Systematic Feature Selection Based on Three-Level Improvements of Fuzzy Dominance Three-Way Neighborhood Rough Sets

Xianyong Zhang, Benwei Chen, and Duoqian Miao

***Abstract*—Feature selection facilitates system processing, and it relies on knowledge granulation and uncertainty measurement. Focusing on ordered decision systems, the fuzzy dominance neighborhood (FDN) granulation and corresponding condition entropy have recently yielded an outstanding algorithm for feature selection, FDNCE-FS (fuzzy dominance neighborhood condition entropy-based feature selection). However, there is room for improvement. Accordingly, three-level improvements of knowledge granulation, information enrichment, and heterogeneity fusion are proposed here, and $2 \times 2 \times 2 = 8$ heuristic algorithms of feature selection are systematically established. First, FDN granulation is improved to fuzzy dominance three-way neighborhood (FD3N) granulation through three-way decision on fuzzy dominance degrees, and FD3N rough sets are modeled to offer better dependency. Second, the FDN condition entropy is improved to FD3N condition entropy by reinforcing the interaction factor and class information, and corresponding measure systems are constructed. Third, FD3N dependency is fused with four types of condition entropy to produce four combined measures, and eight uncertainty measures hierarchically emerge due to the three-level improvements. Fourth, these systematic measures have granulation nonmonotonicity, and they enable heuristic algorithms for feature selection; thus, the current FDNCE-FS method is improved to seven new selection algorithms: FHN-FS, RHN-FS, RFHN-FS, HTWN-FS, FHTWN-FS, RHTWN-FS, and RFHTWN-FS. Finally, the relevant FD3N granulation, uncertainty measurement, and feature selection are validated by data-based experiments, and the seven novel algorithms are shown to outperform FDNCE-FS in terms of classification performance. This study provides new insights into uncertainty modeling, information fusion, and feature selection through granular computing and three-way decision.**

***Index Terms*—Feature selection, fuzzy dominance three-way neighborhood rough set (FD3NRS), granular computing (GrC), ordered decision system, three-way decision (3WD), uncertainty measurement.**

Manuscript received 30 October 2023; revised 23 May 2024; accepted 29 July 2024. Date of publication 2 August 2024; date of current version 4 September 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62376198 and Grant 61976158, in part by the Natural Science Foundation of Sichuan Province of China under Grant 2024NSFSC0486 and Grant 2024NSFSC0443, in part by the Sichuan Science and Technology Program of China under Grant 2022ZYD0001, and in part by the Humanities and Social Sciences Project of the Ministry of Education of China under Grant 23YJA630114. Recommended by Associate Editor Y. Yao.
(Corresponding author: Xianyong Zhang.)

Xianyong Zhang and Benwei Chen are with the School of Mathematical Sciences, Sichuan Normal University, Chengdu 610066, China (e-mail: xianyongzh@sina.com.cn; benweich@sina.com.cn).

Duoqian Miao is with the Department of Computer Science and Technology, Tongji University, Shanghai 201804, China (e-mail: dqmiao@tongji.edu.cn).
Digital Object Identifier 10.1109/TFUZZ.2024.3437367

I. Introduction

FEATURE selection aids data analysis in complex and intelligent systems. It captures effective features to reduce the data dimensions and improve recognition abilities. Feature selection methods have been researched for data mining [1], [2], [3], [4], information processing [5], [6], [7], machine learning [8], [9], [10], etc. In particular, feature selection is closely related to attribute reduction in rough set theory [11], so rough sets and attribute reducts provide strong support for such methods.

Rough sets involve bidirectional cognition in granular computing (GrC) and three-way decision (3WD), and they support reasoning methods for imprecise, inconsistent, and incomplete information. Rough sets are applied in GrC [12] and 3WD [13], [14]. Classical rough sets aid general information/decision systems, in which sample orders are not available. In practice, attribute values may have ordered structures, and the order information is valuable for data mining. Accordingly, ordered information/decision systems (OISs/ODSs) have emerged, and they are handled by dominance-based rough sets (DRSs) [15]. For OISs/ODSs, the dominance degrees between objects determine dominance relations and knowledge granulation, so these grounded measures enable DRSs to recognize preferred decision classes.

When processing numerical data in OISs/ODSs, small measurement fluctuations and data noise easily affect dominance degrees and relations; thus, DRSs are sensitive in uncertainty processing, and related studies on robustness employ attribute reduction [16], [17], [18]. In particular, DRSs have been effectively extended by introducing fuzzy sets and neighborhood granulation. For example, Greco et al. [19] proposed dominance-based fuzzy rough sets by introducing fuzzy logic into dominance relations; Hu et al. [20] introduced fuzzy dominance degrees (FDDs) to obtain fuzzy dominance rough sets (FDRSs); Chen et al. [21] proposed dominance-based neighborhood rough sets (DNRSs) by using neighborhood dominance relations. In contrast, DNRSs further consider noise, but their neighborhood boolean relations hinder the effective measurement of dominance degrees. To address this issue, Sang et al. [22] proposed fuzzy dominance neighborhood rough sets (FDNRSs) by using fuzzy dominance neighborhood (FDN) granulation on FDDs; the corresponding fuzzy dominance neighborhood condition entropy (FDNCE) led to an effective reduction algorithm called

FDNCE-FS. By observation, FDNs and FDNRSs can handle only a one-way modification of uncertain granulation with a central value of 0.5, and they ignore the impact of noise at the extreme values 0 and 1 [22]; thus, we can incorporate two-way corrections for extreme $(0, 1)$ valuations and certain granulation. Overall, this article presents fuzzy dominance three-way neighborhood (FD3N) granulation to model fuzzy dominance three-way neighborhood rough sets (FD3NRSs), and this new model improves FDNRSs to motivate more robust uncertainty measurement and feature selection.

Feature selection depends on uncertainty measurement based on knowledge granulation, and the quality metric determines the resulting optimization efficiency and learning performance. Uncertainty measures include the algebraic measure [11], information measure [23], and algebra–information combination [24]. The latter's heterogeneous fusion ability accelerates the reduction optimization [25], [26], [27], [28]. For selection measures in ODSs, algebraic measures concern knowledge granularity, regional metrics, and rough dependencies [17], [18], [20]; fuzzy information measures include the fuzzy information/condition/joint entropy [29] and fuzzy ranking condition entropy [30], FDNCE [22]; however, fusion measures are rare. Accordingly, information reduction is important, and metric granulation monotonicity/nonmonotonicity underlies monotonic/nonmonotonic feature selection methods, such as the monotonic reduction algorithms DCE-FS [31] and FDCE-FS [30] and nonmonotonic algorithms NDCE-FS [21] and FDNCE-FS [22].

Regarding feature selection for ODSs, there are some research limitations and development thoughts as follows, in terms of uncertainty measures. 1) Knowledge granulation underlies interaction measurement, and it can incorporate FDDs [20]. FDDs introduce the log–sigmoid transfer functions for stability and applicability. On this basis, FDNs and FDNRSs can address noise sensitivity, but their one-way tolerance granulation is not satisfactory [22]. Furthermore, we plan to formulate FD3Ns and FD3NRSs for better robustness. 2) Typical information measurement can use two internal factors

$$P = \frac{|condition{-}granule \ \cap \ decision{-}class|}{|condition{-}granule|},$$
$$FP = \frac{|condition{-}granule \ \cap \ decision{-}class|^2}{|condition{-}granule| \times |decision{-}class|}; \qquad (1)$$

here P means the condition probability, while FP reinforces both the interaction information and class cardinality to improve quantification [32], [33]. For ODSs, condition entropies (CEs) are commonly used, and relevant formulas consider P but never consider FP [22], [29], [30]; thus, we aim to improve FDNCE [22] by replacing P with FP. Furthermore, it would be worthwhile to fuse FDNCE and its promotional CEs with model dependency to produce more powerful measures—DCEs (i.e., dependency-fused CEs)—as shown by various algebra–information fusion cases [26], [27], [28]. 3) For ODSs, monotonic feature selection is common [29], [30], while nonmonotonic reduction is rare; in practice, the latter becomes necessary and valuable for ordering systematicity and complex contexts. Recently, nonmonotonic FDNCE-FS [22] outperformed three comparison methods: DCE-FS [31], FDCE-FS [30], and NDCE-FS [21]. Thus, we still consider nonmonotonic selection for theoretical nonmonotonicity, but we also emphasize and utilize practical monotonicity trends for the optimized design of selection algorithms.

According to the above considerations, this article uses uncertainty measurement and feature selection in ODSs to improve the methods of recent studies discussed in [22]. Concretely, we perform three-level improvements to FDD granulation, CE enrichment, and DCE fusion, and $2 \times 2 \times 2$-systematic selection algorithms are established from a nonmonotonic perspective. Our research framework is shown in Fig. 1, and the relevant content involves 3WD, GrC, and their combination [34], [35], [36]. The main abbreviations are given in Table I.

1) At the knowledge granulation level, three-way revised FDDs on $(0, 0.5, 1)$ are proposed to improve one-way revised FDDs on 0.5 [22] and initial FDDs [20], yielding FD3N to improve FDN for better granulation robustness.
2) At the single-view level of condition-decision interaction, FD3N-based FD3NRSs are modeled to improve FDN-based FDNRSs [22], and a corresponding algebraic dependency $\gamma_B^\delta(D^{\geq})$ arises to aid this improvement. In addition, three new CEs are used to improve the existing FDNCE on (P, FDN) [22] by using 2-D FP deepening and FD3N extensions; moreover, metric systems of CE, informational and joint entropies, and mutual information are established.
3) At the double-view level of algebra–information fusion, $\gamma_B^\delta(D^{\geq})$ is combined with four CEs via $\times$, and thus four DCEs for 2-D $(P, FP) \times (\text{FDN}, \text{FD3N})$ emerge.
4) The above three-level improvements motivate $2 \times 2 \times 2 = 8$ uncertainty measures (where only FDNCE exists [22]), and all measures achieve granulation nonmonotonicity and potential monotonicity. A total of eight nonmonotonic selection algorithms are systematically constructed, and seven new algorithms improve 1 current method FDNCE-FS (i.e., HN-FS) [22].
5) Finally, both the uncertainty measures of FDDs, CEs, and DCEs and the heuristic algorithms of feature selection are validated by data-based experiments, and our new algorithms achieve better classification performance.

Regarding contributions, this study deeply provides both hierarchical improvements of uncertainty measurement and systematic algorithms of feature selection, in terms of ODSs.

The rest of this article is organized as follows. Section II implements three-way FDD corrections to establish FD3Ns and FD3NRSs. Section III adopts the FDN and FD3N to construct and research CE-central information systems and combinational DCEs. Section IV relies on three-level measurement improvements to design $2 \times 2 \times 2$ selection algorithms. Section V validates the uncertainty measurement and feature selection via data-based experiments. Finally, Section VI concludes this article.

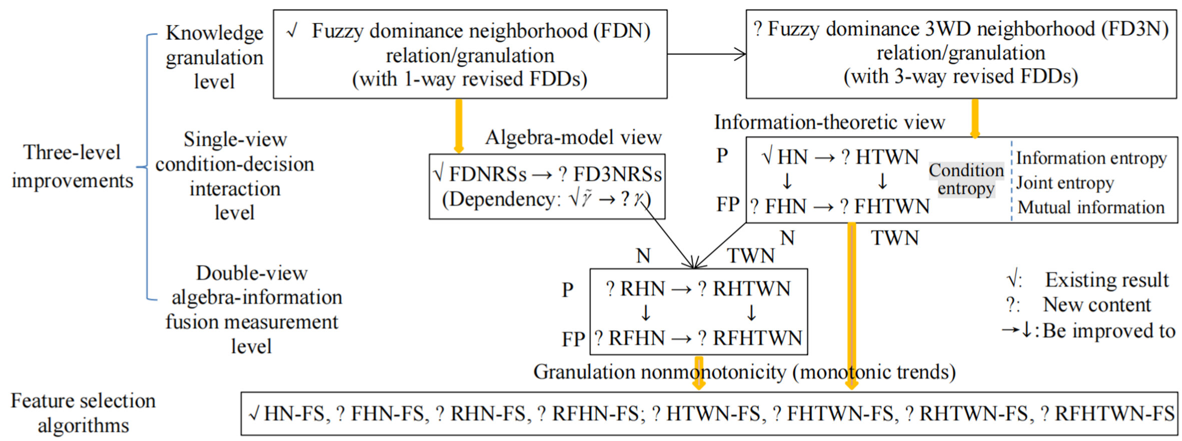

Fig. 1. Three-level improvements and systematic feature selection.

