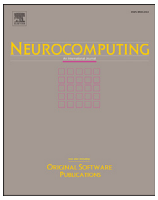
Neurocomputing 683 (2026) 133435

Contents lists available at ScienceDirect

Neurocomputing

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



# Leader-aware feature selection and fusion in interval-valued multi-source systems via entropy-weighted neighborhood rough sets

Hao Yuan, Weihua Xu*

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

ARTICLE INFO

Communicated by W. Ding

*Keywords:*
Multi-source decision-making
Feature selection
Information fusion
Leader-aware fusion
Neighborhood rough set
Entropy weighting

ABSTRACT

Effective information integration and uncertainty management are becoming increasingly difficult due to the proliferation of heterogeneous multi-source data in complex decision-making environments. Interval-valued representations are well suited to modeling data imprecision and variability, but they also introduce substantial computational and structural complexity, particularly for feature selection and information fusion across sources. To address these two closely related challenges, this paper proposes a novel leader-aware framework that synergistically integrates entropy-weighted neighborhood rough set. Our methodology is structured around three core components. First, we introduce a mixed Hausdorff distance metric to accurately quantify dissimilarities between interval-valued data, thereby enabling the construction of an adaptive neighborhood rough set model that more effectively captures complex interval structures. Second, by leveraging information entropy, we dynamically evaluate and update attribute significance based on approximation granularity. This entropy-weighted process then employs cross-source linear aggregation to derive a global feature ranking, thereby highlighting the most discriminative attributes in a principled manner. Finally, we perform leader-aware information fusion at the multi-source level. Source-specific weights are assigned according to approximation quality, and a consistency-based rule then selects the source that best matches an elected "leader," ensuring robust and interpretable fusion outcomes. Extensive experiments on nine UCI datasets with four classifiers show that the proposed method consistently outperforms alternative approaches in terms of feature selection quality, classification accuracy, and fusion effectiveness, especially at high feature retention rates.

## 1. Introduction

The Decision Information System (DIS) [1] is designed to support efficient decision-making in volatile environments by integrating heterogeneous data, embedding intelligent algorithms, and enabling visual interaction. It has been applied in organizational optimization [2], health interventions [3], and land-use zoning [4], providing a structured and transparent basis for complex decisions.

Granular Computing (GrC) [5] offers a paradigm for handling complex problems by reasoning with information granules. Under this paradigm, rough-set-based [6] and neighborhood models have been developed for pattern discovery and knowledge extraction from complex data. Subsequent studies have further enriched their theory and application [7], and demonstrated their usefulness in classification [8], prediction [9], and decision support [10]. These models have also been embedded into intelligent decision-making systems in various domains [11]. To better cope with high-dimensional and heterogeneous data, a series of neighborhood and fuzzy extensions have been proposed, such as fuzzy neighborhood multigranulation rough sets [12], the unified granular-ball learning model of Pawlak rough set and neighborhood rough set [13], heterogeneous neighborhood rough sets for mixed-type data [14], fuzziness-based three-way decision with neighborhood rough sets [15], and fuzzy multi-neighborhood rough sets combined with binary whale optimization for imbalanced data [16]. These works show that neighborhood-based rough models are effective tools for uncertainty handling and feature selection.

In real-world applications, data are typically collected from multiple sources rather than a single channel, making multi-source and heterogeneous information the norm [17]. Compared with single-source analysis, multi-source decision information systems provide more comprehensive and reliable descriptions of complex systems, thereby improving decision robustness. However, heterogeneity among sources naturally induces inconsistency and uncertainty [18], which may significantly affect decision quality in practice [19]. Interval-valued representation has

* Corresponding author.
*Email addresses:* sanjiu123456789@email.swu.edu.cn (H. Yuan), chxuwh@swu.edu.cn (W. Xu).

https://doi.org/10.1016/j.neucom.2026.133435
Received 19 January 2026; Received in revised form 22 February 2026; Accepted 20 March 2026
Available online 28 March 2026

thus been widely adopted as an effective modeling tool for incomplete information, limited measurements, and inherently fuzzy indicators. Unlike single-valued measurements such as blood pressure [20], temperature [21], or unemployment rate [22], interval-valued data explicitly capture both variability and imprecision.

Recently, multi-source interval-valued information systems have attracted increasing attention. Zhao et al. [23] proposed an optimal scheduling method for source–load uncertainty based on a hybrid robust multi-interval optimization model. Zhang et al. [24] studied dynamic information fusion for multi-source incomplete interval-valued systems. Liu et al. [25] constructed a link prediction model for interval-valued oil prices using multi-source information. Li et al. [26] developed an information fusion method for multi-source incomplete interval-valued data based on conditional entropy. These studies have advanced the modeling and fusion of multi-source interval-valued data, but most focus on improving aggregation rules or handling incompleteness under a fixed feature set and typically construct a synthetic fused source by averaging or combining all available sources. As a result, feature selection and information fusion are still treated as separate stages, and source reliability differences are often weakly modeled or ignored, which can lead to information blurring and suboptimal decisions.

In this work, a different strategy is adopted. Feature selection is first conducted over the entire multi-source interval-valued information system, and then the most representative original source is directly selected as the final output. This design aims to explicitly couple feature selection with fusion and to avoid the dilution effect caused by naive averaging of all sources. In this context, information entropy [27] is a natural tool. As a classical measure of uncertainty and information content, entropy has been widely used for attribute importance evaluation [28], feature extraction [29], and information fusion [30] in multi-source and interval-valued environments. By combining neighborhood rough sets with entropy, this paper builds an integrated framework that performs entropy-guided feature selection and leader-aware information fusion while explicitly modeling source reliability and consistency.

The framework of this paper is illustrated in Fig. 1. The main contributions are structured around the following three key innovations, which are subsequently validated through comprehensive experiments:

1. A mixed Hausdorff distance-based adaptive neighborhood rough set model for interval-valued multi-source data, which jointly measures boundary and center differences and employs a double-layer median radius to align neighborhood granularity with intra-class density.
2. An entropy-based feature weighting and cross-source aggregation mechanism, where granular conditional entropy quantifies attribute significance and source reliability, and linear fusion across sources yields a global, redundancy-reduced feature ranking.
3. A consistency-driven leader-aware fusion strategy that constructs a weighted “leader” profile and selects the original source with the highest consensus with this leader, thereby avoiding information blurring and enhancing interpretability.
4. Extensive empirical evaluation on nine UCI datasets with multiple classifiers, demonstrates statistically significant improvements over representative baselines in both feature selection and information fusion, particularly at high feature retention rates.

These three components form a coherent pipeline: the mixed Hausdorff distance and adaptive neighborhoods provide interval-aware granules; on this basis, entropy-based weighting evaluates attributes and sources; and the leader-aware fusion strategy uses these weights to select the most representative source under a consistency criterion. In this way, feature selection and information fusion are tightly integrated into a unified neighborhood–entropy framework.

The remainder of this paper is organized as follows. Section 2 formalizes interval-valued multi-source information systems, introduces the mixed Hausdorff distance, and develops an adaptive neighborhood rough set model. Section 3 presents the entropy-based feature weighting and leader-aware fusion strategy. Section 4 reports comparative experiments on nine UCI datasets. Section 5 concludes the paper and outlines future research directions.

## 2. Preliminaries

In this section, we review the concept of interval-valued multi-source information systems, and combined with the mixed Hausdorff distance,

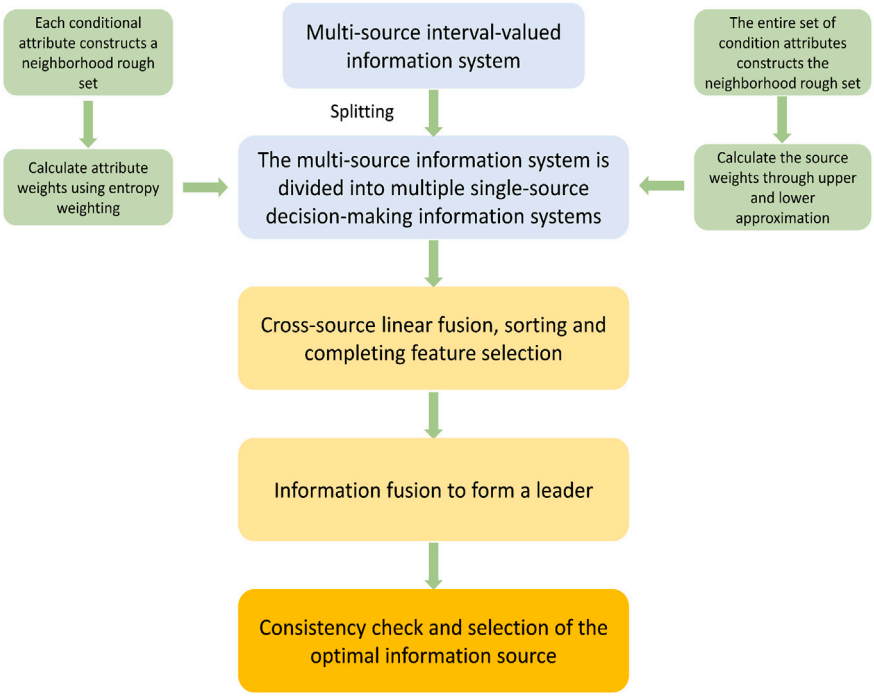



**Fig. 1.** The framework of this paper.

we construct an adaptive neighborhood rough set model for interval values.

### 2.1. Interval-valued multi-source decision information systems

Interval-valued multi-source decision information systems is a model used to handle complex decision-making problems. It integrates data from multiple information sources and represents the uncertainty of the data using interval values, thereby helping decision-makers make more accurate decisions. In a multi-source information system, each source provides different attribute values for the same set of value samples, but the class labels are consistent. Let $S = (U, A \cup \{d\})$ be a multi-source decision information system, where $U = \{x_1, x2, \dots, x_n\}$ is the universe, $A$ is a nonempty finite set of conditional attributes or features, $\{d\} = \{d_1, d2, \dots d_r\}$ is the decision attribute used to classify or make decisions about objects. For each $x \in U$, $a_i$ has s sources, each $a_i = \{a_i^1, a_i^2, \dots a_i^s\}$. Therefore, the multi-source decision information system can be defined as $S = (U, A \cup \{d\}) = \left\{U, \left\{a_i^j | i = 1, 2, \dots m; j = 1, 2, \dots, s\right\} \cup \{d\}\right\}$. When the values of each attribute in the information system are all independent intervals, it can be presented as $\left[a_i^{t-}, a_i^{t+}\right]$. For example, the value of object $x$ under the attribute $a_k$ in the t source is $\left[a_k(x)_l^t, a_k(x)_r^t\right]$, where $a_k(x)_l \le a_k(x)_r$.

### 2.2. Adaptive neighborhood rough set

Given a decision-making information system in a multi-source decision-making system with interval values, let $x$ be an object of $U$, $B$ be the subset of $A$, we quantify the comprehensive distance between two intervals using the mixed Hausdorff distance, which captures both endpoint differences and center deviations. This distance is defined as follows:

$$d_B(x, y) = \sqrt{\frac{1}{|B|} \sum_{ak \in B} d_{ak}(x, y)} \tag{1}$$

where

$$\begin{cases} d_{a_k}(x, y) = \triangle_1 + \triangle_2 \\ \triangle_1 = \frac{1}{2} \max\left(\left|a_k(x)_l - a_k(y)_l\right|, \left|a_k(x)_r - a_k(y)_r\right|\right) \\ \triangle_2 = \frac{1}{2}\left|\frac{a_k(x)_l + a_k(x)_r}{2} - \frac{a_k(y)_l + a_k(y)_r}{2}\right| \end{cases} \tag{2}$$

$\triangle_1$ describes the sensitivity of the shape between two intervals, directly focusing on the maximum difference between the boundary points of the two intervals (that is, the left endpoint and the right endpoint). $\triangle_2$ reflects the difference in the central positions of the two intervals in the numerical space, and measures the distance between the centers of the two intervals. The mixed Hausdorff distance integrates the endpoint discrepancy with the center deviation, capturing subtle boundary variations while maintaining robust overall position estimation. This dual design resists noise and outliers, significantly enhancing sensitivity and reliability for interval-valued data. In this study, we adopt equal weighting (0.5/0.5) for $\triangle_1$ and $\triangle_2$ as a principled default configuration based on the assumption of information complementarity. However, this weighting scheme is flexible and can be adjusted according to the intrinsic properties of specific datasets—for instance, assigning higher weight to $\triangle_1$ in boundary-sensitive applications or to $\triangle_2$ in noise-dominated environments—to further optimize performance for particular domain requirements.

