AIMS Mathematics, 11(8): 26186–26223.
DOI:10.3934/math.20261049
Received: 26 February 2026
Revised: 28 July 2026
Accepted: 11 August 2026
Published: 24 August 2026

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

Research article

Composite uncertainty measurements in relative weighted fuzzy rough sets for feature selection

Hongyuan Gou[1,2,*], **Meng Chen**[3] **and Benwei Chen**[4]

[1] College of Computer Science, Chengdu University, Chengdu 610106, China

[2] Key Laboratory of Digital Innovation of Tianfu Culture, Sichuan Provincial Department of Culture and Tourism, Chengdu, China

[3] Department of Computer Sciences and Mathematics, North Sichuan Medical College, Nanchong 637000, China

[4] School of Mathematical Sciences, China West Normal University, Nanchong 637009, China

* **Correspondence:** Email: gouhongyuan@cdu.edu.cn.

Abstract: Feature selection is strongly influenced by the appropriateness of distance modeling and the reliability of uncertainty evaluation. Existing fuzzy rough set-based approaches often suffer from insufficient sensitivity to data dispersion and limited ability to reflect class discrimination. To overcome these issues, this paper proposes a two-dimensional enhanced fuzzy rough set framework that refines distance measurement and uncertainty quantification. Feature variability is quantified via statistical dispersion. A relative distance scheme and a knowledge-granulation weighted fuzzy rough model are developed accordingly. Furthermore, the correlation information between samples and class centers is incorporated to enhance the discrimination capability of similarity evaluation. On this basis, three novel uncertainty measurements, namely relative fuzzy dependency degree (RFD), relative fuzzy composite entropy (RFCE), and relative fuzzy dependency-based composite entropy (RFDCE), are developed, together with their corresponding quantification principles and granulation-monotonicity properties. Corresponding feature selection algorithms are then designed. Experimental results demonstrate that the proposed methods achieve superior feature selection capability and classification accuracy compared with fuzzy rough set models and competing techniques.

Keywords: feature selection; weighted fuzzy rough model; relative distance; distance measurement; uncertainty measurements; granulation-monotonicity
Mathematics Subject Classification: 28D20, 28E10

1. Introduction

In large-scale data scenarios, high-dimensional datasets often contain a large amount of redundant information, which increases data processing costs and extends algorithmic optimization time. Meanwhile, redundant features may reduce the robustness and prediction performance of learning models. Therefore, effectively selecting valuable features and eliminating irrelevant or repetitive information has become an important problem in current data analysis. To address this problem, feature selection has been gradually regarded as an essential research direction and an efficient solution for reducing the negative effects caused by redundant attributes. The main purpose of feature selection is to obtain a compact feature subset with strong discriminative ability from original high-dimensional data, thereby improving the efficiency of subsequent learning tasks. By analyzing the inherent relationships and hidden characteristics within data, feature selection approaches can effectively remove noisy and redundant attributes. This procedure reduces computational complexity and enhances model transparency and practical applicability, which has been demonstrated in many recent studies [1, 2]. However, existing uncertainty measures still face challenges in accurately characterizing complex data uncertainty and maintaining effective feature discrimination, which motivates the development of more reliable uncertainty evaluation approaches.

During the feature selection process, two key uncertainty factors should be considered, including the effectiveness of similarity or distance measurement and the reliability of uncertainty evaluation. These two factors provide important foundations for constructing intelligent optimization strategies, especially in feature subset evaluation and selection processes. From the perspective of modeling approaches, fuzzy rough set models and their extended frameworks have become widely adopted methodologies. Due to the complex distribution characteristics of real-world data and the limitations of traditional modeling methods, improved fuzzy rough set models can achieve a better trade-off between evaluation accuracy and computational efficiency. Consequently, fuzzy rough set-based feature selection approaches with advanced uncertainty measures have received increasing attention and have been applied in many fields, including pattern recognition [3–5], data mining [6–8], intelligent clustering [9, 10] and classification [11, 12], machine learning [13–15], and medical applications [16, 17].

Uncertainty quantification forms the theoretical basis for feature selection within fuzzy rough set systems, and the accuracy of such quantitative indicators directly governs the overall performance of feature screening algorithms. By combining the knowledge granulation mechanism of rough sets with fuzzy membership functions, fuzzy rough set models establish fuzzy relational structures to quantify uncertain information, which can precisely depict ambiguous, overlapping conceptual patterns embedded in raw data. The derived quantitative indicators offer a data-distribution consistent standard to judge feature significance, making the retained attributes more distinguishable and logically interpretable. Moreover, fuzzy rough set frameworks support parameter adjustment and weight optimization to tailor membership mappings and uncertainty indicators, which strengthens their capacity to perceive noisy samples and ambiguous concept boundaries. Such optimization further elevates the generalization capacity and decision stability of feature selection methods applied to complicated decision tasks [18–20]. Existing studies on feature selection powered by advanced fuzzy rough uncertainty measurements have verified their outstanding robustness and practical value [21–23]. Accurate uncertainty evaluation ensures stable decision outputs for high-uncertainty tasks such as

medical diagnosis [24,25] and pattern recognition [26]. Meanwhile, it drives the evolution of rough set theory toward fuzzy granulation, turning uncertainty measurement into a hot research direction within this discipline [27]. Owing to the fundamental supporting value of uncertainty quantification, metric-guided feature selection has become a mainstream research hotspot in fuzzy rough set research [28–30]. As representative examples, Xu et al. [30] designed an uncertainty metric combining relative dependency and complementary mutual information, which guides feature screening by analyzing sample similarity distribution and decision attribute correlations. Long et al. [31] put forward a three-dimensional fuzzy information indicator to characterize the dynamic changes of feature relevance and redundancy, where feature interactive relationships are fully considered. On this basis, they developed a heuristic search algorithm for feature selection guided by the proposed measure. Yang et al. [32] introduced class separability and constructed a rational linear kernel-weighted model to strengthen the discriminative representativeness of selected subsets. Wan et al. [33] established a joint uncertainty model paired with a two-tier feature evaluation strategy, enhancing the robustness and correlation characterization of feature selection from the dual perspectives of sample density and feature statistics. Xie et al. [34] constructed weighted fuzzy partitions with adaptive neighborhood radii and developed uncertainty indicators from three mutually complementary dimensions to implement efficient feature selection.

Uncertainty measurement subsequently serves as an essential criterion for feature selection in fuzzy rough sets and their extensions, including weighted fuzzy rough sets (WFRSs) [35], probabilistic fuzzy rough sets [36], and fuzzy rough soft sets [37]. It enables feature selection approaches to handle uncertain and fuzzy data conditions effectively. Meanwhile, it provides an important mechanism for balancing feature reduction and classification robustness. In particular, for complex tasks such as fault diagnosis and pattern recognition [24], uncertainty measurement determines whether learning models can discover meaningful discriminative information from highly redundant feature spaces. Therefore, uncertainty measurement acts as an essential connection between theoretical models and practical applications. However, existing extended fuzzy rough set models generally suffer from insufficient integration between distance evaluation and uncertainty quantification. In these models, distance calculation strategies usually ignore the discrete characteristics and distribution diversity of real-world data. As a result, the obtained measurements may not accurately reflect the influence of features on classification uncertainty. This limitation can cause important fuzzy features to be incorrectly removed or redundant attributes to be unnecessarily retained. Consequently, the feature selection models may show limited adaptability when processing complex high-dimensional datasets. To overcome these limitations, this study establishes a two-dimensional collaborative enhancement framework. The framework combines "data distribution-aware distance optimization" with "hierarchical structure matching of uncertainty measurement". This framework provides a promising solution for addressing the above challenges.

To resolve the aforementioned dilemma, this study adopts fuzzy rough set theory as the fundamental research framework and concentrates on refining the measurement mechanism of the weighted fuzzy rough set model. In particular, targeted improvement strategies are developed within a two-dimensional collaborative framework that integrates distance measurement optimization and deepening of uncertainty measurement. The overall research idea and technical framework are illustrated in Figure 1. For clarity and ease of reference, the principal abbreviations and symbols used throughout this paper are systematically summarized in Table 1.

(1) At the distance measurement level, the standard deviation of feature data is introduced as a key quantitative indicator to characterize the heterogeneity of discrete data distributions. By revising the conventional sample distance computation scheme, a relative distance measure is formulated, which allows for a more accurate representation of data distribution characteristics and fuzzy membership relationships of boundary samples. As a result, distance information can more effectively support the construction of fuzzy approximation relations and subsequent uncertainty quantification.

(2) At the uncertainty measurement level, relative weighted fuzzy rough upper and lower approximations, together with the relative fuzzy dependency degree (RFD), are defined on the basis of relative distance. Furthermore, we take two factors into account. One factor is the weight differences among multiple decision classes. The other is the boundary fuzziness captured by dual rough approximations. With these two combined factors, we construct a relative fuzzy composite entropy called RFCE. This measure evaluates uncertainty from three levels: the global level (class weights), the local level (per-class membership aggregation), and the microscopic boundary level (the ratio of lower to upper approximations). These three angles together give a more complete picture of the system's uncertainty. In addition, we propose another relative fuzzy dependency-based composite entropy measure named RFDCE. This measure builds upon the fuzzy dependency degree. It captures uncertainty in a more sensitive and thorough manner. As a result, it provides a solid theoretical foundation for later tasks such as feature selection and model evaluation.

(3) A comprehensive theoretical investigation is performed to demonstrate the fundamental properties of the proposed uncertainty measures, particularly the granulation-monotonicity property. This analysis guarantees the logical coherence of the measurement framework and supports reliable hierarchical information propagation. Based on these newly constructed uncertainty measures, this study further develops three feature selection algorithms, namely RFD-FS, RFCE-FS, and RFDCE-FS. These algorithms establish an effective connection between uncertainty evaluation improvement and feature selection implementation.

(4) The proposed feature selection algorithms RFD-FS, RFCE-FS, and RFDCE-FS show superior and stable performance on 16 UCI datasets. When combined with k-nearest neighbor (KNN) and support vector machine (SVM) classifiers, RFCE-FS and RFDCE-FS achieve higher average classification accuracy and more optimal results than existing methods. Statistical tests further confirm that these improvements are significant, demonstrating the robustness and effectiveness of the proposed algorithms.

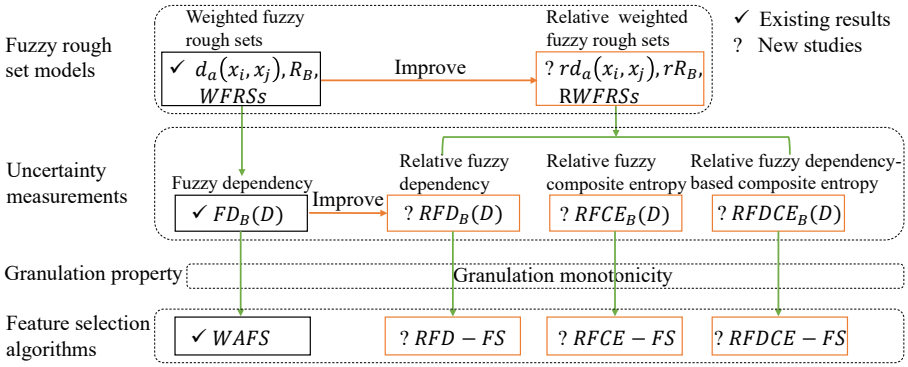


Figure 1. Research framework of uncertainty measurement and feature selection based on hierarchical fusion in fuzzy rough sets.

Table 1. Main term abbreviations in this paper.

Item	Abbreviation	Full name	Notation
Contexts	WFRSs	Weighted fuzzy rough sets	–
	RWFRSs	Relative weighted fuzzy rough sets	–
Uncertainty measures	R_B	Fuzzy similarity relation	R_B
	rR_B	Relative fuzzy similarity relation	rR_B
	FD	Fuzzy dependency degree	$FD_B(D)$
	RFD	Relative fuzzy dependency degree	$RFD_B(D)$
	RFCE	Relative fuzzy composite entropy	$RFCE_B(D)$
	RFDCE	Relative fuzzy dependency-based composite entropy	$RFDCE_B(D)$
Feature selection algorithms	WAFS	Feature selection algorithm based on WFRSs	–
	RFD-FS	Feature selection algorithm based on RFD	–
	RFCE-FS	Feature selection algorithm based on RFCE	–
	RFDCE-FS	Feature selection algorithm based on RFDCE	–

This study focuses on addressing two major limitations of classical fuzzy rough sets and their weighted extensions. The first limitation is the insufficient integration between weight determination and uncertainty evaluation. The second limitation is the inconsistency between distance representation and real data distribution characteristics. To overcome these problems, this study proposes a dual-dimensional collaborative optimization strategy. The proposed strategy improves the capability of removing redundant features and enhances classification reliability in feature selection tasks with complex fuzzy data.

The remainder of this paper is organized as follows. Section 2 reviews the basic concepts and related uncertainty measures of the WFRSs and their related properties. Section 3 investigates relative distance and the corresponding relative weighted fuzzy rough sets (RWFRSs) and related uncertainty measures, with emphasis on their construction definitions, computational algorithms, size relationships, granulation monotonicity, and illustrative examples. Section 4 develops heuristic feature selection algorithms based on the proposed uncertainty measures and feature significance criteria. Section 5 conducts extensive experiments to validate measurement computation, monotonicity properties, and feature selection performance. Finally, Section 6 concludes the paper.

2. Basic notions

Based on Refs. [35, 38], this section introduces weighted fuzzy approximation operators and their corresponding uncertainty measures, together with related theoretical properties and illustrative examples.

In fuzzy rough set theory, a decision system is denoted as (U, A, D), where the universe U is a finite set of samples $\{x_1, x_2, \ldots, x_n\}$. The conditional attribute set $A = \{a_1, a_2, \ldots, a_m\}$ consists of real-valued attributes describing the samples in U, while D represents the decision attribute defined on the same universe. Let $D_1, D_2, \cdots, D_r$ denote the r indiscernibility blocks (i.e., crisp equivalence classes) induced by the decision attribute D such that $U/D = \{D_1, D_2, \cdots, D_r\}$.

Let $U \times U$ denote the Cartesian product of U, and let R be a fuzzy set defined on $U \times U$, i.e., $R(\cdot) : U \times U \to [0, 1]$, where $(x_i, x_j) \to R(x_i, x_j)$ for any $(x_i, x_j) \in U \times U$. In this case, R is referred to as a fuzzy binary relation on U. Given a nonempty attribute subset $B \subseteq A$, a fuzzy similarity relation R_B on U can be induced, which satisfies reflexivity and symmetry. The fuzzy similarity relation generated by attribute $a \in B$ is defined as

$$R_a(x_i, x_j) = \exp(-p \times d_a(x_i, x_j)) \, (d_a(x_i, x_j) = |a(x_i) - a(x_j)|), \tag{2.1}$$

where $a(x)$ denotes the value of sample x on attribute a and p is a corresponding parameter. Let R_B represent the fuzzy similarity relation induced by attribute subset B, as defined in

$$R_B = \bigcap_{a \in B} R_a. \tag{2.2}$$

Accordingly, the fuzzy similarity matrix $R_B = (r_{ij})_{n \times n}$ can be obtained, where $r_{ij} = R_B(x_i, x_j)$. Obviously, $r_{ij} = 1$ when $x_i = x_j$, and $r_{ij} = r_{ji}$ holds, implying that R_B is both reflexive and symmetric.

For any $x_i \in U$, the fuzzy similarity class associated with x_i under relation R_B is defined as $\widetilde{[x_i]}_{R_B}(x_j) = R_B(x_i, x_j)$, $x_j \in U$. Clearly, $\widetilde{[x_i]}_{R_B}(x_j)$ constitutes a fuzzy set on U. When the fuzzy similarity relation R_B degenerates into a crisp relation, the corresponding fuzzy neighborhood $\widetilde{[x_i]}_{R_B}(x_j)$ also reduces to a crisp one. Then, for a specific decision class D_t ($\forall t \in \{1, \cdots, r\}$), the fuzzy decision class $\widetilde{D}_t(x_i)$ of sample x_i is defined as

$$\widetilde{D}_t(x_i) = \frac{|\widetilde{[x_i]}_{R_A} \cap D_t|}{|\widetilde{[x_i]}_{R_A}|}, \tag{2.3}$$

where $D_t(x_j) = 1$ if $x_j \in D_t$ and $D_t(x_j) = 0$ otherwise. Moreover, $|[x_i]_{R_A} \cap D_t| = \Sigma_{x_j \in U} \left(\widetilde{[x_i]}_{R_A}(x_j) \cap D_t(x_j) \right)$, which characterizes the dependency between the original attribute set A and the decision attribute D. Here, $|\cdot|$ denotes the cardinality of a set.

Traditional fuzzy rough set models assume that all conditional attributes are equally important for classification. In fact, real data attributes are heterogeneous, with different information capacities and influences on decision results. Therefore, a distance-weighting method is adopted to quantify the importance of attributes, so as to better fit the data distribution in similarity evaluation and obtain better classification results.

Definition 1. *[35] Consider a decision table (U, A, D), where $U/D = \{D_1, D_2, \ldots, D_r\}$. For any object $x_i \in U$, the distance between x_i and $D_t \in U/D$ is defined as follows:*

$$\begin{aligned} S(x_i, D_t) &= \min\{\Delta_A(x_i, x_j) | x_i \neq x_j, x_j \in D_t\}, \\ \Delta_B(x_i, x_j) &= 1 - R_B(x_i, x_j). \end{aligned} \tag{2.4}$$

Here, $\Delta_B(x_i, x_j)$ denotes the distance function between objects x_i and x_j. Clearly, $S(x_i, D_t)$ represents the minimum relative distance from x_i to all samples belonging to D_t. Let

$$\alpha(x_i, D_t) = \exp(-S(x_i, D_t)). \tag{2.5}$$

This function transforms the distance value $S(x_i, D_t)$ into the interval $(0, 1]$, and the resulting value $\alpha(x_i, D_t)$ is termed the importance degree of x_i with respect to D_t. As $S(x_i, D_t)$ decreases, the corresponding importance degree $\alpha(x_i, D_t)$ increases; in particular, when $S(x_i, D_t) = 0$, $\alpha(x_i, D_t)$ attains its maximum value, i.e., $\alpha(x_i, D_t) = 1$.

