Feature Selection Using Fuzzy Neighborhood Entropy-Based Uncertainty Measures for Fuzzy Neighborhood Multigranulation Rough Sets

Lin Sun, *Member*, *IEEE*, Lanying Wang, Weiping Ding*, *Senior Member*, *IEEE*, Yuhua Qian, *Member*, *IEEE*, and Jiucheng Xu

***Abstract*—For heterogeneous datasets containing numerical and symbolic feature values, feature selection based on fuzzy neighborhood multigranulation rough sets (FNMRS) is a very significant step to preprocess data and improve its classification performance. This paper presents an FNMRS-based feature selection approach in neighborhood decision systems. First, some concepts of fuzzy neighborhood rough sets and neighborhood multigranulation rough sets are given, and then the FNMRS model is investigated to construct uncertainty measures. Second, the optimistic and pessimistic FNMRS models are built by using fuzzy neighborhood multigranulation lower and upper approximations from algebra view, and some fuzzy neighborhood entropy-based uncertainty measures are developed in information view. Inspired by both algebra and information views based on the FNMRS model, the fuzzy neighborhood pessimistic multigranulation entropy is proposed. Third, the Fisher score model is utilized to delete irrelevant features to decrease the complexity of high-dimensional datasets, and then a forward feature selection algorithm is provided to promote the performance of heterogeneous data classification. Experimental results on twelve datasets show that the presented model is effective for selecting important features with higher stability of classification in neighborhood decision systems.**

***Index Terms*—Fuzzy neighborhood rough sets, neighborhood multigranulation rough sets, feature selection, neighborhood entropy, uncertainty measure.**

I. Introduction

RECENTLY, feature selection is a widely used technique as the step of data preprocessing in the fields of granular computing and artificial intelligence [1]-[4], and the process of handling complex data can hoist the performance and reduce the complexity of computation in some cases [5]-[7].

This work was supported in part by the National Natural Science Foundation of China under Grants 61772176, 61976082, 61976120, and 61672332, in part by the Plan for Scientific Innovation Talent of Henan Province under Grant 184100510003, in part by the Young Scholar Program of Henan Province under Grant 2017GGJS041, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20191445, and in part by the Six Talent Peaks Project of Jiangsu Province under Grant XYDXXJS-048, and sponsored by Qing Lan Project of Jiangsu Province. (*Corresponding author: Weiping Ding.*)

L. Sun, L. Wang, and J. Xu are with the College of Computer and Information Engineering, Henan Normal University, Xinxiang 453007, China (e-mail:linsunok@gmail.com, wanglanying5@126.com, jiuchxu@gmail.com).

W. Ding is with the School of Information Science and Technology, Nantong University, Nantong 226019, China (e-mail: dwp9988@163.com).

Y. Qian is with the Institute of Big Data Science and Industry, Shanxi University, Taiyuan 030006, China (e-mail: jinchengqyh@126.com).

A. Related Works

Rough sets and fuzzy sets as two powerful tools of feature selection have been reported more and more frequently on the uncertainty for data classification [8]; however, rough set models [9] can only handle symbolic datasets [10]. The continuous numerical datasets need to be divided into several intervals by different discretization methods, which results in loss of information [11]. Therefore, researchers have presented a variety of rough set models to solve this issue, where the most popular models are neighborhood rough sets (NRS) and fuzzy rough sets (FRS) [12]-[15]. Hu et al. [16] devised the NRS-based feature selection model to deal with mixed datasets. However, there still exist some limitations on the NRS model. For instance, it is difficult to characterize the fuzziness of objects under the fuzzy background [10],[17]. Dubois and Prade [18] studied rough fuzzy sets and FRS. Jensen and Shen [19] presented an FRS-based attribute reduction method to reduce data redundancy. However, the nearest samples are always used in FRS to calculate fuzzy upper and lower approximations of the decision, which leads to more risk-taking by datasets with noise. It becomes the main drawback of FRS.

At present, the technologies and growth have been made on feature selection based on rough sets for data classification [20]; however, most models are based on a single binary relation [7], which may limit applications and lead to an increase in complexity [21]. To overcome this drawback, multigranulation rough sets (MRS) have become a booming direction for feature selection. Qian et al. [21] investigated a pessimistic MRS decision model for attribute reduction. Yao and She [22] constructed a multigranulation space. However, the MRS models are not suitable for handling continuous numerical data toward real-world applications and cannot describe mixed data from neighborhood systems with numerical and symbolic values [23]. Thus, it is necessary to introduce the NRS into MRS. Lin et al. [24] introduced the MRS model based on the neighborhood relation to achieve attribute reduction. Sun et al. [7] presented the attribute reduction model with measures based on neighborhood multigranulation rough sets (NMRS). Since there exist still noises, uncertainty, and fuzziness in information systems, the research on fuzzy multigranulation rough sets (FMRS) is very meaningful [25]. Zhang et al. [26] provided the multiple hesitant fuzzy tolerance relations in the

multigranulation framework. However, there are few reports of the combination of NMRS and FMRS.

Over the last decades, uncertainty measures have rapid development for feature selection from algebra and information views [5],[7],[27]. Fan et al. [28] employed the max-decision NRS model to enlarge the positive region and dependency degree for reduction. Wang et al. [29] constructed the distance measure-based FRS model for fuzzy dependency and attribute significance to decision systems. Hu et al. [23] developed the matrix-based feature selection method to achieve the uncertainty of boundary regions in the NMRS model. All in all, the above references just discussed feature selection from the algebra view. Unfortunately, the significance of features based on algebra view can only state the influence of features included in feature subset [23],[28],[29]. Information entropy [30] and some deformations are widely utilized in feature selection. Zhang et al. [1] presented FRS-based information entropy to select features in fuzzy information systems. Xu et al. [14] studied the fuzzy neighborhood conditional entropy to evaluate feature significance in the FNRS. Zeng et al. [31] proposed optimistic and pessimistic multigranulation entropy for feature selection. So sadly, information view-based feature significance merely interprets the consequences on features from uncertainty classification [1],[14],[31],[32]. It would be a great topic to combine the two views for feature selection and promote the measure quality of uncertainty for neighborhood decision systems. Wang et al. [32] researched rough reduction and relative reduction from algebra and information views. Sun et al. [27] proposed the neighborhood entropy-based attribute reduction method combining algebra with information views. Sun et al. [7] constructed the NMRS-based attribute reduction method to deal with mixed and incomplete datasets in terms of algebra and information views. However, this study has rarely been reported from the perspective of the FMRS.

B. Our Work

In the FRS method, the most basic unit is the fuzzy information granules, and the degrees of membership of samples are calculated by using Min-Max operations [33]-[35]. Nevertheless, when information systems are noisy, there are some risks in the calculation of fuzzy approximation, which may affect the computation of membership degree and result in an increasing classification error rate [10],[36],[37]. Wang et al. [10] constructed the feature selection method in fuzzy neighborhood rough sets (FNRS). Yue et al. [36] developed the fuzzy neighborhood covering data classification. However, the parameters exerted the great influence on the classification performance of their proposed algorithms. Inspired by their contributions, to overcome the disadvantages and explain the sample decisions based on fuzzy information granules in FRS, the FNRS model is further investigated in this paper. Furthermore, from what we can tell, few reports for feature selection are in the news to integrate FNRS with NMRS in neighborhood decision systems. Inspired by this fact, it would be better to study fuzzy neighborhood multigranulation rough sets (FNMRS) and design the FNMRS-based feature selection model with heterogeneous datasets. Unfortunately, few innovations for feature selection have been offered from both algebra and information views for FNMRS in neighborhood decision systems. Therefore, motivated by this observation, it is necessary to study uncertainty measures based on FNMRS from these two views, and then a forward search algorithm for feature selection can be designed for heterogeneous datasets in neighborhood decision systems. In general, the summing up of our main contributions is as follows.

(1) The notions of the FNRS and optimistic and pessimistic NMRS models are given in neighborhood decision systems.

(2) To deal with heterogeneous datasets, FNRS is combined with MRS to construct FNMRS in neighborhood systems. Then the definitions of the optimistic and pessimistic FNMRS methods are provided in neighborhood decision systems.

(3) To better discuss the measures on the basis of algebra and information views for feature selection in the FNMRS model, some uncertainty measures based on fuzzy neighborhood entropy are investigated in detail. By using fuzzy neighborhood pessimistic multigranulation entropy, a novel feature subset selection algorithm is presented.

The remainder is organized as follows: Section II reviews the concepts of FNRS and NMRS. Section III constructs the FNMRS model and presents some fuzzy neighborhood entropy-based measures. The heuristic feature subset selection method is developed in Section IV. Section V provides the experimental analysis on twelve datasets. In Section VI, this paper ends with the conclusion and our future work.

II. PRELIMINARIES

A. Fuzzy Neighborhood Rough Sets

Let $NDS = <U, AT, D, V, f, \Delta, \delta>$ be a neighborhood decision system. $U = \{x_1, x_2, \cdots, x_m\}$ is the set of samples. AT and D are the sets of conditional attributes and decision classes. $V = \bigcup_{a\in AT} V_a$. f: $U \times \{AT \bigcup D\} \to V$ is the map function, and $f(a, x)$ is the attribute value of x on attribute a. Δ is a function of the distance. $0 \le \delta \le 1$ is a parameter of the neighborhood radius. The decision system is simplified to $NDS = <U, AT, D, f>$.