**Property 1.** The mixed Hausdorff distance $d_{a_k}(x, y)$ satisfies the axioms of a metric:

(1) $d_{a_k}(x, y) \ge 0$, and $d_{a_k}(x, y) = 0$ if and only if $x = y$;
(2) $d_{a_k}(x, y) = d_{a_k}(y, x)$;
(3) $d_{a_k}(x, z) \le d_{a_k}(x, y) + d_{a_k}(y, z)$.

**Proof.** (1)–(2) are straightforward from the definitions of $\triangle_1$ and $\triangle_2$. We only prove (3).

Let

$$u_1 = a_k(x)_l - a_k(y)_l, \qquad v_1 = a_k(x)_r - a_k(y)_r,$$
$$u_2 = a_k(y)_l - a_k(z)_l, \qquad v_2 = a_k(y)_r - a_k(z)_r.$$

Then

$$a_k(x)_l - a_k(z)_l = u_1 + u_2,$$
$$a_k(x)_r - a_k(z)_r = v_1 + v_2.$$

Hence

$$\triangle_1(x, z) = \frac{1}{2}\max\left(|u_1 + u_2|, |v_1 + v_2|\right),$$
$$\triangle_2(x, z) = \frac{1}{4}\left|(u_1 + v_1) + (u_2 + v_2)\right|.$$

Let

$$a = \max(|u_1|, |v_1|), \qquad b = \max(|u_2|, |v_2|).$$

Then $|u_1| \le a$, $|v_1| \le a$, $|u_2| \le b$ and $|v_2| \le b$, and

$$|u_1 + u_2| \le |u_1| + |u_2| \le a + b,$$
$$|v_1 + v_2| \le |v_1| + |v_2| \le a + b.$$

Therefore

$$\max(|u_1 + u_2|, |v_1 + v_2|) \le a + b = \max(|u_1|, |v_1|) + \max(|u_2|, |v_2|).$$

Moreover, by the triangle inequality of the absolute value,

$$|u_1 + v_1 + u_2 + v_2| \le |u_1 + v_1| + |u_2 + v_2|.$$

Combining the above inequalities yields

$$\triangle_1(x, z) + \triangle_2(x, z) \le \triangle_1(x, y) + \triangle_2(x, y) + \triangle_1(y, z) + \triangle_2(y, z).$$

Hence, $d_{a_k}(x, z) \le d_{a_k}(x, y) + d_{a_k}(y, z)$. □

Compared with the classical Hausdorff distance, defined as $d_H(x, y) = \max\left(\left|a_k(x)_l - a_k(y)_l\right|, \left|a_k(x)_r - a_k(y)_r\right|\right)$, the proposed mixed Hausdorff distance $d_{mix}$ offers a more comprehensive and discriminative measure by integrating both boundary differences and central positional shifts between intervals. While the classical Hausdorff metric focuses solely on extreme endpoint deviations, it fails to account for the overall location and structural dissimilarity of intervals—often leading to counterintuitive or oversimplified distance assessments. A compelling illustration of this limitation can be observed with the following three intervals: $A = [19, 28]$, $B = [28, 37]$, $C = [46, 55]$. Using the classical Hausdorff measure, we find: $d_H(A, B) = d_H(A, C) = 9$. This suggests that intervals $A$ and $B$ are as distant as $A$ and $C$ which clearly contradicts geometric and intuitive reasoning. In reality, interval $C$ is substantially farther from $A$ than $B$ is in terms of both central position and overall displacement in the numerical space. In contrast, the mixed Hausdorff distance, yields: $d_{mix}(A, B) = 13.5, d_{mix}(A, C) = 31.5$. These results correctly reflect the actual dissimilarity: $A$ is much closer to $B$ than $C$. The key advantage of $d_{mix}$ lies in its dual sensitivity: it responds not only to boundary misalignments but also to shifts in central tendency. This makes it especially robust in the presence of noise and outliers, as it avoids overemphasizing spurious endpoint variations. Moreover, $d_{mix}$ provides a smoother and more nuanced distance profile, better capturing the intrinsic geometry of interval-valued data—a critical enhancement for applications in classification, clustering, and information fusion within uncertain or imprecise environments.

**Definition 1.** Let $Cl \in \{d\}$ be any class. For an object $x$ within the $Cl$ class, its neighborhood radius is set as the double-layer median of the distances between all objects within the $Cl$ class, which is divided into the inner-layer median and the outer-layer median, it can be defined as follows:

$$\delta(Cl) = \underset{x \in Cl}{\text{med}} \left( \underset{y \in Cl}{\text{med}} d_B(x, y) \right) \tag{3}$$

For any class $Cl$, we first compute, for each sample $x \in Cl$, the median of its distances to all samples within the same class, denoted as $med(x)$; then take the median of all $med(x)$ values, achieving the double-layer median estimation. This dual-stage robust estimation mechanism outperforms single-layer median or mean-based alternatives by effectively neutralizing the influence of both local outlier clusters and extreme sample fluctuations, ensuring the neighborhood radius remains stable and accurately reflects the typical intra-class density. Therefore, given a decision-making information system $DIS = (U, A \cup \{d\})$, for each $x \in Cl$, $B \subseteq A$, the adaptive neighborhood rough set over $B$ in class $Cl$ to relation $N$ is defined as follows:

$$N(x) = \left\{ y \in U | d_B(x, y) \leq \delta(Cl), x \in Cl \right\} \tag{4}$$

**Definition 2.** The lower and upper approximations of $Cl$ with respect to $B$ in relation to $N$ are defined as follows:

$$\begin{aligned} \underline{N(Cl)} &= \{x \in U | N(x) \subseteq Cl\} \\ \overline{N(Cl)} &= \{x \in U | N(x) \cap Cl \neq \phi\} \end{aligned} \tag{5}$$

The double-layer median method for determining neighborhood radius provides a robust and adaptive approach to data analysis. This technique operates by first calculating the median distance between each sample and its class counterparts to create a set of median values, and then taking the median of this collection as the final neighborhood radius. The double-layer median computation offers two significant advantages: it effectively eliminates outlier interference while automatically adjusting the radius based on class density. Importantly, the method demonstrates an inherent scaling property where classes with more samples naturally produce larger radius - ensuring optimal coverage where dense populations require broader radius to capture sufficient neighbors, while sparse classes maintain tighter radius to avoid noise. This intelligent self-adjustment makes the approach particularly robust for handling skewed distributions and anomalous data, significantly improving model performance across diverse datasets, and providing a reliable foundation for subsequent analysis. The method's combination of statistical rigor and adaptive intelligence makes it both scientifically sound and practically valuable for modern data processing challenges.

**Example 1.** Consider a Multi-source Medical Diagnosis System (MMDS) for early screening of diabetic retinopathy. The system evaluates a cohort of 6 patients $U = \{x_1, x_2, \ldots, x_6\}$, with each patient assessed based on five quantitative features ($a_1$: macular thickness, $a_2$: nerve fiber layer thickness, $a_3$: vascular morphology index, $a_4$: hard exudate area, $a_5$: microaneurysm count), extracted from their retinal fundus images. These assessments are performed independently by three different medical imaging centers (information sources $S_1$, $S_2$, $S_3$), leading to interval-valued measurements that encapsulate uncertainties from device variations and interpretive differences. The numerical value of each feature is normalized to the range [0, 1], indicating the severity level of the corresponding sign, where a value closer to 1 signifies a higher risk. The decision attribute $d$ represents the final clinical diagnosis (1 = diseased, 2 = healthy). Presented as Tables 1–3.

In Table 1, we calculate the distances of pairs of objects under the attribute $a_5$, the distance matrix is

$$\begin{bmatrix} 0. & 0.535 & 0.325 & 0.515 & 0.11 & 0.645 \\ 0.535 & 0. & 0.21 & 0.04 & 0.425 & 0.21 \\ 0.325 & 0.21 & 0. & 0.23 & 0.215 & 0.4 \\ 0.515 & 0.04 & 0.23 & 0. & 0.405 & 0.17 \\ 0.11 & 0.425 & 0.215 & 0.405 & 0. & 0.545 \\ 0.645 & 0.21 & 0.4 & 0.17 & 0.545 & 0. \end{bmatrix}$$

**Table 1**
Source 1.

| U | $a_1^1$ | $a_2^1$ | $a_3^1$ | $a_4^1$ | $a_5^1$ | $d$ |
|---|---|---|---|---|---|---|
| $x_1$ | [0.12,0.78] | [0.05,0.94] | [0.33,0.59] | [0.21,0.88] | [0.67,0.91] | 1 |
| $x_2$ | [0.03,0.45] | [0.27,0.99] | [0.72,0.97] | [0.15,0.63] | [0.08,0.54] | 1 |
| $x_3$ | [0.44,0.81] | [0.66,0.89] | [0.11,0.35] | [0.52,0.95] | [0.30,0.72] | 2 |
| $x_4$ | [0.76,0.99] | [0.12,0.41] | [0.58,0.84] | [0.70,0.93] | [0.13,0.47] | 2 |
| $x_5$ | [0.25,0.69] | [0.39,0.77] | [0.86,0.98] | [0.01,0.38] | [0.55,0.83] | 1 |
| $x_6$ | [0.61,0.97] | [0.53,0.92] | [0.19,0.68] | [0.34,0.79] | [0.02,0.28] | 2 |

**Table 2**
Source 2.

| U | $a_1^2$ | $a_2^2$ | $a_3^2$ | $a_4^2$ | $a_5^2$ | $d$ |
|---|---|---|---|---|---|---|
| $x_1$ | [0.08,0.63] | [0.14,0.71] | [0.41,0.88] | [0.28,0.66] | [0.73,0.96] | 1 |
| $x_2$ | [0.22,0.57] | [0.30,0.95] | [0.63,0.92] | [0.05,0.49] | [0.12,0.61] | 1 |
| $x_3$ | [0.51,0.85] | [0.59,0.81] | [0.09,0.43] | [0.48,0.90] | [0.21,0.69] | 2 |
| $x_4$ | [0.68,0.94] | [0.12,0.52] | [0.77,0.99] | [0.62,0.87] | [0.04,0.35] | 2 |
| $x_5$ | [0.18,0.74] | [0.44,0.83] | [0.90,1.00] | [0.11,0.55] | [0.60,0.89] | 1 |
| $x_6$ | [0.72,0.98] | [0.46,0.91] | [0.23,0.65] | [0.37,0.82] | [0.01,0.22] | 2 |

**Table 3**
Source 3.

| U | $a_1^3$ | $a_2^3$ | $a_3^3$ | $a_4^3$ | $a_5^3$ | $d$ |
|---|---|---|---|---|---|---|
| $x_1$ | [0.15,0.81] | [0.09,0.68] | [0.38,0.79] | [0.19,0.62] | [0.70,0.93] | 1 |
| $x_2$ | [0.29,0.62] | [0.33,0.88] | [0.67,0.95] | [0.10,0.51] | [0.15,0.58] | 1 |
| $x_3$ | [0.47,0.89] | [0.61,0.87] | [0.13,0.48] | [0.55,0.93] | [0.27,0.74] | 2 |
| $x_4$ | [0.79,0.99] | [0.07,0.46] | [0.54,0.82] | [0.66,0.91] | [0.09,0.40] | 2 |
| $x_5$ | [0.20,0.77] | [0.40,0.80] | [0.84,0.97] | [0.03,0.42] | [0.58,0.86] | 1 |
| $x_6$ | [0.65,0.96] | [0.50,0.89] | [0.17,0.61] | [0.31,0.78] | [0.00,0.25] | 2 |

Under the decision attribute, objects are divided into two classes, where $Cl_1 = \{x_1, x_2, x_5\}$, $Cl_2 = \{x_3, x_4, x_6\}$. Through the distance matrix, we can compute the neighborhood radius of each category, $\delta(Cl_1) = 0.11$, $\delta(Cl_2) = 0.17$. Therefore, we obtain $N(x_1) = \{x_1, x_5\}$, $N(x_2) = \{x_2, x_4\}$, $N(x_3) = \{x_3\}$, $N(x_4) = \{x_2, x_4\}$, $N(x_5) = \{x_1, x_5\}$, $N(x_6) = \{x_4, x_6\}$.