TABLE I
MAIN ABBREVIATIONS AND LABELS OF THIS ARTICLE

Items	Abbreviations/symbols	Corresponding expressions/meanings.
Methodologies	3WD, GrC	Three-way decision, granular computing.
Data contexts	ODSs/OISs	Ordered decision/information systems.
Rough set models	DRSs	Dominance-based rough sets [15].
	FDRSs	Fuzzy dominance rough sets [20].
	DNRSs	Dominance-based neighborhood rough sets [21].
	FDNRSs	Fuzzy dominance neighborhood rough sets [22].
	FD3NRSs	Fuzzy dominance three-way neighborhood rough sets.
Uncertainty measures	FDDs	Fuzzy dominance degrees.
	FDN (from N)	Fuzzy dominance neighborhood (from one-way correctional FDDs) [22].
	FD3N (from TWN)	Fuzzy dominance three-way neighborhood (from three-way correctional FDDs).
	CEs, DCEs	Condition entropies, dependency-fused condition entropies.
	FDNCE	Fuzzy dominance neighborhood condition entropy [22].
	H, FH, RH, RFH	CE $H^{\delta}(D\|B)$, CE $FH^{\delta}(D\|B)$, DCE $RH^{\delta}(D\|B)$, DCE $RFH^{\delta}(D\|B)$.
Feature selection algorithms	HN-FS, FHN-FS, RHN-FS, RFHN-FS	Heuristic algorithms on (H, N), (FH, N), (RH, N), (RFH, N).
	HTWN-FS, FHTWN-FS, RHTWN-FS, RFHTWN-FS	Heuristic algorithms on (H, TWN), (FH, TWN), (RH, TWN), (RFH, TWN).
	FDNCE-FS [22], DCE-FS [31], FDCE-FS [30], NDCE-FS [21]	Four existing heuristic algorithms.

II. FUZZY DOMINANCE THREE-WAY NEIGHBORHOOD ROUGH SETS

A decision system is defined as $DS = (U, AT = C \cup D, V, f)$ [11]. Here, $U = \{x_1, x_2, \ldots, x_n\}$ denotes the nonempty finite universe, $C = \{c_k | k = 1, 2, \ldots, r\}$ and $D = \{d\}$ represent sets of condition and decision attributes, respectively, $V = \bigcup_{a \in C \cup D} V_a$ determines the value range, and $f : U \times (C \cup D) \to V$ is the mapping function. Furthermore, the ODS is established by adding an order

$$\preceq_a: x \preceq_a y \Leftrightarrow f(x, a) \leq f(y, a) \quad (x, y \in U, a \in C \cup D).$$

An attribute subset $B \subseteq C$ induces a dominance relation

$$R_B = \{(x, y) \in U \times U \mid f(x, a) \leq f(y, a), \forall a \in B\} \quad (2)$$

and there are two types of knowledge granules

$$\text{(Dominating class)} \quad [x]_B^+ = \{y \in U \mid (x, y) \in R_B\}$$

$$\text{(Dominated class)} \quad [x]_B^- = \{y \in U \mid (y, x) \in R_B\}. \quad (3)$$

For the decision part of $D = \{d\}$, a value set $V_D = \{d_1, \ldots, d_M\}$ provides a preference sorting $d_1 < \cdots < d_M$, and decision classes $D_k = \{x \in U \mid f(x, d) = d_k\}$ $(k = 1, \ldots, M)$ generate an equivalent partition $U/D = \{D_k \mid k = 1, \ldots, M\}$ and relevant ordering $D_1 \preceq \cdots \preceq D_M$. Upward and downward preference decision classes of D_k become

$$D_k^{\geq} = \cup_{k' \geq k} D_{k'}, D_k^{\leq} = \cup_{k' \leq k} D_{k'}. \quad (4)$$

DRSs mainly describe $D_k^{\geq}, D_k^{\leq}$ via $[x]_B^+, [x]_B^-$ [15].

According to ODSs, DRSs underlie uncertainty modeling and dependency learning. In terms of noise treatments, DNRSs supplement DRSs by introducing the distance measurement and neighborhood granulation [21], but they face object-ranking issues due to dependency. FDNRSs exhibit corresponding improvements based on fuzziness and tolerance [22], and their neighborhood connotation is related to the 0.5-centered correction and uncertainty of FDD, thus including only one-way granulation. Furthermore, extreme FDDs 0,1 and certainty trends are worth incorporating to obtain three-way granulation. Then, the three-way neighborhood mechanism naturally motivates FD3NRSs, and this new model further improves DNRSs [21] and FDNRSs [22].

A. FDD Improvements

Fuzzy dominance relations adopt fuzzy expressions, and this differentiates them from general dominance relations [see (2)].

They can be concretized by FDDs [20], i.e.,

$$D_B^{\prec}(x_i, x_j) = \min_{a \in B} D_a^{\prec}(x_i, x_j)$$

$$D_a^{\prec}(x_i, x_j) = \frac{1}{1 + e^{-K[f(x_j,a) - f(x_i,a)]}} \tag{5}$$

where $K \in \mathbb{N}^+$; let $K = 10$. For $\beta \in [0.4, 0.5), \alpha \in (0.5, 0.6]$, FDDs are revised according to [22] as

$$N_B^{\preceq}(x_i, x_j) = \begin{cases} 0.5, & \text{if } D_B^{\preceq}(x_i, x_j) \in [\beta, \alpha] \\ D_B^{\preceq}(x_i, x_j), & \text{otherwise.} \end{cases} \tag{6}$$

$N_B^{\preceq}(x_i, x_j)$ concerns two operations. Suppose $\bigwedge_{a \in B}$ represents the joint integration of attributes and minimum values, while $\mathcal{A}_{0.5}^E$ represents the equivalent amendment based on the central value 0.5. Operational commutativity, $\bigwedge_{a \in B} \circ \mathcal{A}_{0.5}^E = \mathcal{A}_{0.5}^E \circ \bigwedge_{a \in B}$, holds, so we focus on a fundamental correction to the single attribute

$$N_a^{\prec}(x_i, x_j) = \begin{cases} 0.5, & \text{if } D_a^{\prec}(x_i, x_j) \in [\beta, \alpha] \\ D_a^{\prec}(x_i, x_j), & \text{otherwise.} \end{cases} \tag{7}$$

We can add the properties of $N_a^{\prec}(x_i, x_j)$: $N_a^{\prec}(x_i, x_j) \in (0, 1)$, $N_a^{\prec}(x_i, x_i) = 0.5$, and $N_a^{\prec}(x_i, x_j) + N_a^{\prec}(x_j, x_i) = 1$, which accord with those of $D_a^{\prec}(x_i, x_j)$. $N_a^{\prec}(x_i, x_j) + N_a^{\prec}(x_j, x_i) = 1$ determines $N_a^{\prec}(x_i, x_j) = 0.5 \Leftrightarrow N_a^{\prec}(x_j, x_i) = 0.5$, so this case implies $D_a^{\prec}(x_i, x_j) \in [\beta, \alpha] \Leftrightarrow D_a^{\prec}(x_j, x_i) \in [\beta, \alpha] \Leftrightarrow D_a^{\prec}(x_i, x_j) \in [1 - \alpha, 1 - \beta]$; thus, we offer a symmetrical condition $\alpha + \beta = 1$. This parametric requirement naturally applies to both $N_a^{\prec}(x_i, x_j)$ and $N_B^{\preceq}(x_i, x_j)$, so we alternatively use $N_a^{\delta}(x_i, x_j)$ and $N_B^{\delta}(x_i, x_j)$ by setting $\alpha = 0.5 + \delta, \beta = 0.5 - \delta$ with only the neighborhood threshold $\delta \in [0, 0.1]$.

Definition 1: One-way neighborhood FDDs of x_j over x_i on attribute $a \in C$ and subset $B \subseteq C$ are, respectively

$$N_a^{\delta}(x_i, x_j) = \begin{cases} 0.5, \text{ if } D_a^{\prec}(x_i, x_j) \in [0.5 - \delta, 0.5 + \delta] \\ D_a^{\prec}(x_i, x_j), \text{ otherwise} \end{cases} \tag{8}$$

$$N_B^{\delta}(x_i, x_j) = \bigwedge_{a \in B} N_a^{\delta}(x_i, x_j) \tag{9}$$

which constitute matrices $\mathbf{N}_a^{\delta} = [N_a^{\delta}(x_i, x_j)]_{n \times n}$ and $\mathbf{N}_B^{\delta} = [N_B^{\delta}(x_i, x_j)]_{n \times n}$, respectively.

$N_a^{\delta}(x_i, x_j), N_B^{\delta}(x_i, x_j)$ are individual attribute-driven and parameter-driven expressions, where $\delta = 0$ causes degradation for $D_a^{\prec}(x_i, x_j), D_B^{\preceq}(x_i, x_j)$. Therefore, $N_B^{\delta}(x_i, x_j)$ simplifies $N_B^{\preceq}(x_i, x_j)$ [22], and it has a better granulating mechanism and calculation superiority.

Next, our 3WD strategy of neighborhood extension is clarified by schematic diagrams. FDDs $D_{a_1}^{\prec}(x_i, x_j)$ $(i, j \in \{1, \ldots, 10\})$ in Table II are depicted in Fig. 2, where the horizontal and vertical directions are related to x_j and x_i, respectively. 1) Some FDD values are very close to 0.5, and the relevant objects can be considered to have no differences due to noise effects. Thus, $N_B^{\preceq}(x_i, x_j)$ [22] and $N_B^{\delta}(x_i, x_j)$ are proposed. The basic strategy of the one-way neighborhood is shown in Fig. 2(a), which carries equivalent $N_{a_1}^{\prec}(x_i, x_j)$ (with $\beta = 0.4, \alpha = 0.6$) and $N_{a_1}^{\delta}(x_i, x_j)$ (with $\delta = 0.1$). 2) The other two cases of 0 and 1 can be similarly processed, and thus we propose a systematic strategy of three-way neighborhoods. As shown by Fig. 2(b) with small γ and large ϕ, when the FDDs are in $[0, \gamma]$ (or $[\phi, 1]$), the two samples have very different values, so their FDDs can be considered to be 0 (or 1).

By the above analysis, $N_a^{\prec}(x_i, x_j)$ can be extended to

$$\mathrm{TWN}_a^{\prec}(x_i, x_j) = \begin{cases} 0.5, & \text{if } D_a^{\prec}(x_i, x_j) \in [\beta, \alpha] \\ 1, & \text{if } D_a^{\prec}(x_i, x_j) \in [\phi, 1] \\ 0, & \text{if } D_a^{\prec}(x_i, x_j) \in [0, \gamma] \\ D_a^{\prec}(x_i, x_j), & \text{otherwise} \end{cases} \tag{10}$$

where $\beta \in [0.4, 0.5]$, $\alpha \in [0.5, 0.6]$, $\phi \in [0.9, 1]$, and $\gamma \in [0, 0.1]$. Similar to the symmetrical determination $\alpha + \beta = 1$, we obtain $\gamma + \phi = 1$, so $\mathrm{TWN}_a^{\prec}(x_i, x_j)$ is simplified to

TABLE II
ORDERED DECISION SYSTEM FOR EXAMPLE DEMONSTRATION

U	c_1	c_2	c_3	c_4	d
x_1	0.28	0.28	0.36	0.27	1
x_2	0.25	0.31	0.38	0.36	1
x_3	0.60	0.42	0.68	0.74	1
x_4	0.48	0.47	0.59	0.51	2
x_5	0.42	0.51	0.57	0.54	2
x_6	0.55	0.58	0.70	0.60	2
x_7	0.78	0.71	0.68	0.80	2
x_8	0.75	0.78	0.69	0.81	3
x_9	0.83	0.80	0.98	0.82	3
x_{10}	0.85	0.91	0.96	0.88	3

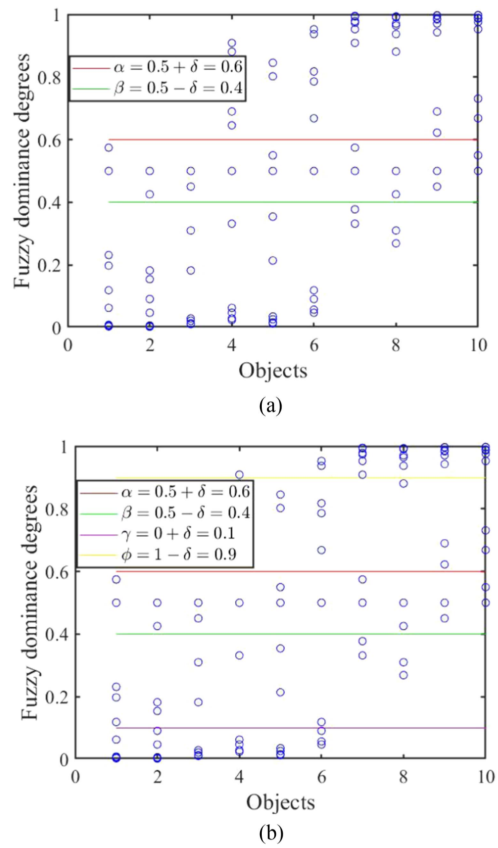

Fig. 2. Value distribution and neighborhood granulation of FDDs. (a) One-way cut and revision on 0.5. (b) Three-way cuts and revisions on (0, 0.5, 1).