Definition 2. *[35][WFRSs] Weighted fuzzy rough lower and upper approximations of decision class $D_t \in U/D$ with respect to an attribute subset $B \subseteq C$ are given by*

$$\begin{aligned} \underline{N}_B(D_t)(x_i) &= \min_{x_j \in U} \max\{\alpha(x_i, D_t)(1 - R_B(x_i, x_j)), \widetilde{D}_t(x_j)\}, \\ \overline{N}_B(D_t)(x_i) &= \max_{x_j \in U} \min\{1 - \alpha(x_i, D_t)(1 - R_B(x_i, x_j)), \widetilde{D}_t(x_j)\}. \end{aligned} \tag{2.6}$$

The fuzzy positive region and dependency degree of D with respect to B in the relative fuzzy rough set are defined as

$$\begin{aligned} POS_B(D)(x_i) &= \max\{\underline{N}_B(D_t)(x_i), t = 1, 2, \ldots, r\}, \\ FD_B(D) &= \frac{1}{|U|} \sum_{x_i \in U} POS_B(D)(x_i). \end{aligned} \tag{2.7}$$

Theorem 1. *[35] If $B_1 \subseteq B_2 \subseteq A$, then $POS_{B_1}(D) \subseteq POS_{B_2}(D)$, $FD_{B_1}(D) \leq FD_{B_2}(D)$.*

$\underline{N}_B(D_t)(x_i)$ and $\overline{N}_B(D_t)(x_i)$ separately describe the definite and potential membership levels of object x_i to fuzzy decision class D_t with weighted attribute constraints. $POS_B(D)(x_i)$ picks the maximum lower approximation across all decision partitions to quantify the separability of each sample, and $FD_B(D)$ averages positive region values over the whole universe U to quantitatively assess the discriminative capacity of attribute subset B.

3. Uncertainty measurement of relative weighted fuzzy rough sets

Traditional fuzzy rough set models treat all conditional attributes as equally important for classification tasks. In practical datasets, nevertheless, each attribute carries distinct information volume and delivers unequal contributions to decision outcomes. Typical cases include image recognition and medical diagnosis: Color features and texture descriptors exert vastly different impacts on image category discrimination, while diverse clinical indicators hold disparate contribution weights for disease evaluation. To tackle this limitation, this work develops a modeling scheme based on relative distance. It calculates attribute relative importance via object distance gaps on single attributes and further defines matched fuzzy approximation operators. The proposed scheme can faithfully reflect the intrinsic data distribution patterns and build a unified evaluation standard for follow-up uncertainty quantification and feature selection tasks. This section elaborates on the established RWFRSs model in Section 3.1 and explores the associated uncertainty quantification approach in Section 3.2.

3.1. Relative weighted fuzzy rough sets

When examining the degree of association between data samples (e.g., through a distance metric), the original metric $d_a(x_i, x_j)$ is able to describe the intrinsic relationship between samples x_i and x_j. Nevertheless, this measure is highly sensitive to the internal fluctuation of the dataset A itself. Specifically, when the dispersion level of the data in A, as measured by the standard deviation $\mathrm{std}_A(x_i, x_j)$, is large, the scale of the original distance may become unstable and even difficult to compare directly across different datasets. To alleviate the influence of such data variability and to improve the universality and comparability of the measure, it is necessary to perform a rescaling (or normalization) operation on the original distance. Accordingly, a readjusted measure $rd_a(x_i, x_j)$ is constructed by incorporating the standard deviation of the attribute set A.

Definition 3. *Let a be an attribute and A be an attribute set with $a \in A$. The relative distance between samples x_i and x_j is defined as*

$$rd_a(x_i, x_j) = \frac{d_a(x_i, x_j)}{1 + \mathrm{std}_A(x_i, x_j)}, \tag{3.1}$$

where $d_a(x_i, x_j)$ denotes the Euclidean distance between x_i and x_j (i.e., $d_a(x_i, x_j) = |a(x_i) - a(x_j)|$) and $\mathrm{std}_A(x_i, x_j)$ represents the standard deviation of distances between samples x_i and x_j in the attribute set A.

The term $\mathrm{std}_A(x_i, x_j)$ quantifies the dispersion degree or fluctuation magnitude of data points x_i and x_j over the data in A. According to Eq (3.1), a relatively large $\mathrm{std}_A(x_i, x_j)$ will raise the denominator $1+\mathrm{std}_A(x_i, x_j)$ correspondingly, which adaptively compresses the value of the original distance $d_a(x_i, x_j)$ to mitigate the bias from intense data fluctuations. Conversely, when $\mathrm{std}_A(x_i, x_j)$ is small with mild data volatility, the denominator approaches 1, and the adjusted robust distance $rd_a(x_i, x_j)$ approximates the original distance $d_a(x_i, x_j)$, preserving the intrinsic pairwise discrepancy. By incorporating the fluctuation metric $\mathrm{std}_A(x_i, x_j)$ into the denominator, the final distance rd_a breaks the sole dependence on raw d_a values and actively suppresses the interference induced by data oscillations within A. This modification yields a more stable, cross-scenario comparable distance metric. In particular, this normalization mechanism unifies the scale of distance results derived from heterogeneous datasets A with distinct fluctuation levels, enabling direct horizontal comparison. Fundamentally, the substitution of d_a with rd_a realizes data-adaptive rescaling of the original distance via the standard deviation of local data dispersion on A, which calibrates distance magnitude and weakens the confounding effect of data variability.

In this work, the pairwise distance values between samples are calibrated using the global standard deviation calculated from all conditional attributes, rather than allocating individual scaling factors for each single independent attribute. Such an implementation exactly realizes the adaptive normalization mechanism defined by the rd_a formula. The proposed calibration scheme establishes a uniform measurement criterion for similarity computation among cross-type heterogeneous feature dimensions, which supports the stable formation of integrated fuzzy neighborhood granules in granular computing. Furthermore, adopting the global standard deviation metric can mitigate the measurement scaling bias that arises from outlier data points on discrete and sparse single attributes. Simultaneously, this metric preserves the native distance difference characteristics among different attributes, thereby sufficiently ensuring the credibility of downstream uncertainty quantification and model evaluation results.

Based on the above relative distance, the relative fuzzy similarity relations $rR_a(x_i, x_j) = \exp(-p \times rd_a(x_i, x_j))$ and $rR_B = \bigcap_{a\in B} rR_a$ are constructed. Suppose that rR_A denotes the relative fuzzy similarity relation defined on U. For a given decision class D_t ($\forall t \in \{1, \cdots, r\}$), the corresponding relative fuzzy decision class $r\widetilde{D}_t(x_i)$ of an object x_i is defined as

$$r\widetilde{D}_t(x_i) = \frac{|\widetilde{[x_i]}_{rR_A} \cap D_t|}{|\widetilde{[x_i]}_{rR_A}|}. \tag{3.2}$$

Definition 4. *Given a decision information table (U, A, D), where $U/D = \{D_1, D_2, \ldots, D_r\}$, let rR_B be a fuzzy similarity relation induced by attribute subset $B \subseteq A$. For any $x_i \in U$, the distance between x_i and $D_t \in U/D$ is defined as*

$$rS(x_i, D_t) = \min\{1 - rR_B(x_i, x_j) | x_j \in D_t, x_i \neq x_j\}. \tag{3.3}$$

Clearly, $rS(x_i, D_t)$ represents the minimum relative distance between x_i and all samples belonging to D_t. Let

$$r\alpha(x_i, D_t) = \exp(-rS(x_i, D_t)). \tag{3.4}$$

This mapping function transforms the distance value $rS(x_i, D_t)$ into the interval $(0, 1]$, and the resulting quantity $r\alpha(x_i, D_t)$ is referred to as the importance degree of x_i with respect to D_t. The smaller the value of $rS(x_i, D_t)$, the larger the corresponding importance degree $r\alpha(x_i, D_t)$ of x_i relative to D_t; in the extreme case where $rS(x_i, D_t) = 0$, the importance degree reaches its maximum, namely $r\alpha(x_i, D_t) = 1$.

Definition 5 (RWFRSs)**.** *The relative fuzzy rough lower and upper approximations of decision class $D_t \in U/D$ with respect to an attribute subset $B \subseteq A$ are defined as*

$$\begin{aligned} \underline{RN}_B(D_t)(x_i) &= \min_{x_j\in U} \max\{r\alpha(x_i, D_t)(1 - rR_B(x_i, x_j)), r\widetilde{D}_t(x_j)\}, \\ \overline{RN}_B(D_t)(x_i) &= \max_{x_j\in U} \min\{1 - r\alpha(x_i, D_t)(1 - rR_B(x_i, x_j)), r\widetilde{D}_t(x_j)\}. \end{aligned} \tag{3.5}$$

The fuzzy lower approximation corresponding to D_t characterizes the certain degree of sample affiliation to D_t, where a higher value of lower approximation corresponds to a weaker uncertainty level. In general, normal samples tend to yield a relatively high lower approximation value for their corresponding decision classes.

Definition 6 (RFPR and RFD)**.** *The relative fuzzy positive region (RFPR) and the corresponding dependency degree of D with respect to B in the relative fuzzy rough set framework are defined as*

$$\begin{aligned} RPOS_B(D)(x_i) &= \max\{\underline{RN}_B(D_t)(x_i), t = 1, 2, \ldots, r\}, \\ RFD_B(D) &= \frac{1}{|U|} \sum_{x_i\in U} RPOS_B(D)(x_i). \end{aligned} \tag{3.6}$$

A greater value of $RPOS_B(D)(x_i)$ means that the feature subset B has a stronger discrimination ability for distinguishing the sample x_i. On the basis of the definition of $RPOS_B(D)$, the function $RFD_B(D)$ is proposed to quantify the approximation performance of the feature subset B for the decision set D, thereby serving as a quantitative index for evaluating feature subset quality.

From the perspective of classification performance, classical fuzzy rough set models produce coarse boundary partitions owing to the neglect of attribute distance in decision-making. The introduced distance weight strategy adaptively calibrates inter-object distances by attribute importance, which clusters samples with vital high-weight attributes tightly and isolates dissimilar samples by weakening low-weight feature interference. This scheme optimizes boundary division to improve classification accuracy and suppresses noise and redundant attribute disturbance to enhance model robustness and generalization ability. These core merits motivate the construction of the proposed RWFRSs model.

Compared with the traditional formula in [35] that builds fuzzy similarity relations directly on raw distances, our data-adaptive relative distance-driven fuzzy approximation space mitigates the negative effect of discrete data fluctuations on fuzzy dual approximation calculation. Accordingly, the weighted fuzzy rough dependency degree and corresponding uncertainty measurement possess excellent anti-interference performance and stable cross-subset measurement ability for heterogeneous data.

Theorem 2. *The relative weighted fuzzy rough approximation operators possess several important theoretical properties.*

(1) For $\forall x_i \in U$, $\underline{RN}_B(D_t)(x_i) \in [0,1]$, $\overline{RN}_B(D_t)(x_i) \in [0,1]$.
(2) $\underline{RN}_B(\sim D_t) = \sim \overline{RN}_B(D_t)$, *and* $\overline{RN}_B(\sim D_t) = \sim \underline{RN}_B(D_t)$.
(3) $\underline{RN}_B(D_t) \subseteq r\widetilde{D}_t \subseteq \overline{RN}_B(D_t)$.
(4) If $B_1 \subseteq B_2 \subseteq A$, *then* $\underline{RN}_{B_1}(D_t) \subseteq \underline{RN}_{B_2}(D_t)$, $\overline{RN}_{B_2}(D_t) \subseteq \overline{RN}_{B_1}(D_t)$.

Proof. (1) Since $r\alpha(x_i, D_t) \in (0,1]$ and $1 - rR_B(x_i, x_j) \in [0,1)$, their product naturally satisfies $r\alpha(x_i, D_t)(1 - rR_B(x_i, x_j)) \in [0,1)$. At the same time, $r\widetilde{D}_t(x_j) \in [0,1]$. Based on Eq (3.5), we know that $\underline{RN}_B(D_t)(x_i) \in [0,1]$. By the same token, it can be shown that the output of the upper approximation operator also satisfies $\overline{RN}_B(D_t)(x_i) \in [0,1]$.
(2)

$$\begin{aligned}
\underline{RN}_B(\sim D_t)(x_i) &= \min_{x_j \in U} \max \left\{ r\alpha(x_i, D_t)(1 - rR_B(x_i, x_j)), 1 - r\widetilde{D}_t(x_j) \right\} \\
&= \min_{x_j \in U} \left\{ 1 - \min \left\{ 1 - r\alpha(x_i, D_t)(1 - rR_B(x_i, x_j)), r\widetilde{D}_t(x_j) \right\} \right\} \\
&= 1 - \max_{x_j \in U} \min \left\{ 1 - r\alpha(x_i, D_t)(1 - rR_B(x_i, x_j)), r\widetilde{D}_t(x_j) \right\} \\
&= 1 - \overline{RN}_B(D_t)(x_i) \\
&= \sim \overline{RN}_B(D_t)(x_i).
\end{aligned}$$

Similarly, $\overline{RN}_B(\sim D_t)(x_i) = \sim \underline{RN}_B(D_t)(x_i)$.
(3) The reflexivity of the relation rR_B guarantees that for $\forall x_i \in U$, we have $rR_B(x_i, x_i) = 1$. With this in mind, we can perform the following bounding argument on the lower approximation:

$$\begin{aligned}
\underline{RN}_B(D_t)(x_i) &= \min_{x_j \in U} \max \left\{ r\alpha(x_i, D_t)(1 - rR_B(x_i, x_j)), r\widetilde{D}_t(x_j) \right\} \\
&\leq \max \left\{ r\alpha(x_i, D_t)\,(1 - rR_B(x_i, x_i))\,, r\widetilde{D}_t(x_i) \right\} \\
&= r\widetilde{D}_t(x_i).
\end{aligned}$$

That is, $\underline{RN}_B(D_t) \subseteq r\widetilde{D}_t$. Similarly, we can bound the upper approximation from below:

$$\begin{aligned}\overline{RN}_B(D_t)(x_i) &= \max_{x_j \in U} \min\left\{1 - r\alpha(x_i, D_t)(1 - rR_B(x_i, x_j)), r\widetilde{D}_t(x_j)\right\} \\ &\geq \min\left\{1 - r\alpha(x_i, D_t)\left(1 - rR_B(x_i, x_i)\right), r\widetilde{D}_t(x_i)\right\} \\ &= r\widetilde{D}_t(x_i).\end{aligned}$$

This shows that the $r\widetilde{D}_t$ is a subset of the upper approximation, i.e., $\overline{RN}_B(D_t) \supseteq r\widetilde{D}_t$. Hence, $\underline{RN}_B(D_t) \subseteq r\widetilde{D}_t \subseteq \overline{RN}_B(D_t)$.

(4) From $B_1 \subseteq B_2 \subseteq A$, we have for $\forall x_i, x_j \in U$, $rR_{B_2}(x_i, x_j) \leq rR_{B_1}(x_i, x_j)$. Then, $1 - rR_{B_1}(x_i, x_j) \leq 1 - rR_{B_2}(x_i, x_j)$. Multiplying by $r\alpha(x_i, D_t) > 0$ preserves the inequality:

$$r\alpha(x_i, D_t)(1 - rR_{B_1}(x_i, x_j)) \leq r\alpha(x_i, D_t)(1 - rR_{B_2}(x_i, x_j)).$$

We obtain

$$\max\{r\alpha(x_i, D_t)(1 - rR_{B_1}(x_i, x_j)),\ r\widetilde{D}_t(x_j)\} \leq \max\{r\alpha(x_i, D_t)(1 - rR_{B_2}(x_i, x_j)),\ r\widetilde{D}_t(x_j)\},$$

thus it follows that

$$\min_{x_j \in U} \max\{r\alpha(x_i, D_t)(1 - rR_{B_1}(x_i, x_j)),\ r\widetilde{D}_t(x_j)\} \leq \min_{x_j \in U} \max\{r\alpha(x_i, D_t)(1 - rR_{B_2}(x_i, x_j)),\ r\widetilde{D}_t(x_j)\}.$$

According to Eq (3.5), this is exactly $\underline{RN}_{B_1}(D_t)(x_i) \leq \underline{RN}_{B_2}(D_t)(x_i)$ $(\forall x_i \in U)$, that is, $\underline{RN}_{B_1}(D_t) \subseteq \underline{RN}_{B_2}(D_t)$. By following a completely parallel line of reasoning, the monotonicity of the upper approximation operator, $\overline{RN}_{B_1}(D_t) \subseteq \overline{RN}_{B_2}(D_t)$, also holds true.

In particular, for the boundary cases:

- If $r\widetilde{D}_t(x_j) = 0$ for all x_j (i.e., $D_t = \emptyset$), then

$$\underline{RN}_B(\emptyset)(x_i) = \min_{x_j \in U} r\alpha(x_i, D_t)(1 - rR_B(x_i, x_j)),$$

$$\overline{RN}_B(\emptyset)(x_i) = \max_{x_j \in U} \min\{1 - r\alpha(x_i, D_t)(1 - rR_B(x_i, x_j)),\ 0\} = 0.$$

  The lower approximation remains monotone with respect to B, and the upper approximation is identically 0, so $\overline{RN}_{B_2}(\emptyset) = \overline{RN}_{B_1}(\emptyset) = \emptyset$, trivially satisfying the inclusion.
- If $r\widetilde{D}_t(x_j) = 1$ for all x_j (i.e., $D_t = U$), then

$$\underline{RN}_B(U)(x_i) = \min_{x_j \in U} \max\{r\alpha(x_i, U)(1 - rR_B(x_i, x_j),\ 1\} = 1,$$

$$\overline{RN}_B(U)(x_i) = \max_{x_j \in U} \min\{1 - r\alpha(x_i, U)(1 - rR_B(x_i, x_j)),\ 1\} = \max_{x_j \in U}\{1 - r\alpha(x_i, U)(1 - rR_B(x_i, x_j))\}.$$

  The lower approximation is U, so the inclusion holds as equality. For the upper approximation, since $rR_{B_2} \leq rR_{B_1}$, we have

$$1 - r\alpha(x_i, U)(1 - rR_{B_2}(x_i, x_j)) \leq 1 - r\alpha(x_i, U)(1 - rR_{B_1}(x_i, x_j)),$$

hence

$$\overline{RN}_{B_2}(U)(x_i) = \max_{x_j \in U}\{1-r\alpha(x_i,U)(1-rR_{B_2}(x_i,x_j))\} \leq \max_{x_j \in U}\{1-r\alpha(x_i,U)(1-rR_{B_1}(x_i,x_j))\} = \overline{RN}_{B_1}(U)(x_i).$$

Thus, $\overline{RN}_{B_2}(U) \subseteq \overline{RN}_{B_1}(U)$.