Definition 1 ([10],[12]): Given $NDS = <U, AT, D, f>$, $B \subseteq AT$ induces the fuzzy binary relation R_B on the universe U, any x, $y \in U$, and then R_B is the fuzzy similarity relation when it satisfies the following reflexive and symmetric properties, respectively:

(1) $R_B(x, x) = 1$, where any $x \in U$;

(2) $R_B(x, y) = R_B(y, x)$, where any $x, y \in U$.

Definition 2: Given $NDS = <U, AT, D, f>$, any $a \in AT$, the fuzzy neighborhood radius parameter is α ($0 < \alpha \le 1$) that describes the similarity of samples; then, for any x, $y \in U$, the fuzzy neighborhood similarity relation between two samples x and y in regard to a is denoted by

$$R_a = \begin{cases} 0, & |f(a,x) - f(a,y)| > \alpha \\ 1 - |f(a,x) - f(a,y)|, & |f(a,x) - f(a,y)| \le \alpha \end{cases}. \quad (1)$$

The matrix of fuzzy neighborhood similarity is $[x]_a(y) = R_a(x, y)$, and then $[x]_B^{\alpha}(y) = \min_{a\in B}([x]_a(y))$ with any $B \subseteq AT$ [14].

Definition 3: Given $NDS = <U, AT, D, f>$, $B \subseteq AT$, for any x, $y \in U$, the parameterized fuzzy neighborhood information

granule of x in regard to B is denoted by

$$\alpha_B(x)=[x]_B^{\alpha}(y)=\begin{cases}0, & R_B(x,y)<1-\alpha\\ R_B(x,y), & R_B(x,y)\geq 1-\alpha\end{cases}. \quad (2)$$

Property 1: Given $NDS = <U, AT, D, f>$, $P, Q \subseteq AT$, for any $x \in U$, the following properties are obtained:

(1) For any $p \in P$, $R_P = \bigcap_{p\in P} R_p$.

(2) If $P \subseteq Q$, then $R_Q \subseteq R_P$.

(3) If $P \subseteq Q$, then $\alpha_Q(x) \subseteq \alpha_P(x)$.

Proof. We provide a detailed proof of this property, which can be found in our supplementary file.

Definition 4: Given $NDS = <U, AT, D, f>$, $U = \{x_1, x_2, \cdots, x_m\}$, $A = \{A_1, A_2, \cdots, A_t\}$, $A \subseteq AT$ and $U/D = \{D_1, D_2, \cdots, D_l\}$; then, the fuzzy decisions of samples exported by D is denoted by

$$FD = \{FD_1^T, FD_2^T, \cdots, FD_l^T\}, \quad (3)$$

where $FD_j = \{FD_j(x_1), FD_j(x_2), \cdots, FD_j(x_m)\}$ is a fuzzy rough set of the decision equivalence class of samples, and $j = 1, 2, \cdots, l$. When $k = 1, 2, \cdots, m$, $FD_j(x_k)$ is the membership degree of $x_k \in U$ on FD_j and expressed [14] as

$$FD_j(x_k)=\frac{\left|[x_k]_A(y)\cap D_j\right|}{\left|[x_k]_A(y)\right|}, \quad (4)$$

where $[x_k]_A(y)$ is the fuzzy neighborhood similarity degree, $y \in U$, $D_j \in U/D$ and $j = 1, 2, \cdots, l$.

Definition 5: Given $NDS = <U, AT, D, f>$, $B \subseteq AT$, any $X \subseteq U$, $\alpha_B(x)$ is the parameterized fuzzy neighborhood information granule for $x \in U$; then the fuzzy neighborhood lower and upper approximations of X in relation to B are denoted, respectively, by

$$\underline{FN_B^{\alpha}}(X) = \{x \in U \mid \alpha_B(x) \subseteq X\}, \quad (5)$$

$$\overline{FN_B^{\alpha}}(X) = \{x \in U \mid \alpha_B(x) \cap X \neq \varnothing\}. \quad (6)$$

Definition 6: Suppose that any $D_j \in U/D = \{D_1, D_2, \cdots, D_l\}$, the fuzzy neighborhood positive region and its dependency degree of D with regard to B are denoted, respectively, as

$$POS_B^{\alpha}(D) = \bigcup_{j=1}^{l} \underline{FN_B^{\alpha}}(D_j), \quad (7)$$

$$d_B^{\alpha}(D)=\frac{\left|POS_B^{\alpha}(D)\right|}{\left|U\right|}. \quad (8)$$

B. Neighborhood Multigranulation Rough Sets

Definition 7 ([27]): Given $NDS = <U, AT, D, f>$, $B \subseteq AT$ and $x, y \in U$, then the neighborhood relation and neighborhood class are denoted, respectively, by

$$NR_{\delta}(B) = \{(x,y) \in U \times U \mid \Delta_B(x,y) \leq \delta\}, \quad (9)$$

$$n_B^{\delta}(x) = \{y \in U \mid \Delta_B(x,y) \leq \delta\}, \quad (10)$$

where $\Delta_B(x,y)=\sqrt{\sum_{k=1}^{|B|}(f(a_k,x)-f(a_k,y))^2}$ is the Euclidean distance function, $a_k \in B$, and $|\ |$ represents the cardinality.

Definition 8: Given $NDS = <U, AT, D, f>$, $X \subseteq U$, $A \subseteq AT$ and $A = \{A_1, A_2, \cdots, A_t\}$; then, the optimistic neighborhood multigranulation lower and upper approximations of X in relation to A_i are denoted, respectively, by

$$\bigcup_{A_i\subseteq A}\underline{apr_{A_i}^{O}}(X) = \{x \in U \mid n_{A_1}^{\delta}(x) \subseteq X \vee n_{A_2}^{\delta}(x) \subseteq X \vee \cdots \vee n_{A_t}^{\delta}(x) \subseteq X\}, \quad (11)$$

$$\bigcup_{A_i\subseteq A}\overline{apr_{A_i}^{O}}(X) = \sim(\bigcup_{A_i\subseteq A}\underline{apr_{A_i}^{O}}(\sim X)), \quad (12)$$

where $n_{A_i}^{\delta}(x)$ represents the neighborhood class of x on $A_i \subseteq A$ and $i = 1, 2, \cdots, t$. Then, in neighborhood decision systems, $(\bigcup_{A_i\subseteq A}\underline{apr_{A_i}^{O}}(X), \bigcup_{A_i\subseteq A}\overline{apr_{A_i}^{O}}(X))$ is named as an optimistic NMRS model.

Definition 9: Assume that $U/D = \{D_1, D_2, \cdots, D_l\}$, the optimistic positive region and its dependency degree of D in relation to A in the optimistic NMRS model are denoted, respectively, by

$$POS_A^{O}(D) = \bigcup_{j=1}^{l}\{\bigcup_{i=1}^{t}\underline{apr_{A_i}^{O}}(D_j)\}, \quad (13)$$

$$d_A^{O}(D)=\frac{|POS_{A_i}^{O}(D)|}{|U|}, \quad (14)$$

where $A_i \subseteq A$, $i = 1, 2, \cdots, t$, $D_j \in U/D$ and $j = 1, 2, \cdots, l$.

Definition 10: Given $NDS = <U, AT, D, f>$, any $X \subseteq U$, $A \subseteq AT$ and $A = \{A_1, A_2, \cdots, A_t\}$; then, the pessimistic neighborhood multigranulation lower and upper approximations of X with respect to A_i are denoted, respectively, by

$$\bigcap_{A_i\subseteq A}\underline{apr_{A_i}^{P}}(X) = \{x \in U \mid n_{A_1}^{\delta}(x) \subseteq X \wedge n_{A_2}^{\delta}(x) \subseteq X \wedge \cdots \wedge n_{A_t}^{\delta}(x) \subseteq X\}, \quad (15)$$

$$\bigcap_{A_i\subseteq A}\overline{apr_{A_i}^{P}}(X) = \sim(\bigcap_{A_i\subseteq A}\underline{apr_{A_i}^{P}}(\sim X)). \quad (16)$$

Thus, in neighborhood decision systems, $(\bigcap_{A_i\subseteq A}\underline{apr_{A_i}^{P}}(X), \bigcap_{A_i\subseteq A}\overline{apr_{A_i}^{P}}(X))$ is named as a pessimistic NMRS model.

Definition 11: Assume that $U/D = \{D_1, D_2, \cdots, D_l\}$, the pessimistic positive region and its dependency degree of D with respect to A in the pessimistic NMRS model are denoted, respectively, by

$$POS_A^{P}(D) = \bigcup_{j=1}^{l}\{\bigcap_{i=1}^{t}apr_{A_i}^{P}(D_j)\}, \quad (17)$$

$$d_A^{P}(D)=\frac{|POS_A^{P}(D)|}{|U|}, \quad (18)$$

where $A_i \subseteq A$, $i = 1, 2, \cdots, t$, $D_j \in U/D$ and $j = 1, 2, \cdots, l$.

III. Uncertainty Measures of The FNMRS Model Based on Fuzzy Neighborhood Entropy

A. The FNMRS Model

Definition 12: Given $NDS = <U, AT, D, f>$, any $X \subseteq U$, $A = \{A_1, A_2, \cdots, A_t\}$ and $A \subseteq AT$; then, the fuzzy neighborhood optimistic multigranulation lower and upper approximations of X in relation to A_i are denoted, respectively, by

$$\bigcup_{A_i\subseteq A}\underline{apr_{A_i}^{O,\alpha}}(X)=\{x\in U\mid \alpha_{A_1}(x)\subseteq X\vee\alpha_{A_2}(x)\subseteq X\vee\cdots\vee\alpha_{A_t}(x)\subseteq X\},\tag{19}$$

$$\bigcup_{A_i\subseteq A}\overline{apr_{A_i}^{O,\alpha}}(X)=\sim(\bigcup_{A_i\subseteq A}\underline{apr_{A_i}^{O,\alpha}}(\sim X)),\tag{20}$$

where $\alpha_{A_i}(x)$ is the parameterized fuzzy neighborhood granule of $x\in U$, $A_i\subseteq A$ and i = 1, 2, ⋯, t. Then, in neighborhood decision systems, $(\bigcup_{A_i\subseteq A}\underline{apr_{A_i}^{O,\alpha}}(X),\ \bigcup_{A_i\subseteq A}\overline{apr_{A_i}^{O,\alpha}}(X))$ is named as an optimistic FNMRS model.