$\mathrm{TWN}_a^{\delta,\theta}(x_i, x_j)$ by setting $\delta = 0.5 - \beta = \alpha - 0.5 \in [0, 0.1]$ and $\theta = 1 - \phi = \gamma - 0 \in [0, 0.1]$. For unification and convenience, $\theta = \delta$ is stipulated to yield a simplified form of FDDs.

Definition 2: The three-way neighborhood FDDs of x_j over x_i on attribute $a \in C$ and on subset $B \subseteq C$ are, respectively

$$\mathrm{TWN}_a^{\delta}(x_i, x_j) = \begin{cases} 0.5, & \text{if } D_a^{\prec}(x_i, x_j) \in [0.5-\delta, 0.5+\delta] \\ 1, & \text{if } D_a^{\prec}(x_i, x_j) \in [1-\delta, 1] \\ 0, & \text{if } D_a^{\prec}(x_i, x_j) \in [0, \delta] \\ D_a^{\prec}(x_i, x_j), & \text{otherwise} \end{cases} \tag{11}$$

$$\mathrm{TWN}_B^{\delta}(x_i, x_j) = \bigwedge_{a \in B} \mathrm{TWN}_a^{\delta}(x_i, x_j) \tag{12}$$

which induce matrices $\mathbf{TWN}_a^{\delta} = [\mathrm{TWN}_a^{\delta}(x_i, x_j)]_{n\times n}$, $\mathbf{TWN}_B^{\delta} = [\mathrm{TWN}_B^{\delta}(x_i, x_j)]_{n\times n}$, respectively.

Three-way correctional FDDs, i.e., $\mathrm{TWN}_a^{\delta}(x_i, x_j)$ and $\mathrm{TWN}_B^{\delta}(x_i, x_j)$, are proposed to extend and improve current one-way correctional FDDs, i.e., $N_B^{\prec}(x_i, x_j)$ [22]. Here, 3WD is considered. Regarding dominance orders, the value 0.5 and its neighborhood represent the greatest uncertainty and most noncommittal decision; in contrast, 0, 1 and their neighborhoods reflect the maximal certainty related to positive decision and negative decision, and the two cases correspond to the largest and smallest values. To achieve equilibrium and simplicity, we require that the 0-driven neighborhood and 1-driven neighborhood have the same threshold δ to match the 0.5-driven neighborhood range 2δ. In the exponential range $D_B^{\prec}(x_i, x_j) \in (0, 1)$, 0 and 1 serve as two limit values that are never realized in practice, and this contradicts the subjective requirements of complete certainty on $0, 1$. $N_B^{\prec}(x_i, x_j) \in (0, 1)$ has a similar constraint and defect regarding $0, 1$ realizability. In contrast, our new $\mathrm{TWN}_B^{\delta}(x_i, x_j)$ breaks through the intrinsic bottleneck of exponential characterizations because it can reach the values 0,1, so this further supports the superiority of our 3WD correction.

B. Uncertainty Modeling

Definition 3: The fuzzy three-way neighborhood dominating and dominated classes of $x_i \in U$ on $B \subseteq C$ are, respectively

$$\begin{aligned} \mathrm{TWN}_B^{\delta+}(x_i) &= \frac{\mathrm{TWN}_B^{\delta}(x_i, x_1)}{x_1} + \cdots + \frac{\mathrm{TWN}_B^{\delta}(x_i, x_n)}{x_n} \\ \mathrm{TWN}_B^{\delta-}(x_i) &= \frac{\mathrm{TWN}_B^{\delta}(x_1, x_i)}{x_1} + \cdots + \frac{\mathrm{TWN}_B^{\delta}(x_n, x_i)}{x_n}. \end{aligned} \tag{13}$$

Proposition 1: 1) $\mathrm{TWN}_{B_1}^{\delta+}(x_i) \cap \mathrm{TWN}_{B_2}^{\delta+}(x_i) = \mathrm{TWN}_{B_1\cup B_2}^{\delta+}(x_i)$. 2) If $B_1 \subseteq B_2$, then $\mathrm{TWN}_{B_1}^{\delta+}(x_i) \supseteq \mathrm{TWN}_{B_2}^{\delta+}(x_i)$. 3) If $\delta_1 \le \delta_2$, then $\mathrm{TWN}_B^{\delta_1+}(x_i) \nsubseteq \mathrm{TWN}_B^{\delta_2+}(x_i)$, $\mathrm{TWN}_B^{\delta_1+}(x_i) \nsupseteq \mathrm{TWN}_B^{\delta_2+}(x_i)$.

Based on three-way neighborhood FDDs, fuzzy knowledge granules must adhere to fuzzy sets and memberships, and they offer attribute subset monotonicity and neighborhood threshold nonmonotonicity. A measure with function $\mathcal{M}_B^{\delta}$ has attribute/granulation monotonicity if it satisfies $B_1 \subseteq B_2 \Rightarrow \mathcal{M}_{B_1}^{\delta} \le \mathcal{M}_{B_2}^{\delta}$ or $B_1 \subseteq B_2 \Rightarrow \mathcal{M}_{B_1}^{\delta} \ge \mathcal{M}_{B_2}^{\delta}$; otherwise, $\mathcal{M}_B^{\delta}$ has attribute/granulation nonmonotonicity. δ-parameter monotonicity and nonmonotonicity can be defined similarly. Furthermore, we propose FD3NRSs for improved approximation, dependency, and properties. Before modeling, we naturally define the decision part under fuzziness. Fuzzy dominating and dominated decision classes of $x_i \in U$ on decision attribute $D = \{d\}$ are defined as

$$\begin{aligned} M_D^{+}(x_i) &= \frac{M_D(x_i, x_1)}{x_1} + \cdots + \frac{M_D(x_i, x_n)}{x_n}, \\ M_D^{-}(x_i) &= \frac{M_D(x_1, x_i)}{x_1} + \cdots + \frac{M_D(x_n, x_i)}{x_n}, \end{aligned} \tag{14}$$

where $M_D(x_i, x_j) = \begin{cases} 0, & \text{if } D(x_i) < D(x_j), \\ 1, & \text{otherwise.} \end{cases}$

$M_D^{+}(x_i), M_D^{-}(x_i)$ adhere to fuzzy decision granulation for later information measurement. For uncertainty modeling, we use fuzzy upward and downward decision classes

$$\begin{aligned} D_k^{\ge}(x_j) &= \begin{cases} 0, & \text{if } x_j \notin D_k^{\ge} \text{ or } D(x_j) < d_k \\ 1, & \text{if } x_j \in D_k^{\ge} \text{ or } D(x_j) \ge d_k \end{cases} \\ D_k^{\le}(x_j) &= \begin{cases} 0, & \text{if } x_j \notin D_k^{\le} \text{ or } D(x_j) > d_k \\ 1, & \text{if } x_j \in D_k^{\le} \text{ or } D(x_j) \le d_k. \end{cases} \end{aligned} \tag{15}$$

Definition 4 (FD3NRSs): The fuzzy lower and upper approximations and the subsequent dependency of upward preference decision classes $D_k^{\ge}$ are represented as

$$\begin{aligned} \underline{\mathrm{TWN}_B^{\delta}}(D_k^{\ge})(x_i) &= \inf_{x_j \in U} \max(|\mathrm{TWN}_B^{\delta+}(x_i, x_j) - \mathrm{TWN}_B^{\delta+}(x_j, x_i)|, D_k^{\ge}(x_j)) \\ &= \inf_{x_j \in U} \max(|1 - 2\mathrm{TWN}_B^{\delta+}(x_i, x_j)|, D_k^{\ge}(x_j)) \\ \overline{\mathrm{TWN}_B^{\delta}}(D_k^{\ge})(x_i) &= \sup_{x_j \in U} \min(1 - |\mathrm{TWN}_B^{\delta+}(x_i, x_j) - \mathrm{TWN}_B^{\delta+}(x_j, x_i)|, D_k^{\ge}(x_j)) \\ &= \sup_{x_j \in U} \min(1 - |1 - 2\mathrm{TWN}_B^{\delta+}(x_i, x_j)|, D_k^{\ge}(x_j)) \end{aligned}$$

$$\gamma_B^{\delta}(D^{\ge}) = \frac{\sum_{k=1}^{M}\sum_{i=1}^{n} \underline{\mathrm{TWN}_B^{\delta}}(D_k^{\ge})(x_i)}{\sum_{k=1}^{M}\sum_{i=1}^{n} D_k^{\ge}(x_i)}. \tag{16}$$

Proposition 2: 1) $\underline{\mathrm{TWN}_B^{\delta}}(D_k^{\ge}) \subseteq D_k^{\ge} \subseteq \overline{\mathrm{TWN}_B^{\delta}}(D_k^{\ge})$. 2) The dependency $\gamma_B^{\delta}(D^{\ge}) \in [0, 1]$ has B-attribute nonmonotonicity and δ-parameter nonmonotonicity.

FD3NRSs improve the current FDNRSs [22] not only in terms of the extended neighborhood granulation but also through the formal structure of approximation cognition. According to [22], the relevant $N_B^{\delta+}(x_i)$ can be provided, and $D_k^{\ge}$-based FDNRSs satisfy

$$\underline{N_B^{\delta}}(D_k^{\ge})(x_i) = \inf_{x_j \in U} \max(1 - N_B^{\delta+}(x_i, x_j), D_k^{\ge}(x_j))$$

$$\overline{N_B^\delta}(D_k^\geq)(x_i) = \sup_{x_j \in U} \min(N_B^{\delta+}(x_i,x_j), D_k^\geq(x_j))$$

$$\widetilde{\gamma}_B^\delta(D^\geq) = \frac{\sum_{k=1}^M \sum_{i=1}^n \underline{N_B^\delta}(D_k^\geq)(x_i)}{\sum_{k=1}^M \sum_{i=1}^n D_k^\geq(x_i)}. \tag{17}$$

However, this definition cannot derive an ideal squeeze rule for bidirectional approximation and cognition. Such a squeeze rule can be blocked by $N_B^{\delta+}(x_i,x_i) = 0.5$, and it is related to a false setting $N_B^{\delta+}(x_i,x_i) \neq 1$. We obtain $1 - |N_B^{\delta+}(x_i,x_i) - N_B^{\delta+}(x_i,x_i)| = 1 - |0.5 - 0.5| = 1$ to motivate the corresponding improvements, so we convert $N_B^{\delta+}(x_i,x_j)$ to $1 - |N_B^{\delta+}(x_i,x_j) - N_B^{\delta+}(x_j,x_i)|$, which yields greater symmetry and information enrichment. Thus, (17) is first updated to an improved formula, and then, new and transitional approximations on $N_B^{\delta+}$ induce approximations on $\mathrm{TWN}_B^{\delta+}$ in turn, as shown in (16). As an improved result, $\gamma_B^\delta(D^\geq) \in [0,1]$ corrects $\widetilde{\gamma}_B^\delta(D^\geq)$, and the previous dependency may exceed 1 because its numerator may be greater than its denominator.