## 3. Entropy-based weight update feature ranking method and information fusion

In this section, the upper approximation obtained from the adaptive neighborhood rough set in the previous section is taken as the granule, which contains all objects that may belong to this class. For each attribute, first, the upper approximation constructed by this attribute is taken as the granule, and its class conditional entropy is calculated. Based on this, the attribute weights are updated. Subsequently, within the "source internal - source external" dual-loop framework, the attribute weights within the same information source are first aggregated, then cross-source fusion is conducted, and finally, the unified feature weights of the multi-source decision information system are derived.

### 3.1. Calculation of feature weights under a single information source

Given a decision-making information system $DIS = (U, A \cup \{d\})$, for each $a \in A$, $Cl \in \{d\}$, the probability of a granule of class $Cl$ under attribute $a$ can be defined as follows:

$$P_k = \frac{|G_k|}{|U|} \tag{6}$$

where

$$G_k = \overline{N(Cl)} = \{x \in U | N(x) \cap Cl \neq \phi\} \tag{7}$$

Each class is associated with one distinct granule; therefore, the number of granules is equal to the number of classes.

For each granule $G_k$ generated by attribute $a$, we tally the class frequencies of its members; the resulting empirical distribution directly

provides the conditional probability $P(Cl|G_k)$ - that is, the normalized proportion of samples falling into granule $G_k$ that belong to class $Cl$, which can be defined as follows:

$$P(Cl|G_k) = \frac{|G_k| \cap Cl}{|G_k|} \tag{8}$$

After calculating the conditional probability $P(c|G_k)$ for each class within each granule $G_k$ generated by attribute $a$, we further compute the conditional entropy $H(a)$. This step quantifies the uncertainty of the class distribution within the granules given attribute $a$.

**Definition 3.** Let $a \in A$, $Cl \in \{d\}$, the weighted conditional entropy of the approximate granule in the generated neighborhood can be defined as:

$$H(a) = -\sum_{k=1}^{m} P_k \sum_{Cl\in\{d\}} \frac{W_c}{Z} P(Cl|G_k) log_2 P(Cl|G_k) \tag{9}$$

where

$$\begin{cases} W_c = \dfrac{1}{|Cl|} \\ Z = \displaystyle\sum_{Cl\in\{d\}} W_c \end{cases} \tag{10}$$

**Property 2.** For any attribute $a$, the weighted conditional entropy $H(a)$ satisfies:

(1) **Non-negativity:** $H(a) \geq 0$, and $H(a) = 0$ if and only if each granule contains objects from a single decision class.
(2) **Monotonicity:** Let the information granules generated by attributes $a_1$ and $a_2$ be$\{G_k^{(1)}\}$ and $\{G_k^{(2)}\}$, respectively. If for all $k$,

$$\max_{Cl} P\left(Cl \mid G_k^{(1)}\right) \geq \max_{Cl} P\left(Cl \mid G_k^{(2)}\right),$$

then the discrimination ability of $a_1$ is strictly superior to that of $a_2$, and$H(a_1) \leq H(a_2)$.
(3) **Upper bound:**

$$H(a) \leq \log_2 \left|\{d\}\right|,$$

where $\left|\{d\}\right|$ is the number of decision classes.

When calculating the conditional entropy of the granule classes, we adopted a class-weighted strategy. $W_c$ represents the reciprocal of the frequency of objects belonging to the class $Cl$ within that granule. Classes with fewer objects have a greater weight, which avoids the underestimation of entropy for classes with few samples. This weighted conditional entropy condenses the internal class uncertainty of attribute $a$ into a numerical value: the lower the entropy, the higher the purity of the granule, and the stronger the discriminative power of the attribute; the higher the entropy, the more mixed the classes, and the lower the attribute value. By class-weighting to suppress the risk of small sample classes being underestimated, the entropy value becomes the sole basis for reverse weighting and feature selection in the subsequent intra-source - inter-source double loop, achieving precise amplification of key information and reducing the weights of less important features.

After calculating the conditional entropy of the granule classes under attribute $a$, we use the exponential decay function as the basis for weight updates to achieve the adjustment of attribute weights. Specifically, for each attribute $a$, the weight update formula is defined as follows:

$$W_a = exp(-\alpha \cdot H(a)) \tag{11}$$

Here, initially, it is set by default that the weight of each attribute is the same and equal to 1. $\alpha$ is a positive adjustment parameter, serving as a scaling factor whose value theoretically does not affect the final feature ranking (as intra-source normalization eliminates scaling effects), but plays a key regulatory role in the morphology of the intermediate weight distribution: larger $\alpha$ values (e.g., 1.0) amplify entropy differences, forming a steep weight distribution to enhance feature discriminability; smaller $\alpha$ values (e.g., 0.1) reduce entropy differences, resulting in a flatter distribution to improve algorithm robustness. This process ensures that attributes that provide more information for distinguishing classes (i.e., those with lower entropy values) receive higher weights, while attributes with weaker distinguishing ability (i.e., those with higher entropy values) receive lower weights. In this way, we can adjust the influence of each attribute to optimize the feature selection process in the multi-source decision information system.

**Example 2 (Continued from Example 1).** With the result of Example 1, we can obtain two granules $P_1 = \overline{N(Cl_1)} = \{x_1, x_2, x_4, x_5\}$, $P_2 = \overline{N(Cl_2)} = \{x_2, x_3, x_4, x_6\}$, then we compute $P_1 = \frac{|G_1|}{|U|} = \frac{2}{3}$, $P_2 = \frac{|G_2|}{|U|} = \frac{2}{3}$. For granule $G_1$, the probability of each category is $P(c = Cl_1|G_1) = 0.75$, $P(c = Cl_2|G_1) = 0.25$. Weighted probability for categories and normalized values are $W_{Cl_1} = 0.25$, $W_{Cl_2} = 0.75$. then we can compute $H(a)_{Cl_1} = -\frac{2}{3}\times(0.25\times0.75\times log_2(0.75)+0.75\times0.25\times log_2(0.25)) = 0.30187$. Similarly, we can calculate that $H(a)_{Cl_2} = 0.30187$, $H(a) = 2 \times 0.30187 = 0.60374$. Finally, set $\alpha = 0.1$, the weight of attribute $a_5$ in the source 1 is $W_{a_1}^1 = exp(-0.1 * 0.60374) = 0.9414$.

### 3.2. Feature weight fusion in multi-source information systems

In multi-source decision-making scenarios, different information sources often collect the same batch of objects using heterogeneous devices, different sampling frequencies, or different quantization precisions, resulting in significant differences in value ranges, distribution patterns, and noise levels of the features from each source. Therefore, a simple concatenation or mean merging would mask the true discriminative information. To balance the "source reliable differences" and the "feature discrimination difference" within a unified framework, we propose a hierarchical fusion mechanism that decouples the local weight calibration and weight aggregation into three sequential steps.

For a multi-source decision information system denoted as $S = (U, A \cup \{d\}) = \{U, \{a_i^j \mid i = 1, \ldots, n; j = 1, \ldots, s\} \cup \{d\}\}$, the fusion of feature weights proceeds as follows. First, within-source normalization is required. The initial weight of each attribute, derived from the adaptive neighborhood rough set model, is normalized via the $L1$ norm within its respective information source. This ensures that the weights of all features within a single source sum to unity, providing a standardized basis for subsequent cross-source comparison and fusion, which can be defined as follows:

$$W_a^t = \frac{W_a^t}{\sum_{a\in A} W_a^t} \tag{12}$$

$W_a^t$ is the weight of the attribute $a$ in the t-th source information system. This operation eliminates the internal bias of the source caused by differences in scale and sample size, making the feature weights of different sources comparable.

Second, in the section on the reliability assessment of information sources, we utilize the upper and lower approximations in the rough set theory to quantify the approximate quality of each information source, and thereby determine its weight $r_p$ in the multi-source decision-making information systems.

**Definition 4.** For the $p$ th information source ($S^p$), the formula to calculate its weight in multi-source decision-making information systems can be defined as follows:

$$r_p = \frac{\left|\underline{N_A(S^p)}\right|}{|U|}\left(1 - \frac{\left|\overline{N_A(S^p)}\right| - \left|\underline{N_A(S^p)}\right|}{\left|\overline{N_A(S^p)}\right|}\right) \tag{13}$$

where

$$\underline{N_A(S^p)} = \bigcup_{Cl\in\{d\}} \underline{N_A(Cl)},$$
$$\overline{N_A(S^p)} = \bigcup_{Cl\in\{d\}} \overline{N_A(Cl)}.$$

$\underline{N_A(S^p)}$ indicates the size of the lower approximation set of the entire condition attribute set $A$ under the $p$ information source, which determines the number of objects belonging to a certain class. $\overline{N_A(S^p)}$ represents the size of the upper approximation set, that is, the number of objects that may belong to a certain class. This formula calculates the ratio of the size of the lower approximation under the entire condition attribute set to the total number of objects in the universe to evaluate the basic coverage capability of the $p$ information source. Then, by calculating the difference between the upper approximation and the lower approximation, it assesses the accuracy of the information source. The higher the weight $r_p$, the more fully and with a smaller boundary the lower approximation of the information source covers in the decision task. Therefore, its contribution in the cross-source linear combination process of condition attributes is greater. Conversely, if the approximation quality of an information source is low, its weight in the linear combination process will also decrease accordingly. In addition, the weights of the information sources need to be normalized.

$$\lambda_p = \frac{r_p}{\sum_{p=1}^{s} r_p} \tag{14}$$

Finally, with the results of the weights of each condition attribute under each information source and the weights of information sources in the multi-source decision-making information systems, we further perform a cross-source linear combination of attribute weights.

**Definition 5.** For each $a \in A$, the weight of attribute $a$ in multi-source information systems can be calculated as follows:

$$W_a^f = \sum_{p=1}^{s} \lambda_p W_a^p \tag{15}$$

where $W_a^f$ is the final weight of attribute $a$ in multi-source information systems, $W_a^p$ is the weight of attribute $a$ under the $p$th information source. This formula ensures that the dominant role of highly reliable sources is explicitly amplified, the noise characteristics of low-reliability sources are naturally suppressed, while retaining all the effective discriminative information of all sources.

The time complexity analysis of Algorithm 1 is as follows: With the core parameters of data source quantity $s$, condition attribute quantity $m$, and total number of domain objects $n$, the overall complexity of the algorithm is dominated by three key steps: Firstly, the pre-computation of the distance matrix requires calculating the $n \times n$ mixed Hausdorff distance for each of the $m$ attributes of $s$ data sources, with a complexity of $O(s \cdot m \cdot n^2)$; Secondly, the feature weight calculation stage requires calculating the neighborhood granules and conditional entropy for each attribute of each data source across $C$ categories, with a cumulative complexity of $O(s \cdot m \cdot s^2)$; Finally, the source weight evaluation is achieved through upper and lower approximation sets, with the adaptive radius estimation and set operations for each data source involving $C$ categories, resulting in a total complexity of $O(s \cdot n^2)$ (as $c \ll n$, the influence of categories can be ignored). Subsequent steps such as cross-source fusion ($O(s \cdot m)$) and sorting ($O(m \log m)$) have lower complexity, and the overall time complexity of the algorithm is $O(s \cdot m \cdot n^2)$. The complexity of each stage of the algorithm is presented in the Table 4.

**Example 3.** We can compute the weight of each attribute in each source Tables 1–3, obtaining $W_a^1 = [1, 1, 0.9229, 0.9414, 0.9414]$, $W_a^2 = [1, 0.9414, 0.9311, 1, 0.9414]$, and $W_a^3 = [1, 1, 0.9702, 1, 1]$. We need to normalize the weights of attributes within the same source. After normalization, we obtain $W_a^1 = [0.2080, 0.2080, 0.1920, 0.1958, 0.1958]$, $W_a^2 = [0.2077, 0.1955, 0.1934, 0.2077, 0.1955]$, and $W_a^3 = [0.2011, 0.2011, 0.1952, 0.2011, 0.2011]$. Next, we compute the weight of each information source using Eq. (13) and normalize it using Eq. (14), resulting in $\lambda = [0.33, 0.33, 0.33]$. Finally, we realize the cross-source linear combination using Eq. (15), obtaining the final attribute weight of the multi-source information sources as $W_a^f = [0.2056, 0.2016, 0.1935, 0.2016, 0.1975]$. By sorting $W_a^f$ in descending order, we obtain the sorted result $[a_1, a_2, a_4, a_5, a_3]$.