Definition 13: Given $NDS=<U, AT, D, f>$, any $X\subseteq U$, $A=\{A_1, A_2, \cdots, A_t\}$, $A\subseteq AT$ and $U/D=\{D_1, D_2, \cdots, D_l\}$; then, the fuzzy neighborhood optimistic positive region and its dependency degree of D in relation to A in the optimistic FNMRS model are denoted, respectively, by

$$POS_A^{O,\alpha}(D)=\bigcup_{j=1}^{l}\{\bigcup_{i=1}^{t}\underline{apr_{A_i}^{O,\alpha}}(D_j)\},\tag{21}$$

$$d_A^{O,\alpha}(D)=\frac{\left|POS_A^{O,\alpha}(D)\right|}{|U|},\tag{22}$$

where $A_i\subseteq A$, i = 1, 2, ⋯, t, $D_j\in U/D$ and j = 1, 2, ⋯, l.

Definition 14: Given $NDS=<U, AT, D, f>$, any $X\subseteq U$, $A=\{A_1, A_2, \cdots, A_t\}$ and $A\subseteq AT$; then, the fuzzy neighborhood pessimistic multigranulation lower and upper approximations of X in relation to A_i are denoted, respectively, by

$$\bigcap_{A_i\subseteq A}\underline{apr_{A_i}^{P,\alpha}}(X)=\{x\in U\mid \alpha_{A_1}(x)\subseteq X\wedge\alpha_{A_2}(x)\subseteq X\wedge\cdots\wedge\alpha_{A_t}(x)\subseteq X\},\tag{23}$$

$$\bigcap_{A_i\subseteq A}\overline{apr_{A_i}^{P,\alpha}}(X)=\sim(\bigcap_{A_i\subseteq A}\underline{apr_{A_i}^{P,\alpha}}(\sim X)),\tag{24}$$

where $x\in U$, $A_i\subseteq A$ and i = 1, 2, ⋯, t.

Then, in neighborhood decision systems, $(\bigcap_{A_i\subseteq A}\underline{apr_{A_i}^{P,\alpha}}(X),\ \bigcap_{A_i\subseteq A}\overline{apr_{A_i}^{P,\alpha}}(X))$ is named as a pessimistic FNMRS model.

Definition 15: Given $NDS=<U, AT, D, f>$, any $X\subseteq U$, $A=\{A_1, A_2, \cdots, A_t\}$, $A\subseteq AT$ and $U/D=\{D_1, D_2, \cdots, D_l\}$; then, the fuzzy neighborhood pessimistic positive region and its dependency degree of D in relation to A in the pessimistic FNMRS model are denoted, respectively, by

$$POS_A^{P,\alpha}(D)=\bigcup_{j=1}^{l}\{\bigcap_{i=1}^{t}\underline{apr_{A_i}^{P,\alpha}}(D_j)\},\tag{25}$$

$$d_A^{P,\alpha}(D)=\frac{\left|POS_A^{P,\alpha}(D)\right|}{|U|},\tag{26}$$

where $A_i\subseteq A$, i = 1, 2, ⋯, t, $D_j\in U/D$ and j = 1, 2, ⋯, l.

Proposition 1: Given $NDS=<U, AT, D, f>$ with any $X\subseteq U$, $R\subseteq S\subseteq AT$, the following properties hold:

(1) $POS_R^{O,\alpha}(X)\subseteq POS_S^{O,\alpha}(X)$ and $d_R^{O,\alpha}(X)\le d_S^{O,\alpha}(X)$.

(2) $POS_S^{P,\alpha}(X)\subseteq POS_R^{P,\alpha}(X)$ and $d_S^{P,\alpha}(X)\le d_R^{P,\alpha}(X)$.

Proof. We provide a detailed proof of this proposition, which can be found in our supplementary file.

B. Fuzzy Neighborhood Entropy-Based Uncertainty Measures

To efficiently analyze decision-making from the viewpoint of multigranulation, Sun et al. [7] and You et al. [38] adapted the conservative decision strategy [21]. Motivated by this fact, to further develop the pessimistic FNMRS model, some uncertainty measures based on fuzzy neighborhood pessimistic multigranulation entropy are provided, and the corresponding properties are constructed in the FNMRS model.

Definition 16: Given $NDS=<U, AT, D, f>$, $B\subseteq AT$, $U=\{x_1, x_2, \cdots, x_m\}$, for any $x_k\in U$, k = 1, 2, ⋯, m, and $n_B^\delta(x_k)$ is the neighborhood class of x_k with regard to B; then the neighborhood entropy of B is denoted by

$$NE(B)=-\sum_{k=1}^{m}\frac{\left|n_B^\delta(x_k)\right|}{|U|}\log_2\frac{\left|n_B^\delta(x_k)\right|}{|U|}.\tag{27}$$

Definition 17: Given $NDS=<U, AT, D, f>$, $B\subseteq AT$ and $U/D=\{D_1, D_2, \cdots, D_l\}$; the neighborhood joint entropy of B and D is denoted by

$$NE(B,D)=-\frac{\left|n_B^\delta(x_k)\cap D_j\right|}{|U|}\sum_{k=1}^{m}\sum_{j=1}^{l}\log_2\frac{\left|n_B^\delta(x_k)\cap D_j\right|}{|U|},\tag{28}$$

where $x_k\in U$, k = 1, 2, ⋯, m, $D_j\in U/D$ and j = 1, 2, ⋯, l.

Definition 18 ([14]): Assume that there are two fuzzy sets M and N on the universe U; $|M\cap N|$ is the size of samples whose membership degree on M is less than or equal to that of N.

Definition 19: Given $NDS=<U, AT, D, f>$, $A\subseteq AT$ and $U=\{x_1, x_2, \cdots, x_m\}$; then the fuzzy neighborhood entropy of A is denoted by

$$FNE_\alpha(A)=-\sum_{k=1}^{m}\frac{\left|\alpha_A(x_k)\right|}{|U|}\log_2\frac{\left|\alpha_A(x_k)\right|}{|U|},\tag{29}$$

where $\alpha_A(x_k)$ is the parameterized fuzzy neighborhood granule, $x_k\in U$ and k = 1, 2, ⋯, m.

Definition 20: Given $NDS=<U, AT, D, f>$, $A\subseteq AT$ and $U/D=\{D_1, D_2, \cdots, D_l\}$; the fuzzy neighborhood joint entropy of A and D is denoted by

$$FNE_\alpha(A,D)=-\sum_{k=1}^{m}\sum_{j=1}^{l}\frac{\left|\alpha_A(x_k)\cap FD_j\right|}{|U|}\log_2\frac{\left|\alpha_A(x_k)\cap FD_j\right|}{|U|},\tag{30}$$

where $x_k\in U$, k = 1, 2, ⋯, m, FD_j is a fuzzy rough set of the decision equivalence class of samples, j = 1, 2, ⋯, l, and $|\alpha_A(x_k)\cap FD_j|$ is the size of samples whose membership degree of $\alpha_A(x_k)$ is less than or equal to that of FD_j.

Proposition 2: Given $NDS=<U, AT, D, f>$ with any $R\subseteq S\subseteq AT$, $FNE_\alpha(R,D)\le FNE_\alpha(S,D)$ holds.

Proof. We provide a detailed proof of this proposition, which can be found in our supplementary file.

Definition 21: Given $NDS=<U, AT, D, f>$, $A\subseteq AT$, $A=\{A_1, A_2, \cdots, A_t\}$, $A_i\subseteq A$, $U=\{x_1, x_2, \cdots, x_m\}$ and $U/D=\{D_1, D_2, \cdots, D_l\}$; the fuzzy neighborhood pessimistic multigranulation entropy (FNPME) of D and A_i is denoted by

$$FNPME_\alpha(A_i,D)=-\sum_{i=1}^{t}(1-d_{A_i}^{P,\alpha}(D))\times\sum_{j=1}^{l}\sum_{k=1}^{m}\frac{\left|\alpha_{A_i}(x_k)\cap FD_j\right|}{|U|}\log_2\frac{\left|\alpha_{A_i}(x_k)\cap FD_j\right|}{|U|},\tag{31}$$

where $d_{A_i}^{P,\alpha}(D)$ is the fuzzy neighborhood pessimistic dependency degree of D in relation to A_i, $i = 1, 2, \cdots, t$, $\alpha_{A_i}(x_k)$ is the parameterized fuzzy neighborhood granule, $x_k \in U$ with $k = 1, 2, \cdots, m$, $\left|\alpha_A(x_k) \cap FD_j\right|$ is the nonzero number of samples with membership degree of $\alpha_{A_i}(x_k)$, which is less than or equal to that of FD_j, and $j = 1, 2, \cdots, l$.