C. Example Illustration

Example 1: Table II shows an ODS (with size order) from [22]. For $B_1 = \{c_1\}, \delta = 0.05$, we obtain the correctional condition matrix

$$\mathbf{TWN}_{B_1}^{0.05} = \begin{bmatrix} 0.50 & 0.43 & 0.77 & 0.88 & 0.80 & 0.94 & 1.00 & 1.00 & 1.00 & 1.00 \\ 0.57 & 0.50 & 0.82 & 0.91 & 0.85 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ 0.23 & 0.18 & 0.50 & 0.69 & 0.50 & 0.82 & 1.00 & 1.00 & 1.00 & 1.00 \\ 0.12 & 0.09 & 0.31 & 0.50 & 0.35 & 0.67 & 1.00 & 0.94 & 1.00 & 1.00 \\ 0.20 & 0.15 & 0.50 & 0.65 & 0.50 & 0.79 & 1.00 & 1.00 & 1.00 & 1.00 \\ 0.06 & 0.00 & 0.18 & 0.33 & 0.21 & 0.50 & 0.91 & 0.88 & 0.94 & 1.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.09 & 0.50 & 0.43 & 0.62 & 0.67 \\ 0.00 & 0.00 & 0.00 & 0.06 & 0.00 & 0.12 & 0.57 & 0.50 & 0.69 & 0.73 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.06 & 0.38 & 0.31 & 0.50 & 0.50 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.33 & 0.27 & 0.50 & 0.50 \end{bmatrix}.$$

We obtain fuzzy classes of condition and decision, such as

$$\mathrm{TWN}_{B_1}^{0.05+}(x_1) = \frac{0.50}{x_1} + \frac{0.43}{x_2} + \frac{0.77}{x_3} + \cdots + \frac{1.00}{x_9} + \frac{1.00}{x_{10}}$$
$$D_3^\geq(x_j) = \frac{0}{x_1} + \frac{0}{x_2} + \frac{0}{x_3} + \cdots + \frac{1}{x_{10}}.$$

In terms of nonzero memberships, the FD3NRS model satisfies

$$\underline{\mathrm{TWN}_{B_1}^{0.05}}(D_3^\geq):\ \frac{0.15}{x_8} + \frac{0.24}{x_9} + \frac{0.34}{x_{10}}$$
$$\overline{\mathrm{TWN}_{B_1}^{0.05}}(D_3^\geq):\ \frac{0.13}{x_4} + \frac{0.24}{x_6} + \frac{0.85}{x_7} + \frac{1}{x_8} + \frac{1}{x_9} + \frac{1}{x_{10}}$$
$$\gamma_{B_1}^{0.05}(D^\geq) = \frac{10 + 5.02 + 0.73}{10 + 7 + 3} = \frac{15.75}{20} = 0.79.$$

III. Uncertainty Measures Based on FD3N Granulation

In the above section, the FD3N granulation improves the current FDN granulation [22]. Regarding condition and decision interactions, FD3NRSs with algebraic dependency already improve current FDNRSs [22]. Next, information measures are systematically developed. Furthermore, their CEs are combined with dependency, so powerful fusion measures, DCEs, are generated to support follow-up feature learning.

FD3N granulation underlies the improvements to information measures, and here, we additionally reinforce the information structures for further improvement. We first review, extend, and upgrade FDNCE [22] via $N_B^{\delta+}(x_i)$. To support granular interactions, suppose the conditional classes $N_B^{\delta+}(x_i), \mathrm{TWN}_B^{\delta+}(x_i)$, and decisional class $M_D^+(x_i)$ produce three types of fuzzy granulation structures

$$\begin{aligned} N_B^{\delta+} &= \left[N_B^{\delta+}(x_1), N_B^{\delta+}(x_2), \ldots, N_B^{\delta+}(x_n)\right] \\ \mathrm{TWN}_B^{\delta+} &= \left[\mathrm{TWN}_B^{\delta+}(x_1), \ldots, \mathrm{TWN}_B^{\delta+}(x_n)\right] \\ M_D^+ &= \left[M_D^+(x_1), \ldots, M_D^+(x_n)\right]. \end{aligned} \tag{18}$$

Definition 5 ([22]): FDNCE of B on D is

$$H^\delta(D|B) = -\frac{1}{n}\sum_{i=1}^n \log_2 \frac{|N_B^{\delta+}(x_i) \cap M_D^+(x_i)|}{|N_B^{\delta+}(x_i)|}. \tag{19}$$

The current $H^\delta(D|B)$ provides two development opportunities: information enrichment and system deepening, which accord with the replacement $P \to FP$ in (1). 1) $H^\delta(D|B)$ mainly concerns the conditional probability $P_B^\delta(x_i) = \frac{|N_B^{\delta+}(x_i) \cap M_D^+(x_i)|}{|N_B^{\delta+}(x_i)|}$. The core interaction $|N_B^{\delta+}(x_i) \cap M_D^+(x_i)|$ can be enhanced by the square function, while the other granular cardinality $|M_D^+(x_i)|$ can be complemented for class information completeness. That is, we adopt more powerful forms

$$\mathrm{FP}_B^\delta(x_i) = \frac{|N_B^{\delta+}(x_i) \cap M_D^+(x_i)|^2}{|M_D^+(x_i)||N_B^{\delta+}(x_i)|} \leq P_B^\delta(x_i)$$
$$\mathrm{FH}^\delta(D|B) = -\frac{1}{n}\sum_{i=1}^n \log_2 FP_B^\delta(x_i) \geq H^\delta(D|B) \tag{20}$$

which refer to relevant existing notions in other environments [32], [33]. 2) $H^\delta(D|B)$ and $\mathrm{FH}^\delta(D|B)$ concern only CEs, and other notions (such as information entropy and mutual information) can be comprehensively mined to establish measure systems. The above two assumptions are made, and we next generalize $N_B^{\delta+}$ to $\mathrm{TWN}_B^{\delta+}$. Note that all following results for $\mathrm{TWN}_B^{\delta+}$ have corresponding descriptions in $N_B^{\delta+}$. Next, the corresponding symbols of FH^δ and $\mathrm{FP}_B^\delta(x_i)$ are uniformly utilized for the formal $\mathrm{TWN}_B^{\delta+}$ and potential $N_B^{\delta+}$, and specific contexts can provide effective identifications.

Definition 6: Regarding $\mathrm{TWN}_B^{\delta+}$, CE of B on D is

$$\mathrm{FH}^\delta(D|B) = -\frac{1}{n}\sum_{i=1}^n \log_2 \frac{|\mathrm{TWN}_B^{\delta+}(x_i) \cap M_D^+(x_i)|^2}{|\mathrm{TWN}_B^{\delta+}(x_i)||M_D^+(x_i)|}. \tag{21}$$

Information entropies of B and D are

$$\mathrm{FH}^\delta(B) = -\frac{1}{n}\sum_{i=1}^n \log_2 \frac{|\mathrm{TWN}_B^{\delta+}(x_i)|}{n}$$

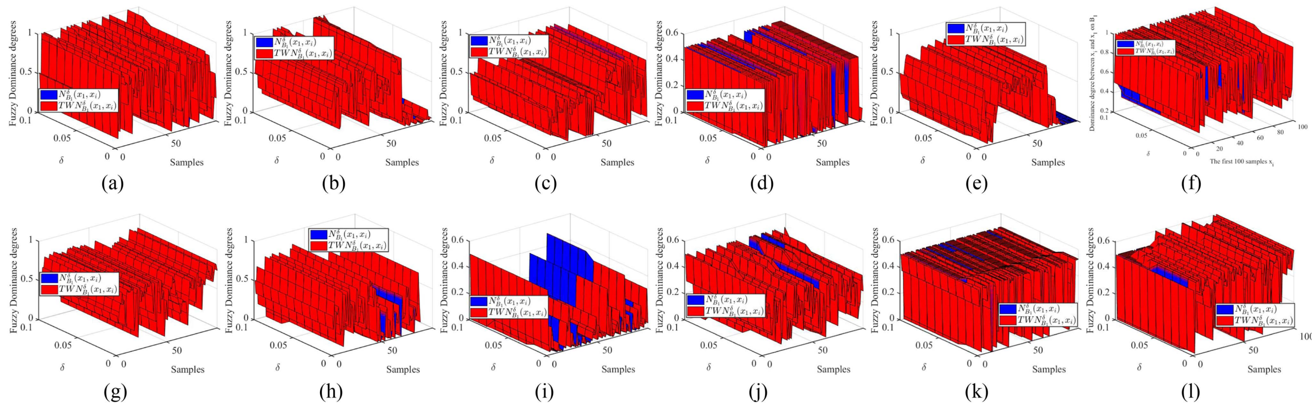

Fig. 3. FDDs where x_i dominates x_1 on B_1 and δ. (a) WPBC. (b) Seeds. (c) Wine. (d) BCW. (e) Breast. (f) Climate. (g) Glass. (h) Chemical. (i) WDBC. (j) Sports. (k) Hill. (l) DARWIN.

Algorithm 1: Calculating Condition Entropy $\mathrm{FH}^{\delta}(D|B)$.

Input: ODS $(U, C \cup D, V, f)$, $\delta \in [0, 0.1]$, $B \subseteq C$.
Output: $FH^{\delta}(D|B)$.

for $x_i \in U$ **do**
  for $x_j \in U$ **do**
    for $a \in B$ **do**
      Compute $TWN_a^{\delta}(x_i, x_j)$ by Eq. (11);
    end
    Calculate $TWN_B^{\delta}(x_i, x_j) = \bigwedge_{a \in B} TWN_a^{\delta}(x_i, x_j)$ by Eq. (12);
  end
  Compute $TWN_B^{\delta +}(x_i)$, $M_D^{+}(x_i)$ by Eqs. (13) (14);
end
Initialize $FH^{\delta}(D|B) = 0$;
for $x_i \in U$ **do**
  By Eq. (21), $FH^{\delta}(D|B) \leftarrow FH^{\delta}(D|B) - \frac{1}{n}\log_2 FP_B^{\delta}(x_i)$.
end
Return $FH^{\delta}(D|B)$.

$$\mathrm{FH}^{\delta}(D) = -\frac{1}{n}\sum_{i=1}^{n}\log_2 \frac{|M_D^{+}(x_i)|}{n} \tag{22}$$

and the joint entropy and mutual information of B on D are

$$\mathrm{FH}^{\delta}(B, D) = -\frac{1}{n}\sum_{i=1}^{n}\log_2 \frac{|\mathrm{TWN}_B^{\delta +}(x_i) \cap M_D^{+}(x_i)|^2}{n|M_D^{+}(x_i)|}$$
$$\mathrm{FH}^{\delta}(B; D) = -\frac{1}{n}\sum_{i=1}^{n}\log_2 \frac{|\mathrm{TWN}_B^{\delta +}(x_i)||M_D^{+}(x_i)|^2}{n|\mathrm{TWN}_B^{\delta +}(x_i) \cap M_D^{+}(x_i)|^2}. \tag{23}$$

Theorem 1: 1) $\mathrm{FH}^{\delta}(D|B) = \mathrm{FH}^{\delta}(B, D) - \mathrm{FH}^{\delta}(B)$,
2) $\mathrm{FH}^{\delta}(B; D) = \mathrm{FH}^{\delta}(B) + \mathrm{FH}^{\delta}(D) - \mathrm{FH}^{\delta}(B, D)$, and
3) $\mathrm{FH}^{\delta}(B; D) = \mathrm{FH}^{\delta}(D) - \mathrm{FH}^{\delta}(D|B)$.

Definition 6 establishes a system of information measurement, and Theorem 1 (proved in Appendix A) reveals systematic relationships. As an example, the promotional CE is calculated in Algorithm 1.

$\mathrm{FH}^{\delta}(D|B)$ improves $H^{\delta}(D|B)$, and they can be fused with algebraic dependency for greater robustness. $\gamma_B^{\delta}(D^{\geq})$ improves $\widetilde{\gamma}_B^{\delta}(D^{\geq})$ so it is considered, and the fusion operation directly adopts $\times$, as is widely used [26], [27], [28]. Next, $H^{\delta}(D|B), FH^{\delta}(D|B)$ are each combined with $\gamma_B^{\delta}(D^{\geq})$, and CEs and fusion measures are used for the two contexts of $\mathrm{TWN}_B^{\delta +}$ and $N_B^{\delta +}$.

Definition 7: Define two DCEs

$$\mathrm{RH}^{\delta}(D|B) = \gamma_B^{\delta}(D^{\geq}) \times H^{\delta}(D|B),$$
$$\mathrm{RFH}^{\delta}(D|B) = \gamma_B^{\delta}(D^{\geq}) \times \mathrm{FH}^{\delta}(D|B). \tag{24}$$

Theorem 2: $H^{\delta}(D|B)$, $\mathrm{FH}^{\delta}(D|B)$, $\mathrm{RH}^{\delta}(D|B)$, and $\mathrm{RFH}^{\delta}(D|B)$ have attribute granulation nonmonotonicity and neighborhood parameter nonmonotonicity.

Theorem 2 is proved in Appendix B from a theoretical view, and it can be practically verified by experiments (such as experimental Fig. 4).

IV. $2 \times 2 \times 2$ Feature Selection Algorithms Based on Three-Level Improvements

Thus far, three-level improvements have been completed. As shown in Fig. 1, we improve FDNRSs to FD3NRSs at the knowledge level; at the interaction level, the algebraic model improvement is from FDNRSs to FD3NRSs (from $\widetilde{\gamma}_B^{\delta}(D^{\geq})$ to $\gamma_B^{\delta}(D^{\geq})$), while the information measure improvement is from $H^{\delta}(D|B)$ to the $\mathrm{FH}^{\delta}(D|B)$ system. At the fusion level, algebraic-informational DCEs $\mathrm{RH}^{\delta}(D|B), \mathrm{RFH}^{\delta}(D|B)$ emerge by adding $\gamma_B^{\delta}(D^{\geq})$. The three-level improvements induce $2 \times 2 \times 2$ uncertainty measures on FDDs, CEs, and DCEs, and the latter measures further motivate feature selection here.

