly because it requires FKG construction and interaction-aware subset evaluation. Nevertheless, on high-dimensional datasets such as GLIOMA, colon, and Toxicity, its runtime remains markedly lower than that of several accuracy-oriented methods. These include FFS-MCC, KNCMI, and CMIM. Although it is still slower than graph-based INF-FS and the lightweight W-MGMN baseline, CASR-FS shows a substantially more moderate runtime profile than several strong but expensive alternatives.

Overall, the runtime of CASR-FS can be regarded as moderate among the compared methods. Although it is not the most efficient algorithm, it remains computationally less demanding than several other competitive methods on challenging high-dimensional datasets. Combined with the classification results in Section 4.4.1, this suggests that CASR-FS provides a reasonable balance between computational cost and classification performance.

Table 10
Average ranking of classification accuracy of the nine algorithms.

Classifier	KND-UFS	W-MGMN	LFSACIE	GLSFS	FFS-MCC	INF-FS	KNCMI	CMIM	CASR-FS
KNN	6.75	7.61	7.57	5.46	3.43	5.82	3.50	3.36	1.50
SVM	7.82	7.21	6.64	5.14	3.86	5.82	3.46	3.32	1.71
DT	8.07	8.11	7.11	4.25	3.68	4.79	4.57	2.96	1.46

Table 11
Results of the Friedman test.

Classifier	Friedman statistic	χ_F^2	p-value
KNN	20.96	69.12	7.34×10^{-12}
SVM	16.47	62.60	1.44×10^{-10}
DT	32.12	79.73	5.54×10^{-14}

Table 12
P-values of the Wilcoxon signed-rank test.

	KND-UFS	W-MGMN	LFSACIE	GLSFS	FFS-MCC	INF-FS	KNCMI	CMIM
KNN	<0.01	<0.01	<0.01	<0.01	0.01	<0.01	<0.01	<0.01
SVM	<0.01	<0.01	<0.01	<0.01	0.02	<0.01	<0.01	<0.01
DT	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.06

4.4.5. Statistical testing

To assess whether the observed performance differences are statistically significant, we performed a two-stage statistical analysis using non-parametric tests. First, a Friedman test was conducted to determine whether statistically significant differences exist among the compared algorithms. The corresponding statistics are defined as

$$F_F = \frac{(T-1)\chi_F^2}{T(s-1)-\chi_F^2}, \tag{16}$$

$$\chi_F^2 = \frac{12T}{s(s+1)}\left(\sum_{i=1}^{s} R_i^2 - \frac{s(s+1)^2}{4}\right), \tag{17}$$

where T and s are the numbers of experimental datasets and algorithms, respectively, and R_i represents the average ranking value of the classification accuracy results of algorithm i under different classifiers.

As shown in Table 10, CASR-FS achieved the best average rank among the three classifiers. The results of the Friedman test, detailed in Table 11, show p-values well below the significance level of $\alpha = 0.05$ for all classifiers. The null hypothesis that all algorithms perform equally is therefore rejected, indicating that further pairwise analysis is warranted.

Following the Friedman test, we employed the Wilcoxon signed-rank test for pairwise comparisons between CASR-FS and each competing method. The resulting p-values are presented in Table 12. For KND-UFS, W-MGMN, LFSACIE, GLSFS, INF-FS, and KNCMI, the p-values are below 0.01 across the reported classifiers, indicating statistically significant differences in favor of CASR-FS.

Compared with FFS-MCC, the difference remains statistically significant on all three classifiers, although the margins are weaker on KNN ($p = 0.01$) and SVM ($p = 0.02$) than for most other baselines. Compared with CMIM, CASR-FS is statistically significant on KNN and SVM, but not on DT ($p = 0.06$). Therefore, the null hypothesis cannot be rejected for the CASR-FS versus CMIM comparison on DT at $\alpha = 0.05$. This indicates that, although CASR-FS often attains better average performance, the observed pairwise differences are not uniformly significant against the strongest baselines on every classifier.

5. Conclusion and future work

This paper presented CASR-FS, a fuzzy-rough feature selection method for class-imbalanced interval-valued data. CASR-FS is built on a unified objective in which class-balanced dependency evaluation, feature interaction modeling, and adaptive conditional redundancy control are integrated within a single selection criterion. Specifically, the method combines a macro-averaged dependency measure, an FJMI-based collaborative gain term, and an adaptive redundancy penalty, so that the selected subset can balance class sensitivity, discriminative ability, complementarity, and redundancy control.