**Algorithm 1** Feature ranking in Multi-Source Interval-Valued decision systems.

**Input:** Multi-source interval-valued decision-making system $S = (U, A \cup \{d\})$ with $s$ sources
**Output:** Feature ranking list based on final weight vector $\mathbf{W}^f$

**for** source $p = 1$ to $s$ **do**
  **for** attribute $a \in A$ **do**
    Compute hybrid-Hausdorff distances between all objects
    **for** each class $Cl$ **do**
      Set adaptive radius $\delta(Cl)$
      $\delta(Cl) = \underset{x\in Cl}{\text{med}}(\underset{y\in Cl}{\text{med}}\, d_B(x, y))$
    **end for**
    Build adaptive neighborhood granules
    Compute weighted conditional entropy $H_a^p$
    Attribute weight $W_a^p \leftarrow \exp(-\alpha\, H_a^p)$ and normalize, $\alpha = 0.1$
  **end for**
  Compute lower/upper approximations with full attribute set $A$
  Source reliability Eq. (18).
**end for**
Normalize source weights $\lambda_p = r_p / \sum_{q=1}^{s} r_q$
**for** attribute $a \in A$ **do**
  Final weight $W_a^f = \sum_{p=1}^{s} \lambda_p W_a^p$
**end for**
Sort attributes by $W_a^f$ descending
**return** Attributes ordering

**Table 4**
Time complexity analysis.

| Stage | Description | Time Complexity |
|---|---|---|
| 1. Distance Matrices | Compute mixed Hausdorff distances for all objects per attribute and source | $O(s \cdot m \cdot n^2)$ |
| 2. Feature Weights | Calculate neighborhood granules and conditional entropy for each attribute and source | $O(s \cdot m \cdot n^2)$ |
| 3. Source Weights | Evaluate source reliability via lower/upper approximations | $O(s \cdot n^2)$ |
| 4. Cross-Source Fusion | Linearly combine attribute weights across sources | $O(s \cdot m)$ |
| 5. Final Sorting | Sort attributes by final fused weights | $O(m \log m)$ |
| Overall Algorithm | | $O(s \cdot m \cdot n^2)$ |

### 3.3. Information fusion

In our previous work, we used the conditional entropy method to analyze and rank the attributes in the interval-valued multi-source decision information system, achieving an initial feature selection. We evaluated the influence degree of each attribute on the decision outcome by calculating the conditional entropy, rather than simply measuring the importance of the features. This approach helps identify the key attributes that have a significant impact on the decision outcome, and reduces the weights of those attributes with less influence, thereby more accurately guiding the selection of features and the decision-making process. The current research focus lies on the following: for the selected

key attributes, how to select the optimal data source from the multiple candidate information sources associated with them.

Let $S = (U, A' \cup \{d\})$ be a multi-source interval-valued decision information system, where $a_i^j | i = 1, 2, \dots m'; j = 1, 2, \dots, s$ represents a set of attributes from multiple sources. We applied the method of averaging the endpoints to convert the interval values in all information sources into point values, which can be defined as follows:

$$a_i^p = \frac{a_i^{p-} + a_i^{p+}}{2} \tag{16}$$

During the data processing stage, we first convert the interval-valued data into a real-valued representation (by taking the midpoint of the interval). Based on the weights of each information source calculated in the previous subsection, we calculate the weighted value of all information sources for each selected attribute and establish it as the "leader" data that represents the overall characteristics. Finally, we identify the single most representative source—-the one whose leader vector exhibits the highest consensus with the overall source leader—-and designate it as the optimal source.

We use $a_i^*$ to denote the data of the elected leader, which can be defined as

$$a_i^* = \sum_{p=1}^{s} \lambda_p \cdot a_i^p \tag{17}$$

where $s$ stands for the number of sources, $\lambda_p$ represents the weight assigned to the $p$th source, and $a_i^p$ indicates the value of the $i$th attribute within the $p$th source. Take the midpoint of the interval values of the same attribute in all sources and calculate the weighted average to generate a "weighted average data table" across sources. Subsequently, calculate the distance between each source and this weighted average data table to determine the representativeness score of each source. Finally, the source with the lowest score (i.e., the highest consistency) is identified as the optimal information source.

**Definition 6.** The consensus degree is defined as follows:

$$DC_p = \sum_{i=1}^{|U|} \sum_{j=1}^{m'} \left| a_j^p(x_i) - a_j^*(x_i) \right| \tag{18}$$

Here, $j$ runs over every retained attribute, $p$ labels the current source, and $a_j^*(x_i)$ is the source-leader data. Fusing these elements produces an outcome that best mirrors the overall data distribution across the multi-source information system, prompting us to designate the corresponding source as the most representative one. The specific steps for this process are illustrated in Algorithm 2.

**Example 4 (Example 3 continued).** In Example 2, the attributes were already ordered. Now we retain the top-3 attributes and remove the attributes $a_5$, and $a_3$. Next, we need to identify the most representative information source from the three available sources. First, using Eq. (16), we convert the interval values into floating-point numbers, and the results of this conversion are shown in Table 5. Then, applying Eq. (17), we calculate the leader source, as shown in Table 6. By measuring the degree of consensus between each source and the leader source, we select Source 3 as the optimal information source based on the highest alignment.

## 4. Experimental analysis

To comprehensively evaluate the performance of the proposed method, we conducted experiments on nine publicly available benchmark datasets from the UCI Machine Learning Repository. The selection of these datasets was strategic, covering a wide range of features in terms of instance quantity, feature dimension, and class number. This diversity ensures a robust assessment of the algorithm's scalability, stability, and effectiveness in different application domains and data complexity. The datasets are listed in Table 7.

**Algorithm 2** Information fusion method in multi-source decision information systems with interval-source.

**Input:** Multi-source interval-valued decision system $S' = (U, A \cup \{d\})$ with $s$ sources, and the weight of information sources $\lambda = [\lambda_1, \lambda_2, \dots, \lambda_s]$
**Output:** An optimal source;
1: create zero matrix: $AP$.
2: **for** $a_i^{p-}$ and $a_i^{p+}$ in $S'$ **do**
3: $AP \leftarrow a_i^p = \frac{a_i^{p-} + a_i^{p+}}{2}$
4: **end for**
5: create zeros matrix: $S^*$.
6: **for** $j = 1$ to $m$ **do**
7: **for** $i = 1$ to $|U|$ **do**
8: **for** $p = 1$ to $s$ **do**
9: $a_j^*(x_i) = \lambda_p * a_i^p$
10: **end for**
11: **end for**
12: $S^* \leftarrow a_j^*(x_i)$
13: **end for**
14: **for** $p = 1$ to $s$ **do**
15: Calculation of $DC_p$
16: **end for**
17: Select the optimal source based on $DC_p$
18: **return** A optimal source

**Table 5**
The outcome of the floating-point conversion.

| $U$ | $S_1$ | | | $S_2$ | | | $S_3$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | $a_1^1$ | $a_2^1$ | $a_4^1$ | $a_1^2$ | $a_2^2$ | $a_4^2$ | $a_1^3$ | $a_2^3$ | $a_4^3$ |
| $x_1$ | 0.45 | 0.495 | 0.545 | 0.355 | 0.425 | 0.47 | 0.48 | 0.385 | 0.405 |
| $x_2$ | 0.24 | 0.63 | 0.39 | 0.395 | 0.625 | 0.27 | 0.455 | 0.605 | 0.305 |
| $x_3$ | 0.625 | 0.775 | 0.735 | 0.68 | 0.7 | 0.39 | 0.68 | 0.74 | 0.674 |
| $x_4$ | 0.875 | 0.265 | 0.815 | 0.81 | 0.35 | 0.745 | 0.89 | 0.265 | 0.785 |
| $x_5$ | 0.47 | 0.58 | 0.195 | 0.46 | 0.635 | 0.33 | 0.485 | 0.6 | 0.225 |
| $x_6$ | 0.0.79 | 0.725 | 0.565 | 0.85 | 0.685 | 0.595 | 0.805 | 0.695 | 0.545 |

**Table 6**
Leader source.

| $S^*$ | $a_1$ | $a_3$ | $a_4$ |
|---|---|---|---|
| $x_1$ | 0.4283 | 0.435 | 0.4733 |
| $x_2$ | 0.3633 | 0.62 | 0.3216 |
| $x_3$ | 0.6616 | 0.7383 | 0.7216 |
| $x_4$ | 0.8583 | 0.2833 | 0.7816 |
| $x_5$ | 0.4716 | 0.605 | 0.25 |
| $x_6$ | 0.815 | 0.7016 | 0.5683 |

The algorithms discussed in this article, including the proposed algorithm and the comparison algorithms, were all implemented using Python 3.8. The development environment used was PyCharm 2024. The testing was carried out on a computer with the following specifications: a 4.7GHz AMD Ryzen 7 6800H processor, a Radeon series graphics processor, 16.0 GB of memory, and a 64-bit Windows 11 operating system.

### 4.1. Diversification and intervalization of datasets

Data in multi-source decision information systems typically originate from multiple measurement sources or represent the aggregated outcomes of repeated experiments. While discrepancies exist among these source-specific datasets, they are usually subtle and tend to follow a normal distribution. Because the datasets we obtained from the UCI repository are single-source, we must transform them to better reflect

**Table 7**
Description of data sets.

| NO. | Date sets | Abbreviation | Instance | Feature | Class |
|---|---|---|---|---|---|
| 1 | Wine | Wine | 178 | 13 | 2 |
| 2 | Optical Recognition of Handwritten Digits | OROHD | 5620 | 64 | 10 |
| 3 | Rice (Cammeo and Osmancik) | Rice | 3810 | 7 | 2 |
| 4 | Parkinson's Disease Classification | PDC | 756 | 754 | 2 |
| 5 | Letter Recognition | LR | 20,000 | 16 | 26 |
| 6 | Statlog (Landsat Satellite) | Statlog | 6435 | 36 | 6 |
| 7 | Waveform Database Generator (Version 1) | Wave | 5000 | 21 | 3 |
| 8 | Period Changer | PC | 90 | 1177 | 2 |
| 9 | Pen-Based Recognition of Handwritten Digits | PROHD | 10,992 | 16 | 10 |

real-world conditions. Starting from the raw data, we first generate several related tables whose values are drawn from normal distributions, thereby mimicking the multiplicity of actual measurement sources. Afterwards, the observations in each simulated source are converted into interval form. The steps of this procedure are outlined in the remainder of this section.

(1) First, each dataset can be defined as a decision table $S = (U, \{A = \{a_i, \ldots, a_m\}\} \cup \{d\})$. To create multiple related data sources, each data source should contain a similar data table to the original one, and the original data table needs to undergo certain processing. The detailed steps are presented below: Initially, create 20 random arrays that conform to a normal distribution, where the mean is 0 and the standard deviation is 0.1. By integrating these random arrays with the original data table, 20 relevant data sources can be produced. This approach can preserve the correlation of each data source while guaranteeing their distinctiveness. The specific steps are as follows:

$$a_i^p = a_i^p \left(1 - \mathcal{N}_k\left(0, 0.1^2\right)\right), \tag{19}$$

where $a_i^p$ denotes the value of the attribute $a_i$ in the $p$th source, and $\aleph_p$ denotes the $p$th generated normal distribution array.

(2) To ensure data consistency, this study uniformly applied the maximum-minimum normalization method to process 20 interrelated data sources. There were significant differences in the measurement scales of each attribute. After maximum-minimum normalization, the data were transformed into the interval [0,1], avoiding analytical deviations caused by scale differences. Normalization can ensure that all features have the same initial weight in the model, thereby improving the generalization ability of the model.

(3) In response to the demand for interval values, this study further processed the dataset after completing the maximum-minimum normalization. Specifically, for each normalized data point, we subtracted 0.1 times itself as the lower endpoint of the interval $a_i^- = a_i(1 - 0.1)$, and added 0.1 times itself as the upper endpoint of the interval $a_i^+ = a_i(1 + 0.1)$. This processing method aims to introduce a certain range of uncertainty for each data point, in order to better reflect measurement errors or data fluctuations in real situations. The interval values generated in this way can not only reflect the central tendency of the data, but also cover the possible fluctuation range, providing more comprehensive and robust data support for subsequent analysis.