As shown in Table III, CEs and DCEs generate 4 measures (i.e., $H^{\delta}(D|B)$, $\mathrm{FH}^{\delta}(D|B)$, $\mathrm{RH}^{\delta}(D|B)$, and $\mathrm{RFH}^{\delta}(D|B)$), while knowledge granulation concerns 2 cases (i.e., FDNs related to N and FD3Ns related to TWN). Thus, 4×2 cases emerge to support the relevant uncertainty measurement and feature

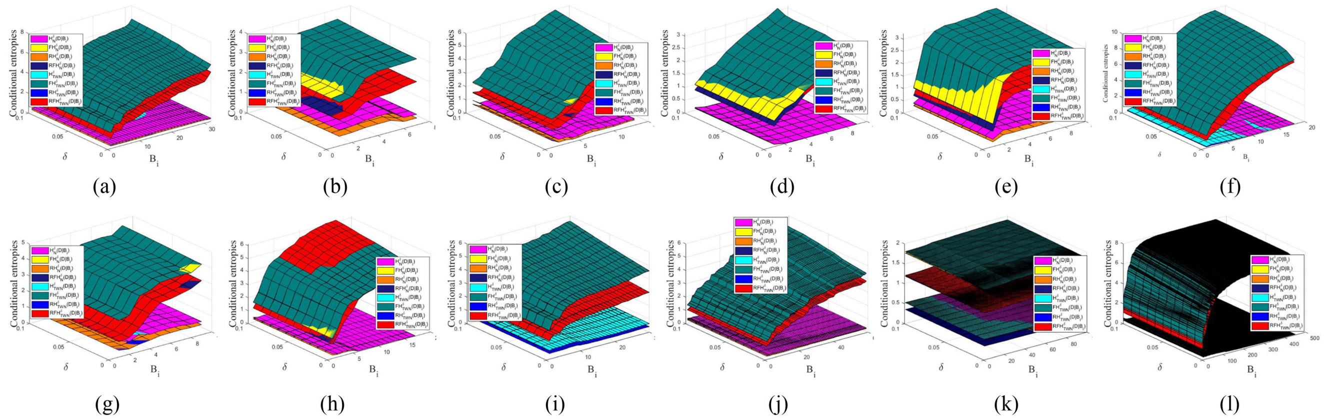

Fig. 4. 3-D surfaces of eight measures on B-attribute and δ-rate chains. (a) WPBC. (b) Seeds. (c) Wine. (d) BCW. (e) Breast. (f) Climate. (g) Glass. (h) Chemical. (i) WDBC. (j) Sports. (k) Hill. (l) DARWIN.

TABLE III
EIGHT COMBINATION MEASURES AND SELECTION ALGORITHMS FROM THREE-LEVEL IMPROVEMENTS

FDDs \ CEs	$H^{\delta}(D\|B)$ (H)	$FH^{\delta}(D\|B)$ (FH)	$RH^{\delta}(D\|B)$ (RH)	$RFH^{\delta}(D\|B)$ (RFH)
N	HN-FS ($H_N^{\delta}(D\|B)$)	FHN-FS ($FH_N^{\delta}(D\|B)$)	RHN-FS ($RH_N^{\delta}(D\|B)$)	RFHN-FS ($RFH_N^{\delta}(D\|B)$)
TWN	HTWN-FS ($H_{TWN}^{\delta}(D\|B)$)	FHTWN-FS ($FH_{TWN}^{\delta}(D\|B)$)	RHTWN-FS ($RH_{TWN}^{\delta}(D\|B)$)	RFHTWN-FS ($RFH_{TWN}^{\delta}(D\|B)$)

selection. The 8 relevant measures are abbreviated as

$$H_{\heartsuit}^{\delta}(D|B), \mathrm{FH}_{\heartsuit}^{\delta}(D|B), \mathrm{RH}_{\heartsuit}^{\delta}(D|B), \mathrm{RFH}_{\heartsuit}^{\delta}(D|B) \tag{25}$$

where $\heartsuit \in \{N, \mathrm{TWN}\}$, and the corresponding algorithms for feature selection become

$$\text{HN-FS}, \text{FHN-FS}, \text{RHN-FS}, \text{RFHN-FS},$$
$$\text{HTWN-FS}, \text{FHTWN-FS}, \text{RHTWN-FS}, \text{RFHTWN-FS} \tag{26}$$

which are shown in Fig. 1 and Table I. HN-FS denotes the existing algorithm FDNCE-FS [22], while the other 7 algorithms are novel and arise from the three-level improvements.

A. Algorithm Construction

Definition 8: On $H_{\heartsuit}^{\delta}(D|B)$, $R \subseteq C$ is a reduct if
1) $H_{\heartsuit}^{\delta}(D|R) \le H_{\heartsuit}^{\delta}(D|C)$, and
2) $\forall r \in R, H_{\heartsuit}^{\delta}(D|(R-\{r\})) > H_{\heartsuit}^{\delta}(D|R)$.

Significances of external attribute $b \in C - B$ and internal attribute $b \in B$ on subset $B \subseteq C$ are, respectively

$$\mathrm{Sig}^{+}(b, B, D) = H_{\heartsuit}^{\delta}(D|B) - H_{\heartsuit}^{\delta}(D|(B \cup \{b\}))$$
$$\mathrm{Sig}^{-}(b, B, D) = H_{\heartsuit}^{\delta}(D|(B - \{b\})) - H_{\heartsuit}^{\delta}(D|B). \tag{27}$$

$H_{\heartsuit}^{\delta}(D|B)$ involves two CEs on N, TWN, and its granulation nonmonotonicity implies nonmonotonic feature selection. The relevant reduct definition and algorithm framework refer to FDNCE-FS [22]. The corresponding Algorithm 2 contains HN-FS and HTWN-FS, and it mainly includes three blocks. The "for" loop in Steps 2–7 adopts $\mathrm{Sig}^{-}(c, C, D) > 0$ to collect the initial attributes. The "while" and "for" loops in Steps 9–14 use the maximal $\mathrm{Sig}^{+}(b, R, D)$ to add optimal features and quickly obtain the selection sufficiency. The "for" loop in Steps 15–20 deletes redundant attributes to meet the reduction necessity. Finally, an effective reduct R is obtained.

Algorithm 2: Feature Selection on CE $H_{\heartsuit}^{\delta}(D|B)$.

```
   Input: ODS = (U, C ∪ D, V, f), δ ∈ [0, 0.1].
   Output: A reduct R ⊆ C on H♡^δ(D|B).
1  Initialize R = ∅, and calculate H♡^δ(D|C);
2  for c ∈ C do
3      Compute Sig⁻(c, C, D) by Eq. (27);
4      if Sig⁻(c, C, D) > 0 then
5          R ← R ∪ {c}.
6      end
7  end
8  Calculate H♡^δ(D|R);
9  while H♡^δ(D|R) > H♡^δ(D|C) do
10     for b ∈ C − R do
11         Compute Sig⁺(b, R, D) by Eq. (27), and then
           orderly select maximal feature:
               b• = arg max_{b∈C−R} Sig⁺(b, R, D).
12     end
13     R ← R ∪ {b•};
14 end
15 Let R• = R;
16 for r ∈ R• do
17     if H♡^δ(D|(R − {r}) ≤ H♡^δ(D|R) then
18         R ← R − {r};
19     end
20 end
21 Return R.
```

Algorithm 3: Feature Selection on Measure $\Diamond^{\delta}_{\heartsuit}(D|B)$.

Input: $ODS = (U, C \cup D, V, f)$, $\delta \in [0, 0.1]$.
Output: A reduct $R \subseteq C$ on $\Diamond^{\delta}_{\heartsuit}(D|B)$.
Initialize $R = \emptyset$ and $\Diamond^{\delta}_{\heartsuit}(D|R) = 0$;
while $|\Diamond^{\delta}_{\heartsuit}(D|R) - \Diamond^{\delta}_{\heartsuit}(D|C)|/|\Diamond^{\delta}_{\heartsuit}(D|C)| > 5\%$ **do**
  Sequentially select maximal feature:
$$a^{\bullet} = arg\max_{a \in C-R} \Diamond^{\delta}_{\heartsuit}(D|(R \cup \{a\})).$$
  Let $R \leftarrow R \cup \{a^{\bullet}\}$;
end
Return R.

The remaining measures $\mathrm{FH}^{\delta}_{\heartsuit}(D|B)$, $\mathrm{RH}^{\delta}_{\heartsuit}(D|B)$, $\mathrm{RFH}^{\delta}_{\heartsuit}(D|B)$ can be unified by

$$\Diamond^{\delta}_{\heartsuit}(D|B), \quad \text{where } \Diamond \in \{\mathrm{FH}, \mathrm{RH}, \mathrm{RFH}\}.$$

Although these improved measures are also nonmonotonic, they follow monotonic trends of attribute granulation, as observed in practical cases. For this case, we adopt the idea of nonmonotonic reduction by approximatively simulating classical monotonic reduction.

Definition 9: On $\Diamond^{\delta}_{\heartsuit}(D|B)$, $R \subseteq C$ is a reduct if

1) $\Diamond^{\delta}_{\heartsuit}(D|R) = \Diamond^{\delta}_{\heartsuit}(D|C)$ and
2) $\forall r \in R$, $\Diamond^{\delta}_{\heartsuit}(D|(R - \{r\})) \neq \Diamond^{\delta}_{\heartsuit}(D|R)$.

The approximation condition of $\Diamond^{\delta}_{\heartsuit}(D|R)$ to $\Diamond^{\delta}_{\heartsuit}(D|C)$ is defined via tolerance threshold $t \in [0, 1]$ as follows:

$$|\Diamond^{\delta}_{\heartsuit}(D|R) - \Diamond^{\delta}_{\heartsuit}(D|C)| \leq t\Diamond^{\delta}_{\heartsuit}(D|C). \tag{28}$$

According to theoretical nonmonotonicity and approximative monotonicity, the reducts on $\Diamond^{\delta}_{\heartsuit}(D|B)$ are related to metric preservation, but their Algorithm 3 adds maximal metric attributes to satisfy the approximation condition. For (28), $|\Diamond^{\delta}_{\heartsuit}(D|R) - \Diamond^{\delta}_{\heartsuit}(D|C)|/\Diamond^{\delta}_{\heartsuit}(D|C)$ reflects the relative approximation ratio, and its tolerance parameter is set to $t = 5\%$ for feasibility. Tolerant and approximate reducts are pursued by referring to the relevant framework [29], [37], and they resolve the difficulty of measure preservation. Algorithm 3 yields the six new selection algorithms in Table III.

B. Example Validation

Example 2: We continue Example 1 with Table II and $\delta = 0.05$. Tables IV and V record the detailed processes of HN-FS and RFHTWN-FS and clarify Algorithms 2 and 3, respectively. Finally, the eight algorithms yield reducts $\{c_1, c_2, c_3, c_4\}$, $\{c_4\}$, $\{c_1, c_2, c_3, c_4\}$, $\{c_2, c_3, c_4\}$, $\{c_1, c_2, c_3, c_4\}$, $\{c_4\}$, $\{c_1, c_2, c_3, c_4\}$, and $\{c_2, c_3, c_4\}$. In contrast, the existing HN-FS and FDNCE-FS [22] never remove attributes, while new algorithms (such as RFHTWN-FS) may obtain fewer features.

V. Data-Based Experiments on Uncertainty Measurement and Feature Selection

Here, data-based experiments are made to validate the improvements to both uncertainty measurement and feature selection. A total of 12 datasets from the UCI machine learning repository,[1] described in Table VI, are used as the ODS $(U, C \cup D, V, f)$. After min–max normalization on $f(x_i, c_k) \to \hat{f}(x_i, c_k)$, a noise addition mechanism (with $r_{ik} = 0.1$) is used for noise-oriented verification [22], i.e.,

$$\hat{f}'(x_i, c_k) = \begin{cases} \hat{f}(x_i, c_k) + r_{ik}, & \text{if } \hat{f}(x_i, c_k) + r_{ik} \in [0, 1] \\ \hat{f}(x_i, c_k), & \text{otherwise.} \end{cases}$$

A. Calculation Verification of Uncertainty Measurement

For the uncertainty measures, we first calculate and validate the bottom FDDs. Concretely, we show observation values where x_i (with an upper bound of 100) dominates x_1 on $B = \{c_1\}$, and the relevant 3-D graphs on (x_i, δ) are given in Fig. 3, where δ comes from (29). The FDDs on N, TWN have the same trends in the main part. The correctional FDDs can reach a certainty value 0 (or 1) to resist noise and enhance reasoning, so they are reasonable and effective for GrC and 3WD.