Property 2: Given $NDS = <U, AT, D, f>$ with any $A \subseteq AT$, $A_i \subseteq A = \{A_1, A_2, \cdots, A_t\}$, and $U = \{x_1, x_2, \cdots, x_m\}$; then, one has

$$FNPME_\alpha(A_i, D) = \sum_{i=1}^{t}(1 - d_{A_i}^{P,\alpha}(D)) \times FNE_\alpha(A_i, D) \geq 0 .$$

Proof. We provide a detailed proof of this property, which can be found in our supplementary file.

Remark 1: From Definition 21 and Property 2, $d_A^{P,\alpha}(D)$ denotes the fuzzy neighborhood pessimistic dependency degree from algebra view, and $FNE_\alpha(A, D)$ represents the fuzzy neighborhood joint entropy from information view. Thus, FNPME is able to measure the uncertainty of neighborhood decision systems from both algebra and information views.

IV. Feature Selection Method for The FNMRS Model in Neighborhood Decision Systems

A. FNPME-Based Feature Selection

Proposition 3: Given $NDS = <U, AT, D, f>$ with any $R \subseteq S \subseteq AT$, $FNPME_\alpha(R, D) \leq FNPME_\alpha(S, D)$ holds.

Proof. We provide a detailed proof of this proposition, which can be found in our supplementary file.

Remark 2: As a significant aspect of uncertainty measure, the monotonicity plays an important role in feature selection. Proposition 3 indicates that as the feature subset is going to increase, FNPME is in a monotonic increasing trend; then the degree of decision-making will be higher. Thus, the monotonic increase of FNPME leads us to select the greedy algorithm for feature selection in neighborhood decision systems.

Definition 22: Given $NDS = <U, AT, D, f>$, $A' \subseteq A \subseteq AT$, $A = \{A_1, A_2, \cdots, A_t\}$, if $FNPME_\alpha(A', D) = FNPME_\alpha(A, D)$ and there exists $FNPME_\alpha(A', D) > FNPME_\alpha(A' - A_i, D)$ for any $A_i \subseteq A'$ and $i = 1, 2, \cdots, t$, then A' is a reduct of A with respect to D.

Definition 23: Given $NDS = <U, AT, D, f>$, $A' \subseteq A \subseteq AT$, $A = \{A_1, A_2, \cdots, A_t\}$, $A_i \subseteq A'$, and $i = 1, 2, \cdots, t$, the internal significance of attribute subset A_i with respect to D is denoted by

$$Sig^{inner}(A_i, A', D) = FNPME_\alpha(A', D) - FNPME_\alpha(A' - A_i, D) . \quad (32)$$

Definition 24: Given $NDS = <U, AT, D, f>$, $A' \subseteq A \subseteq AT$ and $A = \{A_1, A_2, \cdots, A_t\}$, for any $A_i \subseteq A'$, $i = 1, 2, \cdots, t$, if $Sig^{inner}(A_i, A', D) > 0$, then A_i in A' is necessary; otherwise A_i is unnecessary. If each A_i in A' is necessary, then A' is independent.

Definition 25: Given $NDS = <U, AT, D, f>$, $A \subseteq AT$ and $A = \{A_1, A_2, \cdots, A_t\}$, for any $A_i \subseteq AT$, $i = 1, 2, \cdots, t$, if $Sig^{inner}(A_i, AT, D) > 0$, then A_i is called a core of A_i with respect to AT.

Definition 26: Given $NDS = <U, AT, D, f>$, $A' \subseteq A \subseteq AT$ and $A = \{A_1, A_2, \cdots, A_t\}$, $A_i \subseteq A - A'$, $i = 1, 2, \cdots, t$; then, the external significance of attribute subset A_i with respect to D is denoted by

$$Sig^{outer}(A_i, A', D) = FNPME_\alpha(A' \cup A_i, D) - FNPME_\alpha(A', D) . \quad (33)$$

Remark 3: In $NDS = <U, AT, D, f>$, $A = \{A_1, A_2, \cdots, A_t\}$, for any $A_i \subseteq A' \subseteq A \subseteq AT$, and $i = 1, 2, \cdots, t$, when calculating $Sig^{inner}(A_i, A', D)$, one only obtains $FNPME_\alpha(A' - A_i, D)$ each time since $FNPME_\alpha(A', D)$ may be a constant. Therefore, similarly one just calculates $FNPME_\alpha(A' \cup A_i, D)$ for $Sig^{outer}(A_i, A', D)$, where any $A_i \subseteq A - A'$.

B. Feature Selection Algorithm

In this section, an FNPME-based feature selection (FNPME-FS) method is demonstrated in Algorithm 1.

Algorithm 1. FNPME-FS

Input: $NDS = <U, AT, D, f>$ with a fuzzy neighborhood radius parameter α.
Output: An optimal feature subset RED.

1. Initialize $RED = \emptyset$ and $B = \emptyset$.
2. Calculate $FNPME_\alpha(AT, D)$.
3. **FOR** $i = 1$ to t do
4. Compute $Sig^{inner}(A_i, AT, D)$.
5. **IF** $Sig^{inner}(A_i, AT, D) > 0$
6. then let $RED = RED \cup A_i$.
7. **ENDIF**
8. **ENDFOR**
9. Let $B = AT - RED$.
10. **WHILE** $FNPME_\alpha(RED, D) \neq FNPME_\alpha(AT, D)$
11. **FOR** $j = 1$ to $|B|$
12. Calculate $FNPME_\alpha(RED \cup A_j, D)$.
13. Select A_j that satisfies $\max\{A_j \in R | FNPME_\alpha(RED \cup A_j, D)\}$, and if multiple feature subsets exhibit the maximum value, this front is selected.
14. **ENDFOR**
15. Let $RED = RED \cup A_j$ and $B = B - A_j$, and calculate $FNPME_\alpha(RED, D)$.
16. **ENDWHILE**
17. **FOR** $k = 1$ to $|RED|$ do
18. Select $B_k \in RED$.
19. Calculate $FNPME_\alpha(RED - \{B_k\}, D)$.
20. **IF** $FNPME_\alpha(RED - \{B_k\}, D) \geqslant FNPME_\alpha(RED, D)$
21. then let $RED = RED - \{B_k\}$.
22. **ENDIF**
23. **ENDFOR**
24. **RETURE** the optimal feature subset RED.

For this FNPME-FS method, the worst influence on complexity is the calculation of the parameterized fuzzy neighborhood granule. The buckets sorting algorithm [39] was first introduced to decrease the time complexity of computing the parameterized fuzzy neighborhood granule, and then its time complexity is reduced to $O(mn)$, where m and n respectively express the sizes of objects and features. Thus, the computational complexity of the FNPME is approximately $O(n)$. Because two loops are in Steps 3 to 8 and Steps 10 to 16 of FNPME-FS, respectively, its time complexity is $O(n^3m)$ at worst. Assume that the size of the selected feature subset is r, when calculating the fuzzy neighborhood granules, one need to take the candidate feature subsets into account. Thus, the computational complexity of the fuzzy neighborhood granule is close to $O(rm)$. In this FNPME-FS algorithm, $n - r$ and m are the running times in the inner and outer loops, respectively. Thus, the worst complexity of FNPME-FS is closely $O(rm(n-r)n)$. Since $r \ll n$ for the most part, the total computational time complexity of implementing FNPME-FS is approximate to $O(mn)$. Moreover, this space complexity of FNPME-FS is $O(mn)$.

V. Experimental Analyses

A. Data Sets and Experimental Design

To test the feasibility and stability of our developed model, our total experiments can be divided into four portions, and twelve public datasets including seven low-dimensional UCI datasets and five high-dimensional microarray gene expression datasets are selected, where seven UCI datasets could be downloaded at http://archive.ics.uci.edu/ml/datasets.php, and five gene expression datasets could be downloaded at http://portals.broadinstitute.org/cgi-bin/cancer/datasets.cgi. Table I lists all datasets.

TABLE I
TWELVE DATASETS IN EXPERIMENTS

No.	Datasets	Samples	Features	Classes
1	Glass	214	10	6
2	Heart	270	13	2
3	Ionosphere	351	33	2
4	Sonar	208	60	2
5	Wdbc	569	31	2
6	Wine	178	13	3
7	Wpbc	198	34	2
8	Breast	84	9216	5
9	Colon	62	2000	2
10	DLBCL	77	5469	2
11	Leukemia	72	7129	2
12	Prostate	136	12600	2

All experiments are executed on MATLAB 2016a under Intel (R) i5 CPU @ 3.20 GHz & 4.0 GB RAM; WEKA 3.8 is employed to verify classification efficiency under three classifiers KNN, CART and C4.5 whose default values of all parameters are selected; and ten-fold cross-validation is adopted for classification of the selected feature subset in our following subsections. The classifier is learned under ten-fold cross-validation; then it can be set by the training datasets, and the accuracy of classification is obtained on the testing datasets.

B. Performance on Different Fuzzy Neighborhood Parameters

The second subsection focuses on classification performance of different fuzzy neighborhood parameters and selects an appropriate parameter for each dataset. To effectively assess the classification effectiveness, the reduction rate [7],[40] is employed to appraise the reduction performance of feature selection and select the right fuzzy neighborhood parameters.

As far as five gene expression datasets with high dimensions are concerned, in order to efficiently reduce time cost, one usually adopts heuristic strategy to independently calculate the scores of each gene in metrics of a certain criteria, and then Fisher score [27] is used as the extraction strategy of genes for preliminarily reducing dimensions because it has many merits: less computation, strong operability, high precision, and effective decreasing computational complexity. It is employed as the common feature correlation criterion to calculate the jth Fisher score value [39]. By calculating its Fisher score value for each gene, a feature subset will be formed by the first m genes. The Fisher score method is demonstrated as Algorithm 2.