### 4.2. Classification accuracy of the optimal data

In this section, we retain the proportion $k$ of the sorted features, with $k$ ranging from 0.1 to 0.9 and the step size being 0.1. We evaluated the classification accuracy for the KNN, Naive Bayes, Decision Tree, and SVM classifiers, and all results were obtained through 5-fold cross-validation. The classification accuracy results of the nine datasets shown in Table 7 are depicted in Fig. 2. Across most datasets and classifiers, a clear regularity can be observed: as the retention ratio $k$ increases, the classification accuracy improves, but the rate of improvement gradually decreases and eventually tends to saturate. Take the dataset Wine as an example. When the retention ratio $k$ increases from 0.1 to 0.4, in all classifiers, the classification accuracy significantly improves by 15%–20%. However, when $k$ continues to increase, the fluctuation range of the classification accuracy is only within 5%. A similar pattern appears in the Wave dataset under the KNN classifier: when $k$ increases from 0.1 to 0.3, the accuracy rises from 54% to 70.28%; from 0.4 to 0.6, it further increases from 73.12% to 79.10%; while from 0.7 to 0.9, the gain narrows to 79.52%–81.24%.

This behavior is closely related to our feature ordering strategy. During the data pre-processing stage, we sorted the features based on the granule information entropy. Granule information entropy is an indicator that measures the importance of features, and it can reflect the information content of the features in the data and their contribution to the classification results. Through this sorting method, the features that have a significant impact on classification are placed at the front, while the features with less impact are placed at the back. When the retention ratio $k$ is small, only the features that have a significant impact on classification are retained. These features contain highly relevant effective information for the classification task, so the classification accuracy significantly improves. As $k$ increases, features with less impact on classification are also retained. Although these features do carry a certain amount of discriminative signal, their marginal contribution to classification performance is markedly inferior to that of the salient features; consequently, the attendant gain in accuracy diminishes progressively as they are incorporated. From a methodological perspective, this phenomenon not only verifies that the entropy-based ranking can effectively concentrate discriminative information in the top-ranked features, but also reflects a typical "diminishing marginal utility" of added features.

In datasets PDC and PC, the classification accuracy fluctuates as the retention ratio $k$ increases. This deviation from the monotonic trend mainly reflects the interaction effects among features and the limitations of one-dimensional ranking. On the one hand, there are both collaborative and interfering effects between features. The collaborative effect can improve the performance of the classifier, while the interference effect reduces the performance, resulting in unstable accuracy. On the other hand, although the granule information entropy sorting can filter out important features, it is difficult to capture the complex relationships between features, and it may misjudge redundant features, further exacerbating the fluctuations. Therefore, these two datasets highlight that, in scenarios with strong feature dependence, purely univariate ranking may need to be complemented by methods that explicitly model feature interactions.

Fig. 3 shows the differences in the weights of the same features from different information sources in four datasets (Wine, OROHD, Wave, PC). The heatmap clearly displays the variations in feature weights across different sources, highlighting the necessity of cross-source linear combination of feature weights. For instance, the weight of feature $A7$ in the Wine dataset and the weight of $A13$ in the PC dataset fluctuate significantly among different information sources, indicating that a single source may not fully capture the importance of the feature. This weight disparity underscores the importance of integrating multi-source information to obtain more accurate feature weights. The proposed cross-source linear aggregation smooths out source-specific biases while preserving genuinely discriminative patterns, so that our method can more effectively fuse multi-source data, enhancing the accuracy of feature selection and the reliability of decision-making. Overall, these observations confirm that the entropy-based ranking and cross-source weighting jointly contribute to stable and competitive classification performance on a wide range of datasets.

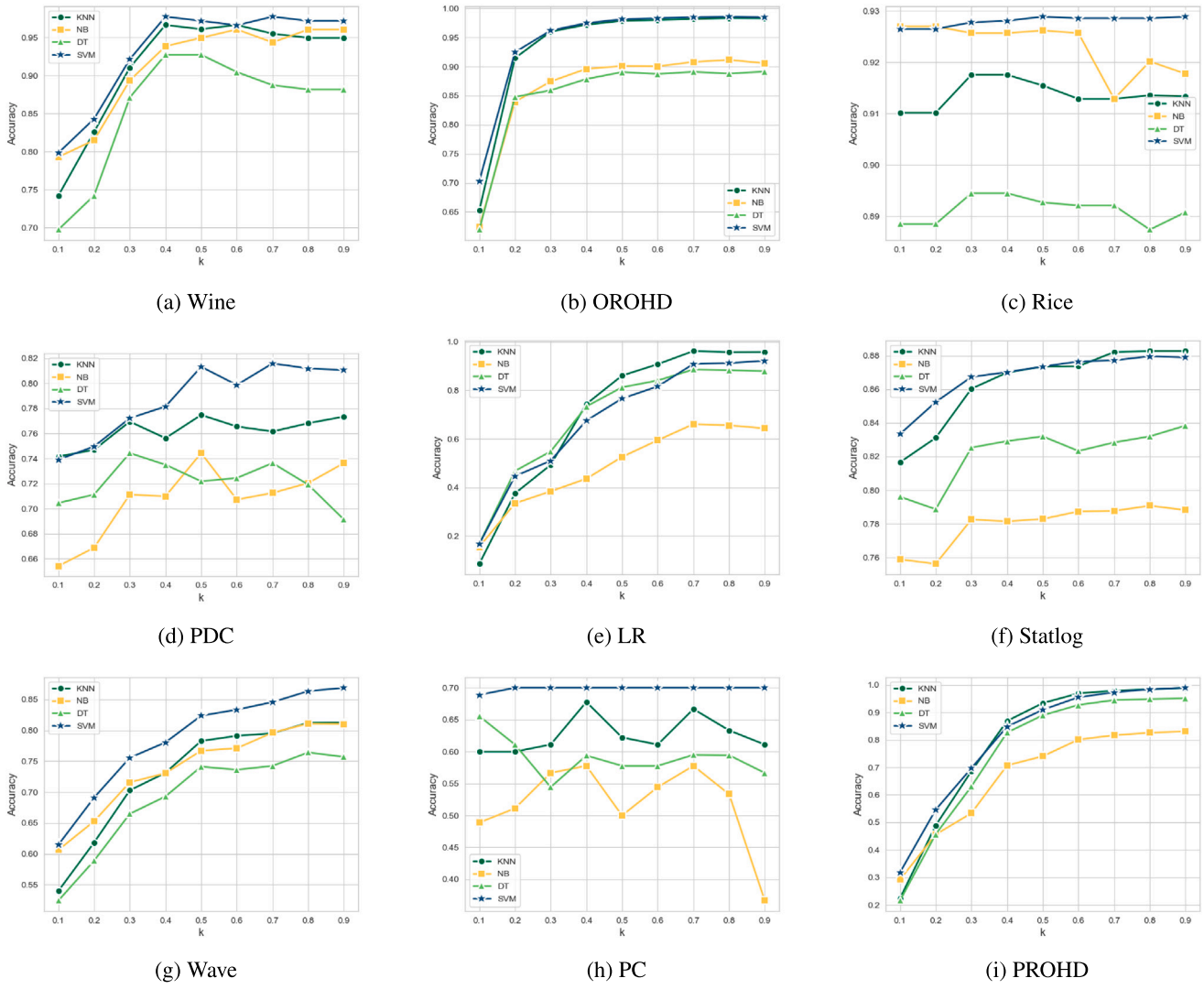



**Fig. 2.** Classification accuracy of different feature retention rates k under KNN, NB, DT, and SVM classifiers.

### *4.3. A comparison of the optimal data*

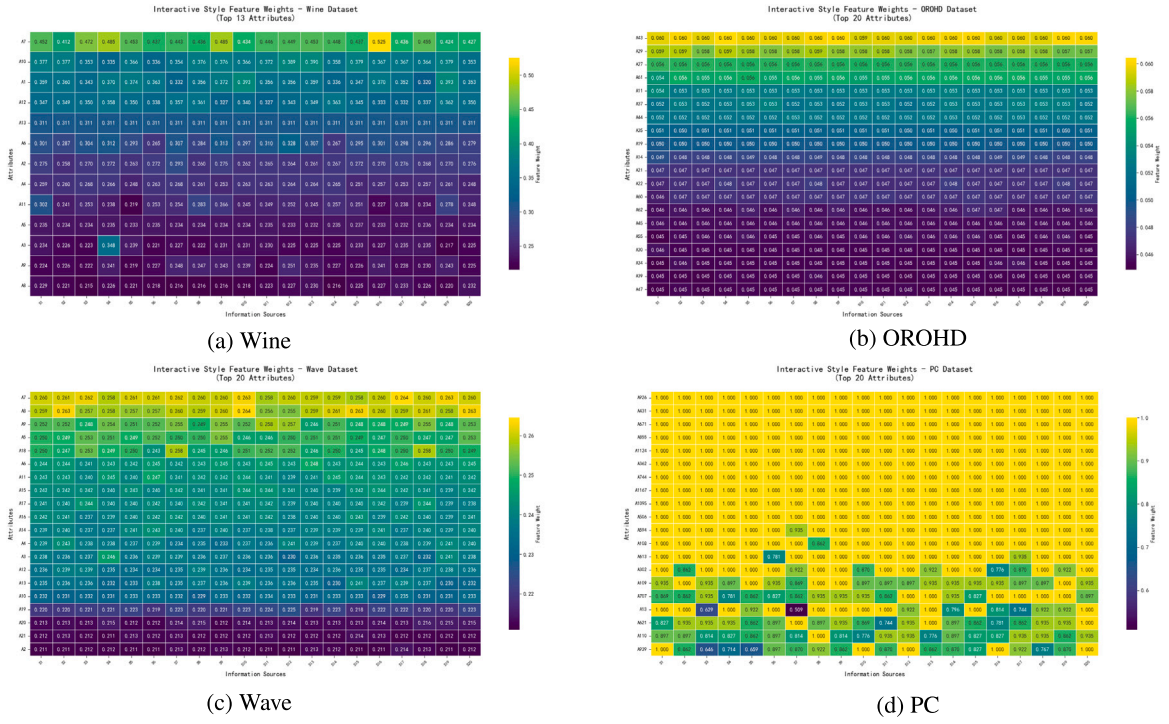
Unlike traditional methods that routinely fuse information through arithmetic averaging, this study proposes a leader-aware fusion framework in which the "leader" is identified as the optimal source among multiple information providers. Rather than being generated by simple averaging, the leader is constructed by jointly evaluating the contribution degree of each source and its consistency with the feature-selection outcome. This leader-aware strategy ensures that the selected source not only represents the overall performance of the information set but also maximizes consensus with the leader, thereby enhancing model stability and objectivity.

To verify the effectiveness of the proposed fusion strategy, we employed the k-nearest neighbor (KNN) classifier and compared the classification accuracy achieved using the optimal information sources selected by our method with that of the leader for different values of the attribute proportion k. As shown in Tables 8 and 9, as k increases, the two methods exhibit highly consistent classification accuracies. Even for datasets PDC and PC, the classification accuracy of the optimal information source exceeds that of the leader, further demonstrating the effectiveness of our method in selecting representative information sources. Further experiments indicate that, regardless of the value of k, there is no significant difference in classification performance between the optimal source and the leader, highlighting the robustness of our method in diverse information-fusion contexts. It is worth noting that, although traditional methods may perform poorly in scenarios with limited attributes and large sample sizes due to insufficient fusion conditions, the method proposed in this study effectively alleviates this problem by jointly considering feature consistency and contribution, thereby improving classification accuracy and enhancing the scalability of the system.

### *4.4. Effective comparative experiment for feature ranking*

In this section, we further verify the effectiveness of the granule information entropy sorting method by comparing it with several representative single-source feature ranking approaches. Since there are few methods for feature ranking in multi-source information systems, we revert to the single-source setting: the original UCI datasets are used without source partition, our method is adapted by disabling the cross-source fusion step, and its ranking performance is evaluated in this single-source scenario. The three feature ranking methods used for comparison are Principal Component Analysis (PCA) [31], Laplacian Score [32], and Random Forest [33].