Then, we consider middle CEs and top DCEs, and eight relevant measures are obtained from

$$\begin{aligned} B &: B_1 = \{c_1\} \subset \cdots \subseteq B_r = \{c_1, c_2, \ldots, c_r\} = C \\ \delta &: \delta_0 = 0 < \delta_1 = 0.01 < \cdots < \delta_{10} = 0.1. \end{aligned} \tag{29}$$

The 3-D values on (B, δ) are displayed in Fig. 4. Fig. 4 supports the theoretical granulation nonmonotonicity and actual monotonicity trends, and these two characteristics are respectively related to the mathematical results in Theorem 2 and the design of Algorithms 2 and 3. For example, the four measures $\mathrm{FH}^{\delta}_{N}(D|B)$, $\mathrm{FH}^{\delta}_{\mathrm{TWN}}(D|B)$, $\mathrm{RFH}^{\delta}_{N}(D|B)$, and $\mathrm{RFH}^{\delta}_{\mathrm{TWN}}(D|B)$ always have monotonically increasing trends, while the other four measures exhibit a trend of monotonically increasing or decreasing in different cases.

B. Classification Comparison of Feature Selection

Now, we turn to the analysis of feature selection and algorithm comparison. As given in Fig. 1 and Table III, the systematic $2 \times 2 \times 2 = 8$ selection algorithms come from the three-level improvements and measures. Their 4×2 structure is given in (26), and the relevant symbols are identified by combinations of four CE/DCE labels (i.e., H, FH, RH, and RFH) and two FDD labels (i.e., N and TWN). Relevant algorithmic contrasts can be obtained from parallel and structural perspectives, and these two types of observations reveal the overall algorithm optimization and three improvement points, respectively. Moreover, the algorithm HN-FS is actually the existing FDNCE-FS [22], and this recent algorithm has outperformed three contrasted approaches: DCE-FS [31], FDCE-FS [30], and NDCE-FS [21]. Therefore, our goal is to show the superiority of the new algorithms over HN-FS/FDNCE-FS; thus, our novel algorithms can

[1][Online]. Available: http://archive.ics.uci.edu/ml

TABLE IV
EXECUTION PROCESS OF FEATURE SELECTION HN-FS/FDNCE-FS ON $\delta = 0.05$

Initialization step	Procedural subset R	Whole feature set C	Condition entropy $H_N^{0.05}(D\|C)$	Significance $\mathrm{Sig}^-(c, C, D)\ (c \in C)$	Selected features
1)	$\emptyset$	$\{c_1, c_2, c_3, c_4\}$	0.1208	[0.0003,0.0213,0.0186,0.0074]	c_1, c_2, c_3, c_4
Initialization R	$\{c_1, c_2, c_3, c_4\}$				
Addition step	**Incremental subset R**	**Sufficient requirement $H_N^{0.05}(D\|R) \le H_N^{0.05}(D\|C)$**	**Rest features set $C - R$**	**Significance $Sig^+(b, R, D)\ (b \in C - R)$**	**Added feature $b^\bullet$**
1)	$\{c_1, c_2, c_3, c_4\}$	$0.1208 \le 0.1208$	–	–	–
Interim R ($R^\bullet$)	$\{c_1, c_2, c_3, c_4\}$				
Deletion step	**Decremental subset R**	**Contrast value $H_N^{0.05}(D\|R)$**	**Target feature $r \in R^\bullet$ and rest set $R - \{r\}$**	**Redundant deletion condition $H_N^{0.05}(D\|(R - \{r\})) \le H_N^{0.05}(D\|R)$**	**Deleted feature**
1)	$\{c_1, c_2, c_3, c_4\}$	0.1208	$c_1, \{c_2, c_3, c_4\}$	$0.1282 > 0.1208$	–
2)	$\{c_1, c_2, c_3, c_4\}$	0.1208	$c_2, \{c_1, c_3, c_4\}$	$0.1394 > 0.1208$	–
3)	$\{c_1, c_2, c_3, c_4\}$	0.1208	$c_3, \{c_1, c_2, c_4\}$	$0.1421 > 0.1208$	–
4)	$\{c_1, c_2, c_3, c_4\}$	0.1208	$c_4, \{c_1, c_2, c_3\}$	$0.1211 > 0.1208$	–
Last reduct R	$\{c_1, c_2, c_3, c_4\}$				

TABLE V
EXECUTION PROCESS OF FEATURE SELECTION RFHTWN-FS ON $\delta = 0.05$

Addition step	Incremental subset R	Sufficient requirement $\frac{\|RFH_{TWN}^{0.05}(D\|R) - RFH_{TWN}^{0.05}(D\|C)\|}{\|RFH_{TWN}^{0.05}(D\|C)\|} \le 5\%$ (Yes/No)	Rest attributes in $C - R$	CEs $RFH_{TWN}^{0.05}(D\|(R \cup \{a\}))$	Added attribute $a^\bullet$
1)	$\emptyset$	$\frac{\|0-0.9445\|}{0.9445} = 1 \le 5\%$ (No)	c_1, c_2, c_3, c_4	$[0.6661, 0.6515, 0.6839, \mathbf{0.7331}]$	c_4
2)	$\{c_4\}$	$\frac{\|0.7311-0.9445\|}{0.9445} = 22.59\% \le 5\%$ (No)	c_1, c_2, c_3	$[0.7859, 0.7856, \mathbf{0.8554}]$	c_3
3)	$\{c_3, c_4\}$	$\frac{\|0.8554-0.9445\|}{0.9445} = 9.43\% \le 5\%$ (No)	c_1, c_2	$[0.8993, \mathbf{0.9174}]$	c_2
4)	$\{c_2, c_3, c_4\}$	$\frac{\|0.9174-0.9445\|}{0.9445} = 2.87\% \le 5\%$ (Yes)	–	–	–
Last reduct R	$\{c_2, c_3, c_4\}$				

TABLE VI
UCI DATASETS AND THEIR DETAILS

No.	Dataset	Abbreviation	Objects	Features	Classes
(a)	Wisconsin Prognostic Breast Cancer	WPBC	198	32	2
(b)	Seeds	Seeds	210	7	3
(c)	Wine	Wine	178	13	3
(d)	BCW	BCW	683	9	2
(e)	Breast Tissue	Breast	106	9	6
(f)	Climate Model Simulation Crashe	Climate	540	18	2
(g)	Glass	Glass	214	9	7
(h)	Chemical	Chemical	88	18	2
(i)	Wisconsin Diagnostic Breast Cancer	WDBC	569	30	2
(j)	Sports articles for objectivity	Sports	1000	59	2
(k)	Hill valley	Hill	606	100	2
(l)	DARWIN	DARWIN	174	450	2

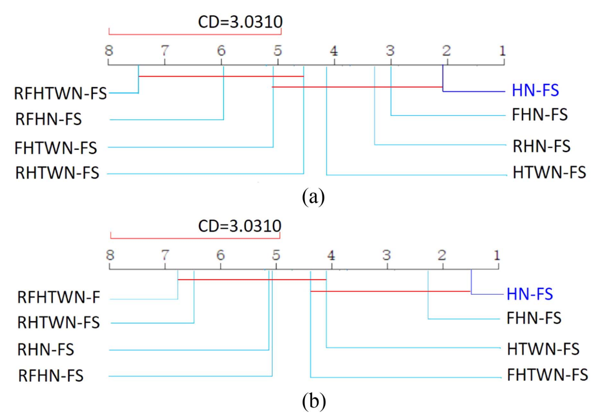

Fig. 5. Nemenyi's test figures of eight algorithms on classification accuracies. (a) KNN. (b) SVM.

achieve direct and indirect improvement verification. For the evaluation indexes, classification accuracies are mainly used, and we employ two classifiers KNN (K=3) and SVM, which denote the K-Nearest Neighbor and Support Vector Machine respectively.

For the eight algorithms, based on (29), Tables VII and VIII show the δ-optimal accuracies (and corresponding selected numbers) for KNN and SVM, respectively. Tables VII and VIII show the main results for the algorithm comparisons (where bold entities reflect maximums), and their last 2 lines concern the 12-dataset statistics on average accuracies (average features) and maximum frequencies. For the selected features, all eight algorithms can effectively reduce the attributes, and FHN-FS, RFHN-FS, HTWN-FS, FHTWN-FS, and RFHTWN-FS can obtain smaller average lengths than HN-FS/FDNCE-FS. Furthermore, Tables VII and VIII can be statistically analyzed to determine the algorithmic improvement. Friedman's test and Nemenyi's test are performed for the eight algorithms and 12 datasets with the statistical threshold 0.05. By calculation, $\tau_F = 10.5860 > F_{0.05}(7, 77) = 2.1310$ holds for the KNN while $\tau_F = 14.6667 > F_{0.05}(7, 77) = 2.1310$ holds for the SVM, so these algorithms have significant differences. In addition, the critical difference is $CD_{0.05} = 3.0310$, and the figures for Nemenyi's test are depicted in Fig. 5. 1) By the KNN-based Table VII, the corresponding rankings based on average accuracies are

$$\text{RFHTWN-FS} \succ \text{RFHN-FS} \succ \text{FHTWN-FS} \succ \text{RHTWN-FS} \succ$$

TABLE VII
δ-OPTIMAL KNN CLASSIFICATION ACCURACIES (AND FEATURE SELECTION LENGTHS) OF 8 SELECTION ALGORITHMS

No	Datasets	HN-FS	FHN-FS	RHN-FS	RFHN-FS	HTWN-FS	FHTWN-FS	RHTWN-FS	RFHTWN-FS
(a)	WPBC	0.7879 (14)	0.7879 (13)	0.7424 (32)	0.7929 (15)	**0.8081** (17)	0.7980 (14)	0.7424 (32)	0.7929 (15)
(b)	Seeds	0.6810 (1)	0.9048 (2)	0.9667 (7)	**0.9762** (4)	0.6810 (1)	0.9143 (3)	0.9667 (7)	**0.9762** (4)
(c)	Wine	0.8371 (3)	0.9719 (6)	0.9775 (13)	**0.9831** (11)	0.9157 (4)	0.9775 (7)	0.9775 (13)	**0.9831** (11)
(d)	BCW	0.9722 (4)	**0.9839** (7)	0.9780 (9)	**0.9839** (7)	0.9795 (8)	**0.9839** (7)	0.9780 (9)	**0.9839** (7)
(e)	Breast	0.7453 (1)	0.6132 (1)	0.8208 (9)	**0.8302** (4)	0.7453 (1)	0.6132 (1)	0.8208 (9)	**0.8302** (4)
(f)	Climate	0.9259 (3)	0.9315 (15)	0.9463 (18)	**0.9574** (15)	0.9259 (3)	0.9444 (13)	0.9463 (18)	**0.9574** (15)
(g)	Glass	0.7196 (3)	0.7617 (4)	0.8037 (9)	**0.8084** (7)	0.7897 (4)	0.7617 (4)	0.8037 (9)	**0.8084** (6)
(h)	Chemical	0.9886 (2)	**1.0000** (7)	**1.0000** (16)	**1.0000** (10)	**1.0000** (1)	**1.0000** (7)	**1.0000** (16)	**1.0000** (10)
(i)	WDBC	0.8190 (3)	0.9649 (11)	**0.9842** (29)	0.9666 (13)	0.7926 (3)	0.9754 (12)	**0.9842** (29)	0.9719 (14)
(j)	Sports	0.8620 (7)	0.8820 (18)	0.6350 (54)	**0.8830** (22)	0.8860 (24)	**0.8830** (20)	0.6350 (54)	**0.8830** (22)
(k)	Hill	0.7855 (26)	0.7756 (2)	0.7393 (1)	**0.7987** (2)	0.7822 (29)	**0.7987** (2)	0.7393 (1)	**0.7987** (2)
(l)	DARWIN	0.8678 (193)	0.8448 (40)	0.7989 (449)	0.9080 (53)	0.9368 (33)	0.8966 (15)	0.8333 (434)	**0.9540** (18)
(a-l)	Average	0.8327 (21.7)	0.8685 (10.5)	0.8661 (53.8)	0.9074 (13.6)	0.8536 (10.7)	0.8789 (8.8)	0.8689 (52.6)	**0.9116** (10.7)
	Win	0	2	2	9	2	4	2	10