To set an appropriate dimension of each gene expression dataset, through employing the Fisher score method, the classification accuracy under classifier KNN is acquired under seven different dimensions (10, 50, 100, 200, 300, 400, and 500). Fig. 1 reveals the variation trend of accuracy via the selected gene subsets under different dimensions for five gene expression datasets. As seen in Fig. 1, as the size of gene subset changes, so does the accuracy in most cases. Then, the optimal balance of the selected gene subset and the accuracy are found to obtain right dimension for the next feature selection of each gene expression dataset. Thus, from Fig. 1, the f values of three datasets Colon, DLBCL, and Prostate are set as 100 dimensions, respectively, the 200 dimensions are appropriate for dataset Breast, and the 300 dimensions are for dataset Leukemia.

Algorithm 2. Fisher score

Input: The matrix $X \in R^{m \times n}$ of original gene expression dataset with m objects and n genes, and the expected size of selected genes is f.

Output: The selected gene subset S.

1. **FOR** each gene
2. Calculate the Fisher score value by Eq. (16) in [39].
3. **ENDFOR**
4. Use the radix sorting algorithm in [41] to sort the Fisher score values in descending order.
5. Select the first f genes with larger value and put the gene sequence into the set T.
6. Obtain the gene subset S after dimension reduction by using T.
7. **RETURE** the selected gene subset S.

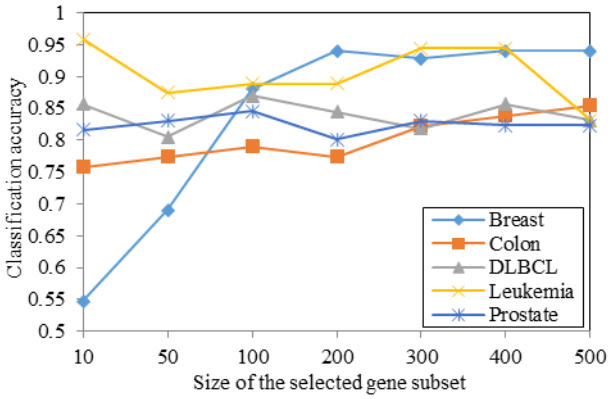


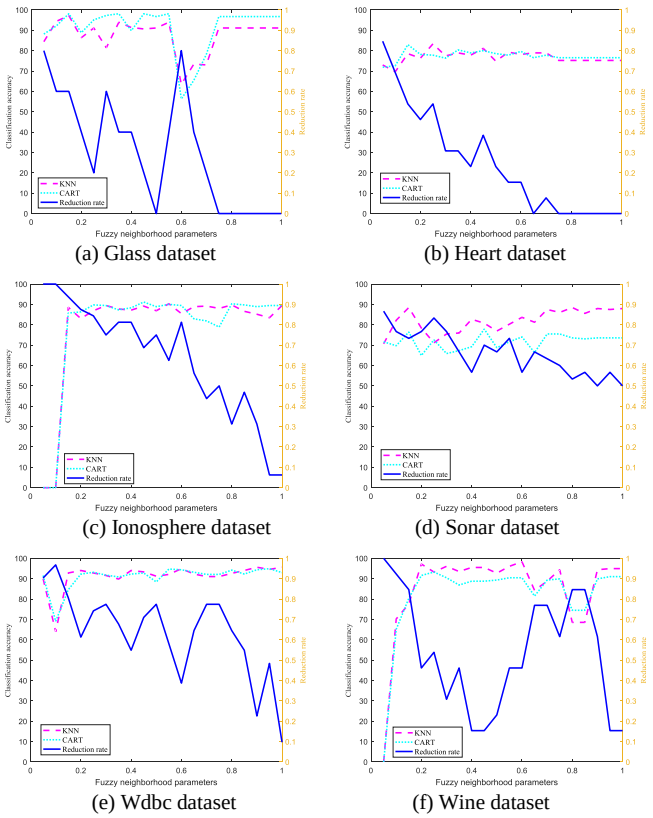
Fig. 1. Classification accuracy via the size of gene subset on five gene datasets.

To simulate multigranulation scenarios in selecting proper feature subset, the first point to be done is that each dataset will start with the first feature and form a feature subset for every two features on each low-dimensional UCI datasets and four features on each gene expression dataset with high dimensions in the following experiments. When the size of feature subset cannot be divided exactly, the final remaining features form a feature subset. Next, this accuracy of the selected feature subset by using FNPME-FS on all datasets is acquired with different fuzzy neighborhood parameter (α) values. The curves of the classification accuracy via the reduction rate for twelve datasets are displayed in Fig. 2 as the value of fuzzy neighborhood parameter increases, where the horizontal axis is the value of α, and it will be set in steps of 0.05 from 0 to 1; the two vertical axes, respectively, delegate the classification accuracy and the reduction rate; the classification accuracy of seven UCI datasets under two classifiers KNN and CART is illustrated in Figs. 2(a)-2(g); and the accuracy for five gene datasets on three classifiers KNN, CART and C4.5 is shown in Figs. 2(h)-2(l).

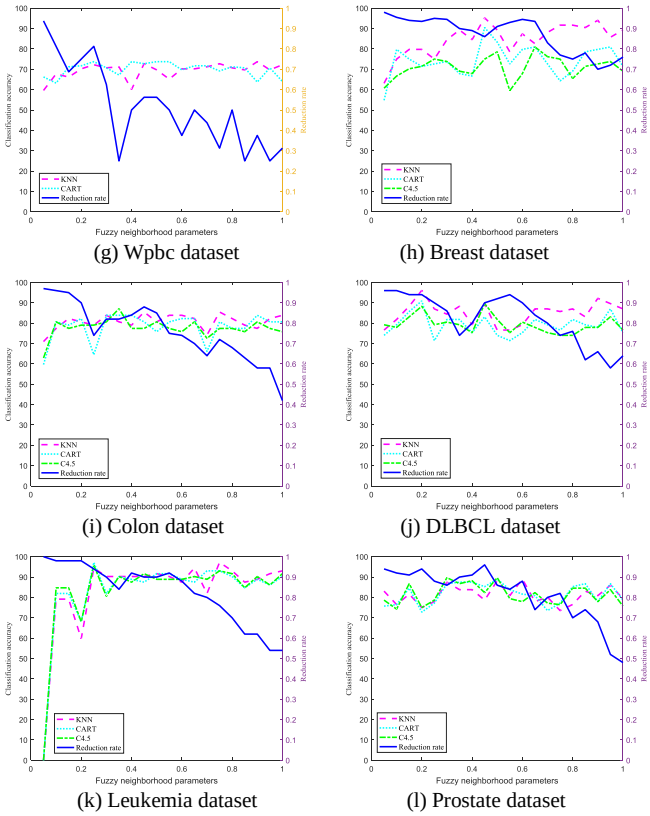
In what follows, the influence of the parameter α on our method will be analyzed in detail. The α can be considered as a threshold controlling the feature search. Fig. 2 shows that when

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2020.2989098, IEEE Transactions on Fuzzy Systems

the fuzzy neighborhood parameter value becomes large, the accuracy and the reduction rate are all changing, namely, the reduction rate reduces as the α increases in most cases, and the accuracy will appear the different changes. As seen in Fig. 2, the different thresholds indeed make a certain impact on the classification performance of our algorithm. Fortunately, all test datasets can exhibit the high classification accuracy in a wider range of α. For seven UCI datasets, the values of α are set to 0.15-0.55 under the classifier KNN and 0.2-0.45 under the classifier CART, respectively. For five gene expression datasets, the values of α are set as 0.2-0.6 under the classifier KNN, 0.1-0.4 under classifier CART, and 0.1-0.65 under classifier C4.5, respectively. Thus, these datasets exhibit the relatively satisfied accuracy in terms of larger reduction rate (namely fewer features selected in feature selection), and achieve a wide area of classification accuracy in terms of feasibility and stability for their respective value domains of α. Generally speaking, our experimental results demonstrate the efficiency of the developed model, which can achieve an optimal subset of features for complex data classification. Especially, the selected optimal value of α is different among twelve datasets. Therefore, one indeed needs to train the α before reducing the given datasets in the experiments. As described in Fig. 2, the optimal α values have been trained and then can be selected from between 0.15 and 0.55 for the UCI datasets and between 0.1 and 0.65 for the gene datasets with a step of 0.05, respectively. In particular, the significant standard of selecting parameters is to maximally balance classification accuracy and reduction rate; then Tables II and III illustrate the optimal fuzzy neighborhood parameters selected from Fig. 2 for each dataset.



(a) Glass dataset (b) Heart dataset (c) Ionosphere dataset (d) Sonar dataset (e) Wdbc dataset (f) Wine dataset



(g) Wpbc dataset (h) Breast dataset (i) Colon dataset (j) DLBCL dataset (k) Leukemia dataset (l) Prostate dataset

Fig. 2. Classification accuracy via reduction rate under different fuzzy neighborhood parameter values on twelve datasets.