To implement feature ranking using PCA, we first perform PCA on each dataset to obtain the explained variance of every principal component and the corresponding loading vectors. The importance of a feature is then computed as a weighted sum of its loadings over all principal




(a) Wine (b) OROHD

(c) Wave (d) PC

**Fig. 3.** Heat map.

**Table 8**
Comparison of the optimal and the leader accuracy on KNN($k$ = 0.1–0.5).

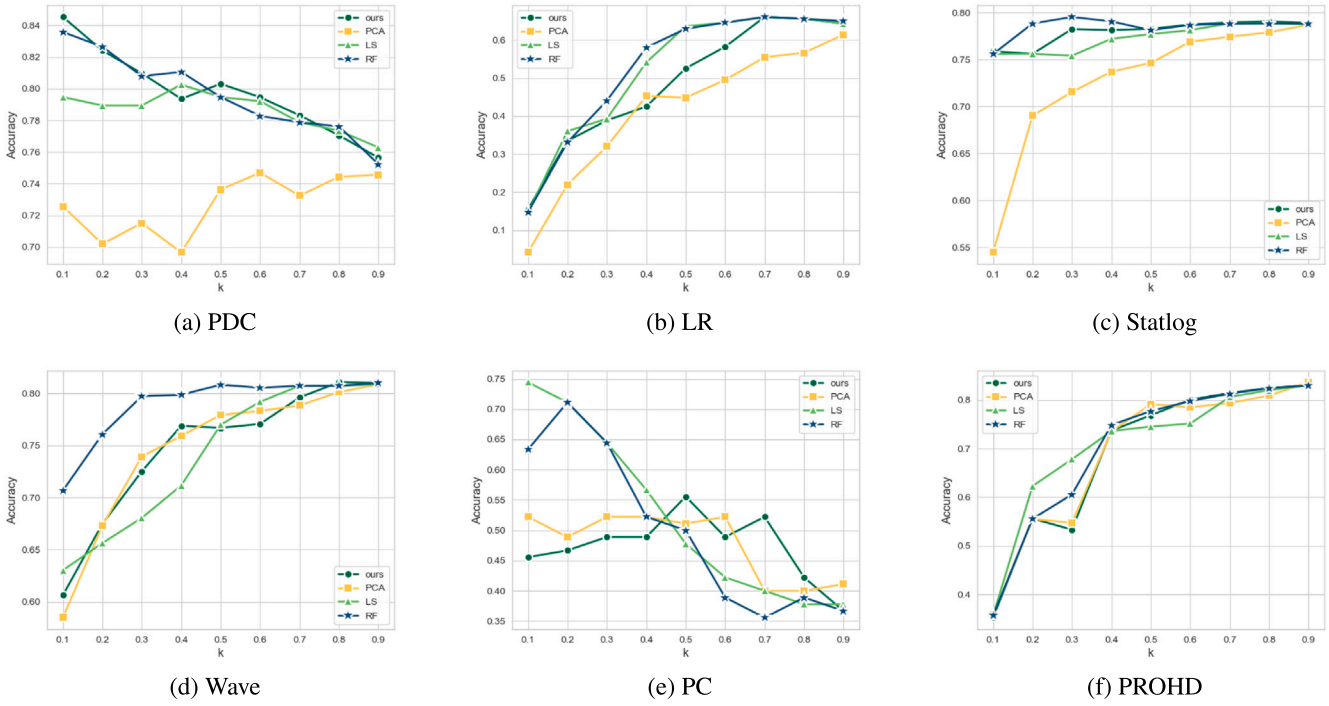
| Data | k | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0.1 | | 0.2 | | 0.3 | | 0.4 | | 0.5 | |
| | Optimal leader | | Optimal leader | | Optimal leader | | Optimal leader | | Optimal leader | |
| Wine | 74.21 ± 9.08 | 76.59 ± 3.21 | 82.62 ± 7.50 | 81.02 ± 2.13 | 91.06 ± 8.83 | 90.06 ± 4.12 | 96.67 ± 4.44 | 94.90 ± 5.23 | 96.11 ± 4.16 | 96.01 ± 2.14 |
| OROHD | 65.20 ± 2.14 | 64.94 ± 2.34 | 91.49 ± 1.52 | 89.91 ± 2.36 | 96.01 ± 0.79 | 96.00 ± 1.24 | 97.21 ± 0.58 | 96.73 ± 4.32 | 97.88 ± 0.47 | 97.48 ± 0.87 |
| Rice | 91.02 ± 2.20 | 91.44 ± 2.14 | 91.02 ± 2.20 | 89.44 ± 4.12 | 91.76 ± 1.82 | 91.44 ± 1.46 | 91.76 ± 1.82 | 91.44 ± 1.46 | 91.55 ± 1.82 | 91.44 ± 1.46 |
| PDC | 74.17 ± 4.61 | 71.24 ± 1.23 | 74.70 ± 3.61 | 72.14 ± 3.24 | 76.95 ± 3.68 | 73.87 ± 1.34 | 75.63 ± 1.94 | 73.76 ± 2.36 | 77.48 ± 4.53 | 76.85 ± 1.24 |
| LR | 8.53 ± 0.91 | 9.30 ± 2.11 | 37.56 ± 1.16 | 35.46 ± 1.36 | 49.29 ± 0.60 | 48.21 ± 2.36 | 74.37 ± 1.05 | 74.36 ± 0.36 | 86.06 ± 0.48 | 86.04 ± 3.21 |
| Statlog | 81.66 ± 2.41 | 81.30 ± 1.34 | 83.11 ± 2.29 | 83.09 ± 1.96 | 86.03 ± 1.32 | 86.06 ± 4.36 | 86.99 ± 1.10 | 86.79 ± 3.14 | 87.35 ± 1.13 | 87.34 ± 2.14 |
| Wave | 54.00 ± 0.80 | 53.93 ± 1.24 | 61.84 ± 0.71 | 61.79 ± 0.64 | 70.28 ± 0.62 | 70.09 ± 1.54 | 73.12 ± 1.32 | 73.53 ± 1.68 | 78.28 ± 1.18 | 77.80 ± 1.36 |
| PC | 60.00 ± 8.89 | 58.17 ± 2.36 | 60.00 ± 8.16 | 59.61 ± 2.17 | 61.11 ± 7.03 | 62.11 ± 3.25 | 67.78 ± 9.94 | 64.11 ± 7.89 | 62.22 ± 10.18 | 61.83 ± 9.47 |
| PROHD | 22.37 ± 2.77 | 22.36 ± 1.23 | 48.88 ± 11.80 | 48.46 ± 9.63 | 68.63 ± 0.63 | 68.43 ± 2.34 | 86.98 ± 1.49 | 86.84 ± 2.87 | 93.38 ± 0.83 | 93.37 ± 1.45 |

**Table 9**
Comparison of the optimal and the leader accuracy on KNN($k$ = 0.6–0.9).

| Data | k | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 0.6 | | 0.7 | | 0.8 | | 0.9 | |
| | Optimal leader | | Optimal leader | | Optimal leader | | Optimal leader | |
| Wine | 96.67 ± 4.08 | 96.11 ± 2.14 | 95.52 ± 2.85 | 95.41 ± 2.31 | 94.97 ± 3.69 | 95.27 ± 2.14 | 94.97 ± 3.69 | 95.27 ± 2.14 |
| OROHD | 98.02 ± 0.42 | 97.56 ± 1.02 | 98.19 ± 0.53 | 98.16 ± 0.96 | 98.35 ± 0.36 | 98.33 ± 0.45 | 98.31 ± 0.42 | 98.27 ± 1.34 |
| Rice | 91.29 ± 1.82 | 91.44 ± 1.46 | 91.29 ± 1.82 | 91.44 ± 1.46 | 91.36 ± 1.82 | 91.44 ± 1.46 | 91.34 ± 1.82 | 91.44 ± 1.46 |
| PDC | 76.56 ± 3.09 | 75.52 ± 1.24 | 76.16 ± 2.75 | 76.17 ± 1.87 | 76.82 ± 1.51 | 75.99 ± 2.45 | 77.35 ± 2.19 | 76.44 ± 4.36 |
| LR | 90.75 ± 0.46 | 90.65 ± 2.34 | 96.16 ± 0.23 | 95.96 ± 2.34 | 95.72 ± 0.35 | 95.81 ± 0.86 | 95.77 ± 0.21 | 95.87 ± 0.78 |
| Statlog | 87.37 ± 0.97 | 87.37 ± 2.14 | 88.21 ± 1.09 | 88.21 ± 3.21 | 88.27 ± 1.29 | 88.25 ± 2.04 | 88.27 ± 1.11 | 88.27 ± 2.41 |
| Wave | 79.10 ± 1.38 | 78.83 ± 1.47 | 79.52 ± 1.14 | 79.17 ± 0.97 | 81.22 ± 1.76 | 80.98 ± 1.74 | 81.24 ± 0.98 | 81.25 ± 0.75 |
| PC | 61.11 ± 9.30 | 61.89 ± 2.01 | 66.67 ± 6.09 | 63.72 ± 1.36 | 63.33 ± 9.03 | 61.33 ± 1.87 | 61.11 ± 1.11 | 61.11 ± 2.17 |
| PROHD | 96.93 ± 0.45 | 96.92 ± 1.36 | 97.90 ± 0.027 | 97.85 ± 1.37 | 98.91 ± 0.28 | 98.44 ± 1.48 | 98.91 ± 0.26 | 98.94 ± 1.85 |

components, where the weight is the product of the absolute loading and the explained variance, and features are finally sorted in descending order of this score. This way, the influence of a feature on all principal components is taken into account rather than only on the first few. For all four methods (our entropy–granule ranking and the three baselines), we keep the top $k$ proportion of features, where $k$ ranges from 0.1 to 0.9 with a step size of 0.1. The corresponding classification accuracies under the Naive Bayes classifier on the last six datasets are reported in Fig. 4.

Our method shows a generally steady growth in accuracy as the feature retention ratio $k$ increases, and tends to be particularly stable and effective when $k$ is relatively large (e.g., 0.7–0.9). Intuitively, once most




(a) PDC (b) LR (c) Statlog

(d) Wave (e) PC (f) PROHD

**Fig. 4.** Comparison of ranking effectiveness on Naive Bayes.

of the features with low entropy (high importance) are included, further adding low-ranked features brings only marginal benefits and rarely harms performance. In contrast, although PCA sometimes achieves comparable results, its improvement is usually less obvious when $k$ is small. Laplacian Score performs well on certain datasets, but its behavior at high $k$ is less stable than that of our method. Random Forest is fairly stable overall, yet its accuracy is often slightly lower than ours when only a small proportion of features is retained. Across multiple datasets, our method exhibits better or at least comparable stability to the other three approaches, especially at higher $k$ values, and can reach or surpass their accuracy. In dataset 8 (PC), for example, all methods show noticeable fluctuations, but our approach still maintains a relatively high accuracy when $k$ is large, which indicates good robustness to noisy or redundant features. These observations are consistent with our design: ranking features by granule information entropy directly reflects their contribution to reducing class uncertainty, so the most informative features are more likely to appear at the top of the list.

Based on the above analysis, we can conclude that our method is effective in feature ranking. It achieves high accuracy, stability, and robustness on multiple datasets, particularly when more top-ranked features are retained, and it can make better use of key features for classification. Therefore, even in the single-source setting, the proposed entropy-based ranking performs competitively against well-established baselines under different feature retention ratios, which supports its applicability as a general feature selection tool.

**Table 10**
Description of ablation variant models.

| Variants | Description |
|---|---|
| HD-C | Replaces the Mixed Hausdorff distance with Classical Hausdorff distance. |
| AN-F | Removes the adaptive neighborhood mechanism;fixes radius to $\delta = 0.1$. |
| SRA-E | Eliminates source reliability assessment;applies equal weights to all sources. |
| CC-M | Disregards the consistency check; uses the mean of all sources for fusion. |

**Table 11**
Average accuracy on KNN of ablation Variants (k = 0.8).

| Dataset | Full | HD-C | AN-F | SRA-E | CC-M |
|---|---|---|---|---|---|
| Data1 | 94.95 ± 3.29 | 94.95 ± 3.29 | 94.90 ± 5.10 | 95.03 ± 2.75 | **95.26 ± 3.04** |
| Data2 | **98.38 ± 0.54** | 98.19 ± 0.43 | 97.78 ± 0.41 | 98.01 ± 3.10 | 98.09 ± 2.05 |
| Data3 | **91.44 ± 1.80** | 91.36 ± 1.95 | 91.36 ± 1.95 | 91.34 ± 0.21 | 91.34 ± 2.05 |
| Data4 | **78.68 ± 1.84** | 77.48 ± 1.78 | 77.81 ± 1.87 | 76.42 ± 1.43 | 77.62 ± 1.28 |
| Data5 | **95.72 ± 0.35** | 95.21 ± 1.24 | 93.83 ± 0.41 | 95.27 ± 2.52 | 95.06 ± 0.30 |
| Data6 | **88.27 ± 0.01** | 88.17 ± 1.27 | 87.72 ± 1.19 | 88.01 ± 1.34 | 88.04 ± 1.28 |
| Data7 | **81.22 ± 1.76** | 81.00 ± 2.11 | 80.00 ± 1.38 | **81.22 ± 1.76** | 81.06 ± 1.28 |
| Data8 | **64.00 ± 5.14** | 63.33 ± 9.03 | 63.33 ± 11.11 | 63.33 ± 9.03 | 63.33 ± 11.13 |
| Data9 | 98.43 ± 0.28 | 97.98 ± 0.25 | 98.31 ± 0.14 | 98.45 ± 0.25 | **98.47 ± 0.28** |
| Average | **87.90** | 87.52 | 87.23 | 87.45 | 87.59 |

### *4.5. Ablation experiment*

In the previous subsection, we validated the effectiveness of the feature ranking method based on entropy weighting. Building on this, this subsection further evaluates the contributions of each key component through systematic ablation experiments. Specifically, four variant models were constructed by removing or modifying one component at a time, and their configurations are summarized in Table 10. Under the KNN classifier with a feature retention ratio of $k = 0.8$, which is close to the accuracy peak in Fig. 2 and thus provides a representative and relatively demanding test case, the classification accuracies of the four variant models and the full model are shown in Table 11. The experimental results demonstrate that the full model achieves optimal or near-optimal performance on most datasets, with the best overall mean, confirming the positive role of each component in enhancing model accuracy and stability.