Therefore, Theorem 2 (4) holds for all possible decision classes, with no exceptional cases. □

Theorem 3 (RFPR and RFD granulation monotonicity)**.** *If $B_1 \subseteq B_2 \subseteq A$, then $RPOS_{B_1}(D) \subseteq RPOS_{B_2}(D)$ and $RFD_{B_1}(D) \leq RFD_{B_2}(D)$.*

Proof. Since $B_1 \subseteq B_2 \subseteq A$, according to Theorem 2, we obtain that $\underline{RN}_{B_1}(D_t)(x_i) \leq \underline{RN}_{B_2}(D_t)(x_i)$, then $\max\{\underline{RN}_{B_1}(D_t)(x_i)\} \leq \max\{\underline{RN}_{B_2}(D_t)(x_i)\}$. So $RPOS_{B_1}(D)(x_i) \leq RPOS_{B_2}(D)(x_i)$ from Definition 6. Hence, $RPOS_{B_1}(D) \subseteq RPOS_{B_2}(D)$ and $RFD_{B_1}(D) \leq RFD_{B_2}(D)$ hold. □

The above characteristics guarantee the rationality and consistency of the designed approximation operators, while maintaining the monotonic granularity property: Expanding the attribute set enlarges the lower approximation and shrinks the upper approximation, which compresses the uncertain boundary region. Such monotonic variation makes the uncertainty measurements derived from the proposed approximation schemes well suited for attribute evaluation, since supplementing relevant attributes can steadily refine the granular knowledge representation. In Section 4, we will fully exploit this feature to establish a monotonic evaluation standard for feature selection tasks.

3.2. Uncertainty measurements

Dual approximations or approximation accuracy can quantify uncertainty in decision datasets, but these measurements ignore data distribution and fail on imbalanced data. To solve this, Xu et al. [39] proposed composite entropy combining rough approximations and Shannon entropy. This paper integrates dual approximations and Shannon entropy to define fuzzy composite entropy for fuzzy decision systems.

Definition 7 (RFCE)**.** *Given* DS $= (U, A, D)$ *and a conditional attribute subset $B \subseteq A$, the RFCE can be defined as*

$$RFCE_B(D) = -\sum_{t=1}^{r} \frac{|D_t|}{|U|} \log \frac{\sum_{x_i \in U} \underline{RN}_B(D_t)(x_i)}{\sum_{x_i \in U} \overline{RN}_B D_t(x_i)}. \tag{3.7}$$

By jointly considering the cardinalities of decision classes ($|D_t|, t = 1, 2, \ldots, r$) together with the corresponding dual approximations, the uncertainty inherent in the decision system can be described in a more refined manner. In this formulation, the ratio $\frac{\sum_{x_i \in U} \underline{RN}_B(D_t)(x_i)}{\sum_{x_i \in U} \overline{RN}_B D_t(x_i)}$ between the lower and upper approximations is employed to characterize the so-called fuzzy boundary uncertainty of a decision class D_t. The further this ratio deviates from 1, the larger the discrepancy between the states of possibly belonging and definitely belonging, which implies a more indistinct class boundary and hence a higher level of uncertainty.

To quantify this uncertainty, a logarithmic function $\log(\cdot)$ combined with a negative sign is introduced to transform the above ratio into an entropy-based measure, where a larger entropy value corresponds to greater uncertainty. The use of the logarithmic mapping amplifies variations in the ratio

of the upper and lower approximation bases, ensuring that even subtle changes in boundary fuzziness can be effectively captured, thereby enhancing the sensitivity of the uncertainty measurement. The proposed measurements satisfy several important properties.

Theorem 4 (RFCE granulation monotonicity)**.** *If $B_1 \subseteq B_2 \subseteq A$, then $RFCE_{B_2}(D) \leq RFCE_{B_1}(D)$.*

Proof. Since $B_1 \subseteq B_2 \subseteq A$ holds, according to Theorem 2, we have $\underline{RN}_{B_1}(D_t) \leq \underline{RN}_{B_2}(D_t)$, $\overline{RN}_{B_1}(D_t) \geq \overline{RN}_{B_2}(D_t)$, which indicates that $\frac{\sum_{x_i\in U}\underline{RN}_{B_2}(D_t)(x_i)}{\sum_{x_i\in U}\overline{RN}_{B_2}D_t(x_i)} \geq \frac{\sum_{x_i\in U}\underline{RN}_{B_1}(D_t)(x_i)}{\sum_{x_i\in U}\overline{RN}_{B_1}D_t(x_i)}$ and $\log\frac{\sum_{x_i\in U}\underline{RN}_{B_2}(D_t)(x_i)}{\sum_{x_i\in U}\overline{RN}_{B_2}D_t(x_i)} \geq \log\frac{\sum_{x_i\in U}\underline{RN}_{B_1}(D_t)(x_i)}{\sum_{x_i\in U}\overline{RN}_{B_1}D_t(x_i)}$ for $t = 1, 2, \cdots, r$, and consequently $\sum_{t=1}^{r}\frac{|D_t|}{|U|}\log\frac{\sum_{x_i\in U}\underline{RN}_{B_2}(D_t)(x_i)}{\sum_{x_i\in U}\overline{RN}_{B_2}D_t(x_i)} \geq \sum_{t=1}^{r}\frac{|D_t|}{|U|}\log\frac{\sum_{x_i\in U}\underline{RN}_{B_1}(D_t)(x_i)}{\sum_{x_i\in U}\overline{RN}_{B_1}D_t(x_i)}$. Furthermore, based on Eq (3.7), it is quite straightforward to obtain $-\sum_{t=1}^{r}\frac{|D_t|}{|U|}\log\frac{\sum_{x_i\in U}\underline{RN}_{B_2}(D_t)(x_i)}{\sum_{x_i\in U}\overline{RN}_{B_2}D_t(x_i)} \leq -\sum_{t=1}^{r}\frac{|D_t|}{|U|}\log\frac{\sum_{x_i\in U}\underline{RN}_{B_1}(D_t)(x_i)}{\sum_{x_i\in U}\overline{RN}_{B_1}D_t(x_i)}$. Therefore, $\mathrm{RFCE}_{B_2}(D) \leq \mathrm{RFCE}_{B_1}(D)$. □

It is also evident that $RFCE_B(D) \geq 0$ for $\forall B \subseteq A$. In addition, according to the definition of the rough set, we have $\frac{\sum_{x_i\in U}\underline{RN}_B(D_t)(x_i)}{\sum_{x_i\in U}\overline{RN}_BD_t(x_i)} \in [0, 1]$ for all specific decision classes, $\log\frac{\sum_{x_i\in U}\underline{RN}_B(D_t)(x_i)}{\sum_{x_i\in U}\overline{RN}_BD_t(x_i)} \leq 0$ $(t = 1, 2, \ldots, r)$, then $RFCE_B(D) \geq 0$. More specifically, when $\frac{\sum_{x_i\in U}\underline{RN}_B(D_t)(x_i)}{\sum_{x_i\in U}\overline{RN}_BD_t(x_i)} = 0$, $RFCE_B(D) = \infty$; when $\frac{\sum_{x_i\in U}\underline{RN}_B(D_t)(x_i)}{\sum_{x_i\in U}\overline{RN}_BD_t(x_i)} = 1$, $RFCE_B(D) = 0$.

Fuzzy uncertainty measurements, such as relative fuzzy roughness and composite entropy, are widely adopted for the imprecise numerical characterization of uncertainty. The dependency degree, defined as the ratio of the size of the positive region induced by classification to the total number of samples, is commonly used to evaluate the importance of an attribute subset. In contrast, fuzzy composite entropy is primarily utilized to measure the uncertainty of fuzzy similarity classes induced by attribute subsets. By integrating these two perspectives, it becomes possible to formally decompose and subsequently reconstruct the deterministic component and the uncertain component of knowledge within a unified framework. Accordingly, both the fuzzy dependency degree and the composite entropy are considered to construct a comprehensive composite uncertainty measure.

Definition 8 (RFDCE)**.** *The RFDCE of D with respect to B is given as*

$$\begin{aligned} RFDCE_B(D) &= (1 - RFD_B(D)) \times RFCE_B(D) \\ &= \left(1 - \frac{1}{|U|}\sum_{x_i\in U} RPOS_B(D)(x_i)\right) \times \left(-\sum_{t=1}^{r}\frac{|D_t|}{|U|}\log\frac{\sum_{x_i\in U}\underline{RN}_B(D_t)(x_i)}{\sum_{x_i\in U}\overline{RN}_BD_t(x_i)}\right). \end{aligned} \tag{3.8}$$

The $RFDCE_B(D)$ comprehensively quantifies useless fuzzy interference that fails to facilitate classification. Here, $RFCE_B(D)$ gauges the fuzzy boundary degree of decision partitions under feature subset B, and $RFD_B(D)$ reflects the distinguishing power of attributes. The factor $1 - RFD_B(D)$ serves as an inverse weighting coefficient to reduce the penalty imposed by fuzzy class boundaries on attributes with strong discriminative capacity. The multiplication of these two terms measures the overall ineffective uncertainty stemming from overlapping decision boundaries: Attribute subsets with outstanding separability and distinct partition boundaries yield lower index values, whereas irrelevant attributes with severe boundary ambiguity generate higher values. This property directs the algorithm to pick concise feature subsets with minimal fuzzy noise interference.

Theorem 5 (RFDCE granulation monotonicity)**.** *If $B_1 \subseteq B_2 \subseteq A$, then $RFDCE_{B_2}(D) \leq RFDCE_{B_1}(D)$.*

Proof. Given the condition $B_1 \subseteq B_2$, it can be deduced from Theorems 3 and 4 that $RFD_{B_1}(D) \leq RFD_{B_2}(D)$ and $\mathrm{RFCE}_{B_2}(D) \leq \mathrm{RFCE}_{B_1}(D)$, and noting that $1-RFD_{B_2}(D) \leq 1-RFD_{B_1}(D)$. Accordingly, we can obtain $(1-RFD_{B_2}(D))\times\mathrm{RFCE}_{B_2}(D) \leq (1-\mathrm{RFD}_{B_1}(D))\times\mathrm{RFCE}_{B_1}(D)$, which verifies the validity of $\mathrm{RFDCE}_{B_2}(D) \leq \mathrm{RFDCE}_{B_1}(D)$. □

Distinct from existing works treating distance measurement and feature evaluation as independently optimized modules, this study constructs a two-dimensional collaborative enhancement framework with mutually reinforcing distance refinement and uncertainty deepening, achieving systematic advancement from underlying distance metrics to upper-level uncertainty evaluation.

Algorithm 1 Calculating improved uncertainty measurements.

Require: Decision system (U, A, D), parameter p and nonempty feature subset $B \subseteq A$.
Ensure: New uncertainty measurements value: $RFD_B(D)$, $RFCE_B(D)$, $RFDCE_B(D)$.

1: Compute relative distance $rd_a(x_i, x_j)$ and fuzzy similarity relation $rR_a(x_i, x_j) = \exp(-p\times rd_a(x_i, x_j))$ and $rR_B = \cap_{a\in B} rR_a$.
2: Obtain decision classification $U/D = \{D_1, \cdots, D_r\}$ and its fuzzy decision classes $r\widetilde{D}_t$ $(t = 1, \cdots, r)$.
3: Let $RFCEDt = 0$.
4: **for** $t \in \{1, \cdots, r\}$ **do**
5: Let $LowAppDt = 0$, $UpAppDt = 0$.
6: **for** $i \in \{1, \cdots, n\}$ **do**
7: By Definition 5, calculate $\underline{RN}_B(D_t)(x_i)$, $\overline{RN}_B(D_t)(x_i)$.
8: Make the integrated updates:
Let $LowAppDt \leftarrow LowAppDt + \underline{RN}_B(D_t)(x_i)$, $UpAppDt \leftarrow UpAppDt + \underline{RN}_B(D_t)(x_i)$.
9: **end for**
10: Let $RFCEDt \leftarrow RFCEDt - \frac{|D_t|}{|U|}\log\frac{LowAppDt}{UpAppDt}$.
11: **end for**
12: $RFCE_B(D) = RFCEDt$.
13: By Eqs. (3.6) and (3.8), obtain $RFD_B(D)$ and $RFDCE_B(D)$.
14: **return** $RFD_B(D)$, $RFCE_B(D)$, $RFDCE_B(D)$.

For the newly proposed fuzzy rough sets and their associated uncertainty measures, Algorithm 1 outlines the corresponding computational procedure. In Algorithm 1, Steps 1 and 2 are devoted to constructing the conditional granulation and the decision granulation, respectively. Subsequently, Steps 3–12 employ two nested "for" loops to compute the fuzzy lower and upper approximations and to obtain $RFCE_B(D)$. In Step 13, $RFD_B(D)$ and $RFDCE_B(D)$ are calculated, and finally, Step 14 outputs $RFD_B(D)$, $RFCE_B(D)$ and $RFDCE_B(D)$.

The algorithm's time complexity stems from fuzzy similarity matrix construction and approximation computation. Let n, m and r represent the total sample count, feature subset size, and number of decision classes, respectively. Constructing the fuzzy similarity relation for feature subset B costs $O(mn^2)$, while the double loop over all decision classes and samples to compute upper/lower approximations takes $O(rn^2)$. Neglecting low-order linear terms, the overall time complexity equals $O(n^2(m + r))$.

Next, a numerical example is provided to illustrate the relative fuzzy rough dual approximations and the corresponding uncertainty measurements.

Example I. A fuzzy decision table is presented in Table 2 [35], where U denotes the sample set $\{x_1, x_2, x_3, \cdots, x_{10}, x_{11}, x_{12}\}$, $A = \{a_1, a_2\}$ represents the conditional attribute set, and D is the decision attribute with $U/D = \{D_1 = \{x_1, x_2, x_3, x_4\}, D_2 = \{x_5, x_6, x_7, x_8\}, D_3 = \{x_9, x_{10}, x_{11}, x_{12}\}\}$. For convenience in subsequent statistical analysis, the parameter is fixed as $p = 2$.

The relative fuzzy decision classes are derived from the original decision classes, denoted as $\{r\widetilde{D}_1, r\widetilde{D}_2, r\widetilde{D}_3\}$, and the corresponding results are summarized in Table 3. According to Definition 5, based on the given data table, the relative fuzzy approximations associated with each relative fuzzy decision class can be computed. Consequently, the relative weighted fuzzy rough lower and upper approximations are obtained, and the detailed results are explicitly presented in Table 4. Furthermore, the corresponding dependency degree can be calculated using Definition 6 together with Eq (3.6), that is,

$$RFD_A(D) = \frac{1}{|U|}\sum_{x_i \in U} RPOS_A(D)(x_i) = 0.3343 \geq RFD_{\{a_1\}}(D) = 0.2970.$$

These experimental results confirm the monotonicity of $RFD_B(D)$ with respect to attribute inclusion, thereby validating Theorem 3. In addition, the granulation monotonicity of $RPOS_B(D)$ stated in Theorem 3 can also be fully verified. Specifically, from the RFPR values listed in Table 4, it can be observed that for any $x_i \in U$, the relation

$$RPOS_{\{a_1\}}(D)(x_i) \leq RPOS_A(D)(x_i) \text{ holds, that is, } RPOS_{\{a_1\}}(D) \subseteq RPOS_A(D).$$

On the basis of the computational procedures and results of the relative weighted fuzzy rough lower and upper approximations, together with the fuzzy dependency degree $RFD_A(D)$ discussed previously, both fuzzy composite entropy and fuzzy dependency-based composite entropy can be further calculated. The corresponding results demonstrate their monotonic behavior, which confirms the validity of Theorems 4 and 5.

$$RFCE_A(D) = -\sum_{t=1}^{r} \frac{|D_t|}{|U|} \log \frac{\sum_{x_i \in U} \underline{RN}_A(D_t)(x_i)}{\sum_{x_i \in U} \overline{RN}_A D_t(x_i)} = 0.7290 \leq RFCE_{\{a_1\}}(D) = 0.8057,$$

$$RFDCE_A(D) = (1-RFD_A(D))\times RFCE_A(D) = (1-0.3343)\times 0.7290 = 0.4853 \leq RFDCE_{\{a_1\}}(D) = 0.5664.$$

Table 2. Data information table.

U	a_1	a_2	D
x_1	0.38	0.70	1
x_2	0.25	0.72	1
x_3	0.36	0.80	1
x_4	0.42	0.57	1
x_5	0.45	0.30	2
x_6	0.60	0.40	2
x_7	0.48	0.44	2
x_8	0.60	0.29	2
x_9	0.85	0.60	3
x_{10}	0.90	0.60	3
x_{11}	0.81	0.46	3
x_{12}	0.80	0.66	3

Table 3. Relative fuzzy decision classes.

U	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
$r\widetilde{D}_1$	0.4525	0.4755	0.4717	0.3996	0.2866	0.2788	0.3165	0.2580	0.2428	0.2388	0.2421	0.2614
$r\widetilde{D}_2$	0.2933	0.2894	0.2737	0.3416	0.4484	0.4150	0.4146	0.4405	0.2863	0.2811	0.3275	0.2865
$r\widetilde{D}_3$	0.2541	0.2351	0.2546	0.2588	0.2649	0.3062	0.2689	0.3014	0.4710	0.4800	0.4303	0.4521

Table 4. Relative fuzzy double approximations of decision classes.