TABLE VIII
δ-OPTIMAL SVM CLASSIFICATION ACCURACIES (AND FEATURE SELECTION LENGTHS) OF 8 SELECTION ALGORITHMS

No	Datasets	HN-FS	FHN-FS	RHN-FS	RFHN-FS	HTWN-FS	FHTWN-FS	RHTWN-FS	RFHTWN-FS
(a)	WPBC	0.6818 (14)	0.6919 (13)	0.7020 (32)	**0.7172** (15)	0.6970 (20)	0.7071 (14)	0.7020 (32)	0.7020 (15)
(b)	Seeds	0.5619 (1)	0.8381 (2)	**0.9333** (7)	0.9286 (5)	0.5619 (1)	0.8905 (3)	**0.9333** (7)	0.9286 (5)
(c)	Wine	0.6348 (3)	0.9607 (6)	**0.9944** (13)	0.9888 (11)	0.8539 (4)	0.9607 (6)	**0.9944** (13)	0.9888 (11)
(d)	BCW	0.9649 (4)	0.9722 (7)	0.9736 (9)	0.9722 (7)	**0.9766** (7)	0.9722 (7)	0.9736 (9)	0.9722 (7)
(e)	Breast	0.4057 (1)	0.3962 (1)	**0.5943** (9)	0.5660 (4)	0.4057 (1)	0.3962 (1)	**0.5943** (9)	0.5660 (4)
(f)	Climate	0.9148 (3)	0.9148 (15)	**0.9241** (18)	**0.9241** (15)	0.9148 (3)	0.9148 (13)	**0.9241** (18)	**0.9241** (15)
(g)	Glass	0.4813 (3)	0.4860 (4)	0.5654 (9)	0.5561 (7)	0.4813 (3)	0.4860 (4)	0.5654 (9)	**0.5935** (6)
(h)	Chemical	0.9545 (2)	**1.0000** (7)	**1.0000** (16)	**1.0000** (10)	**1.0000** (1)	**1.0000** (7)	**1.0000** (16)	**1.0000** (10)
(i)	WDBC	0.6274 (3)	0.9596 (11)	**0.9789** (29)	0.9613 (13)	0.6274 (3)	0.9649 (12)	0.9736 (29)	0.9613 (13)
(j)	Sports	0.8050 (7)	0.8150 (18)	0.3650 (54)	**0.8340** (22)	0.8310 (24)	0.8190 (20)	0.3650 (54)	0.8320 (22)
(k)	Hill	0.5033 (26)	0.5033 (1)	0.5033 (1)	0.5033 (2)	0.5149 (31)	0.5033 (1)	0.5033 (1)	**0.5165** (2)
(l)	DARWIN	0.9598 (193)	0.9080 (40)	**1.0000** (449)	0.9598 (53)	**1.0000** (193)	0.9080 (40)	**1.0000** (412)	**1.0000** (17)
(a-l)	Average	0.7079 (21.7)	0.7872 (10.4)	0.7945 (53.8)	0.8260 (13.7)	0.7387 (24.3)	0.7936 (10.7)	0.7941 (50.8)	**0.8321** (10.6)
	Win	0	1	8	4	3	1	7	5

$$\text{FHN-FS} \succ \text{RHN-FS} \succ \text{HTWN-FS} \succ \text{HN-FS}.$$

RFHTWN-FS is optimal, RFHN-FS is suboptimal, and HN-FS is last. By the statistics in Fig. 5(a), the first 2 algorithms are significantly superior to HN-FS/FDNCE-FS. 2) By the SVM-based Table VIII, the corresponding rankings based on average accuracies are

$$\text{RFHTWN-FS} \succ \text{RFHN-FS} \succ \text{RHN-FS} \succ \text{RHTWN-FS} \succ$$
$$\text{FHTWN-FS} \succ \text{FHN-FS} \succ \text{HTWN-FS} \succ \text{HN-FS}.$$

RFHTWN-FS and RFHN-FS are also optimal and suboptimal, respectively, while HN-FS is last again. Similarly, by the statistics in Fig. 5(b), the first four algorithms are significantly superior to HN-FS/FDNCE-FS. By the two sets of analysis results, we obtain the order

$$\text{RFHTWN-FS} \succ \text{RFHN-FS} \succ \cdots \succ \text{HTWN-FS} \succ \text{HN-FS}.$$

Our 7 new algorithms exhibit systematicity, robustness, and optimization, and all of them outperform the comparison algorithm HN-FS/FDNCE-FS [22]. In addition, RFHTWN-FS and RFHN-FS constitute the optimal echelon, and their advantages over HN-FS/FDNCE-FS are statistically significant.

The improved algorithms of feature selection benefit from three-level measurement improvements to FDDs, CEs, and DCEs. Finally, the experimental results of statistical accuracies are hierarchically analyzed regarding the 3 improvement points. For this purpose, the average accuracies in Tables VII and VIII are arranged two-dimensionally in Table IX. In this table, $\leftarrow$, $\uparrow$, respectively, indicate the horizontal and longitudinal maximums, while subtables (a) and (b), respectively, correspond to the KNN and SVM. Next, the algorithmic improvements are analyzed to match and justify the three-level improvements; see the main Table IX and auxiliary Fig. 5. 1) HTWN-FS, FHTWN-FS, RHTWN-FS, and RFHTWN-FS outperform HN-FS, FHN-FS, RHN-FS, and RFHN-FS, respectively, thus verifying the FDD improvement: TWN $\succ N$. At the interior level of N or TWN, the combined order for macro-level measure improvements is RFH $\succeq$ RH $\approx$ FH $\succ$ H. 2) Regarding the addition of F, FHN-FS $\succ$ HN-FS, FHTWN-FS $\succ$ HTWN-FS, RFHN-FS $\succ$ RHN-FS, and RFHTWN-FS $\succ$ RHTWN-FS always hold. Thus, we can also infer the CE improvement on FH $\succeq H$ and RFH $\succeq$ RH. At the interior level, the combined order becomes RHTWN $\approx$ RHN $\succ$ HTWN $\succ$ HN. 3) Regarding the fusion of R, RHN-FS, RFHN-FS, RHTWN-FS, and RFHTWN-FS outperform HN-FS, FHN-FS, HTWN-FS, and FHTWN-FS, respectively. Hence, we determine the DCE fusion improvements on RH $\succeq H$ and RFH $\succeq$ FH. At the interior level, the combined order becomes FHTWN $\succ$ FHN $\succ$ HTWN $\succ$ HN. Overall, the three-level measure improvements are validated via

TABLE IX
8-ALGORITHMIC δ-OPTIMIZATION AND DATASET-AVERAGE CLASSIFICATION ACCURACIES ON 2 FDDS AND 4 CES/DCES

FDDs	(a) KNN				(b) SVM			
	H	FH	RH	RFH	H	FH	RH	RFH
N	0.8327	0.8685	0.8661	$\underleftarrow{\mathbf{0.9074}}$	0.7079	0.7872	0.7945 $\uparrow$	$\underleftarrow{\mathbf{0.8260}}$
TWN	0.8536 $\uparrow$	0.8789 $\uparrow$	0.8689 $\uparrow$	$\underleftarrow{\mathbf{0.9116}}$ $\uparrow$	0.7387 $\uparrow$	0.7936 $\uparrow$	0.7941	$\underleftarrow{\mathbf{0.8321}}$ $\uparrow$

systematic comparisons of the selection algorithms. The bottom improvement TWN $\succ N$ is clear, the middle improvement of CE promotion is somewhat clear, and the top improvement of algebra–information fusion is most prominent.

VI. Conclusion

This article advances recent studies of the FDN-driven information measure FDNCE and selection algorithm FDNCE-FS [22], and we obtain both FD3N-driven three-level improvements of uncertainty measurement and $2 \times 2 \times 2$-structural algorithms of feature selection, as shown in Fig. 1. The three-level measurement improvements yield stronger FDDs, dependency/CEs, and DCEs. Thus, FD3NRSs improve FDNRSs [22], and the seven new selection algorithms outperform the current FDNCE-FS [22]. Regarding advantages, this study offers deeper measurement content and greater algorithm robustness in ODSs, so it provides new insights into uncertainty modeling, information fusion, and feature selection in terms of GrC and 3WD. Regarding disadvantages, the relevant constructions are mainly restricted to ODSs, and the proposed algorithms of nonmonotonic feature selection may be affected by tolerance parameters. In the future, three-level improvements (such as those related to FD3Ns, FD3NRSs, and measure construction) can be generalized for interval-valued data, incomplete and fuzzy systems. Systematic algorithms of feature selection are also worth further researching in terms of incremental learning and application in noisy environments.

Appendix A
Theoretical Proof of Theorem 1

Proof: By (21)–(23), we have

$$\begin{aligned}
&\mathrm{FH}^{\delta}(B,D) - \mathrm{FH}^{\delta}(B)\\
&= -\frac{1}{n}\sum_{i=1}^{n}\log_2 \frac{|\mathrm{TWN}_B^{\delta+}(x_i)\cap M_D^{+}(x_i)|^2}{n|M_D^{+}(x_i)|}\\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\log_2 \frac{|\mathrm{TWN}_B^{\delta+}(x_i)|}{n}\\
&= -\frac{1}{n}\sum_{i=1}^{n}\log_2 \frac{|\mathrm{TWN}_B^{\delta+}(x_i)\cap M_D^{+}(x_i)|^2}{n|M_D^{+}(x_i)|}\frac{n}{|\mathrm{TWN}_B^{\delta+}(x_i)|}\\
&= -\frac{1}{n}\sum_{i=1}^{n}\log_2 \frac{|\mathrm{TWN}_B^{\delta+}(x_i)\cap M_D^{+}(x_i)|^2}{|\mathrm{TWN}_B^{\delta+}(x_i)||M_D^{+}(x_i)|} = \mathrm{FH}^{\delta}(D|B)
\end{aligned}$$

$$\begin{aligned}
&\mathrm{FH}^{\delta}(B) + \mathrm{FH}^{\delta}(D) - \mathrm{FH}^{\delta}(B,D)\\
&= -\frac{1}{n}\sum_{i=1}^{n}\log_2 \frac{|\mathrm{TWN}_B^{\delta+}(x_i)|}{n} - \frac{1}{n}\sum_{i=1}^{n}\log_2 \frac{|M_D^{+}(x_i)|}{n}\\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\log_2 \frac{|\mathrm{TWN}_B^{\delta+}(x_i)\cap M_D^{+}(x_i)|^2}{n|M_D^{+}(x_i)|}\\
&= -\frac{1}{n}\sum_{i=1}^{n}\log_2 \frac{|\mathrm{TWN}_B^{\delta+}(x_i)|}{n}\frac{|M_D^{+}(x_i)|}{n}\\
&\qquad \frac{n|M_D^{+}(x_i)|}{|\mathrm{TWN}_B^{\delta+}(x_i)\cap M_D^{+}(x_i)|^2}\\
&= -\frac{1}{n}\sum_{i=1}^{n}\log_2 \frac{|\mathrm{TWN}_B^{\delta+}(x_i)||M_D^{+}(x_i)|^2}{n|\mathrm{TWN}_B^{\delta+}(x_i)\cap M_D^{+}(x_i)|^2} = \mathrm{FH}^{\delta}(B;D).
\end{aligned}$$

By combining the two results, the eventual result $\mathrm{FH}^{\delta}(B;D) = \mathrm{FH}^{\delta}(D) - \mathrm{FH}^{\delta}(D|B)$ naturally holds. ■

Appendix B
Theoretical Proof of Theorem 2

Proof: At first, the nonmonotonicity of $H^{\delta}(D|B)$ can be referenced in [22].