TABLE II
SELECTED FUZZY NEIGHBORHOOD PARAMETERS FOR SEVEN UCI DATASETS

Datasets	KNN	CART
Glass	0.35	0.35
Heart	0.25	0.2
Ionosphere	0.55	0.45
Sonar	0.15	0.45
Wdbc	0.15	0.25
Wine	0.2	0.25
Wpbc	0.25	0.25

TABLE III
SELECTED FUZZY NEIGHBORHOOD PARAMETERS FOR FIVE GENE DATASETS

Datasets	KNN	CART	C4.5
Breast	0.6	0.1	0.65
Colon	0.45	0.1	0.1
DLBCL	0.2	0.2	0.2
Leukemia	0.25	0.25	0.25
Prostate	0.6	0.4	0.4

C. *Classification Results of the UCI Datasets*

This first portion of the subsection pays attention to the size of the selected feature subset by all compared methods and our optimal feature subset selected by FNPME-FS on seven UCI datasets. FNPME-FS is contrasted with five feature selection methods FSNTDJE [40], FNCE [14], IFPR [13], OMGE [31], and PMGE [31]. By using the obtained fuzzy neighborhood parameters, the features of raw data and the average size of the selected feature subset by the above six feature selection models employing ten-fold cross-validation on seven UCI datasets are displayed in Table IV. Then the optimal feature subsets selected by FNPME-FS for all UCI datasets under two classifiers KNN and CART are shown in Table V.

TABLE IV
NUMBER OF SELECTED FEATURES WITH SIX FEATURE SELECTION METHODS

Datasets	Raw	FSNTDJE	FNCE	IFPR	OMGE	PMGE	FNPME-FS
Glass	15	7.5	12.3	10.1	**5.1**	6.4	5.7
Heart	10	8.1	7.4	7.7	6.9	6	**5.5**
Ionosphere	33	9	16	28	8.2	**6.4**	11.6
Sonar	60	14.8	21.7	12.9	**7**	8	17.5
Wdbc	31	11.4	**4.2**	17.4	5	10.7	7
Wine	13	7.2	5.9	6.6	4.4	**4**	6
Wpbc	34	7.6	8.1	10	5.3	**4.8**	5.7
Mean	28	9.37	10.8	13.24	**5.99**	6.61	8.42

TABLE V
OPTIMAL FEATURE SUBSET WITH FNPME-FS FOR SEVEN UCI DATASETS

Datasets	KNN	CART
Glass	{1 2 7 8 9 10}	{1 2 7 8 9 10}
Heart	{1 2 3 4 11 12}	{1 10 11 12 13}
Ionosphere	{7 8 9 10 11 12 13 14 29 30 31 32}	{5 6 7 8 13 14 15 16 29 30}
Sonar	{1 2 3 4 9 10 17 18 19 20 23 24 37 38 45 46}	{1 2 7 8 9 10 15 16 29 30}
Wdbc	{1 2 5 6 27 28}	{1 2 3 4 5 6 27 28}
Wine	{1 2 7 8 11 12 13}	{1 2 7 8 11 12}
Wpbc	{1 2 13 14 23 24}	{1 2 13 14 23 24}

Table IV exhibits the average size of the selected feature subset by the six methods. Compared with OMGE and PMGE, our method selects more features for six datasets Glass, Ionosphere, Sonar, Wdbc, Wine, and Wpbc. The average size of the selected feature subset by FNPME-FS is 5.5 and reaches the minimum on dataset Heart. On the whole, the mean size of the selected feature subset by using FNPME-FS is the minimum against FSNTDJE, FNCE, and IFPR for the seven UCI datasets.

This second portion is to reveal the effectiveness of classification for FNPME-FS, six feature selection methods including the ODP (original data processing), FNCE [14], FSNTDJE [40], IFPR [13], OMGE [31], and PMGE [31], are applied to the comparison of classification accuracy on the optimal subsets of the selected features under two classifiers KNN and CART for seven UCI datasets. Then the average classification accuracy of the selected feature subsets by seven models under classifiers KNN and CART is demonstrated in Tables VI and VII, respectively.

TABLE VI
CLASSIFICATION ACCURACY WITH SEVEN METHODS ON CLASSIFIER KNN

Datasets	ODP	FNCE	FSNTDJE	IFPR	OMGE	PMGE	FNPME-FS
Glass	0.9112	0.8627	0.8592	0.8209	0.6822	0.9252	**0.9393**
Heart	0.7518	0.7213	0.7741	0.7889	0.8037	0.7851	**0.8333**
Ionosphere	0.8245	0.8823	0.8467	0.8020	0.8746	0.8718	**0.9031**
Sonar	0.8653	0.8001	0.6975	0.7598	0.8269	0.7932	**0.8846**
Wdbc	0.9124	**0.9825**	0.9466	0.9311	0.9613	0.9525	0.9279
Wine	0.9195	0.9073	0.9057	0.9429	0.9550	0.9382	**0.9719**
Wpbc	0.6969	0.6572	**0.7374**	0.6840	0.7042	0.6667	0.7222
Mean	0.8402	0.8305	0.8238	0.8185	0.8297	0.8475	**0.8915**

TABLE VII
CLASSIFICATION ACCURACY WITH SEVEN METHODS ON CLASSIFIER CART

Datasets	ODP	FNCE	FSNTDJE	IFPR	OMGE	PMGE	FNPME-FS
Glass	0.9672	0.8475	0.8915	0.9074	0.6869	0.9413	**0.9813**
Heart	0.7851	0.7679	0.7653	0.6666	**0.8333**	0.7851	0.7815
Ionosphere	0.8645	0.8053	0.7620	0.8704	0.8473	0.8917	**0.9117**
Sonar	0.7115	0.7616	0.6707	0.6092	0.7740	0.7088	**0.7789**
Wdbc	0.8595	0.8807	0.9124	0.8994	**0.9490**	0.9402	0.9314
Wine	0.8932	0.8249	0.8027	0.8727	0.8932	0.9157	**0.9326**
Wpbc	0.7121	0.6143	0.6869	0.6658	0.7020	0.7171	**0.7374**
Mean	0.8276	0.7860	0.7845	0.7845	0.8122	0.8428	**0.8649**

As observed from the results of six methods shown in Table IV, Tables VI and VII illustrate the differences among the seven methods. As we have seen, for almost all datasets, the average classification accuracy of FNPME-FS obviously outperforms the other methods, except for two datasets Wdbc and Wpbc on the KNN classifier and two datasets Heart and Wdbc on classifier CART. What is more, the mean classification accuracy of FNPME-FS has been perfected and achieves the highest value under these two classifiers. As seen in Tables IV and VI, the average size of the selected feature subset by FNPME-FS is smaller than those of FNCE, FSNTDJE, and IFPR under the classifier KNN on the whole. Although FNPME-FS cannot select fewer features than OMGE and PMGE, the classification accuracy of FNPME-FS is the highest on five datasets Glass, Heart, Ionosphere, Sonar, and Wine; in particular, the accuracy of FNPME-FS is 1.41%-25.71% larger than FNCE, FSNTDJE, IFPR, OMGE, and PMGE on the Glass dataset; however, 4.4%-7.3% gaps clearly exist between FNPME-FS and the other methods in metrics of the mean classification accuracy under classifier KNN. In the same way, it can be observed from Tables IV and VII, the average classification accuracy of FNPME-FS is larger than ODP, FNCE, FSNTDJE, IFPR, OMGE, and PMGE on almost all datasets under classifier CART, except for two datasets Heart and Wdbc. However, FNPME-FS selects fewer features than OMGE, although its accuracy is 5.18% and 1.76% lower than OMGE on two datasets Heart and Wdbc, respectively. Since no method always performs better than the others under different classifiers and for learning tasks, the FNPME-FS method overall eliminates all the redundant features and exhibits better performance than the other compared models on UCI datasets.

This final part is designed to verify the efficiency of classification of FNPME-FS for five selected UCI datasets from Table I. Five feature selection algorithms compared with FNPME-FS include FNCE [14], FNRS [10], FISEN [10],[42], OMGE [31], and PMGE [31]. In addition, the ODP method is used to compare with the above six methods. The classification results for the five UCI datasets are appraised on two classifiers KNN and CART. Three evaluation metrics are Accuracy (*Acc*), True Positive Rate (*TPR*), and False Positive Rate (*FPR*) [40], which are utilized to measure the effectiveness of classification of all compared approaches. Notably, the higher the *Acc* and *TPR*, the lower the *FPR*, and the better the results [40]. The *Acc*, *TPR* and *FPR* on the selected feature subsets by seven methods under two classifiers KNN and CART are demonstrated in Tables VIII and IX, respectively.

It can be known from Tables VIII and IX that three metrics *Acc*, *TPR* and *FPR* of feature subset selected by FNPME-FS outperforms those of FNCE, FNRS, FISEN, OMGE, and PMGE on most of all datasets under two classifiers KNN and CART. FNPME-FS apparently achieves higher *Acc* than the other five methods, except for two datasets Glass and Wdbc under classifier KNN and two datasets Wdbc and Wpbc under classifier CART. The *TPR* and *FPR* values of FNPME-FS reach the optimal results for most of the five datasets. According to Table VIII under classifier KNN, although FNPME-FS is 1.4% and 5.5% lower than FISEN and FNCE in terms of *Acc* on datasets Glass and Wdbc, respectively, the

mean *Acc* values of FNPME-FS are the highest under classifier KNN. As we know, the lower the *FPR* value, the better the performance. Although FNPME-FS is 4.2% lower than FISEN for dataset Wdbc on metric *TPR*, its *FPR* is the minimum. In the same measure, as seen in Table IX under classifier CART, the mean metrics of *Acc*, *TPR*, and *FPR* of FNPME-FS are the largest on five datasets. The mean of *Acc* is nearly 1%-10% higher than those of the other six methods, and its mean values of *FPR* are mostly 4.1%-31.1% less than the other six models. While comparing FNPME-FS against the other six methods, the values of two metrics *Acc* and *TPR* have been obviously improved. Furthermore, on almost all datasets, FNPME-FS exhibits the minimal value as for *FPR*, except for the Ionosphere and Wpbc datasets under classifier CART. As observed in Tables VIII and IX, FNPME-FS eliminates some important features from some datasets, which result in a decrease in classification accuracy. Overall speaking, our algorithm can eliminate all redundant features and promote the values of two metrics *Acc* and *TPR* for most UCI datasets.