U	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
$\underline{RN}_{\{a_1\}}(D_1)$	0.2706	0.2866	0.2792	0.2580	0.2580	0.2421	0.2580	0.2421	0.2388	0.2388	0.2388	0.2388
$\underline{RN}_{\{a_1\}}(D_2)$	0.2737	0.2737	0.2737	0.2737	0.2737	0.2865	0.2737	0.2758	0.2811	0.2811	0.2811	0.2811
$\underline{RN}_{\{a_1\}}(D_3)$	0.2351	0.2351	0.2351	0.2351	0.2351	0.2541	0.2362	0.2541	0.3475	0.3913	0.3014	0.3014
$RPOS_{\{a_1\}}(D)$	0.2737	0.2866	0.2792	0.2737	0.2737	0.2865	0.2737	0.2758	0.3475	0.3913	0.3014	0.3014
$\underline{RN}_A(D_1)$	0.3165	0.3442	0.3868	0.2788	0.2580	0.2421	0.2580	0.2421	0.2388	0.2388	0.2388	0.2388
$\underline{RN}_A(D_2)$	0.2737	0.2737	0.2737	0.2737	0.3416	0.3184	0.3046	0.3275	0.2811	0.2811	0.2811	0.2811
$\underline{RN}_A(D_3)$	0.2351	0.2351	0.2351	0.2351	0.2588	0.2588	0.2541	0.2649	0.3493	0.3913	0.3014	0.3510
$RPOS_A(D)$	0.3165	0.3442	0.3868	0.2788	0.3416	0.3184	0.3046	0.3275	0.3493	0.3913	0.3014	0.3510
$\overline{RN}_{\{a_1\}}(D_1)$	0.4755	0.4755	0.4755	0.4755	0.4755	0.4755	0.4755	0.4755	0.4755	0.4755	0.4755	0.4755
$\overline{RN}_{\{a_1\}}(D_2)$	0.4484	0.4484	0.4484	0.4484	0.4484	0.4484	0.4484	0.4484	0.4484	0.4484	0.4484	0.4484
$\overline{RN}_{\{a_1\}}(D_3)$	0.4800	0.4800	0.4800	0.4800	0.4800	0.4800	0.4800	0.4800	0.4800	0.4800	0.4800	0.4800
$\overline{RN}_A(D_1)$	0.4755	0.4755	0.4755	0.4755	0.4755	0.4755	0.4755	0.4755	0.4755	0.4755	0.4755	0.4755
$\overline{RN}_A(D_2)$	0.4484	0.4484	0.4484	0.4484	0.4484	0.4484	0.4484	0.4484	0.4484	0.4484	0.4484	0.4484
$\overline{RN}_A(D_3)$	0.4800	0.4800	0.4800	0.4800	0.4800	0.4800	0.4800	0.4800	0.4800	0.4800	0.4800	0.4800

4. Feature selection based on fuzzy uncertainty measurement

In the preceding sections, the traditional fuzzy rough set model was enhanced along two dimensions:

distance optimization and refinement of uncertainty measurement, resulting in the RWFRSs model, as well as the uncertainty measures RFD and RFCE. Moreover, these two measures are further integrated to construct a new uncertainty measure, namely RFDCE. Within this two-dimensional improvement framework, the core uncertainty measures are systematically established, and their granulation monotonicity properties are theoretically and experimentally revealed. The proposed two-dimensional collaborative enhancement framework achieves intrinsic synergy via closed-loop coupling: Distance calibration underpins uncertainty measurements, whose theoretical properties and feature importance outputs iteratively refine the distance space to form a tight coupling rather than mere superposition. In this section, feature selection is performed using the newly proposed uncertainty measures, and the corresponding heuristic algorithms are developed. Accordingly, three feature selection algorithms induced by the uncertainty measures RFD, RFCE, and RFDCE are denoted as RFD-FS, RFCE-FS, and RFDCE-FS, respectively.

Specifically, the heuristic feature selection algorithms are first determined based on uncertainty measurement and feature importance, followed by a detailed illustrative example. Due to differences in monotonicity properties, RFD-FS continues to adopt the feature selection framework proposed by Wang et al. [35], which is referred to as WAFS (feature selection algorithm based on WFRSs). For RFCE-FS and RFDCE-FS, for simplicity, the uncertainty measures or their essential notions are uniformly denoted by a symbol $\heartsuit_B(D)(\heartsuit \in RFCE, RFDCE)$, and the corresponding selection algorithms are collectively represented as $\heartsuit$-FS. Thus, $\heartsuit_B(D)$ and $\heartsuit$-FS serve as generic notations for the measures and algorithms, respectively.

Feature selection is investigated under the decision system (U, A, D). Based on the underlying measure $\heartsuit_B(D)$, feature importance is defined to capture the difference in uncertainty during the incremental process, and the corresponding feature selection algorithms are subsequently designed through heuristic search.

Definition 9. *The significance of an external feature (Sig) $b \in A - B$ with respect to an attribute subset $B \subseteq A$ is defined as*

$$Sig\heartsuit_B(b, D) = \heartsuit_B(D) - \heartsuit_{B\cup\{b\}}(D). \tag{4.1}$$

The quantity $\mathrm{Sig}\heartsuit_B(b, D)$ measures the change in uncertainty during the incremental process $B \overset{+b}{\longmapsto} B\cup\{b\}$, thereby enabling a dynamic evaluation of uncertainty informativeness in the process of attribute reduction. Owing to the monotonicity of the measure $\heartsuit$, the significance indicator is well-defined and nonnegative, providing a consistent criterion for feature ranking. Based on the above definitions, a forward greedy search algorithm for computing an optimal attribute subset can be constructed.

Algorithm 2 ($\heartsuit$-FS) feature selection algorithm grounded in the uncertainty measure $\heartsuit_B(D)$.

Require: Decision system (U, A, D) and parameter p.

Ensure: A subset of features Red.

1: Compute relative fuzzy similarity relation $rR_a(x_i, x_j)$ $(\forall a \in A)$.
2: The fuzzy decision classes $\{r\widetilde{D}_1, r\widetilde{D}_2, \ldots, r\widetilde{D}_r\}$ is obtained through converting crisp labels to fuzzy label form by Eq (3.2).
3: Initialize begin=1, $Red = \{a_*\}$, where $a_* = \arg\min_{a\in A} \heartsuit_{\{a\}}(D)$ (with ties broken by the first one), and compute $\heartsuit_{Red}(D)$.
4: **while** begin=1 **do**
5: **for** $\forall a \in A - Red$ **do**
6: **for** $\forall D_t, t = 1, \ldots, r$ **do**
7: **for** $\forall x \in U$ **do**
8: Compute $rS(x, D_t)$ and $r\alpha(x, D_t)$ according to Definition 4.
9: Compute $\underline{RN}_{Red\cup\{a\}}(D_t)(x)$ and $\overline{RN}_{Red\cup\{a\}}(D_t)(x)$ according to Definition 5.
10: **end for**
11: **end for**
12: Compute $\heartsuit_{Red\cup\{a\}}(D)$, and obtain $\mathrm{Sig}\heartsuit_{Red}(a, D) = \heartsuit_{Red}(D) - \heartsuit_{Red\cup\{a\}}(D)$.
13: **end for**
14: Find $a_0 = \arg\max_{a\in A-Red} \mathrm{Sig}\heartsuit_{Red}(a, D)$ (where the first one is chosen).
15: **if** $\mathrm{Sig}\heartsuit_{Red}(a_0, D) > 0$ **then**
16: Let $Red \leftarrow Red \cup \{a_0\}$, and compute $\heartsuit_{Red}(D)$.
17: **else**
18: begin=0.
19: **end if**
20: **end while**
21: **return** Red.

The feature selection algorithm grounded in the uncertainty measure $\heartsuit_B(D)$ and its corresponding significance is implemented via heuristic search, as described in Algorithm 2. Specifically, when selecting the optimal class center a_0 from the candidate set $A - Red$, the feature significance (or relevance) $\mathrm{Sig}\heartsuit_{Red}(a, D)$ is evaluated for each candidate $a \in A - Red$. The optimal class center a_0 is then determined as the one that maximizes this significance value, that is,

$$a_0 = \arg \max_{a\in A-Red} \mathrm{Sig}\heartsuit_{Red}(a, D). \tag{4.2}$$

Algorithm 2 is composed of 21 steps, including one "while" loop and one triple-nested "for" loop. Specifically, Step 1 is devoted to computing the relative fuzzy similarity relations for each conditional attribute, while Step 2 constructs the fuzzy decision classes. In Step 3, the algorithm initializes the feature subset by evaluating all singleton feature subsets and selecting the one with the minimum uncertainty, i.e., $Red = \{a_*\}$, where $a_* = \arg\min_{a\in A} \heartsuit_{\{a\}}(D)$ (with ties broken by the first one), and then computes $\heartsuit_{Red}(D)$. Subsequently, Steps 4–20 execute a "while" loop, within which the feature selection procedure is implemented through two successive stages. In the first stage, covering Steps 5–14, three nested "for" loops are employed to search for the maximum value of $\mathrm{Sig}\heartsuit_{Red}(a, D)$. In the second stage, consisting of Steps 15–19, a conditional judgment is performed to determine whether

the algorithm should continue updating the subset or terminate. In this updating process, attributes with positive importance values are added to the current subset. Finally, Step 21 outputs the optimized feature subset obtained through the selection process.

Let n, m and r denote the sample size, total conditional attributes, and number of decision classes, respectively. Step 1 costs $O(mn^2)$ to build fuzzy similarity matrices for individual attributes. The pre-selection in Step 3 evaluates all m singleton subsets, each requiring $O(rn^2)$ computations, hence its cost is $O(mrn^2)$. Since the first feature has already been pre-selected in Step 3, the outer greedy loop runs at most $m-1$ iterations, where each iteration scans up to m candidate attributes. Evaluating each candidate demands traversing r decision classes and n samples to calculate neighborhood approximations, bringing an $O(rn^2)$ overhead. Consequently, the total cost is $O(m^2rn^2)$.

In practice, Algorithm ♡-FS encompasses two specific algorithms, namely RFCE-FS and RFDCE-FS. The overall structure of Algorithm 2 is designed by referring to the framework proposed in [35]. Moreover, the WAFS algorithmic framework presented in [35] is applicable to basic uncertainty measures. Therefore, the RFD-FS algorithm corresponding to the uncertainty measure $RFD_B(D)$ also adopts this heuristic framework and is employed as a comparative method in subsequent experimental evaluations.

To illustrate the feature selection procedure, particularly the heuristic algorithm, a concrete example is provided.

Example II. Focusing on feature selection algorithms, the dataset Bupa from UCI [39] is employed, which contains 345 samples, 6 conditional attributes, and 1 decision attribute. First, the dataset is normalized using a max-min normalization scheme, as defined in

$$f(x,c) \leftarrow \frac{f(x,c) - \min_{y\in U} f(y,c)}{\max_{y\in U} f(y,c) - \min_{y\in U} f(y,c)} \in [0,1] \ \ (\forall x \in U, \forall c \in A). \tag{4.3}$$

Subsequently, 20 samples are randomly selected to construct a decision system (U, A, D), which is shown in Table 5. The resulting decision system contains two decision classes, denoted as $D_1 = \{x_1, x_2, x_3, x_9, x_{12}, x_{13}, x_{14}, x_{17}, x_{20}\}$, $D_2 = \{x_4, x_5, x_6, x_7, x_8, x_{10}, x_{11}, x_{15}, x_{16}, x_{18}, x_{19}\}$.

Tables 6 and 7 report the execution processes of the monotonic feature selection algorithms RFCE-FS ($p = 2$) and RFDCE-FS ($p = 2$), respectively. In both tables, as the "Addition" (iteration index) increases from 1 to 4, the selection set *Red* evolves progressively from the empty set $\emptyset$ to $\{a_6\}$, then to $\{a_3, a_6\}$, and finally to $\{a_3, a_5, a_6\}$. At the same time, several indicators, including $RFCE_{\text{Red}}(D)$ or $RFDCE_{\text{Red}}(D)$, the feature difference set $A - Red$, and $RFCE_{\text{Red}\cup\{a\}}(D)$ or $RFDCE_{\text{Red}\cup\{a\}}(D)$ (together with $\text{Sig}RFCE_{\text{Red}}(a, D)$ or $\text{Sig}RFDCE_{\text{Red}}(a, D)$ for feature significance evaluation), are updated accordingly.

When the "Addition" index reaches 4, the conditions $\text{Sig}RFCE_{\text{Red}}(a_k, D) = 0$ (in Table 6) and $\text{Sig}RFDCE_{\text{Red}}(a_k, D) = 0$ (in Table 7) are observed, which no longer satisfy the continuation condition ($\text{Sig}RFCE_{\text{Red}}(a_k, D) > 0$ or $\text{Sig}RFDCE_{\text{Red}}(a_k, D) > 0$). Consequently, the final reduct set obtained by both algorithms is $\{a_3, a_5, a_6\}$. These results clearly demonstrate the feature selection and reduction processes of the RFCE-FS and RFDCE-FS algorithms. Although both methods produce the same final reduct set, they differ in the specific indicator values and computational trajectories during execution. Therefore, both RFCE-FS and RFDCE-FS are shown to be effective feature selection approaches. Their comparative performance will be further analyzed in subsequent experimental studies.

Table 5. A standardized decision system for feature selection demonstration.

U	a_1	a_2	a_3	a_4	a_5	a_6	d
x_1	0.6842	0.4783	0.1060	0.3377	0.0445	0.0250	1
x_2	0.7368	0.0000	0.2053	0.1948	0.0514	0.0250	1
x_3	0.7368	0.4696	0.1854	0.1688	0.0822	0.0250	1
x_4	0.6316	0.3478	0.1325	0.1948	0.1130	0.0250	2
x_5	0.6053	0.2261	0.1060	0.2208	0.0479	0.0250	2
x_6	1.0000	0.4522	0.0993	0.3247	0.0274	0.0500	2
x_7	0.5263	0.3130	0.1391	0.1948	0.0685	0.1500	2
x_8	0.7105	0.4348	0.1325	0.2078	0.1473	0.2000	2
x_9	0.7895	0.2348	0.0596	0.1169	0.0342	0.3000	1
x_{10}	0.6842	0.4435	0.5497	0.5844	0.2123	0.3000	2
x_{11}	0.7368	0.2348	0.0927	0.2597	0.0411	0.3000	2
x_{12}	0.6579	0.3826	0.4834	0.4416	0.3733	0.4000	1
x_{13}	0.6842	0.4087	0.0993	0.1818	0.0582	0.0250	1
x_{14}	0.6842	0.6348	0.2053	0.2208	0.4452	0.0250	1
x_{15}	0.5263	0.3043	0.0927	0.2468	0.0377	0.0250	2
x_{16}	0.5789	0.3130	0.2185	0.2857	0.0993	0.1000	2
x_{17}	0.8684	0.3739	0.6556	0.6753	0.3733	0.3000	1
x_{18}	0.6842	0.2000	0.1722	0.2468	0.1164	0.3500	2
x_{19}	0.8158	0.4000	0.3245	0.4935	0.6781	0.6000	2
x_{20}	0.8684	0.6609	0.3510	0.5195	0.2055	1.0000	1

Table 6. Execution process of monotonic algorithm RFCE-FS ($p = 2$) on Table 5.

Addition	R	$RFCE_{Red}(D)$	$A - R$	$RFCE_{Red\cup\{a\}}(D)$ $SigRFCE_{Red}(a, D)$	Optional feature a_0	Sufficient requirement $SigRFCE_{Red}(a_0, D) > 0$
1	$\emptyset$	1000	$\{a_1, a_2, a_3, a_4, a_5, a_6\}$	[0.3049,0.3049,0.3031,0.3047,0.3041,0.2440] [999.6951,999.6951,999.6969,999.6953,999.6959,999.7560]	a_6	999.756>0(Yes)
2	$\{a_6\}$	0.244	$\{a_1, a_2, a_3, a_4, a_5\}$	[0.2440,0.2440,0.2395,0.2437,0.2435] [0,0,0.004500,0.0003,0.0005]	a_3	0.0045>0(Yes)
3	$\{a_3, a_6\}$	0.2395	$\{a_1, a_2, a_4, a_5\}$	[0.2395,0.2395,0.2395,0.2390] [0,0,0,0.0005]	a_5	0.0005>0(Yes)
4	$\{a_3, a_5, a_6\}$	0.239	$\{a_1, a_2, a_4\}$	[0.2390,0.2390,0.2390] [0,0,0]	a_1	0=0(No)
Last reduct R				$\{a_3, a_5, a_6\}$		

Table 7. Execution process of monotonic algorithm RFDCE-FS ($p = 2$) on Table 5.

Addition	R	$RFDCE_{Red}(D)$	$A - R$	$RFDCE_{Red\cup\{a\}}(D)$ $SigRFDCE_{Red}(a, D)$	Optional feature a_0	Sufficient requirement $SigRFDCE_{Red}(a_0, D) > 0$
1	$\emptyset$	1000	$\{a_1, a_2, a_3, a_4, a_5, a_6\}$	[0.1532,0.1532,0.1522,0.1530,0.1528,0.1173] [999.8468,999.8468,999.8478,999.8470,999.8472,999.8827]	$\{a_6\}$	999.8827>0(Yes)
2	$\{a_6\}$	0.1173	$\{a_1, a_2, a_3, a_4, a_5\}$	[0.1173,0.1173,0.1148,0.1172,0.1171] [0,0,0.002600,0.0001,0.0002]	$\{a_3\}$	0.0026>0(Yes)
3	$\{a_3, a_6\}$	0.1148	$\{a_1, a_2, a_4, a_5\}$	[0.1148,0.1148,0.1148,0.1145] [0,0,0,0.0002]	$\{a_5\}$	0.0002>0(Yes)
4	$\{a_3, a_5, a_6\}$	0.1145	$\{a_1, a_2, a_4\}$	[0.1145,0.1145,0.1145] [0,0,0]	$\{a_1\}$	0=0(No)
Last reduct R				$\{a_3, a_5, a_6\}$		

5. Data experiment analyses

In this section, additional datasets are employed to further investigate the proposed uncertainty measures and feature selection algorithms. Table 8 summarizes the detailed characteristics of the datasets used in the experiments, covering a wide range of sample sizes, numbers of conditional attributes, and decision classes, thereby enabling a comprehensive and systematic experimental evaluation.

Table 8. Basic descriptions of 16 UCI datasets.

No.	Name (abbreviation)	Sample number	Condition attribute number	Decision class number
(1)	Zoo	101	16	7
(2)	Breast_Tissue (BT)	106	9	6
(3)	Seeds	210	7	3
(4)	Bupa	345	6	2
(5)	Glass	214	9	6
(6)	Lymphography (Lym.)	148	18	4
(7)	Haberman (Habe.)	306	3	2
(8)	Vertebral-column-2C (Vert.)	310	6	2
(9)	Parkinsons (Park)	195	22	2
(10)	Transfusion (Trans.)	748	4	2
(11)	Wpbc	198	34	2
(12)	Sonar	208	60	2
(13)	Cmc	1473	9	3
(14)	Biodeg	1055	41	2
(15)	Darwin	174	450	2
(16)	Leaf	340	14	30

To illustrate the two core components of this study, namely uncertainty measurement and heuristic feature selection algorithms, a total of 16 UCI datasets [40] are selected for comprehensive experimental evaluation. The fundamental characteristics of these datasets are summarized in Table 8. The selected datasets, including Zoo, Bupa, Glass, and others, are all sourced from the UCI Machine Learning Repository, which serves as a widely recognized benchmark repository. It should be emphasized that the latter four datasets belong to large-scale data benchmarks. Cmc and Biodeg contain abundant sample instances, Biodeg and Darwin possess high-dimensional feature spaces, and Leaf is characterized by a large number of decision classes. These datasets provide rich and diverse data resources for experiments on uncertainty quantification and reduction algorithms, thereby establishing a solid basis for subsequent validation and analysis. Further investigations into uncertainty measures and feature selection strategies can be conducted through additional experimental studies on such datasets.