Further analysis reveals that each component contributes to system performance through different mechanisms: the mixed Hausdorff distance improves performance by 1.20% on datasets such as dataset PDC, proving its ability to more accurately capture the boundary and central features of interval data; the adaptive neighborhood mechanism delivers

**Table 12**
Classification accuracy of information fusion based on Decision Tree.

| | MinF | MaxF | MeanF | CieF | CeF | NREW (40%) | NREW (60%) | NREW (80%) |
|---|---|---|---|---|---|---|---|---|
| Data1 | 89.35 ± 4.81 | 88.81 ± 6.31 | 91.65 ± 7.85 | 87.11 ± 3.73 | 88.25 ± 5.87 | **92.76 ± 5.71** | 90.49 ± 5.41 | 88.19 ± 4.52 |
| Data2 | 88.90 ± 4.35 | 88.27 ± 2.10 | **91.67 ± 3.54** | 87.31 ± 6.31 | 87.47 ± 1.22 | 87.88 ± 1.25 | 88.75 ± 1.12 | 88.81 ± 0.17 |
| Data3 | 88.16 ± 1.81 | 88.50 ± 2.36 | 88.92 ± 1.57 | 88.82 ± 1.43 | 88.98 ± 2.20 | **89.45 ± 1.24** | 89.21 ± 1.62 | 88.74 ± 0.18 |
| Data4 | 67.15 ± 14.89 | 63.71 ± 19.42 | 62.91 ± 18.98 | 68.08 ± 14.25 | 63.31 ± 18.10 | **73.15 ± 2.62** | 72.45 ± 2.70 | 71.92 ± 1.85 |
| Data5 | 84.27 ± 1.02 | 83.67 ± 0.50 | 83.60 ± 0.30 | 88.21 ± 1.40 | 87.96 ± 3.60 | 73.35 ± 0.95 | 84.04 ± 5.10 | **88.26 ± 0.52** |
| Data6 | 83.06 ± 1.39 | 82.75 ± 1.44 | 82.22 ± 2.35 | 82.31 ± 2.64 | 83.06 ± 2.79 | 82.92 ± 0.94 | 82.33 ± 1.99 | **83.20 ± 2.19** |
| Data7 | 74.88 ± 0.64 | 73.58 ± 0.71 | 75.36 ± 1.40 | 76.21 ± 1.42 | 75.38 ± 0.97 | 69.29 ± 0.53 | 73.62 ± 0.80 | **76.42 ± 0.96** |
| Data8 | 55.56 ± 9.94 | 48.89 ± 12.37 | 51.11 ± 11.33 | 57.78 ± 5.67 | **64.44 ± 10.30** | 59.39 ± 15.95 | 57.78 ± 6.67 | 59.44 ± 10.54 |
| Data9 | **95.97 ± 0.54** | 95.64 ± 0.67 | 95.51 ± 0.85 | 95.24 ± 1.40 | 95.33 ± 0.60 | 82.68 ± 1.13 | 92.71 ± 0.86 | 95.64 ± 0.68 |
| Average | 80.81 ± 4.28 | 79.31 ± 4.87 | 80.32 ± 5.35 | 81.23 ± 4.25 | 81.58 ± 5.07 | 78.99 ± 3.67 | 81.26 ± 2.92 | **82.28 ± 2.40** |

**Table 13**
Classification accuracy of information fusion based on SVM.

| | MinF | MaxF | MeanF | CieF | CeF | NREW (40%) | NREW (60%) | NREW (80%) |
|---|---|---|---|---|---|---|---|---|
| Data1 | **98.33 ± 2.22** | 96.51 ± 1.76 | 97.26 ± 2.08 | 97.21 ± 1.76 | 96.70 ± 2.20 | 97.76 ± 2.08 | 96.63 ± 1.10 | 98.26 ± 2.08 |
| Data2 | 97.90 ± 0.40 | 98.03 ± 0.39 | 97.76 ± 2.08 | 98.31 ± 0.24 | 98.43 ± 0.43 | 97.49 ± 0.71 | 98.33 ± 0.44 | **98.59 ± 0.42** |
| Data3 | 92.63 ± 1.87 | 92.66 ± 1.94 | 92.58 ± 1.79 | 92.86 ± 1.67 | 92.65 ± 1.93 | 92.81 ± 1.84 | 92.86 ± 1.76 | **92.89 ± 1.89** |
| Data4 | 82.28 ± 2.62 | 81.32 ± 2.38 | 82.58 ± 2.37 | **84.37 ± 3.40** | 82.91 ± 2.88 | 78.15 ± 2.01 | 79.87 ± 3.76 | 83.19 ± 3.64 |
| Data5 | 92.45 ± 0.56 | 92.47 ± 0.36 | **92.59 ± 0.63** | 92.51 ± 2.40 | 92.50 ± 1.43 | 67.54 ± 1.24 | 81.60 ± 0.91 | 92.58 ± 0.65 |
| Data6 | 87.98 ± 1.03 | 87.97 ± 1.03 | 88.00 ± 1.03 | 88.02 ± 1.45 | 88.30 ± 1.12 | 87.02 ± 0.97 | 88.65 ± 1.25 | **88.96 ± 1.09** |
| Data7 | 86.24 ± 1.11 | 86.42 ± 1.05 | 86.31 ± 1.40 | 86.32 ± 3.60 | 86.38 ± 0.94 | 78.04 ± 1.09 | 83.34 ± 0.90 | **86.88 ± 1.12** |
| Data8 | 70.00 ± 2.27 | 70.00 ± 2.27 | 70.00 ± 2.27 | 70.00 ± 2.27 | 70.00 ± 2.27 | 70.00 ± 2.27 | 70.00 ± 2.27 | 70.00 ± 2.27 |
| Data9 | 99.31 ± 0.20 | **99.32 ± 0.21** | 99.31 ± 0.20 | 99.14 ± 3.40 | 99.31 ± 0.20 | 84.84 ± 2.02 | 95.51 ± 1.40 | 99.14 ± 0.43 |
| Average | 89.47 ± 11.36 | 89.41 ± 1.27 | 89.59 ± 1.53 | 89.85 ± 2.24 | 89.72 ± 1.52 | 83.73 ± 1.58 | 87.42 ± 1.54 | **90.07 ± 1.51** |

a significant gain of 1.89% on dataset LR, highlighting its adaptability to complex data distributions through dynamic adjustment of the neighborhood radius; the source reliability assessment achieves a 2.26% performance improvement on dataset PDC, underscoring the critical role of weight allocation in identifying information source quality; although the consistency check mechanism slightly underperforms mean fusion on individual datasets, it effectively enhances model stability in most scenarios.

These findings confirm the synergistic effects among the components: the distance metric and neighborhood mechanism form the core foundation of the method, while source weight assessment and consistency checks further enhance the system's robustness. The superior performance of the full model, achieved through the collaborative action of all components, provides a reliable technical pathway and theoretical support for feature selection and information fusion in multi-source information systems.

### 4.6. Comparison with other information fusion methods

In this subsection, we conducted a comparative experiment on the proposed information fusion method with five other methods for information fusion from multiple source systems of interval values. This further verifies the effectiveness of our method in information fusion. The comparison algorithms used are as follows:

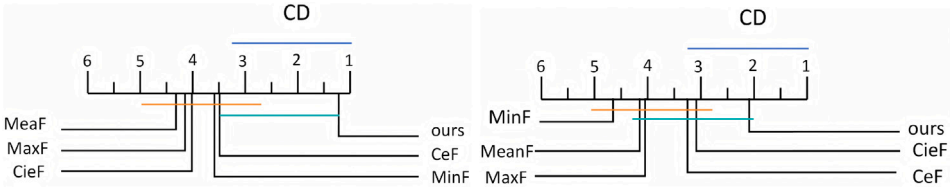
(1) MinF: $a_i^* = \min\left\{a_i^{1-}, a_i^{2-}, \ldots, a_i^{q-}\right\}$, where $a_i^{q-}$ denotes the lower bound of the interval value of the attribute $a_i$ within the $q$th information source.
(2) MaxF: $a_i^* = \max\left\{a_i^{1+}, a_i^{2+}, \ldots, a_i^{q+}\right\}$, where $a_i^{q+}$ denotes the upper bound of the interval value of the attribute $a_i$ within the $q$th information source.
(3) MeanF: $a_i^* = mean\left\{\frac{a_i^{1-}+a_i^{1+}}{2} \cdots, \frac{a_i^{q-}+a_i^{q+}}{2}\right\}$, which uses the midpoint value of the interval as the endpoint.
(4) CieF [34]: Improved Entropy Fusion Method for Multi-Source Incomplete Interval Datasets.
(5) CeF [24]: Dynamic Fusion in Multi-Source Interval-Valued Systems.

The experimental results are presented in Tables 12 and 13, where the best results are highlighted in bold and the runner-up results are underlined to facilitate comparison. Overall, the data demonstrate that our proposed NREW method delivers superior performance and adaptability across a diverse range of classifiers and datasets.

In the context of Decision Tree (DT) classifiers, NREW (80%) emerged as the most robust configuration. It not only achieved the highest average accuracy of 82.28% by securing top ranks on LR, Statlog, and Wave but also demonstrated remarkable consistency by claiming runner-up positions on two additional datasets. While NREW (60%) did not achieve the highest accuracy on any single dataset, it played a crucial intermediary role; it maintained commendable stability with runner-up results on two datasets, yielding an average accuracy significantly higher than the 40% version. In contrast, although the 40% configuration exhibited the poorest overall stability, it surprisingly outperformed others on specific datasets like Wine, Rice, and PDC. This divergence suggests that while low k values act as effective noise filters for datasets with low feature redundancy, the higher retention rate of NREW (80%) is essential for preserving the discriminative information required by more complex data distributions.

The results from Support Vector Machine (SVM) classifiers further corroborate the efficacy of NREW, showcasing a clear performance trajectory dependent on the k value. NREW (80%) dominated almost universally, achieving a peak average accuracy of 90.07% by winning on OROHD, Rice, Statlog, and Wave, while also securing runner-up spots on Wine, PDC, and LR. The NREW (60%) configuration also showed a marked improvement over the 40% version, frequently achieving second-best results and effectively bridging the performance gap. This distinct upward trend implies that SVMs, which rely on constructing hyperplanes in high-dimensional spaces, are particularly sensitive to feature space collapse. The aggressive reduction at $k = 40\%$ likely discards support vectors essential for defining optimal class margins, whereas the 60% and 80% configurations retain sufficient structural information to maximize classification accuracy.

Ultimately, the choice of k serves as a critical regulator between information preservation and noise reduction. A lower k (40%) can exploit specific low-redundancy scenarios but lacks generalizability. A medium




(a) Average Ranking result On DT (b) Average Ranking result On SVM

**Fig. 5.** The result of Friedman test and Nemenyi test.

k (60%) offers a stable compromise, successfully balancing the extremes. However, the higher k (80%) represents the optimal trade-off, combining high accuracy with broad robustness. By consistently delivering superior best and second-best performances across varying conditions, NREW (80%) verifies its capability to enhance feature selection while maintaining the strong generalization required for practical engineering applications.