TABLE VIII
THREE METRICS WITH SEVEN METHODS UNDER CLASSIFIER KNN

Methods	Metrics	Glass	Ionosphere	Wdbc	Wine	Wpbc	Mean
ODP	*Acc*	0.911	0.825	0.912	0.920	0.697	0.853
	TPR	0.667	0.895	0.903	0.873	0.697	0.807
	FPR	0.022	0.568	0.057	0.042	0.498	0.237
FNCE	*Acc*	0.863	0.882	**0.983**	0.907	0.657	0.858
	TPR	0.729	0.916	0.904	0.852	0.657	0.811
	FPR	0.046	0.062	0.175	0.131	0.704	0.223
FNRS	*Acc*	0.925	0.860	0.925	0.932	0.606	0.849
	TPR	0.824	0.875	0.920	0.930	0.606	0.831
	FPR	0.036	0.686	0.080	0.031	0.548	0.276
FISEN	*Acc*	**0.953**	0.880	0.953	0.955	0.651	0.878
	TPR	0.862	0.889	**0.943**	0.915	0.652	0.852
	FPR	0.015	0.568	0.057	0.034	0.225	0.179
OMGE	*Acc*	0.682	0.875	0.961	0.955	0.704	0.835
	TPR	0.673	0.946	0.934	0.930	0.468	0.790
	FPR	0.117	0.054	0.066	0.028	0.179	0.088
PMGE	*Acc*	0.925	0.827	0.953	0.938	0.667	0.862
	TPR	0.778	0.900	0.915	0.887	0.667	0.829
	FPR	0.015	0.451	0.085	0.059	0.504	0.222
FNPME-FS	*Acc*	0.939	**0.903**	0.928	**0.972**	**0.722**	**0.893**
	TPR	**0.939**	**0.978**	0.901	**0.972**	**0.815**	**0.921**
	FPR	**0.007**	**0.022**	**0.056**	**0.025**	**0.016**	**0.025**

TABLE IX
THREE METRICS WITH SEVEN METHODS UNDER CLASSIFIER CART

Methods	Metrics	Glass	Ionosphere	Wdbc	Wine	Wpbc	Mean
ODP	*Acc*	0.967	0.865	0.860	0.893	0.712	0.859
	TPR	0.889	0.895	0.892	0.875	0.712	0.852
	FPR	0.007	0.660	0.108	0.103	0.679	0.311
FNCE	*Acc*	0.848	0.805	0.881	0.825	0.614	0.794
	TPR	0.889	0.892	0.917	0.891	0.696	0.857
	FPR	0.013	0.524	0.085	0.093	0.275	0.198
FNRS	*Acc*	0.946	0.891	0.946	0.898	**0.742**	0.884
	TPR	0.889	0.892	0.915	0.887	0.742	0.865
	FPR	0.007	0.869	0.085	0.093	0.711	0.353
FISEN	*Acc*	0.941	0.864	0.914	0.898	0.722	0.867
	TPR	0.923	0.903	0.920	0.887	0.722	0.871
	FPR	0.005	**0.016**	0.080	0.093	0.742	0.187
OMGE	*Acc*	0.687	0.848	**0.949**	0.893	0.702	0.815
	TPR	0.687	0.917	0.925	0.875	0.702	0.821
	FPR	0.144	0.038	0.075	0.103	**0.053**	0.083
PMGE	*Acc*	0.941	0.892	0.940	0.916	0.717	0.881
	TPR	0.889	0.892	0.906	0.873	0.717	0.855
	FPR	0.007	0.661	0.094	0.056	0.744	0.321
FNPME-FS	*Acc*	**0.981**	**0.900**	0.931	**0.933**	0.737	**0.896**
	TPR	**0.981**	**0.965**	**0.950**	**0.983**	**0.904**	**0.957**
	FPR	**0.004**	0.035	**0.050**	**0.028**	0.093	**0.042**

For these compared feature selection methods, in metrics of time complexity, the rough ranking of eight methods is shown as: O(FNPME-FS) = O(FSNTDJE) = O(FISEN) < O(FNCE) < O(OMGE) = O(PMGE) < O(IFPR) < O(FNRS), where the time complexity of Algorithm A is described as O(A). Assume that there are m samples and n features in decision systems, the complexity of FSNTDJE and FISEN is $O(mn)$ [40],[42], and the complexity of FNCE is no more than $O(mn^2)$ [14]. The complexity of OMGE and PMGE is $O(mn^2\log m)$ [31], which is larger than that of FNPME-FS and FNCE. As for the IFPR and FNRS algorithms, their computational complexity may be $O(m^2n)$ and $O(m^2n/2+n^2)$, respectively. Therefore, our presented FNPME-FS algorithm has the relatively lower computational complexity on UCI datasets.

D. Classification Results of Gene Expression Datasets

In metrics of the average size of genes selected from microarray gene expression datasets with high dimensions under ten-fold cross-validation, this first subsection follows with interest on the effectiveness of classification of FNPME-FS and the optimal selected gene subset by FNPME-FS under the KNN, CART and C4.5classifiers. The ODP method and three feature selection methods FNCE [14], IFPR [13], and FSNTDJE [40] are selected to compare with FNPME-FS. The average number of the selected gene subsets by these four algorithms with the original genes of raw data under ten-fold cross-validation is shown in Table X. Then the optimal selected gene subsets by using the FNPME-FS algorithm for all gene datasets on three classifiers KNN, CART and C4.5 are shown in Table XI.

TABLE X
NUMBER OF SELECTED GENES WITH FOUR FEATURE SELECTION METHODS

Datasets	Raw	FNCE	IFPR	FSNTDJE	FNPME-FS
Breast	9216	13.7	9.5	**7**	10.6
Colon	2000	**4.9**	6	11.5	7
DLBCL	5469	16.3	12.4	7.3	**6.4**
Leukemia	7129	8.7	6.4	7	**6**
Prostate	12600	13.7	9.8	**8**	11.5
Mean	7282.8	11.46	8.82	**8.16**	8.3

TABLE XI
OPTIMAL GENE SUBSET WITH FNPME-FS FOR FIVE GENE DATASETS

Datasets	KNN	CART	C4.5
Breast	{5651 8944 2611 5426 7137 8958 6797 3484 6410 1313 3522}	{2585 109 4085 6959 2779 7515 6098 8303 6648}	{2585 109 4085 7321 4161 6503 9111 4918 3484 4278 5643 4442 5485}
Colon	{249 72 187 83 107 698 1770 1917 994 1247 1902 1808}	{249 1047 1414 1334}	{249 1047 1414 1334}
DLBCL	{3313 409 453 1698 2280 4767}	{3313 409 453 1698 2280 4767}	{3313 409 453 1698 2280 4767}
Leukemia	{758 5565 1268 538 6180 1787}	{758 5565 1268 538 6180 1787}	{758 5565 1268 538 6180 1787}
Prostate	{5370 1656 2903 3363 1414 2134 877 2867 2521 1864 4346 4072}	{3124 5136 1479 1600 3023 930 4644 2521 4072}	{3124 5136 1479 1600 3023 930 4644 2521 4072}

As clearly seen from Table X, for two datasets DLBCL and Leukemia, FNPME-FS is obviously superior against the other four algorithms; however, for datasets Breast and Prostate, FSNTDJE achieves the best classification results, and FNCE selects the minimum of gene subset on dataset Colon. Moreover, the mean size of the selected gene subset by

FNPME-FS is less than that of FNCE and IFPR; however, it is slightly inferior to FSNTDJE, which is only 0.14 fewer than FSNTDJE in metrics of the mean number. In short, FNPME-FS can select a smaller gene subset from each gene dataset.

In this following portion, from the results of Tables X and XI, two classifiers KNN and CART are applied to verify the average accuracy of classification for these reduced results of five methods including ODP, FNCE [14], IFPR [13], FSNTDJE [40], and FNPME-FS under ten-fold cross-validation for five gene datasets, and then the outcomes are provided in Tables XII and XIII, respectively.

TABLE XII
CLASSIFICATION ACCURACY WITH FIVE METHODS ON CLASSIFIER KNN

Datasets	ODP	FNCE	IFPR	FSNTDJE	FNPME-FS
Breast	0.6285	0.6918	0.7833	0.7143	**0.8738**
Colon	0.7903	**0.9231**	0.7574	0.7742	0.8548
DLBCL	0.8701	0.8917	0.9480	0.8051	**0.9611**
Leukemia	0.7344	0.8970	0.8042	0.8327	**0.9444**
Prostate	0.7647	0.8516	0.8095	0.7871	**0.8903**
Mean	0.7576	0.8510	0.8204	0.7826	**0.9048**

TABLE XIII
CLASSIFICATION ACCURACY WITH FIVE METHODS ON CLASSIFIER CART

Datasets	ODP	FNCE	IFPR	FSNTDJE	FNPME-FS
Breast	0.7023	0.7509	0.7312	0.6905	**0.7976**
Colon	0.5967	0.7964	0.7595	0.7097	**0.8064**
DLBCL	0.8181	0.7698	0.8700	0.8054	**0.9091**
Leukemia	0.7556	0.8342	0.7556	0.7222	**0.9722**
Prostate	0.6917	0.8046	0.7610	**0.8492**	0.8224
Mean	0.7128	0.7911	0.7754	0.7554	**0.8615**

From Tables XII and XIII, for almost all gene expression datasets, FNPME-FS exhibits the highest accuracy, except for the Colon dataset on the KNN classifier and dataset Prostate under classifier CART. Although FSNTDJE selects fewer genes for datasets Breast and Prostate in Table X, FNPME-FS achieves the highest accuracy for dataset Breast under two different classifiers, which is only 4.68% lower than that of FSNTDJE for dataset Prostate under classifier CART. Compared with FNCE and IFPR, our method eliminates a lot of redundant genes with noise to improve classification performance. As for the mean accuracy under two classifiers KNN and CART, our method is 5.38%-14.72% and 7.04%-14.87% higher than the other four methods, respectively. Therefore, FNPME-FS achieves the best results of data classification for these five gene datasets.