Each dataset is formulated as a decision system $\mathrm{DS} = (U, A, D)$. Samples containing missing values are removed prior to analysis, and data preprocessing is performed using the max-min normalization defined in Eq (4.3). All experiments are implemented in MATLAB R2018a and executed on a computing platform equipped with an Intel (R) Core (TM) i5-8265U CPU running at 1.80 GHz, 16

GB RAM, and a 64-bit operating system.

This section is organized into two main parts. The first part focuses on experimentally verifying the properties of the uncertainty measures proposed in this work, while the second part demonstrates the effectiveness of the corresponding feature selection algorithms.

5.1. Measurement calculation and verification

The primary objective of this study lies in the deep integration of distance optimization and uncertainty quantification. Within this two-dimensional enhancement framework, the uncertainty indicators RFD, RFCE, and RFDCE are computed and systematically examined. The numerical results obtained from these measures are employed to validate their fundamental properties, with particular emphasis on granulation monotonicity. Guided by this perspective, a series of targeted experimental designs are conducted.

In this paper, fuzzy weighted uncertainty measures are introduced to characterize data uncertainty, including RFD, RFCE, and RFDCE. To facilitate analysis, the dataset scenario is fixed (for instance, the parameter p is held constant), while the attribute subset B is allowed to vary across different subfigures. Under this setting, the behavior of the corresponding uncertainty measures with respect to changes in B and their performance across different datasets are investigated. Specifically, each dataset (such as Zoo or Bupa) is treated as an independent analysis unit, and the variation trends of $RFD_B(D)$, $RFCE_B(D)$, and $RFDCE_B(D)$ are examined as the attribute subset B evolves.

In this experimental design, two factors are involved: the parameter p and the attribute subset B. The parameter p is fixed throughout the experiments, whereas B serves as the variable parameter represented along the horizontal axis, and the corresponding uncertainty values constitute the vertical axis. The attribute subsets B are constructed in a sequential manner to enable progressive observation, as specified in

$$B : B_1 = \{a_1\} \subset B_2 = \{a_1, a_2\} \subset \cdots \subset B_{|A|} = \{a_1, a_2, \cdots, a_{|A|}\} = A. \tag{5.1}$$

According to the corresponding measurement algorithms (e.g., Algorithm 1), the uncertainty values are computed along the attribute change chain. The resulting trends are illustrated in Figure 2, where B varies along the horizontal axis under a fixed p value, and the changes in the uncertainty measures are depicted along the vertical axis.

By analyzing Figure 2, the monotonic behavior of the proposed uncertainty measures with respect to the attribute subset B can be clearly observed and verified. These observations are consistent with the theoretical results established in Theorems 3–5. The experimental outcomes further demonstrate the robustness of the proposed measures under attribute chain variations. For clarity and conciseness, detailed discussions are primarily focused on the dataset (1) Zoo, where representative evidence fragments are provided, as shown in

$$\begin{aligned} &\text{(1) Zoo, } B_1 \subset B_3 \subset B_{16}, \\ &RFD_{B_1}(D) = 0.3705 < RFD_{B_3}(D) = 0.3712 < RFD_{B_{16}}(D) = 0.3933, \\ &RFCE_{B_1}(D) = 0.6952 > RFCE_{B_3}(D) = 0.6749 > RFCE_{B_{16}}(D) = 0.5193, \\ &RFDCE_{B_1}(D) = 0.4376 > RFDCE_{B_3}(D) = 0.4244 > RFDCE_{B_{16}}(D) = 0.3150. \end{aligned} \tag{5.2}$$

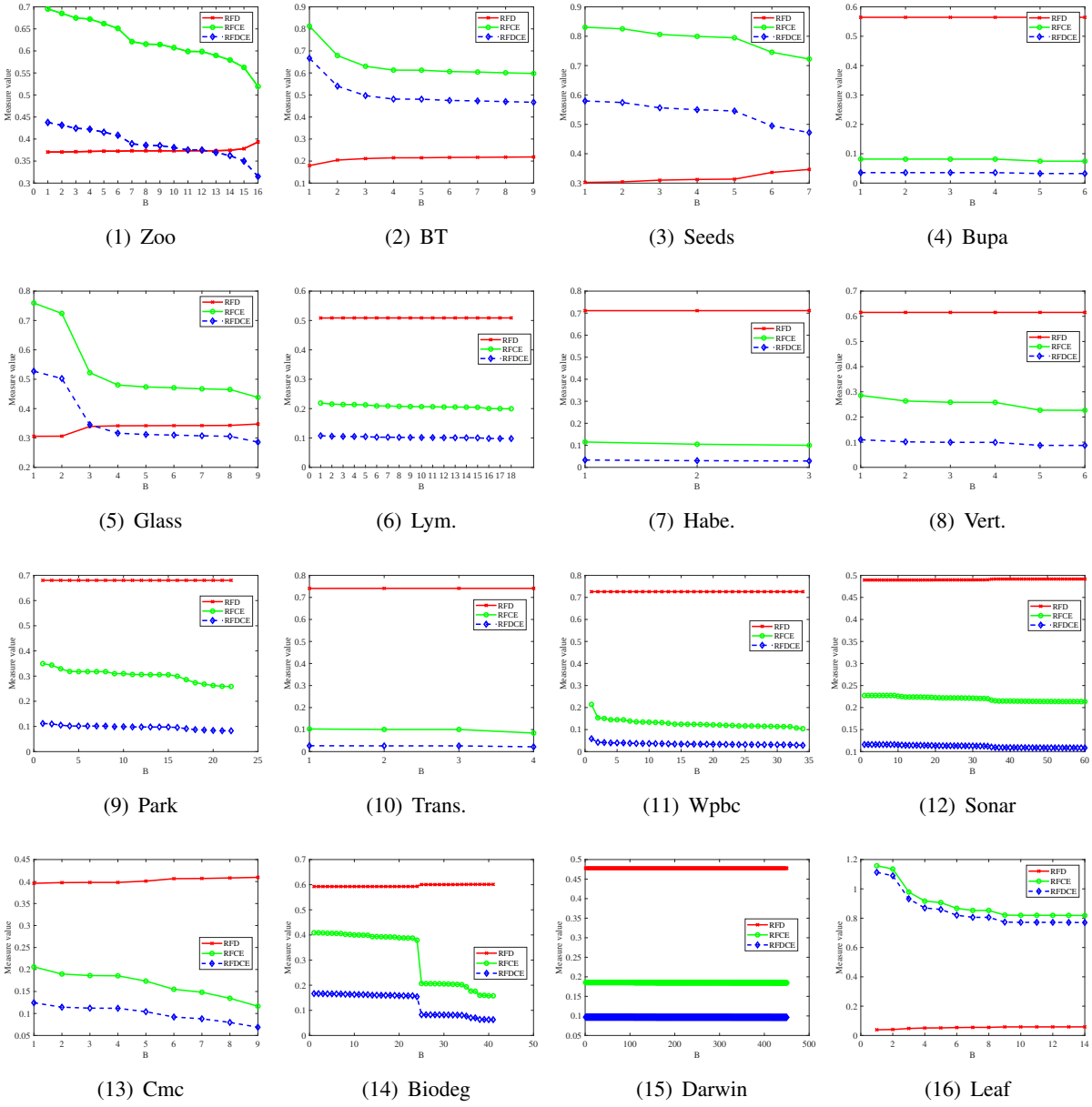


Figure 2. Experimental relative weighted fuzzy uncertainty measures on B chain.

In this part, data-driven experiments are conducted to compute and validate the fuzzy weighted uncertainty measures along the attribute chain B. These measures characterize the uncertainty properties on different datasets, such as Zoo, Bupa, and Glass.

5.2. Classification performances of feature selection

For feature selection, classification capability is commonly regarded as a primary criterion for evaluating algorithmic performance. In practice, this capability is usually assessed by classification accuracy, which is defined as the proportion of correctly predicted samples in supervised learning tasks. A 10-fold cross-validation strategy is adopted in our experiments to obtain reliable evaluation

outcomes. As reported in [35], the WAFS algorithm has been thoroughly verified to achieve competitive classification performance when compared with several existing feature selection methods, including fuzzy self-information-based feature selection algorithm (FSI-FS) [38], fitting fuzzy rough set-based feature selection algorithm (FTFS) [41], and heuristic algorithm based on variable distance parameter (AVDP) [42].

Building upon the feature selection framework described above, we further compare the effectiveness and superiority of different algorithms across multiple datasets. The comparison set includes the existing methods WAFS [35], FSI-FS [38], FTFS [41], and AVDP [42], together with the three newly proposed approaches (i.e., RFD-FS, RFCE-FS, and RFDCE-FS). For comparative evaluation, the feature subsets obtained by the heuristic selection algorithms are assessed using two widely adopted classifiers, namely KNN (with $K = 3$) and SVM. The standard classification accuracy is selected as the primary performance indicator for estimating classification ability and comparing algorithm effectiveness. Through supervised classification experiments, two complementary evaluation perspectives are provided: global comparisons across all algorithms and statistical performance analyses.

5.2.1. Global comparisons of all parallel selection algorithms

For the experiments reported herein, the experimental results of all algorithms on the two classifiers (KNN and SVM) over 16 datasets are computed, including the number of selected features, runtime, and classification accuracy. Table 9 summarizes the 10-fold cross-validation averages of classification accuracy, runtime, and selected feature subset size for all seven algorithms across two classifiers on six representative datasets, enabling a holistic comparison of prediction performance, computational efficiency, and feature sparsity. Overall, all involved feature selection approaches are effective, and this table provides an integrated view of their comprehensive performance when paired with different classifiers.

Subsequently, emphasis is placed on the key performance indicator, namely classification accuracy, to highlight the impact of feature selection on classification outcomes. Tables 10 and 11 summarize the classification accuracies obtained on all datasets using the two classifiers with the seven algorithms, and the average accuracy values over the 16 datasets are also reported. To facilitate intuitive comparison, a unified formatting standard is adopted for all tabular results: The maximum value within each row is highlighted in bold type, and the suboptimal values in the rows of averaged statistical results are labeled in italic font. From the perspective of average accuracy on the KNN classifier shown in Table 10, clear performance differences can be observed among the seven feature selection algorithms, where RFCE-FS and RFDCE-FS achieve the best and second-best results, respectively.

Table 9. Average outcomes of all algorithms across two classifiers derived via 10-fold cross-validation (for six representative datasets).

Dataset	Algorithm	Feature selection		Classification			
		Feature number	Time	KNN		SVM	
				Accuracy	Time	Accuracy	Time
(1) Zoo	FTFS	7.2 ± 0.7888	0.5929 ± 0.037	0.9 ± 0.0816	0.0118 ± 0.0049	0.92 ± 0.1135	0.1923 ± 0.0197
	FSI-FS	11.9 ± 0.5676	1.2764 ± 0.0447	0.8909 ± 0.0877	0.0092 ± 0.0016	0.8909 ± 0.0877	0.1906 ± 0.0217
	AVDP	3 ± 0	0.4024 ± 0.0171	0.8118 ± 0.0737	0.0094 ± 0.0011	0.8118 ± 0.0737	0.1694 ± 0.0089
	WAFS	10.8 ± 0.4216	1.19 ± 0.0721	0.8709 ± 0.0952	0.0089 ± 0.0011	0.8518 ± 0.084	0.1883 ± 0.0191
	RFD-FS	4 ± 0	2.3366 ± 0.1149	0.8709 ± 0.0827	0.0099 ± 0.0018	0.8118 ± 0.0567	0.1797 ± 0.0185
	RFCE-FS	12.7 ± 0.483	2.9774 ± 0.1609	0.9209 ± 0.0787	0.0096 ± 0.0017	0.8709 ± 0.1063	0.1891 ± 0.0213
	RFDCE-FS	12.8 ± 0.4216	2.8873 ± 0.0591	0.9209 ± 0.0787	0.0097 ± 0.0023	0.8909 ± 0.074	0.1853 ± 0.0136
	Average	9.4000 ± 0.3880	1.4898 ± 0.0675	0.8872 ± 0.0832	0.0097 ± 0.0019	0.8685 ± 0.0863	0.1857 ± 0.0192
(3) Seeds	FTFS	1 ± 0	0.0836 ± 0.0034	0.8476 ± 0.0586	0.0058 ± 0.0007	0.8619 ± 0.0726	0.0248 ± 0.0009
	FSI-FS	1 ± 0	0.1593 ± 0.0038	0.8476 ± 0.0586	0.0057 ± 0.0003	0.8619 ± 0.0726	0.0247 ± 0.0008
	AVDP	1 ± 0	0.092 ± 0.015	0.8476 ± 0.0586	0.0058 ± 0.0007	0.8619 ± 0.0726	0.0276 ± 0.0086
	WAFS	7 ± 0	0.3751 ± 0.0089	0.9333 ± 0.0512	0.0058 ± 0.0006	0.9429 ± 0.0492	0.0248 ± 0.0013
	RFD-FS	5.8 ± 0.4216	2.066 ± 0.0271	0.9191 ± 0.0505	0.0057 ± 0.0007	0.9238 ± 0.0559	0.025 ± 0.0011
	RFCE-FS	7 ± 0	2.0887 ± 0.0375	0.9333 ± 0.0512	0.0058 ± 0.0011	0.9429 ± 0.0492	0.0244 ± 0.001
	RFDCE-FS	7 ± 0	2.0857 ± 0.024	0.9333 ± 0.0512	0.0055 ± 0.0003	0.9429 ± 0.0492	0.0247 ± 0.0012
	Average	4.6000 ± 0.0527	0.8747 ± 0.0152	0.8994 ± 0.0539	0.0057 ± 0.0006	0.9101 ± 0.0588	0.0251 ± 0.0020
(6) Lym.	FTFS	4.1 ± 1.1972	0.3933 ± 0.0708	0.6619 ± 0.1553	0.0103 ± 0.0034	0.7357 ± 0.1267	0.0643 ± 0.0097
	FSI-FS	16.4 ± 0.6992	1.6966 ± 0.0805	0.7781 ± 0.1035	0.0082 ± 0.0008	0.8124 ± 0.1163	0.0605 ± 0.0042
	AVDP	3.2 ± 0.6325	0.3665 ± 0.0566	0.6148 ± 0.0773	0.0087 ± 0.0009	0.7305 ± 0.1023	0.0579 ± 0.0047
	WAFS	8.7 ± 2.1628	1.248 ± 0.2045	0.7433 ± 0.1247	0.0093 ± 0.0016	0.7433 ± 0.0981	0.0608 ± 0.0064
	RFD-FS	6.8 ± 1.6865	5.019 ± 0.1919	0.649 ± 0.1208	0.0094 ± 0.0013	0.7038 ± 0.1072	0.0607 ± 0.0046
	RFCE-FS	14.7 ± 1.1595	5.5058 ± 0.1442	0.7848 ± 0.0809	0.0084 ± 0.0005	0.799 ± 0.1291	0.0618 ± 0.0061
	RFDCE-FS	14.7 ± 1.1595	5.614 ± 0.1164	0.7848 ± 0.0809	0.0094 ± 0.0016	0.799 ± 0.1291	0.0642 ± 0.0067
	Average	10.7000 ± 1.0872	2.4995 ± 0.1094	0.7251 ± 0.1025	0.0090 ± 0.0014	0.7687 ± 0.1123	0.0614 ± 0.0058
(8) Vertebral	FTFS	1 ± 0	0.1509 ± 0.0095	0.6742 ± 0.0721	0.0093 ± 0.0013	0.6774 ± 0	0.0242 ± 0.0023
	FSI-FS	1 ± 0	0.3109 ± 0.0162	0.6742 ± 0.0721	0.0095 ± 0.0013	0.6774 ± 0	0.0243 ± 0.0013
	AVDP	2.6 ± 0.8433	0.2507 ± 0.0399	0.6484 ± 0.0514	0.0098 ± 0.0016	0.6774 ± 0	0.0242 ± 0.0017
	WAFS	1.9 ± 0.9944	0.4543 ± 0.1125	0.6387 ± 0.1481	0.0095 ± 0.0008	0.7226 ± 0.0612	0.024 ± 0.0019
	RFD-FS	6 ± 0	4.3036 ± 0.0972	0.771 ± 0.0811	0.01 ± 0.0019	0.8097 ± 0.0386	0.0247 ± 0.0024
	RFCE-FS	5.9 ± 0.3162	4.3153 ± 0.0728	0.7742 ± 0.0761	0.01 ± 0.0016	0.8065 ± 0.0402	0.0236 ± 0.0018
	RFDCE-FS	5.9 ± 0.3162	4.2718 ± 0.0913	0.7742 ± 0.0761	0.0097 ± 0.001	0.8065 ± 0.0402	0.0242 ± 0.0013
	Average	3.7875 ± 0.3088	1.7662 ± 0.0558	0.7157 ± 0.0823	0.0097 ± 0.0014	0.7484 ± 0.0274	0.0242 ± 0.0019
(12) Sonar	FTFS	1 ± 0	0.3889 ± 0.019	0.516 ± 0.1275	0.0056 ± 0.0003	0.5874 ± 0.1309	0.0146 ± 0.0005
	FSI-FS	1 ± 0	0.7543 ± 0.0158	0.516 ± 0.1275	0.0056 ± 0.0003	0.5874 ± 0.1309	0.0144 ± 0.0003
	AVDP	2.2 ± 0.9189	0.643 ± 0.1844	0.6369 ± 0.1022	0.0058 ± 0.0007	0.6617 ± 0.121	0.0147 ± 0.0011
	WAFS	22.2 ± 2.2998	6.9608 ± 0.5216	0.7971 ± 0.0439	0.0054 ± 0.0004	0.8307 ± 0.0667	0.0147 ± 0.0005
	RFD-FS	60 ± 0	62.7878 ± 4.1948	0.8357 ± 0.0569	0.0057 ± 0.0004	0.8795 ± 0.0603	0.0165 ± 0.0005
	RFCE-FS	35.5 ± 2.2236	62.1375 ± 4.227	0.85 ± 0.0667	0.0054 ± 0.0003	0.9088 ± 0.0612	0.0155 ± 0.0007
	RFDCE-FS	35.6 ± 2.1187	60.8519 ± 3.057	0.8502 ± 0.0528	0.0053 ± 0.0004	0.9133 ± 0.0489	0.0157 ± 0.0021
	Average	27.1875 ± 0.9451	24.4389 ± 1.5338	0.7297 ± 0.0793	0.0055 ± 0.0004	0.7810 ± 0.0850	0.0153 ± 0.0010
(15) Darwin	FTFS	1 ± 0	2.0573 ± 0.103	0.5477 ± 0.1009	0.0114 ± 0.0115	0.5118 ± 0.0206	0.0214 ± 0.0114
	FSI-FS	1 ± 0	3.8969 ± 0.0884	0.5477 ± 0.1009	0.0064 ± 0.0007	0.5118 ± 0.0206	0.0157 ± 0.0012
	AVDP	2.2 ± 1.3166	3.4899 ± 1.4356	0.5977 ± 0.1131	0.0063 ± 0.0005	0.6333 ± 0.1697	0.0156 ± 0.0018
	WAFS	2 ± 0.6667	5.5339 ± 1.1587	0.701 ± 0.1027	0.0075 ± 0.004	0.5578 ± 0.0492	0.0158 ± 0.0012
	RFD-FS	147.9 ± 7.7667	1965.4442 ± 79.4725	0.7242 ± 0.0806	0.0073 ± 0.0013	0.798 ± 0.0914	0.0196 ± 0.0013
	RFCE-FS	12.3 ± 2.6687	2205.7221 ± 1577.0637	0.769 ± 0.0978	0.0079 ± 0.003	0.7765 ± 0.0991	0.0189 ± 0.0043
	RFDCE-FS	12.3 ± 2.6687	1708.9425 ± 129.2631	0.769 ± 0.0978	0.0086 ± 0.0055	0.7765 ± 0.0991	0.0169 ± 0.0028
	Average	72.0625 ± 2.8698	772.2715 ± 225.9612	0.6689 ± 0.0946	0.0081 ± 0.0038	0.6369 ± 0.0730	0.0186 ± 0.0034

Table 10. KNN classification accuracies of all selection algorithms.