### *4.7. Statistical analysis*

In this subsection, we conduct a comprehensive statistical comparison of six information fusion algorithms based on their classification performance using SVM and DT classifiers. The evaluation is performed through the following rigorous statistical analysis procedure:

First, we employ the Friedman test to assess whether there exist statistically significant differences among the algorithms across all nine datasets. The Friedman test is particularly suitable for this comparison as it is a non-parametric method designed for multiple comparisons of matched samples. The test is conducted at a significance level of $\alpha = 0.05$, with the null hypothesis stating that all compared algorithms perform equivalently. The Friedman statistic [35] is calculated as follows:

$$\tau_{\mathrm{F}} = \frac{(N-1)\tau_{\chi}^2}{N(k-1)-\tau_{\chi}^2} \tag{20}$$

where

$$\tau_{\chi^2} = \frac{12N}{k(k+1)}\left(\sum_{i=1}^{k} r_i^2 - \frac{k(k+1)^2}{4}\right) \tag{21}$$

Here, $k$ represents the number of algorithms, $N$ represents the number of datasets, and $r_i$ represents the average ranking of a certain method across all datasets. When N is sufficiently large, $\tau_{\mathrm{F}}$ follows an F distribution with (k-1, N-1) degrees of freedom. The Friedman test results for both classifiers showed statistically significant p-values, which were 0.0097 (for the SVM) and 0.0412 (for the decision tree), respectively. Both were lower than the significance threshold $\alpha = 0.05$ that we had set beforehand. This strong statistical evidence led us to decisively reject the null hypothesis, indicating that there are differences among all the algorithms. Furthermore, we conducted a Nemenyi test. The null hypothesis was that all the algorithms were the same. The null hypothesis was rejected when the average ranking gap was greater than the CD critical value. The critical value [36] domain can be computed as follows:

$$CD = q_{\alpha}\sqrt{\frac{k(k+1)}{6N}} \tag{22}$$

where $q_{\alpha} = 2.576$ when $k = 6$, $N = 9$.

The results of Nemenyi's post hoc analysis are shown in Fig. 5. In the Decision Tree classifier, our method demonstrates significant performance advantages over the MinF, MaxF, MeanF, and CieF algorithms; in the SVM classifier, our method also outperforms the MinF algorithm significantly. Regardless of whether it is in the Decision Tree or in SVM classifiers, our method consistently ranks first. This result strongly demonstrates the outstanding performance and wide applicability of our algorithm.

## 5. Conclusion

This study establishes a leader-aware feature selection and fusion framework for interval-valued multi-source systems. By synergizing neighborhood rough set theory with information entropy, we address the dual challenges of uncertainty management and consensus extraction. The methodology makes three key contributions: incorporating a mixed Hausdorff distance to capture interval structures, implementing a double-layer median adaptive radius for robust granularity construction, and designing a consistency-driven "Leader Source" mechanism to prevent information blurring. Experimental results on UCI datasets confirm the framework's superior accuracy and stability, highlighting its potential in critical domains such as medical diagnostics.

Practically, the framework exhibits substantial potential, particularly in critical domains like multi-source medical data fusion. In scenarios involving combined diagnostics (e.g., CT, MRI, and PET) where features are represented as interval values, our method enables adaptive fusion across heterogeneous devices. It evaluates the discriminative importance of specific features and designates the most representative device data as the dominant source, thereby providing physicians with more robust and interpretable decision support.

Despite these advancements, two primary limitations restrict the model's broader applicability. First, the computational complexity of constructing neighborhoods and calculating approximate granularity remains high, specifically exhibiting quadratic growth with sample size, which hinders efficiency in high-dimensional or large-scale scenarios. Second, the current entropy-based feature ranking treats attributes independently, failing to account for feature redundancy and higher-order correlations, which may limit the compactness of the selected subset.

Future research will specifically address these bottlenecks through two concrete directions: 1) To enhance computational efficiency, we will explore acceleration techniques such as Locality-Sensitive Hashing (LSH) and develop incremental learning mechanisms for dynamic data environments; 2) To optimize feature selection, we aim to integrate Graph Neural Networks (GNNs) or dependency analysis to explicitly model complex feature interactions and redundancies, thereby refining the discriminative power of the final subset. In summary, this work provides a solid theoretical foundation for interval-valued fusion, and addressing these future avenues will significantly bolster its scalability in complex decision-making systems.

## CRediT authorship contribution statement

**Hao Yuan:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation. **Weihua Xu:** Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

## Declaration of competing interest

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

## Acknowledgments

This paper is supported in part by the National Natural Science Foundation of China (No. 62376229) and the Natural Science Foundation of Chongqing, China (No. CSTB2023NSCQ-LZX0027).

## Data availability

The link of used data for the research described have been shared in the article.

## References

[1] Q. Hu, D. Yu, J. Liu, C. Wu, Neighborhood rough set based heterogeneous feature subset selection, Inf. Sci. 178 (18) (2008) 3577–3594.
[2] G.M. Marakas, J.A. O'brien, Management Information Systems, vol. 6, McGraw-Hill Irwin New York, NY, USA, 2006.
[3] R. Haux, Health information systems – past, present, future, Int. J. Med. Inform. 75 (3) (2006) 268–281.
[4] T. Bernhardsen, Geographic Information Systems: an Introduction, John Wiley & Sons, 2022.
[5] J.T. Yao, A.V. Vasilakos, W. Pedrycz, Granular computing: perspectives and challenges, IEEE Trans. Cybern. 43 (6) (2013) 1977–1989.
[6] Z. Pawlak, J. Grzymala-Busse, R. Slowinski, W. Ziarko, Rough sets, Commun. ACM 38 (11) (Nov 1995) 88–95.
[7] X. Zhang, X. Shen, Graph-driven feature selection via granular-rectangular neighborhood rough sets for interval-valued data sets, Appl. Soft Comput. 170 (2025) 112716.
[8] L. Chen, L. Zhao, Z. Xiao, Y. Liu, J. Wang, A granular computing based classification method from algebraic granule structure, IEEE Access 9 (2021) 68118–68126.
[9] C. Ma, L. Zhang, W. Pedrycz, W. Lu, The long-term prediction of time series: a granular computing-based design approach, IEEE Trans. Syst. Man Cybern. Syst. 52 (10) (2022) 6326–6338.
[10] X. Zhu, W. Pedrycz, Z. Li, Development and analysis of neural networks realized in the presence of granular data, IEEE Trans. Neural Netw. Learn. Syst. 31 (9) (2020) 3606–3619.
[11] X. Zhang, F. Liu, Probabilistic linguistic three-way decisions: integrating prospect theory with fuzzy possibilistic c-means clustering, Fuzzy Sets and Systems 517 (2025) 109442.
[12] L. Sun, L. Wang, W. Ding, Y. Qian, J. Xu, Feature selection using fuzzy neighborhood entropy-based uncertainty measures for fuzzy neighborhood multigranulation rough sets, IEEE Trans. Fuzzy Syst. 29 (1) (2021) 19–33.
[13] S. Xia, C. Wang, G. Wang, X. Gao, W. Ding, J. Yu, Y. Zhai, Z. Chen, Gbrs: a unified granular-ball learning model of Pawlak rough set and neighborhood rough set, IEEE Trans. Neural Netw. Learn. Syst. 36 (1) (2025) 1719–1733.
[14] K. Yuan, D. Miao, W. Pedrycz, W. Ding, H. Zhang, Ze-hfs: zentropy-based uncertainty measure for heterogeneous feature selection and knowledge discovery, IEEE Trans. Knowl. Data Eng. 36 (11) (2024) 7326–7339.
[15] J. Yang, X. Wang, G. Wang, Q. Zhang, N. Zheng, D. Wu, Fuzziness-based three-way decision with neighborhood rough sets under the framework of shadowed sets, IEEE Trans. Fuzzy Syst. 32 (9) (2024) 4976–4988.
[16] L. Sun, S. Si, W. Ding, X. Wang, J. Xu, Tfsfb: two-stage feature selection via fusing fuzzy multi-neighborhood rough set with binary whale optimization for imbalanced data, Inf. Fusion. 95 (2023) 91–108.
[17] C. Aliferis, I. Guyon, Causal feature selection, Comput. Methods Feature Sel, Chapman Hall/CRC (2007) 79–102.
[18] X. Sun, P. Zeng, A DS evidence theory-based method for multi-source information fusion in advanced systems: synergistic intelligence, in: Proceedings of the 2024 4th International Conference on Big Data, Artificial Intelligence and Risk Management, ICBAR '24, Association for Computing Machinery, New York, NY, USA, 2025, pp. 503–508.
[19] Q. Zhang, P. Zhang, T. Li, Information fusion for large-scale multi-source data based on the Dempster-Shafer evidence theory, Inf. Fusion. 115 (2025) 102754.
[20] A. Blanco-Fernández, N. Corral, G. González-Rodríguez, Estimation of a flexible simple linear model for interval data based on set arithmetic, Comput. Stat. Data Anal. 55 (9) (2011) 2568–2578.
[21] P. D'Urso, J.M. Leski, Fuzzy c-ordered medoids clustering for interval-valued data, Pattern Recognition 58 (2016) 49–67.
[22] S. Dias, P. Brito, Off the beaten track: a new linear model for interval data, Eur. J. Oper. Res. 258 (3) (2017) 1118–1130.
[23] Z. Zhao, J. Wu, B. Wang, R. Wang, Research on source-load uncertainty optimal scheduling based on a hybrid robust multi-interval optimization method, Renew. Energy 251 (2025) 123316.
[24] X. Zhang, X. Chen, W. Xu, W. Ding, Dynamic information fusion in multi-source incomplete interval-valued information system with variation of information sources and attributes, Inf. Sci. 608 (2022) 1–27.
[25] J. Liu, X. Zhao, R. Luo, Z. Tao, A novel link prediction model for interval-valued crude oil prices based on complex network and multi-source information, Appl. Energy. 376 (2024) 124261.
[26] Z. Li, J. Liu, Y. Peng, C.-F. Wen, A novel method to information fusion in multi-source incomplete interval-valued data via conditional information entropy: application to mutual information entropy based attribute reduction, Inf. Sci. 658 (2024) 120011.
[27] C.E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (3) (1948) 379–423.
[28] Q. Chen, Q. Chen, X. Zhao, W. Sun, C. Wang, A node importance ranking method based on the rate of network entropy changes, in: 2021 17th International Conference on Mobility, Sensing and Networking (MSN), 2021, pp. 32–39.
[29] C. Li, K. Noman, Z. Liu, K. Feng, Y. Li, Optimal symbolic entropy: an adaptive feature extraction algorithm for condition monitoring of bearings, Inf. Fusion. 98 (2023) 101831.
[30] B. Sang, L. Yang, W. Xu, H. Chen, T. Li, W. Li, VCOS: multi-scale information fusion to feature selection using fuzzy rough combination entropy, Inf. Fusion 117 (2025) 102901.
[31] A. Malhi, R.X. Gao, Pca-based feature selection scheme for machine defect classification, IEEE Trans. Instrum. Meas. 53 (6) (2004) 1517–1525.
[32] X. He, D. Cai, P. Niyogi, Laplacian score for feature selection, in: Proceedings of the 19th International Conference on Neural Information Processing Systems, NIPS '05, MIT Press, Cambridge, MA, USA, 2005, pp. 507–514.
[33] L. Breiman, Breiman, L. Random forests, Mach. Learn. (2001) 5–32.
[34] W. Xu, K. Cai, D.D. Wang, A novel information fusion method using improved entropy measure in multi-source incomplete interval-valued datasets, Int. J. Approx. Reason. 164 (2024) 109081.
[35] M. Friedman, A comparison of alternative tests of significance for the problem of $m$ rankings, Ann. Math. Stat. 11 (1) (1940) 86–92.
[36] J. Demšar, Statistical comparisons of classifiers over multiple data sets, J. Mach. Learn. Res. 7 (Dec 2006) 1–30.

## Author biography

**Weihua Xu** received the Ph.D. degree in mathematics from the School of Sciences, Xi'an Jiaotong University, Xi'an, China, in 2007, and the M.Sc. degree in mathematics from the School of Mathematics and Information Sciences, Guangxi University, Nanning, China, in 2004. He is currently a Professor with the College of Artificial Intelligence, Southwest University, Chongqing, China. He has published five monographs and more than 240 articles in international journals. His current research interests include granular computing, cognitive computing, and information fusion. Dr. Xu also serves as a Senior Member of the Chinese Association for Artificial Intelligence (CAAI). He is an Associate Editor of the *International Journal of Machine Learning and Cybernetics* and the *Journal of Intelligent and Fuzzy Systems*.

**Hao Yuan** is working toward the B.Sc. degree in intelligent science at Southwest University, Chongqing, China. His research interests include feature selection and granular computing.