Then, the nonmonotonicity of $\mathrm{FH}^{\delta}(D|B)$ is mainly focused on. For $\forall B_1 \subseteq B_2 \subseteq C$ and by (21), we have

$$\begin{aligned}
&\mathrm{FH}^{\delta}(D|B_2) - \mathrm{FH}^{\delta}(D|B_1)\\
&= -\frac{1}{n}\sum_{i=1}^{n}\log_2 \frac{|\mathrm{TWN}_{B_2}^{\delta+}(x_i)\cap M_D^{+}(x_i)|^2}{|\mathrm{TWN}_{B_2}^{\delta+}(x_i)||M_D^{+}(x_i)|}\\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\log_2 \frac{|\mathrm{TWN}_{B_1}^{\delta+}(x_i)\cap M_D^{+}(x_i)|^2}{|\mathrm{TWN}_{B_1}^{\delta+}(x_i)||M_D^{+}(x_i)|}\\
&= -\frac{1}{n}\sum_{i=1}^{n}\log_2 \frac{|\mathrm{TWN}_{B_2}^{\delta+}(x_i)\cap M_D^{+}(x_i)|^2}{|\mathrm{TWN}_{B_2}^{\delta+}(x_i)||M_D^{+}(x_i)|}\\
&\qquad \frac{|TWN_{B_1}^{\delta+}(x_i)||M_D^{+}(x_i)|}{|TWN_{B_1}^{\delta+}(x_i)\cap M_D^{+}(x_i)|^2}\\
&= -\frac{1}{n}\sum_{i=1}^{n}\log_2 \frac{|\mathrm{TWN}_{B_2}^{\delta+}(x_i)\cap M_D^{+}(x_i)|^2}{|\mathrm{TWN}_{B_1}^{\delta+}(x_i)\cap M_D^{+}(x_i)|^2}\frac{|TWN_{B_1}^{\delta+}(x_i)|}{|TWN_{B_2}^{\delta+}(x_i)|}\\
&= -\frac{1}{n}\sum_{i=1}^{n}\log_2 g(x_i)h(x_i)
\end{aligned}\tag{30}$$

where

$$\begin{aligned}
g(x_i) &= \frac{|\mathrm{TWN}_{B_2}^{\delta+}(x_i)\cap M_D^{+}(x_i)|^2}{|\mathrm{TWN}_{B_1}^{\delta+}(x_i)\cap M_D^{+}(x_i)|^2}\\
h(x_i) &= \frac{|\mathrm{TWN}_{B_1}^{\delta+}(x_i)|}{|\mathrm{TWN}_{B_2}^{\delta+}(x_i)|}.
\end{aligned}\tag{31}$$

$B_1 \subseteq B_2 \Rightarrow \mathrm{TWN}_{B_1}^{\delta+}(x_i) \supseteq \mathrm{TWN}_{B_2}^{\delta+}(x_i)$ by Proposition 1’ 2). Thus, $0 \le g(x_i) \le 1$, $h(x_i) \ge 1$, neither $g(x_i)h(x_i) \ge 1$ nor $g(x_i)h(x_i) \le 1$ always holds. The conclusion whether $\mathrm{FH}^{\delta}(D|B_2) - \mathrm{FH}^{\delta}(D|B_1)$ is greater than 0 is uncertain, i.e., neither $\mathrm{FH}^{\delta}(D|B_2) \le \mathrm{FH}^{\delta}(D|B_1)$ nor $\mathrm{FH}^{\delta}(D|B_2) \ge \mathrm{FH}^{\delta}(D|B_1)$ always holds. Therefore, the attribute granulation nonmonotonicity of $\mathrm{FH}^{\delta}(D|B)$ is verified. In contrast, the neighborhood parameter nonmonotonicity on $\delta_1 \le \delta_2$ can be similarly validated due to (30), (31), and Proposition 1’ 3).

At last, $\mathrm{RH}^{\delta}(D|B)$ and $\mathrm{RFH}^{\delta}(D|B)$, respectively, develop $H^{\delta}(D|B)$ and $\mathrm{FH}^{\delta}(D|B)$ by multiplying $\gamma_B^{\delta}(D^{\ge})$, so their nonmonotonicity naturally comes from the nonmonotonicity of factorial measures, where the nonmonotonicity of $\gamma_B^{\delta}(D^{\ge})$ is given in Proposition 2’ 2). ■

Acknowledgment

The authors would like to thank the reviewers and editors for their valuable suggestions, which substantially improve this article.

References

[1] X. Y. Che, D. G. Chen, and J. S. Mi, "A novel approach for learning label correlation with application to feature selection of multi-label data," *Inf. Sci.*, vol. 512, pp. 795–812, 2020.

[2] Z. X. Li, F. P. Nie, D. Y. Wu, Z. X. Hu, and X. L. Li, "Unsupervised feature selection with weighted and projected adaptive neighbors," *IEEE Trans. Cybern.*, vol. 53, no. 2, pp. 1260–1271, Feb. 2023.

[3] B. H. Nguyen, B. Xue, and M. Zhang, "A survey on swarm intelligence approaches to feature selection in data mining," *Swarm Evol. Comput.*, vol. 54, no. 2, 2020, Art. no. 100663.

[4] N. Zhou, Y. Y. Xu, H. Cheng, Z. J. Yuan, and B. D. Chen, "Maximum correntropy criterion-based sparse subspace learning for unsupervised feature selection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 2, pp. 404–417, Feb. 2019.

[5] M. A. Elaziz, H. M. Abu-Donia, R. A. Hosny, S. L. Hazae, and R. A. Ibrahim, "Improved evolutionary-based feature selection technique using extension of knowledge based on the rough approximations," *Inf. Sci.*, vol. 594, pp. 76–94, 2022.

[6] D. Chen and Y. Yang, "Attribute reduction for heterogeneous data based on the combination of classical and fuzzy rough set models," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 5, pp. 1325–1334, Oct. 2014.

[7] S. Y. Xia, H. Zhang, W. H. Li, G. Y. Wang, E. Giem, and Z. Z. Chen, "GBNRS: A novel rough set algorithm for fast adaptive attribute reduction in classification," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 3, pp. 1231–1241, Mar. 2022.

[8] J. Cai, J. W. Luo, S. L. Wang, and S. Yang, "Feature selection in machine learning: A new perspective," *Neurocomputing*, vol. 300, pp. 70–79, 2018.

[9] P. Dhal and C. Azad, "A comprehensive survey on feature selection in the various fields of machine learning," *Appl. Intell.*, vol. 52, pp. 4543–4581, 2021.

[10] X. Wu et al., "FTAP: Feature transferring autonomous machine learning pipeline," *Inf. Sci.*, vol. 593, no. 5, pp. 385–397, 2022.

[11] Z. Pawlak, "Rough set," *Int. J. Comput. Inf. Sci.*, vol. 11, no. 5, pp. 341–356, 1982.

[12] J. T. Yao, A. V. Vasilakos, and W. Pedrycz, "Granular computing: Perspectives and challenges," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1977–1989, Dec. 2013.

[13] X. Y. Zhang, Z. Yuan, and D. Q. Miao, "Outlier detection using three-way neighborhood characteristic regions and corresponding fusion measurement," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 5, pp. 2082–2095, May 2024.

[14] D. C. Liang, W. Pedrycz, and D. Liu, "Determining three-way decisions with decision-theoretic rough sets using a relative value approach," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 47, no. 8, pp. 1785–1799, Aug. 2018.

[15] S. Greco, B. Matarazzo, and R. Slowinski, "Rough set theory for multicriteria decision analysis," *Eur. J. Oper. Res.*, vol. 129, no. 1, pp. 1–47, 2001.

[16] Y. H. Qian, C. Y. Dang, J. Y. Liang, and D. W. Tang, "Set-valued ordered information systems," *Inf. Sci.*, vol. 179, no. 16, pp. 2809–2832, 2009.

[17] M. Inuiguchi, Y. Yoshioka, and Y. Kusunoki, "Variable-precision dominance-based rough set approach and attribute reduction," *Int. J. Approx. Reasoning*, vol. 50, no. 8, pp. 1199–1214, 2009.

[18] K. Dembczynski, S. Greco, and R. SlOwinski, "Rough set approach to multiple criteria classification with imprecise evaluations and assignments," *Eur. J. Oper. Res.*, vol. 198, no. 2, pp. 626–636, 2009.

[19] S. Greco, M. Inuiguchi, and R. Slowinski, "Fuzzy rough sets and multiple-premise gradual decision rules," *Int. J. Approx. Reasoning*, vol. 41, no. 2, pp. 179–211, 2006.

[20] Q. H. Hu, D. R. Yu, and M. Z. Guo, "Fuzzy preference based rough sets," *Inf. Sci.*, vol. 180, no. 10, pp. 2003–2022, 2010.

[21] H. M. Chen, T. R. Li, Y. Cai, C. Luo, and H. Fujita, "Parallel attribute reduction in dominance-based neighborhood rough set," *Inf. Sci.*, vol. 373, pp. 351–368, 2016.

[22] B. B. Sang, H. M. Chen, L. Yang, T. R. Li, and W. H. Xu, "Incremental feature selection using a conditional entropy based on fuzzy dominance neighborhood rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 6, pp. 1683–1697, Jun. 2022.

[23] X. Y. Zhang, H. Yao, Z. Y. Lv, and D. Q. Miao, "Class-specific information measures and attribute reducts for hierarchy and systematicness," *Inf. Sci.*, vol. 563, no. 5, pp. 196–225, 2021.

[24] G. Y. Wang, J. Zhao, J. J. An, and Y. Wu, "A comparative study of algebra viewpoint and information viewpoint in attribute reduction," *Fundam. Inf.*, vol. 68, no. 3, pp. 289–301, 2005.

[25] H. Y. Gou and X. Y. Zhang, "Feature selection based on double-hierarchical and multiplication-optimal fusion measurement in fuzzy neighborhood rough sets," *Inf. Sci.*, vol. 618, pp. 437–467, 2022.

[26] Y. M. Chen, K. S. Wu, X. H. Chen, C. H. Tang, and Q. X. Zhu, "An entropy-based uncertainty measurement approach in neighborhood systems," *Inf. Sci.*, vol. 279, pp. 239–250, 2014.

[27] H. Ge and C. J. Yang, "New measures of uncertainty based on the granularity distribution of approximation sets," *Artif. Intell. Rev.*, vol. 55, pp. 3801–3831, 2022.

[28] Y. M. Zhang, X. Y. Jia, and Z. M. Tang, "Information-theoretic measures of uncertainty for interval-set decision tables," *Inf. Sci.*, vol. 577, pp. 81–104, 2021.

[29] J. H. Dai, Y. J. Yan, Z. W. Li, and B. S. Liao, "Dominance-based fuzzy rough set approach for incomplete interval-valued data," *J. Intell. Fuzzy Syst.*, vol. 34, pp. 423–436, 2018.

[30] Q. H. Hu et al., "Feature selection for monotonic classification," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 1, pp. 69–81, Feb. 2012.

[31] Q. H. Hu, M. Z. Guo, D. R. Yu, and J. F. Liu, "Information entropy for ordinal classification," *Sci. China Inf. Sci.*, vol. 53, no. 6, pp. 1188–1200, 2010.

[32] S. Tsumoto, "Characteristics of accuracy and coverage in rough set rule induction," in *Proc. Int. Conf. Rough Sets*, Berlin, Heidelberg, 2003, pp. 373–380.

[33] L. Sun, X. Y. Zhang, Y. H. Qian, J. C. Xu, and S. G. Zhang, "Feature selection using neighborhood entropy-based uncertainty measures for gene expression data classification," *Inf. Sci.*, vol. 502, pp. 18–41, 2019.

[34] Y. Y. Yao, "Three-way granular computing, rough sets, and formal concept analysis," *Int. J. Approx. Reasoning*, vol. 116, pp. 106–125, 2020.

[35] Y. Y. Yao, "Tri-level thinking: Models of three-way decision," *Int. J. Mach. Learn. Cybern.*, vol. 11, pp. 947–959, 2020.

[36] X. Y. Zhang and Y. Y. Yao, "Tri-level attribute reduction in rough set theory," *Expert Syst. Appl.*, vol. 190, 2022, Art. no. 116187.

[37] X. F. Liu, J. H. Dai, J. L. Chen, and C. C. Zhang, "A fuzzy α-similarity relation-based attribute reduction approach in incomplete interval-valued information systems," *Appl. Soft Comput.*, vol. 109, no. 5, 2021, Art. no. 107593.

Xianyong Zhang received the B.S., M.S., and Ph.D. degrees in mathematics from Sichuan Normal University, Chengdu, China, in 2001, 2004, and 2011, respectively.

He completed two-and-a-half-year Visiting Research with the University of Regina, Regina, SK, Canada, and two-year Postdoctoral work with Tongji University, Shanghai, China. He is currently a Professor and a Doctoral Tutor with the School of Mathematical Sciences, Sichuan Normal University, Chengdu, China. His research interests include uncertainty analysis, intelligent computing, three-way decision, and machine learning.

Benwei Chen received the B.S. degree in mathematics from Jiujiang University, Jiangxi, China, in 2009, and the M.S. degree in mathematics from Southwest Petroleum University, Chengdu, China, in 2012. He is currently working toward the Ph.D. degree in mathematics from Sichuan Normal University, Chengdu.

His research interests include fuzzy sets, rough sets, information processing, and data analysis.

Duoqian Miao received the B.S. and M.S. degrees in mathematics and statistics from Shanxi University, Taiyuan, China, in 1985 and 1991, respectively, and the Ph.D. degree in pattern recognition and intelligent system from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 1997.

He is currently a Professor and a Doctoral Tutor with the Department of Computer Science and Technology, Tongji University, Shanghai, China. His research interests include data mining, granular computing, soft computing, and machine learning.