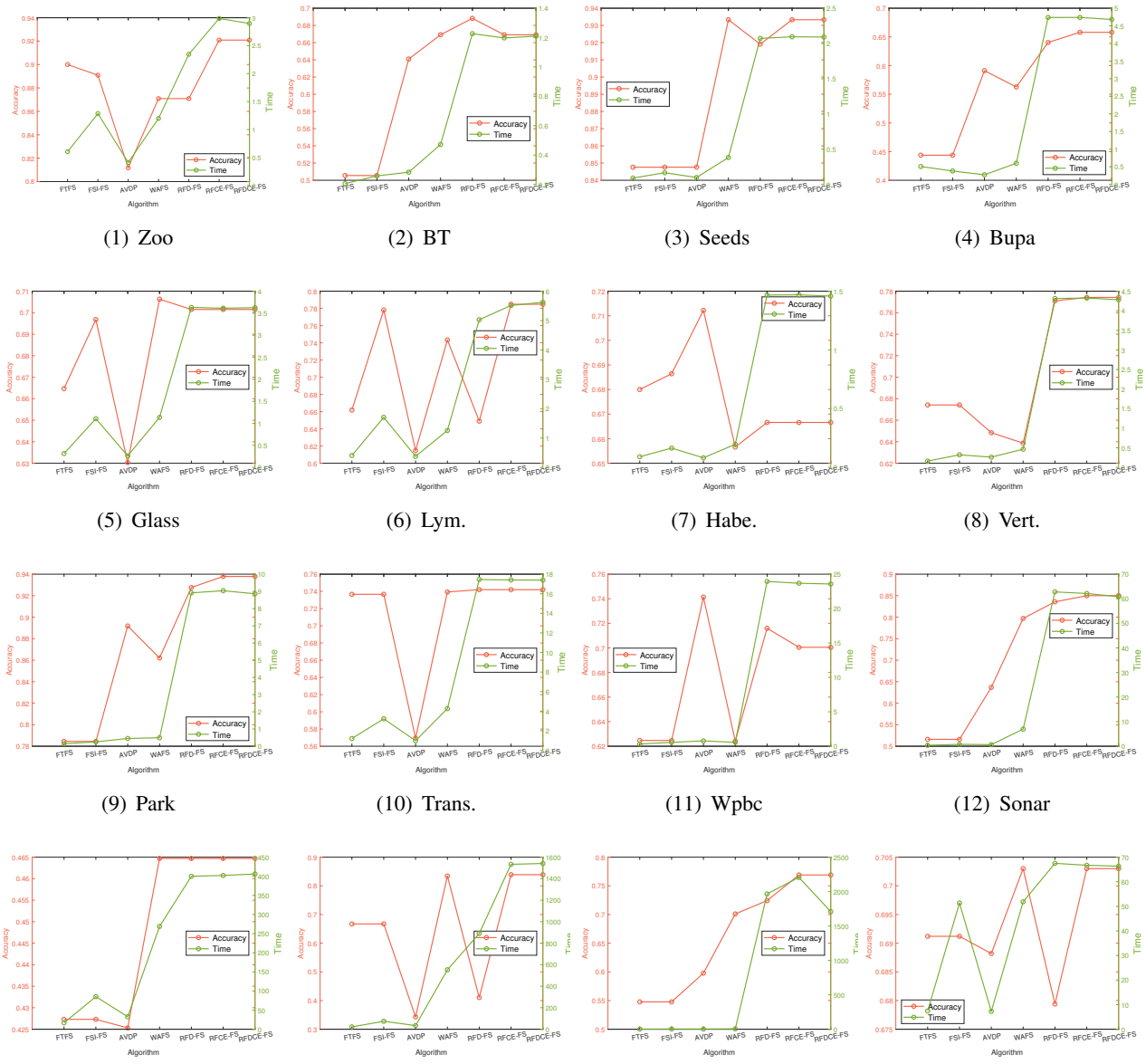
No.	Dataset	FTFS	FSI-FS	AVDP	WAFS	RFD-FS	RFCE-FS	RFDCE-FS
(1)	Zoo	0.9000±0.0816	0.8909±0.0877	0.8118±0.0737	0.8709±0.0952	0.8709±0.0827	**0.9209±0.0787**	**0.9209±0.0787**
(2)	BT	0.5055±0.1163	0.5055±0.1163	0.6409±0.1409	0.6691±0.1127	**0.6882±0.1214**	0.6691±0.1127	0.6691±0.1127
(3)	Seeds	0.8476±0.0586	0.8476±0.0586	0.8476±0.0586	**0.9333±0.0512**	0.9191±0.0505	**0.9333±0.0512**	**0.9333±0.0512**
(4)	Bupa	0.4437±0.0557	0.4437±0.0557	0.5911±0.0785	0.5627±0.0864	0.6403±0.0629	**0.6580±0.0709**	**0.6580±0.0709**
(5)	Glass	0.6647±0.0749	0.6968±0.0862	0.6303±0.1165	**0.7063±0.0818**	0.7015±0.0885	0.7015±0.0885	0.7015±0.0885
(6)	Lym.	0.6619±0.1553	0.7781±0.1035	0.6148±0.0773	0.7433±0.1247	0.6490±0.1208	**0.7848±0.0809**	**0.7848±0.0809**
(7)	Habe.	0.6800±0.0635	0.6864±0.0590	**0.7122±0.0656**	0.6566±0.0688	0.6666±0.0731	0.6666±0.0731	0.6666±0.0731
(8)	Verte.	0.6742±0.0721	0.6742±0.0721	0.6484±0.0514	0.6387±0.1481	0.7710±0.0811	**0.7742±0.0761**	**0.7742±0.0761**
(9)	Park.	0.7845±0.0919	0.7845±0.0919	0.8918±0.0826	0.8621±0.0759	0.9276±0.0497	**0.9379±0.0529**	**0.9379±0.0529**
(10)	Trans.	0.7366±0.0272	0.7366±0.0272	0.5694±0.0724	0.7393±0.0262	**0.7420±0.0310**	**0.7420±0.0310**	**0.7420±0.0310**
(11)	Wpbc	0.6247±0.0983	0.6247±0.0983	**0.7413±0.0847**	0.6242±0.0809	0.7160±0.0705	0.7005±0.0822	0.7005±0.0822
(12)	Sonar	0.5160±0.1275	0.5160±0.1275	0.6369±0.1022	0.7971±0.0439	0.8357±0.0569	0.8500±0.0667	**0.8502±0.0528**
(13)	Cmc	0.4273±0.0023	0.4273±0.0023	0.4253±0.0123	**0.4647±0.0356**	**0.4647±0.0356**	**0.4647±0.0356**	**0.4647±0.0356**
(14)	Biodeg	0.6672±0.0319	0.6672±0.0319	0.3431±0.0133	0.8341±0.0262	0.4105±0.0327	**0.8389±0.0221**	**0.8389±0.0221**
(15)	Darwin	0.5477±0.1009	0.5477±0.1009	0.5977±0.1131	0.7010±0.1027	0.7242±0.0806	**0.7690±0.0978**	**0.7690±0.0978**
(16)	Leaf	0.6912±0.0710	0.6912±0.0710	0.6882±0.0710	**0.7030±0.0562**	0.6794±0.0671	**0.7030±0.0562**	**0.7030±0.0562**
Average		0.6483±0.0768	0.6574±0.0744	0.6494±0.0759	0.7192±0.0760	0.7129±0.0691	*0.7572±0.0673*	**0.7572±0.0664**
Time		0	0	2	4	3	*11*	**12**

Table 11. SVM classification accuracies of all selection algorithms.

No.	Dataset	FTFS	FSI-FS	AVDP	WAFS	RFD-FS	RFCE-FS	RFDCE-FS
(1)	Zoo	**0.9200±0.1135**	0.8909±0.0877	0.8118±0.0737	0.8518±0.0840	0.8118±0.0567	0.8709±0.1063	0.8909±0.0740
(2)	BT	0.4500±0.0906	0.4500±0.0906	0.5927±0.1148	**0.5928±0.0975**	**0.5928±0.0975**	**0.5928±0.0975**	**0.5928±0.0975**
(3)	Seeds	0.8619±0.0726	0.8619±0.0726	0.8619±0.0726	**0.9429±0.0492**	0.9238±0.0559	**0.9429±0.0492**	**0.9429±0.0492**
(4)	Bupa	0.5798±0.0089	0.5798±0.0089	0.5798±0.0089	0.5798±0.0089	**0.5944±0.0352**	0.5856±0.0282	0.5856±0.0282
(5)	Glass	0.6078±0.0788	0.6355±0.0779	0.5699±0.0586	0.6214±0.0766	**0.6403±0.0782**	**0.6403±0.0782**	**0.6403±0.0782**
(6)	Lym.	0.7357±0.1267	**0.8124±0.1163**	0.7305±0.1023	0.7433±0.0981	0.7038±0.1072	0.7990±0.1291	0.7990±0.1291
(7)	Habe.	**0.7351±0.0130**	**0.7351±0.0130**	0.7319±0.0150	0.7319±0.0150	**0.7351±0.0130**	**0.7351±0.0130**	**0.7351±0.0130**
(8)	Verte.	0.6774±0.0000	0.6774±0.0000	0.6774±0.0000	0.7226±0.0612	**0.8097±0.0386**	0.8065±0.0402	0.8065±0.0402
(9)	Park.	0.8250±0.0511	0.8250±0.0511	0.8508±0.0619	0.8150±0.0506	**0.8711±0.0458**	**0.8711±0.0458**	**0.8711±0.0458**
(10)	Trans.	**0.7607±0.0065**	**0.7607±0.0065**	**0.7607±0.0065**	**0.7607±0.0065**	**0.7607±0.0065**	**0.7607±0.0065**	**0.7607±0.0065**
(11)	Wpbc	0.7632±0.0232	0.7632±0.0232	0.7632±0.0232	0.7632±0.0232	**0.7892±0.0542**	0.7839±0.0509	0.7839±0.0509
(12)	Sonar	0.5874±0.1309	0.5874±0.1309	0.6617±0.1210	0.8307±0.0667	0.8795±0.0603	0.9088±0.0612	**0.9133±0.0489**
(13)	Cmc	0.4267±0.0121	0.4267±0.0121	0.4341±0.0383	**0.5319±0.0412**	**0.5319±0.0412**	**0.5319±0.0412**	**0.5319±0.0412**
(14)	Biodeg	0.6900±0.0223	0.6900±0.0223	0.6626±0.0037	0.8303±0.0414	0.6626±0.0037	**0.8759±0.0209**	**0.8759±0.0209**
(15)	Darwin	0.5118±0.0206	0.5118±0.0206	0.6333±0.1697	0.5578±0.0492	**0.7980±0.0914**	0.7765±0.0991	0.7765±0.0991
(16)	Leaf	**0.6647±0.0575**	0.6618±0.0576	0.6471±0.0500	0.6618±0.0608	0.6500±0.0740	**0.6647±0.0558**	**0.6647±0.0558**
Average		0.6748±0.0518	0.6794±0.0495	0.6856±0.0575	0.7211±0.0519	0.7347±0.0537	*0.7592±0.0577*	**0.7607±0.0549**
Time		4	3	1	4	**10**	*9*	**10**

Furthermore, Figures 3 and 4 adopt the algorithms as benchmarks and employ dual y-axes to simultaneously present classification accuracy and execution time. Specifically, the horizontal axis lists the seven feature selection algorithms (FTFS, FSI-FS, AVDP, WAFS, RFD-FS, RFCE-FS, RFDCE-FS). The left vertical axis denotes classification accuracy, with the "Accuracy" curve illustrating the predictive performance of the classifier after feature selection. The right vertical axis denotes computational time, and the "Time" curve characterizes the computational overhead of each algorithm. This measurement accumulates the runtime consumed in two successive phases: the feature selection phase and the classification phase, as recorded in Table 9. By jointly analyzing the accuracy-time curves, the practical applicability and trade-offs of different algorithms on a given dataset can be clearly evaluated.

Finally, Figures 3 and 4 extend this analysis to multiple datasets in a dataset-by-dataset manner. Both figures employ visual curve comparisons to systematically examine the classification performance and time efficiency of the various feature selection algorithms, thereby providing an intuitive and comprehensive comparison across different experimental scenarios.



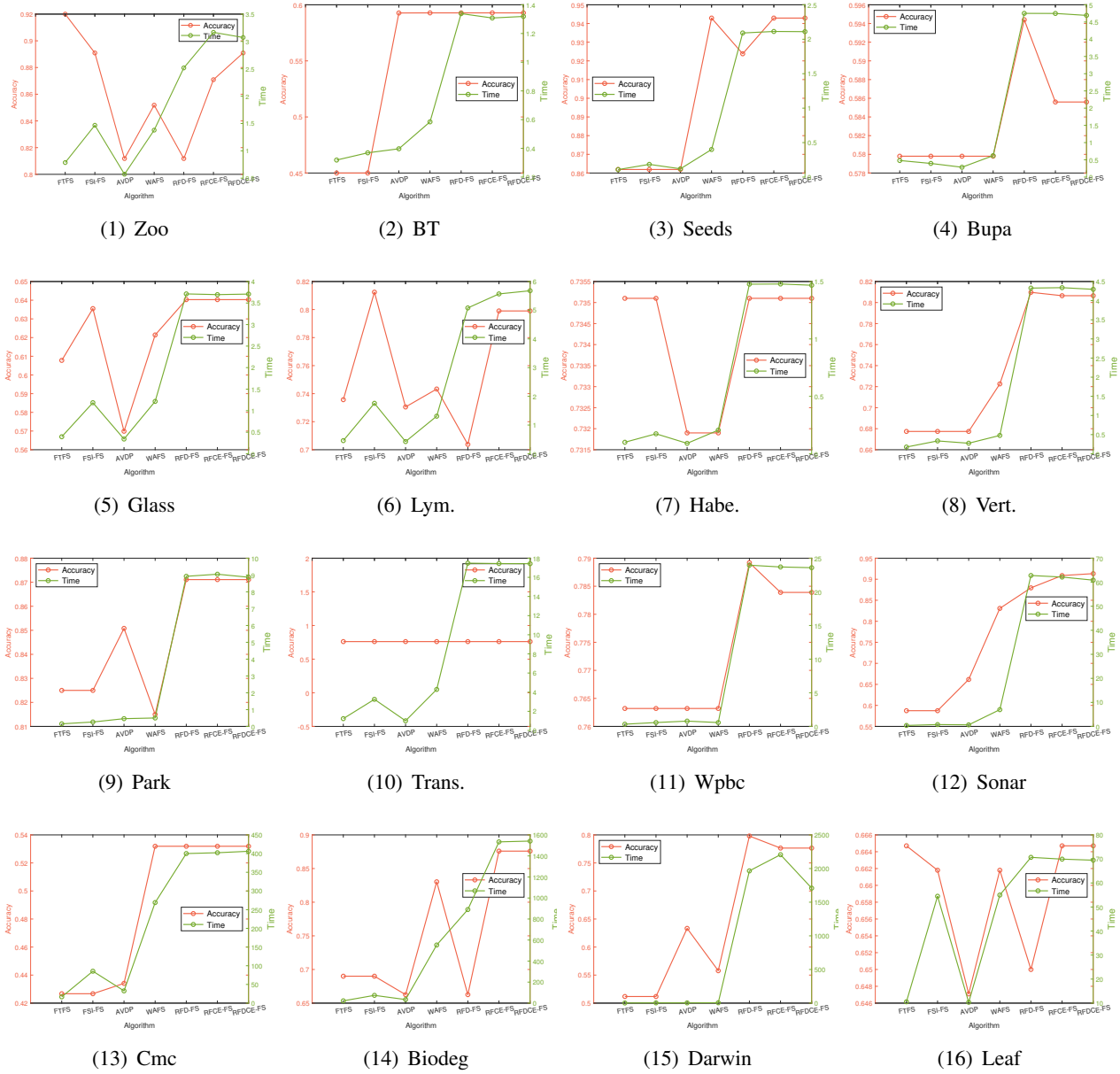
(1) Zoo (2) BT (3) Seeds (4) Bupa

(5) Glass (6) Lym. (7) Habe. (8) Vert.