Experiments on fourteen public datasets with multiple classifiers support the usefulness of combining class-sensitive dependency evaluation, collaborative gain, and redundancy control for interval-valued feature selection. The average accuracy, Macro-F1, and subset-size results indicate that CASR-FS can identify discriminative subsets under class-imbalanced interval-valued settings. Nevertheless, CASR-FS is not uniformly optimal on every dataset and classifier. Therefore, the results should be understood as showing an overall favorable feature-selection framework rather than a universally dominant method.

Although CASR-FS performs well, several directions remain for future work. One is to improve computational efficiency, since fuzzy information-theoretic measures can become costly on large-scale datasets. In particular, object-level fuzzy similarity matrices

require substantial memory when the sample size is large, so sparse storage, blockwise computation, and approximate neighbor-based similarity construction deserve further study. Another limitation is that the current experiments use public real-valued datasets converted into intervals by a fixed ±2std rule; future work should include more naturally interval-valued datasets to further examine external validity. Matrix-based acceleration techniques or incremental learning mechanisms may help improve scalability. Another direction is to extend the current framework from static datasets to dynamic or streaming scenarios, where feature relevance and interactions may evolve over time. In addition, more flexible adaptive weighting schemes for the collaborative and redundancy terms may further improve robustness across different data characteristics.

CRediT authorship contribution statement

Weihua Xu: Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Yuhang Lv:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant 62376229 and in part by the Natural Science Foundation of Chongqing, China under Grant CSTB2023NSCQ-LZX0027.

Data availability

Links to all datasets used in this study are provided in the article.