(9) Park (10) Trans. (11) Wpbc (12) Sonar

(13) Cmc (14) Biodeg (15) Darwin (16) Leaf

Figure 3. KNN classification accuracy and running time curves of algorithms on 16 UCI datasets.



(1) Zoo (2) BT (3) Seeds (4) Bupa

(5) Glass (6) Lym. (7) Habe. (8) Vert.

(9) Park (10) Trans. (11) Wpbc (12) Sonar

(13) Cmc (14) Biodeg (15) Darwin (16) Leaf

Figure 4. SVM classification accuracy and running time curves of algorithms on 16 UCI datasets.

To comprehensively assess experimental outcomes across 16 datasets, we adopt arithmetic mean values for statistical analysis. The final-row "Average" entries in Tables 10 and 11 summarize each algorithm's average classification accuracy over all datasets to intuitively compare overall performance. As shown in Table 10, RFDCE-FS obtains the best average accuracy of 0.7572 ± 0.0664 and outperforms all competitors, which offers quantitative evidence for its superiority. Besides, the bottom-row "Time" tallies how many datasets that each method attains the highest classification accuracy. In certain cases, several algorithms yield identical top-level accuracy. We treat such outcomes favorably: They demonstrate stable prediction capacity for competing approaches under corresponding data distributions, while our method can consistently reach the optimal performance.

Taken together, average accuracy and optimal-dataset counts serve as complementary measurements to deliver trustworthy overall evaluation for different feature selection algorithms.

All statistical classification accuracies reported in Tables 9–12 — with particular emphasis on Table 12 — serve as the basis for a systematic comparison and analysis of the seven feature selection algorithms under the two classifiers. To clearly express the relative performance relationships among algorithms, the symbols $\succ$ and $\succeq$ are introduced to denote strict superiority and non-inferiority (or superiority with equality) in terms of classification accuracy, respectively.

(1) Tables 10 and 12 summarize the average classification accuracies obtained using the KNN classifier. For the overall statistics across the 16 datasets, the optimal and suboptimal results are highlighted using bold and italic fonts, respectively. Accordingly, the seven algorithms yield the following average accuracy rankings:

$$\begin{aligned} &[0.6483 \pm 0.0768, 0.6574 \pm 0.0744, 0.6494 \pm 0.0759, \\ &0.7192 \pm 0.0760, 0.7129 \pm 0.0691, \textit{0.7572} \pm \textit{0.0673}, \mathbf{0.7572} \pm \mathbf{0.0664}], \\ &RFDCE\text{-}FS \succeq RFCE\text{-}FS \succ WAFS \succ RFD\text{-}FS \succ FSI\text{-}FS \succ AVDP \succ FTFS. \end{aligned} \tag{5.3}$$

Among them, RFDCE-FS achieves the best overall performance; RFCE-FS achieves an identical average accuracy but with a slightly larger standard deviation, attaining the second-best ranking based on variance.

(2) With respect to the SVM classifier, the average results over the 16 datasets reported in Tables 11 and 12 exhibit

$$\begin{aligned} &[0.6748 \pm 0.0518, 0.6794 \pm 0.0495, 0.6856 \pm 0.0575, \\ &0.7211 \pm 0.0519, 0.7347 \pm 0.0537, \textit{0.7592} \pm \textit{0.0577}, \mathbf{0.7607} \pm \mathbf{0.0549}], \\ &RFDCE\text{-}FS \succ RFCE\text{-}FS \succ RFD\text{-}FS \succ WAFS \succ AVDP \succ FSI\text{-}FS \succ FTFS. \end{aligned} \tag{5.4}$$

By comparison, RFDCE-FS achieves the best overall performance while RFCE-FS yields the second-best result correspondingly. Specifically, RFDCE-FS attains the maximum classification accuracy on 10 of the total 16 datasets, and RFCE-FS hits the optimal accuracy measurement across nine datasets.

Based on these observations, Eqs (5.3) and (5.4) respectively summarize the performance orderings of the feature selection algorithms under the KNN and SVM classifiers. From these statistical rankings, RFDCE-FS has the overall best performance, and RFCE-FS and RFD-FS also demonstrate superior performance under different classifiers. In contrast, the comparative algorithms FTFS, FSI-FS, AVDP, and WAFS consistently rank lower. In essence, the first three algorithms are newly proposed approaches, whereas the latter four serve as baseline methods with relatively weaker performance.

Table 12. Average classification accuracies of all selection algorithms.

No.	Dataset	FTFS	FSI-FS	AVDP	WAFS	RFD-FS	RFCE-FS	RFDCE-FS
KNN	Average	0.6483±0.0768	0.6574±0.0744	0.6494±0.0759	0.7192±0.0760	0.7129±0.0691	*0.7572±0.0673*	**0.7572±0.0664**
	Time	0	0	2	4	3	*11*	**12**
SVM	Average	0.6748±0.0518	0.6794±0.0495	0.6856±0.0575	0.7211±0.0519	0.7347±0.0537	*0.7592±0.0577*	**0.7607±0.0549**
	Time	4	3	1	4	**10**	*9*	**10**

We conduct a joint assessment covering two core evaluation metrics: the classification accuracy and the number of times each algorithm can attain the optimal performance. The experimental

assessment uncovers an obvious trade-off between the two evaluation indicators. Certain feature selection algorithms can deliver outstanding average classification accuracy yet seldom attain the optimal performance on individual datasets; conversely, other algorithms can frequently acquire optimal outcomes but suffer from relatively low average accuracy values. Among all the compared methods, RFDCE-FS achieves a favorable balance between the two metrics. It maintains a competitive average classification accuracy and attains the optimal performance 12 times under the KNN classification model and 10 times under the SVM classification model, respectively. On the contrary, FTFS and FSI-FS can barely reach the optimal performance benchmark, accompanied by inferior classification accuracy and weak overall stability. The above analytical results offer instructive references for the practical selection of feature selection algorithms. For real-world application scenarios that simultaneously demand stable prediction performance and satisfactory classification accuracy, researchers can refer to the above analytical conclusions to select the appropriate feature selection approach.

In summary, the tables offer a comprehensive and quantitative evaluation of feature selection algorithms by jointly considering multiple datasets, multiple competing methods, and the dual dimensions of classification accuracy and optimal-performance frequency, thereby supplying solid empirical support for informed algorithm selection.

5.2.2. Statistical analysis

Finally, we focus on statistical hypothesis testing to evaluate the statistical significance of performance differences among different feature selection algorithms. We comprehensively analyze classification accuracies obtained under both the KNN and SVM classifiers to deliver more convincing comparative conclusions, as the two distinct classifiers can jointly reflect the generalizability and core performance of feature selection approaches. Based on the experimental outcomes, we summarize the classification accuracies of the seven competing algorithms across all 16 datasets for KNN and SVM separately in Tables 10 and 11, respectively.

To verify whether the observed disparities in average classification accuracy possess statistical significance, we first carry out the Friedman test [43], followed by the Nemenyi test [44] for the results of each classifier. Concretely, the classification accuracies of the seven algorithms across the 16 datasets under KNN and SVM are separately transformed into rank values. The ranking results derived from the two accuracy tables are reported in Tables 13 and 14. During ranking, all algorithms with identical classification accuracy are allocated the same tied rank. The formula of the Friedman test statistic is defined as follows:

$$\tau_{\chi^2} = \frac{12n}{k(k+1)}\left(\sum_{i=1}^{k} r_i^2 - \frac{k(k+1)^2}{4}\right),\ \tau_F = \frac{(n-1)\tau_{\chi^2}}{n(k-1)-\tau_{\chi^2}}, \tag{5.5}$$

where k denotes the number of feature selection algorithms, n represents the number of datasets, and r_i refers to the average rank of the i-th algorithm. The test statistic F_α obeys an F-distribution with degrees of freedom $k-1$ and $(k-1)(n-1)$. In our experiments, the significance level is set to $\alpha = 0.1$, and the corresponding critical value is $F_\alpha(6, 90) = 1.8406$. The test statistics derived from the Friedman test under the KNN and SVM classification models are computed as $\tau_F = 10.5331$ and $\tau_F = 8.2200$, respectively. The two obtained statistic values are both well above the precomputed critical threshold, which demonstrates that there exists a statistically significant discrepancy among the feature selection

approaches involved in the comparative experiments. Consequently, the Nemenyi test is further applied to identify pairwise performance differences.

Table 13. Accuracy rank of seven algorithms on the KNN classifier.

No.	Dataset	FTFS	FSI-FS	AVDP	WAFS	RFD-FS	RFCE-FS	RFDCE-FS
(1)	Zoo	3	4	7	5.5	5.5	1.5	1.5
(2)	BT	6.5	6.5	5	3	1	3	3
(3)	Seeds	6	6	6	2	4	2	2
(4)	Bupa	6.5	6.5	4	5	3	1.5	1.5
(5)	Glass	6	5	7	1	3	3	3
(6)	Lym.	5	3	7	4	6	1.5	1.5
(7)	Habe.	3	2	1	7	5	5	5
(8)	Verte.	4.5	4.5	6	7	3	1.5	1.5
(9)	Park.	6.5	6.5	4	5	3	1.5	1.5
(10)	Trans.	5.5	5.5	7	4	2	2	2
(11)	Wpbc	5.5	5.5	1	7	2	3.5	3.5
(12)	Sonar	6.5	6.5	5	4	3	2	1
(13)	Cmc	5.5	5.5	7	2.5	2.5	2.5	2.5
(14)	Biodeg	4.5	4.5	7	3	6	1.5	1.5
(15)	Darwin	6.5	6.5	5	4	3	1.5	1.5
(16)	Leaf	4.5	4.5	6	2	7	2	2
Average		5.3	5.2	5.3	4.1	3.7	2.2	2.2

Table 14. Accuracy rank of seven algorithms on the SVM classifier.

No.	Dataset	FTFS	FSI-FS	AVDP	WAFS	RFD-FS	RFCE-FS	RFDCE-FS
(1)	Zoo	1	2.5	6.5	5	6.5	4	2.5
(2)	BT	6.5	6.5	5	2.5	2.5	2.5	2.5
(3)	Seeds	6	6	6	2	4	2	2
(4)	Bupa	5.5	5.5	5.5	5.5	1	2.5	2.5
(5)	Glass	6	4	7	5	2	2	2
(6)	Lym.	5	1	6	4	7	2.5	2.5
(7)	Habe.	3	3	6.5	6.5	3	3	3
(8)	Verte.	6	6	6	4	1	2.5	2.5
(9)	Park.	5.5	5.5	4	7	2	2	2
(10)	Trans.	4	4	4	4	4	4	4
(11)	Wpbc	5.5	5.5	5.5	5.5	1	2.5	2.5
(12)	Sonar	6.5	6.5	5	4	3	2	1
(13)	Cmc	6.5	6.5	5	2.5	2.5	2.5	2.5
(14)	Biodeg	4.5	4.5	6.5	3	6.5	1.5	1.5
(15)	Darwin	6.5	6.5	4	5	1	2.5	2.5
(16)	Leaf	2	4.5	7	4.5	6	2	2
Average		5.0	4.9	5.6	4.4	3.3	2.5	2.3

Based on the statistical guidelines presented in Ref. [43], the Nemenyi test adopts a critical value of $q_{0.1} = 2.693$ at the 0.1 significance level, with the corresponding critical difference (CD) calculated as $CD_{0.1} = 2.0568$. For intuitive observation of pairwise performance discrepancies among all candidate algorithms, we visualize the Nemenyi test results via a CD diagram in Figure 5. In this schematic diagram, all compared algorithms are arranged on a number line according to their overall average ranking. Notably, any pair of algorithms linked by a continuous horizontal line has no statistically significant performance difference. As illustrated in Figure 5, RFDCE-FS, RFCE-FS, and RFD-FS achieve a statistically significant performance advantage over the four baseline competitors.

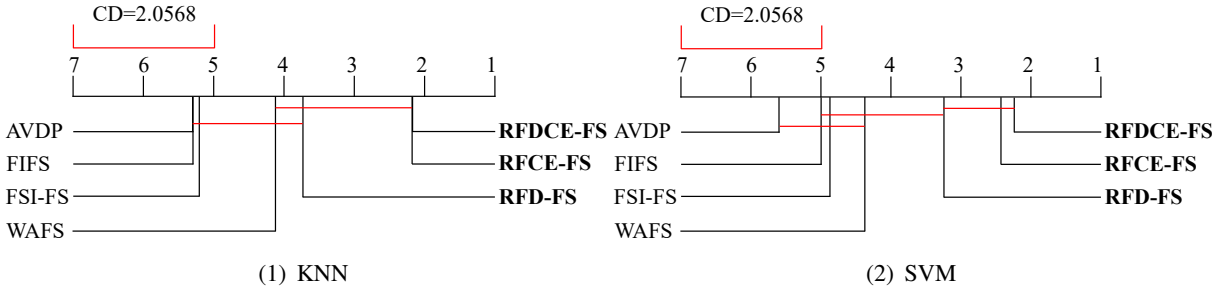


Figure 5. Nemenyi test figures of classification accuracy on two classifiers.

Table 15. Wilcoxon signed-rank test: between RFDCE-FS and three baseline algorithms (KNN).

No.	Dataset	RFDCE-FS vs FTFS				RFDCE-FS vs FSI-FS				RFDCE-FS vs WAFS			
		RFDCE-FS	FTFS	Sign	Abs Rank	RFDCE-FS	FSI-FS	Sign	Abs Rank	RFDCE-FS	WAFS	Sign	Abs Rank
(1)	Zoo	0.9209	0.9000	+	0.0209 4	0.9209	0.8909	+	0.0300 6	0.9209	0.8709	+	0.0500 6
(2)	BT	0.6691	0.5055	+	0.1636 12	0.6691	0.5055	+	0.1636 12	0.6691	0.6691	0	0.0000 –
(3)	Seeds	0.9333	0.8476	+	0.0857 8	0.9333	0.8476	+	0.0857 9	0.9333	0.9333	0	0.0000 –
(4)	Bupa	0.6580	0.4437	+	0.2143 14	0.6580	0.4437	+	0.2143 14	0.6580	0.5627	+	0.0953 11
(5)	Glass	0.7015	0.6647	+	0.0368 5	0.7015	0.6968	+	0.0047 1	0.7015	0.7063	-	0.0048 2.5
(6)	Lym.	0.7848	0.6619	+	0.1229 10	0.7848	0.7781	+	0.0067 3	0.7848	0.7433	+	0.0415 5
(7)	Habe.	0.6666	0.6800	-	0.0134 3	0.6666	0.6864	-	0.0198 5	0.6666	0.6566	+	0.0100 4
(8)	Verte.	0.7742	0.6742	+	0.1000 9	0.7742	0.6742	+	0.1000 10	0.7742	0.6387	+	0.1355 12
(9)	Park.	0.9379	0.7845	+	0.1534 11	0.9379	0.7845	+	0.1534 11	0.9379	0.8621	+	0.0758 9
(10)	Trans.	0.7420	0.7366	+	0.0054 1	0.7420	0.7366	+	0.0054 2	0.7420	0.7393	+	0.0027 1
(11)	Wpbc	0.7005	0.6247	+	0.0758 7	0.7005	0.6247	+	0.0758 8	0.7005	0.6242	+	0.0763 10
(12)	Sonar	0.8502	0.5160	+	0.3342 16	0.8502	0.5160	+	0.3342 16	0.8502	0.7971	+	0.0531 7
(13)	Cmc	0.4647	0.4273	+	0.0374 6	0.4647	0.4273	+	0.0374 7	0.4647	0.4647	0	0.0000 –
(14)	Biodeg	0.8389	0.6672	+	0.1717 13	0.8389	0.6672	+	0.1717 13	0.8389	0.8341	+	0.0048 2.5
(15)	Darwin	0.7690	0.5477	+	0.2213 15	0.7690	0.5477	+	0.2213 15	0.7690	0.7010	+	0.0680 8
(16)	Leaf	0.7030	0.6912	+	0.0118 2	0.7030	0.6912	+	0.0118 4	0.7030	0.7030	0	0.0000 –

Table 16. Wilcoxon signed-rank test: comparison between RFDCE-FS and three baseline algorithms (SVM).

No.	Dataset	RFDCE-FS vs FTFS				RFDCE-FS vs FSI-FS				RFDCE-FS vs WAFS			
		RFDCE-FS	FTFS	Sign	Abs Rank	RFDCE-FS	FSI-FS	Sign	Abs Rank	RFDCE-FS	WAFS	Sign	Abs Rank
(1)	Zoo	0.8909	0.9200	-	0.0291 3	0.8909	0.8909	0	0.0000 –	0.8909	0.8518	+	0.0391 6
(2)	BT	0.5928	0.4500	+	0.1428 10	0.5928	0.4500	+	0.1428 10	0.5928	0.5928	0	0.0000 –
(3)	Seeds	0.9429	0.8619	+	0.0810 7	0.9429	0.8619	+	0.0810 7	0.9429	0.9429	0	0.0000 –
(4)	Bupa	0.5856	0.5798	+	0.0058 1	0.5856	0.5798	+	0.0058 3	0.5856	0.5798	+	0.0058 3
(5)	Glass	0.6403	0.6078	+	0.0325 4	0.6403	0.6355	+	0.0048 2	0.6403	0.6214	+	0.0189 4
(6)	Lym.	0.7990	0.7357	+	0.0633 6	0.7990	0.8124	-	0.0134 4	0.7990	0.7433	+	0.0557 8
(7)	Habe.	0.7351	0.7351	0	0.0000 –	0.7351	0.7351	0	0.0000 –	0.7351	0.7319	+	0.0032 2
(8)	Verte.	0.8065	0.6774	+	0.1291 9	0.8065	0.6774	+	0.1291 9	0.8065	0.7226	+	0.0839 11
(9)	Park.	0.8711	0.8250	+	0.0461 5	0.8711	0.8250	+	0.0461 6	0.8711	0.8150	+	0.0561 9
(10)	Trans.	0.7607	0.7607	0	0.0000 –	0.7607	0.7607	0	0.0000 –	0.7607	0.7607	0	0.0000 –
(11)	Wpbc	0.7839	0.7632	+	0.0207 2	0.7839	0.7632	+	0.0207 5	0.7839	0.7632	+	0.0207 5
(12)	Sonar	0.9133	0.5874	+	0.3259 13	0.9133	0.5874	+	0.3259 13	0.9133	0.8307	+	0.0826 10
(13)	Cmc	0.5319	0.4267	+	0.1052 8	0.5319	0.4267	+	0.1052 8	0.5319	0.5319	0	0.0000 –
(14)	Biodeg	0.8759	0.6900	+	0.1859 11	0.8759	0.6900	+	0.1859 11	0.8759	0.8303	+	0.0456 7
(15)	Darwin	0.7765	0.5118	+	0.2647 12	0.7765	0.5118	+	0.2647 12	0.7765	0.5578	+	0.2187 12
(16)	Leaf	0.6647	0.6647	0	0.0000 –	0.6647	0.6618	+	0.0029 1	0.6647	0.6618	+	0.0029 1

The foregoing statistical results globally verify the superiority of our proposed RFDCE-FS algorithm. To further quantify pairwise performance discrepancies, we adopt the Wilcoxon signed-rank test [45] for pairwise comparisons. We implement pairwise tests between RFDCE-FS and three competing approaches (FTFS, FSI-FS, WAFS) under two classifiers, namely KNN and SVM. Complete rank statistics of all paired comparisons under KNN and SVM are separately summarized in Tables 15 and 16. Among them, the notation "Abs" denotes the absolute value of the difference, whereas "+", "−", and "0" correspond to positive discrepancies, negative discrepancies, and equivalent results, respectively. Given the critical threshold $W_{(16,0.05)} = 29$, we can quantitatively evaluate the statistical significance of performance gaps between every pair of algorithms. Taking the results obtained under the KNN classifier as a typical example, we elaborate on the comparison between RFDCE-FS and FTFS based on the left section of Table 15. After computing the positive and negative rank sums $W_+ = 4 + 12 + 8 + 14 + 5 + 10 + 9 + 11 + 1 + 7 + 16 + 6 + 13 + 15 + 2 = 133$ and $W_- = 3$, we derive the test statistic satisfying $W = \min\{W_+, W_-\} = 3 < W_{(16,0.05)} = 29$. This outcome demonstrates that RFDCE-FS achieves statistically significant performance improvements over FTFS.

6. Conclusions

In this paper, a RWFRSs framework is systematically investigated by integrating distance optimization and uncertainty measurement refinement. The related research ideas and contents are presented in Figure 1. Within this two-dimensional improvement paradigm, three fuzzy weighted uncertainty measures, namely RFD, RFCE, and RFDCE, are formally defined to characterize the

deterministic and uncertain components of knowledge in decision systems. Their theoretical properties, especially granulation monotonicity with respect to attribute subsets, are rigorously analyzed and verified through both mathematical reasoning and experimental validation. Based on the proposed uncertainty measures, corresponding heuristic feature selection algorithms, including RFD-FS, RFCE-FS, and RFDCE-FS, are constructed. These algorithms inherit the monotonic properties of the underlying measures and employ forward heuristic search strategies to efficiently identify informative attribute subsets. Compared with existing methods, the proposed algorithms achieve a more balanced trade-off between uncertainty reduction and feature dependency preservation, enabling more accurate and stable feature evaluation. Extensive experiments conducted on multiple UCI benchmark datasets demonstrate the effectiveness and robustness of the proposed approaches. Experimental results confirm that the fuzzy weighted uncertainty measures exhibit consistent monotonic trends along attribute chains, validating their theoretical properties. Moreover, the proposed feature selection algorithms generally achieve higher classification accuracy with competitive computational efficiency. In particular, RFCE-FS and RFDCE-FS consistently outperform the competing methods in terms of average accuracy and optimal performance frequency. To further ensure the reliability of the experimental conclusions, statistical hypothesis testing is performed based on the Friedman test, Nemenyi post-hoc test, and Wilcoxon signed-rank test. The results indicate that the performance improvements achieved by the proposed algorithms are statistically significant, rather than due to random fluctuations across datasets. This study provides strong statistical support for the superiority of the new uncertainty measures and their induced feature selection strategies.

In summary, this study establishes a unified fuzzy weighted uncertainty measurement framework and corresponding feature selection algorithms with solid theoretical foundations and strong empirical performance. The proposed methods not only enrich the uncertainty modeling mechanisms in fuzzy rough set theory but also offer effective and reliable tools for feature selection in real-world data analysis tasks. Recent advances in granular-ball rough sets, including cost-sensitive three-way granular-ball generation and shadowed granular-ball classifiers, show high efficiency and robustness, providing valuable references for optimizing our framework [46, 47]. Future work will extend the proposed method to large-scale datasets by incorporating granular-ball computing mechanisms and further explore its applicability in complex scenarios including dynamic data environments and multi-label decision systems.

Author contributions

Hongyuan Gou: Conceptualization, Methodology, Formal analysis; Meng Chen: Software, Visualization; Benwei Chen: Validation, Investigation.

Use of Generative-AI tools declaration

The authors declare that they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgements

This work was supported by the Key Laboratory of Digital Innovation of Tianfu Culture, Sichuan Provincial Department of Culture and Tourism, Chengdu, China (TFWH-2024-51), Fundamental Research Funds of the Chengdu University (2081925074), Fundamental Research Funds of the China West Normal University (493145).

Conflict of interest

The authors declare there are no conflicts of interest.

References

1. J. H. Wan, H. M. Chen, T. R. Li, B. B. Sang, Z. Yuan, Feature grouping and selection with graph theory in robust fuzzy rough approximation space, *IEEE Trans. Fuzzy Syst.*, **31** (2023), 213–225. https://doi.org/10.1109/TFUZZ.2022.3185285
2. Z. Yuan, B. Y. Chen, J. Liu, H. M. Chen, D. Z. Peng, P. L. Li, Anomaly detection based on weighted fuzzy-rough density, *Appl. Soft Comput.*, **134** (2023), 109995. https://doi.org/10.1016/j.asoc.2023.109995
3. B. Yu, Z. Zheng, M. Cai, W. Pedrycz, W. Ding, FRCM: a fuzzy rough c-means clustering method, *Fuzzy Sets Syst.*, **480** (2024), 108860. https://doi.org/10.1016/j.fss.2024.108860
4. W. C. Yu, W. Yao, Logical distance-based fuzzy rough set model and its application in feature selection, *Fuzzy Sets Syst.*, **519** (2025), 109543. https://doi.org/10.1016/j.fss.2025.109543
5. N. Gehlot, A. Jena, A. Vijayvargiya, R. Kumar, Surface electromyography based explainable artificial intelligence fusion framework for feature selection of hand gesture recognition, *Eng. Appl. Artif. Intel.*, **137** (2024), 109119. https://doi.org/10.1016/j.engappai.2024.109119
6. X. Y. Zhang, J. H. Wang, J. L. Hou, Matrix-based approximation dynamic update approach to multi-granulation neighborhood rough sets for intuitionistic fuzzy ordered datasets, *Appl. Soft Comput.*, **163** (2024), 111915. https://doi.org/10.1016/j.asoc.2024.111915
7. R. Shang, J. Song, L. Gao, M. Lu, L. Jiao, S. Xu, et al., Double-dictionary learning unsupervised feature selection cooperating with low-rank and sparsity, *Knowl.-Based Syst.*, **304** (2024), 112566. https://doi.org/10.1016/j.knosys.2024.112566
8. L. P. Deng, M. Q. Xiao, Hyperparameter recommendation via automated meta-feature selection embedded with kernel group Lasso learning, *Knowl.-Based Syst.*, **306** (2024), 112706. https://doi.org/10.1016/j.knosys.2024.112706
9. T. Bhadra, S. Mallik, A. Sohel, Z. Zhao, Unsupervised feature selection using an integrated strategy of hierarchical clustering with singular value decomposition: an integrative biomarker discovery method with application to acute myeloid leukemia, *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **19** (2022), 1354–1364. https://doi.org/10.1109/TCBB.2021.3110989
10. Y. J. Cai, H. Yang, J. Y. Zhu, F. P. Nie, Multi-subspace graph clustering joint dimensionality reduction and feature selection, *Pattern Recogn.*, **172** (2026), 112557. https://doi.org/10.1016/j.patcog.2025.112557

11. Q. Li, Z. Liu, J. Dai, Online multi-label streaming feature selection by label semantic categorization considering label structure unevenness and local label interaction, *Eur. J. Oper. Res.*, **329** (2026), 629–640. https://doi.org/10.1016/j.ejor.2025.10.037
12. J. Yang, X. Qin, G. Y. Wang, Q. H. Zhang, S. Li, D. Wu, Attribute reduction for hierarchical classification based on improved fuzzy rough set, *Inf. Sci.*, **677** (2024), 120900. https://doi.org/10.1016/j.ins.2024.120900
13. A. Osborne, A. A. Soladoye, K. O. Usani, A. I. Adekoya, O. Z. Wada, D. B. Olawade, Machine learning prediction of kangaroo mother care in Sierra Leone: a comparative study of feature selection techniques and classification algorithms, *Int. J. Med. Inform.*, **206** (2026), 106166. https://doi.org/10.1016/j.ijmedinf.2025.106166
14. N. Taheri, L. Karttunen, S. Jouttijärvi, A. Piazzi, M. Tucci, K. Miettunen, Enhancing residential load forecasting accuracy through dynamic feature selection and ensemble machine learning models: a real-world scenario in Southern Finland, *Energ. Buildings*, **349** (2025), 116589. https://doi.org/10.1016/j.enbuild.2025.116589
15. A. Saha, N. R. Pal, Group-feature (Sensor) selection with controlled redundancy using neural networks, *Neurocomputing*, **610** (2024), 128596. https://doi.org/10.1016/j.neucom.2024.128596
16. R. Karmakar, A. K. Das, S. K. Biswas, A. Mandal, A. Bhattacharya, D. Saha, et al., Intelligent breast cancer diagnosis: K-means-based oversampling and comparative feature selection for enhanced classification, *Eng. Appl. Artif. Intel.*, **163** (2026), 113109. https://doi.org/10.1016/j.engappai.2025.113109
17. D. Yadav, D. Rani, O. P. Verma, Impact of feature selection and feature engineering in prediction of cardiovascular diseases, *Comput. Biol. Med.*, **197** (2025), 111027. https://doi.org/10.1016/j.compbiomed.2025.111027
18. C. Gao, Q. Wang, Y. P. Chen, Q. Q. Pei, Z. M. Wang, A novel mixed-attribute data anomaly detection method based on granular-ball multi-kernel fuzzy rough sets, *Neurocomputing*, **656** (2025), 131486. https://doi.org/10.1016/j.neucom.2025.131486
19. J. Q. Wang, X. H. Zhang, L. L. Mao, An OWA-based multi-granulation fuzzy rough set model using Choquet integrals and its applications, *Fuzzy Sets Syst.*, **521** (2025), 109595. https://doi.org/10.1016/j.fss.2025.109595
20. X. Y. Zhang, B. W. Chen, D. Q. Miao, Systematic feature selection based on three-level improvements of fuzzy dominance three-way neighborhood rough sets, *IEEE Trans. Fuzzy Syst.*, **32** (2024), 5060–5072. https://doi.org/10.1109/TFUZZ.2024.3437367
21. W. T. Li, C. J. Deng, W. Pedrycz, W. P. Ding, X. C. Hu, C. Zhang, et al., Dominance-based feature selection approach to interval-valued ordered intuitionistic fuzzy data, *Appl. Soft Comput.*, **186** (2026), 114228. https://doi.org/10.1016/j.asoc.2025.114228
22. X. T. Zou, J. H. Dai, Fuzzy rough feature selection via stripped decision β-neighborhood set and misclassification ratio, *Fuzzy Sets Syst.*, **520** (2025), 109544. https://doi.org/10.1016/j.fss.2025.109544
23. C. Z. Wang, Y. Zhang, S. An, T. Q. Deng, Adaptive feature selection based on fuzzy rough set fusion model with class variance, *Pattern Recogn.*, **170** (2026), 112014. https://doi.org/10.1016/j.patcog.2025.112014

24. J. W. Xie, L. Q. Li, Combined variable precision fuzzy rough set and its application in medical diagnosis, *Appl. Soft Comput.*, **184** (2025), 113797. https://doi.org/10.1016/j.asoc.2025.113797

25. S. D. Xian, N. Xu, M. M. Feng, F. Y. Tu, Game pythagorean hesitant fuzzy rough set model and its application in medical decision-making problems, *Commun. Nonlinear Sci. Numer. Simul.*, **152** (2026), 109463. https://doi.org/10.1016/j.cnsns.2025.109463

26. Y. Y. Fan, X. H. Zhang, J. Liu, J. Q. Wang, Parameterized fuzzy β-covering relations-based fuzzy rough set models and their applications to three-way decision and attribute reduction, *Fuzzy Sets Syst.*, **521** (2025), 109596. https://doi.org/10.1016/j.fss.2025.109596

27. B. Xu, C. Z. Wang, S. An, Y. Huang, Feature selection driven by maximum likelihood estimation and fuzzy similarity relation learning, *Fuzzy Sets Syst.*, **527** (2026), 109673. https://doi.org/10.1016/j.fss.2025.109673

28. X. A. Ma, H. Xu, Y. Liu, J. Z. Zhang, Class-specific feature selection using fuzzy information-theoretic metrics, *Eng. Appl. Artif. Intel.*, **136** (2024), 109035. https://doi.org/10.1016/j.engappai.2024.109035

29. H. Y. Gou, X. Y. Zhang, Feature selection based on double-hierarchical and multiplication-optimal fusion measurement in fuzzy rough sets, *Inform. Sciences*, **618** (2022), 437–467. https://doi.org/10.1016/j.ins.2022.10.133

30. J. C. Xu, X. R. Meng, K. L. Qu, Y. H. Sun, Q. C. Hou, Feature selection using relative dependency complement mutual information in fitting fuzzy rough set model, *Appl. Intell.*, **53** (2023), 18239–18262. https://doi.org/10.1007/s10489-022-04445-9

31. C. Long, B. Li, J. H. Wen, A state of charge feature selection method for electric buses via fuzzy information measures, *Appl. Soft Comput.*, **186** (2026), 114190. https://doi.org/10.1016/j.asoc.2025.114190

32. Y. Yang, J. H. Wan, X. P. Li, X. L. Yang, H. M. Chen, K. C. Tan, et al., Rational linear kernelized weighted fuzzy rough attribute selection with class separability, *Fuzzy Sets Syst.*, **521** (2025), 109600. https://doi.org/10.1016/j.fss.2025.109600

33. J. H. Wan, X. P. Li, J. Zhao, M. Li, Z. X. Deng, H. M. Chen, Joint uncertainty model and metric for robust feature selection: a bi-level distribution consideration and feature evaluation approach, *Fuzzy Sets Syst.*, **523** (2026), 109615. https://doi.org/10.1016/j.fss.2025.109615

34. X. Xie, X. Y. Zhang, X. L. Yang, J. L. Yang, Noise-tolerant feature selections based on two-type weight-fuzzy granulations and three-view uncertainty measures, *IEEE Trans. Knowl. Data Eng.*, **37** (2025), 6535–6549. https://doi.org/10.1109/TKDE.2025.3604005

35. C. Z. Wang, C. Y. Wang, Y. H. Qian, Q. K. Leng, Feature selection based on weighted fuzzy rough sets, *IEEE Trans. Fuzzy Syst.*, **32** (2024), 4027–4037. https://doi.org/10.1109/TFUZZ.2024.3387571

36. X. R. Li, L. Q. Li, C. Z. Jia, Weighted multi-granularity fuzzy probabilistic rough set based on semi-overlapping function and its application in three-way decision, *Eng. Appl. Artif. Intel.*, **161** (2025), 112023. https://doi.org/10.1016/j.engappai.2025.112023

37. Q. T. Bui, T. N. Nguyen, H. S. Nguyen, B. Vo, A novel framework for handling uncertainty: intuitionistic fuzzy rough soft sets, *Inform. Sciences*, **722** (2026), 122592. https://doi.org/10.1016/j.ins.2025.122592

38. C. Z. Wang, Y. Huang, W. P. Ding, Z. H. Cao, Attribute reduction with fuzzy rough self-information measures, *Inform. Sciences*, **549** (2021), 68–86. https://doi.org/10.1016/j.ins.2020.11.021
39. W. H. Xu, K. H. Yuan, W. T. Li, W. P. Ding, An emerging fuzzy feature selection method using composite entropy-based uncertainty measure and data distribution, *IEEE Trans. Emerg. Top. Comput. Intel.*, **7** (2023), 76–88. https://doi.org/10.1109/TETCI.2022.3171784
40. D. Dua, C. Graff, UCI machine learning repository, University of California, Irvine, 2019. Available from: `http://archive.ics.uci.edu/ml`.
41. C. Z. Wang, Y. L. Qi, M. W. Shao, Q. H. Hu, D. G. Chen, Y. H. Qian, et al., A fitting model for feature selection with fuzzy rough sets, *IEEE Trans. Fuzzy Syst.*, **25** (2017), 741–753. https://doi.org/10.1109/TFUZZ.2016.2574918
42. C. Z. Wang, Y. Huang, M. W. Shao, X. D. Fan, Fuzzy rough set-based attribute reduction using distance measures, *Knowl.-Based Syst.*, **164** (2019), 205–212. https://doi.org/10.1016/j.knosys.2018.10.038
43. M. Friedman, A comparison of alternative tests of significance for the problem of *m* ranking, *Ann. Math. Statist.*, **11** (1940), 86–92. https://doi.org/10.1214/aoms/1177731944
44. J. Demsar, Statistical comparison of classifiers over multiple data sets, *J. Mach. Learn. Res.*, **7** (2006), 1–30.
45. F. Wilcoxon, Individual comparisons by ranking methods, *Biom. Bull.*, **1** (1945), 80–83.
46. J. Yang, L. Y. Xiaodiao, G. Y. Wang, W. Pedrycz, S. Y. Xia, Q. H. Zhang, et al., A robust three-way classifier with shadowed granular balls based on justifiable granularity, *IEEE Trans. Neur. Net. Lear.*, **36** (2025), 16534–16548. https://doi.org/10.1109/TNNLS.2025.3563889
47. J. Yang, F. Zhao, G. Y. Wang, W. Pedrycz, S. Y. Xia, Y. M. Liu, et al., CS3W-GBG: a cost-sensitive three-way granular-ball generation method, *IEEE Trans. Fuzzy Syst.*, **33** (2025), 3681–3694. https://doi.org/10.1109/TFUZZ.2025.3596066

AIMS Press