References

[1] Y. Li, T. Li, H. Liu, Recent advances in feature selection and its applications, Knowl. Inf. Syst. 53 (2017) 551–577.
[2] Y.W. Liyanage, D.-S. Zois, C. Chelmis, Dynamic instance-wise joint feature selection and classification, IEEE Trans. Artif. Intell. 2 (2) (2021) 169–184.
[3] L. Sun, L. Wang, W. Ding, Y. Qian, J. Xu, Neighborhood multi-granulation rough sets-based attribute reduction using lebesgue and entropy measures in incomplete neighborhood decision systems, Knowl.-Based Syst. 192 (2020) 105373.
[4] J. Dai, W. Wang, J.-S. Mi, Uncertainty measurement for interval-valued information systems, Inf. Sci. 251 (2013) 63–78.
[5] H. Zapata, H. Bustince, S. Montes, B. Bedregal, G.P. Dimuro, Z. Takáč, M. Baczyński, J. Fernández, Interval-valued implications and interval-valued strong equality index with admissible orders, Int. J. Approx. Reason. 88 (2017) 91–109.
[6] X. Qi, H. Guo, Z. Artem, W. Wang, An interval-valued data classification method based on the unified representation frame, IEEE Access 8 (2020) 17002–17012.
[7] Z. Pawlak, R. sets, Int. J. Comput. Inf. Sci. 11 (5) (1982) 341–356.
[8] O.O. Aremu, R.A. Cody, D. Hyland-Wood, P.R. McAree, A relative entropy based feature selection framework for asset data in predictive maintenance, Comput. Ind. Eng. 145 (2020) 106536.
[9] C. Wang-Wang, L. Dian-Qing, T. Xiao-Song, C. Zi-Jun, Probability distribution of shear strength parameters using maximum entropy principle for slope reliability analysis, Rock Soil Mech. 39 (4) (2018) 1469–1478.
[10] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, Int. J. Gen. Syst. 17 (2–3) (1990) 191–209.
[11] B. Sun, W. Ma, H. Zhao, A rough set approach to handling fuzzy interval values group decision making, Decis. Support Syst. 46 (2) (2008) 507–517.
[12] L.A. Zadeh, Probability measures of fuzzy events, J. Math. Anal. Appl. 23 (2) (1968) 421–427.
[13] B. Sang, L. Yang, H. Chen, W. Xu, X. Zhang, Fuzzy rough feature selection using a robust non-linear vague quantifier for ordinal classification, Expert Syst. Appl. 230 (2023) 120480.
[14] Q. Hu, D. Yu, Z. Xie, J. Liu, Fuzzy probabilistic approximation spaces and their information measures, IEEE Trans. Fuzzy Syst. 14 (2) (2006) 191–201.
[15] C. Wang, Y. Qian, W. Ding, X. Fan, Feature selection with fuzzy-rough minimum classification error criterion, IEEE Trans. Fuzzy Syst. 30 (8) (2022) 2930–2942.
[16] R. Jensen, Q. Shen, Fuzzy–rough attribute reduction with application to web categorization, Fuzzy Sets Syst. 141 (3) (2004) 469–485.
[17] W. Li, H. Zhou, W. Xu, X.-Z. Wang, W. Pedrycz, Interval dominance-based feature selection for interval-valued ordered data, IEEE Trans. Neural Netw. Learn. Syst. 34 (10) (2022) 6898–6912.
[18] D. Chen, L. Zhang, S. Zhao, Q. Hu, P. Zhu, A novel algorithm for finding reducts with fuzzy rough sets, IEEE Trans. Fuzzy Syst. 20 (2) (2011) 385–389.
[19] L. Dong, R. Wang, D. Chen, Incremental feature selection with fuzzy rough sets for dynamic data sets, Fuzzy Sets Syst. 467 (2023) 108503.
[20] X. Zhang, X. Shen, Graph-driven feature selection via granular-rectangular neighborhood rough sets for interval-valued data sets, Appl. Soft Comput. 170 (2025) 112716.
[21] W. Xu, Z. Yuan, Z. Liu, Feature selection for unbalanced distribution hybrid data based on k-nearest neighborhood rough set, IEEE Trans. Artif. Intell. 5 (1) (2024) 229–243.
[22] W. Xu, K. Yuan, W. Li, W. Ding, An emerging fuzzy feature selection method using composite entropy-based uncertainty measure and data distribution, IEEE Trans. Emerg. Top. Comput. Intell. 7 (1) (2022) 76–88.
[23] J. Wan, H. Chen, Z. Yuan, T. Li, X. Yang, B. Sang, A novel hybrid feature selection method considering feature interaction in neighborhood rough set, Knowl.-Based Syst. 227 (2021) 107167.
[24] J. Wan, H. Chen, T. Li, Z. Yuan, J. Liu, W. Huang, Interactive and complementary feature selection via fuzzy multigranularity uncertainty measures, IEEE Trans. Cybern. 53 (2) (2023) 1208–1221.
[25] H. Peng, F. Long, C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, IEEE Trans. Pattern Anal. Mach. Intell. 27 (8) (2005) 1226–1238.
[26] F. Fleuret, Fast binary feature selection with conditional mutual information, J. Mach. Learn. Res. 5 (Nov) (2004) 1531–1555.
[27] G. Roffo, S. Melzi, U. Castellani, A. Vinciarelli, M. Cristani, Infinite feature selection: a graph-based feature filtering approach, IEEE Trans. Pattern Anal. Mach. Intell. 43 (12) (2020) 4396–4410.
[28] Z. Chen, C. Wu, Y. Zhang, Z. Huang, B. Ran, M. Zhong, N. Lyu, Feature selection with redundancy-complementariness dispersion, Knowl.-Based Syst. 89 (2015) 203–217.
[29] Z. Zeng, H. Zhang, R. Zhang, C. Yin, A novel feature selection method considering feature interaction, Pattern Recognit. 48 (8) (2015) 2656–2666.

[30] X. Tang, Y. Dai, Y. Xiang, Feature selection based on feature interactions with application to text categorization, Expert Syst. Appl. 120 (2019) 207–216.
[31] X.-A. Ma, C. Ju, Fuzzy information-theoretic feature selection via relevance, redundancy, and complementarity criteria, Inf. Sci. 611 (2022) 564–590.
[32] J. Wan, X. Li, P. Zhang, H. Chen, X. Ouyang, T. Li, K.C. Tan, FFS-MCC: fusing approximation and fuzzy uncertainty measures for feature selection with multi-correlation collaboration, Inf. Fusion 120 (2025) 103101.
[33] F. Pukelsheim, The three sigma rule, Am. Stat. 48 (2) (1994) 88–91.
[34] W. Xu, Y. Zhang, Y. Qian, A novel unsupervised feature selection for high-dimensional data based on FCM and k-nearest neighbor rough sets, IEEE Trans. Neural Netw. Learn. Syst. 36 (6) (2025) 10889–10898.
[35] W. Xu, Q. Bu, Matrix-based incremental feature selection method using weight-partitioned multigranulation rough set, Inf. Sci. 681 (2024) 121219.