The third part continues to test the efficiency of classification of FNPME-FS, and five feature selection methods including FNCE [14], IFPR [13], FSNTDJE [40], MIBARK [43], and EGGS [44] are selected for the comparison results of the Colon, DLBCL, Leukemia, and Prostate datasets from Table I. According to the last results on different fuzzy neighborhood parameters, all the used models are implemented under ten-fold cross-validation. Tables XIV and XV demonstrate the average values of accuracy of classification for these four gene expression datasets with seven methods including the ODP method under two classifiers KNN and C4.5, respectively.

As seen from the average classification results shown in Tables XIV and XV, our algorithm achieves the largest values for the DLBCL, Leukemia, and Prostate datasets on two classifiers KNN and C4.5. For dataset Colon in Table XIV under classifier KNN, FNCE exhibits the highest accuracy, which is 6.83% higher than that of FNPME-FS. What is more, the mean accuracy of FNPME-FS is 2.18%-21.97% higher than the other six methods. Similarly, from Table XV, FSNTDJE obtains the accuracy of 91.19%, which is 10.54% higher than our algorithm for dataset Colon under classifier C4.5. However, the classification accuracy of FNPME-FS under classifier C4.5 is 5.2%-10.52%, 4.1%-22.5%, and 11.08%-31.58% higher than the other six methods on three datasets DLBCL, Leukemia, and Prostate, respectively. In addition, the mean accuracy of FNPME-FS exhibits the largest results. In general, FNPME-FS can efficiently delete the redundant genes and is superior against the other six compared models on these gene datasets.

TABLE XIV
CLASSIFICATION ACCURACY WITH SEVEN METHODS ON CLASSIFIER KNN

Datasets	ODP	FNCE	IFPR	FSNTDJE	MIBARK	EGGS	FNPME-FS
Colon	0.7903	**0.9231**	0.7574	0.7742	0.7700	0.6493	0.8548
DLBCL	0.8701	0.8917	0.9480	0.8051	0.7654	0.8540	**0.9611**
Leukemia	0.7344	0.8970	0.8042	0.8327	0.8285	0.6292	**0.9444**
Prostate	0.7647	0.8516	0.8095	0.7871	0.5127	0.6394	**0.8903**
Mean	0.7898	0.8908	0.8297	0.7997	0.7191	0.6929	**0.9126**

TABLE XV
CLASSIFICATION ACCURACY WITH SEVEN METHODS ON CLASSIFIER C4.5

Datasets	ODP	FNCE	IFPR	FSNTDJE	MIBARK	EGGS	FNPME-FS
Colon	0.7419	0.7419	0.8342	**0.9119**	0.8221	0.6464	0.8065
DLBCL	0.7922	0.8181	0.8312	0.7890	0.7780	0.8264	**0.8832**
Leukemia	0.8143	0.8754	0.9044	0.9173	0.8341	0.7333	**0.9583**
Prostate	0.6400	0.7426	0.7058	0.7715	0.5665	0.5913	**0.8823**
Mean	0.7471	0.7945	0.8189	0.8474	0.7501	0.6993	**0.8825**

E. Statistical Analysis

In this subsection, to systematically explore the statistical performance on classification accuracy of all compared algorithms, the Friedman test and corresponding post-hoc tests will be conducted. The Friedman statistic is expressed [45] as

$$\chi_F^2 = \frac{12N}{k(k+1)}\left(\sum_{j=1}^{k} R_j^2 - \frac{k(k+1)^2}{4}\right) \text{ and } F_F = \frac{(N-1)\chi_F^2}{N(k-1)-\chi_F^2}, \quad (34)$$

where N and k is the number of datasets and algorithms, respectively; R_j ($j = 1, 2, \ldots, k$) denotes the mean rank of a certain method on all datasets; and F_F denotes an F-distribution under ($k - 1$) and ($k - 1$)($N - 1$) freedom degrees. Then the critical difference is described [46] as

$$CD_\alpha = q_\alpha \sqrt{\frac{k(k+1)}{6N}}, \quad (35)$$

where α describes the significance level of the Bonferroni-Dunn test and q_α denotes a critical value [46].

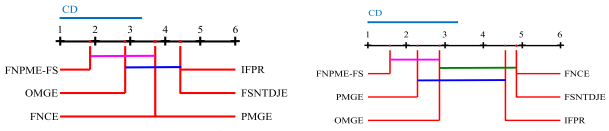
Following the statistical test provided in [5],[7],[40], for all datasets, the mean values of ranking are acquired through averaging all levels of the accuracy of classification. The best-ranking value under the accuracy metric is set to 1; the second is the rank of 2, and so on. For the classification accuracy on seven datasets in Tables VI and VII, the Friedman tests are achieved by the comparison of FNPME-FS with FSNTDJE, FNCE, IFPR, OMGE, and PMGE. When all algorithms are equivalent in metrics of accuracy of classification, one can establish the null hypothesis of the Friedman test [3]. Then the rankings of six algorithms can be easily calculated, and their mean rankings are obtained under two classifiers KNN and CART. It follows that the values of

χ_F^2 and F_F can be computed. Table XVI demonstrates the results of the mean ranking of six algorithms and the values of χ_F^2 and F_F under classifiers KNN and CART.

TABLE XVI
STATISTICAL TEST OF SIX METHODS UNDER CLASSIFIERS KNN AND CART

Classifiers	Mean rankings						χ_F^2	F_F
	FSNTDJE	FNCE	IFPR	OMGE	PMGE	FNPME-FS		
KNN	4.43	3.71	4.43	2.86	3.71	1.86	9.83	2.35
CART	4.86	4.86	4.57	2.86	2.29	1.57	21.02	9.03

By computing, when the significance level equals 0.1, one has the value of $F(5, 30)$ in the F-distribution, which is 2.05, and the value of $\chi_F^2(5)$ on χ_F^2 distribution is 9.24. Then, according to Table XVI, the values of χ_F^2 and F_F under classifiers KNN and CART are larger than those of $\chi_F^2(5)$ and $F(5, 30)$. It follows from Table 5 in [46] that $CD = 2.394$. Thus, all null hypotheses are rejected, and the six algorithms are different under classifiers KNN and CART. To more objectively compare the differences of several algorithms more intuitively, a graph [46] was introduced to connect the methods, which are not clearly different from each other, and then the critical values among all algorithms can be obviously illustrated in these graphs. Fig. 3 shows the comparison of FNPME-FS with the other five algorithms under classifiers KNN and CART, where the top line is the critical value, the coordinate axis shows the mean ranking of each method, and the mean ranking of the left-hand side is the lowest. It is noted that the horizontal lines are used to connect groups of methods with no significant difference in classification performance [46]. As seen in Fig. 3, the significant differences of six algorithms are obvious. In Fig. 3(a), under the classifier KNN, FNPME-FS performs clearly better than IFPR and FSNTDJE. Although there is no significant difference between OMGE, PMGE, and FNCE, it can be concluded that the performance of classification of FNPME-FS is better than the others. Similarly, as shown in Fig. 3(b) on the CART, FNPME-FS outperforms FNCE, FSNTDJE and IFPR, and is similar to OMGE and PMGE. However, in the same group, FNPME-FS excels against the two compared methods. To sum up, our model outperforms the other five algorithms as the whole.



(a) KNN classifier (b) CART classifier

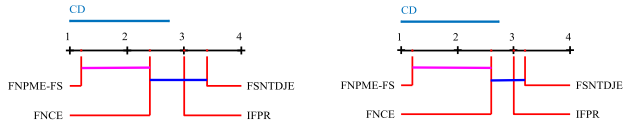
Fig. 3. Accuracy comparison with six methods on classifiers KNN and CART.

The next statistical results and analyses are executed on four different algorithms FNCE, IFPR, FSNTDJE, and FNPME-FS for five gene expression datasets from Tables XII and XIII under two classifiers KNN and CART. The average rankings of four algorithms and the values of χ_F^2 and F_F under classifiers KNN and CART are calculated and illustrated in Table XVII.

As we can see from Table XVII, the values of χ_F^2 and F_F are larger than those of $\chi_F^2(3) = 6.25$ and $F(3, 12) = 2.61$, and then one has $CD = 1.74$. Thus, all null hypotheses are rejected, and the four algorithms are different under two classifiers KNN and CART. In the same way, the comparisons of the four algorithms under two different classifiers are demonstrated in Fig. 4. The differences between FNPME-FS and other algorithms can obviously be exhibited in Fig. 4. FNPME-FS is statistically superior to IFPR and FSNTDJE under classifiers KNN and CART, which show an excellent classification effect on the two criteria. Therefore, our FNPME-FS method is the best when compared with FNCE, IFPR and FSNTDJE.

TABLE XVII
STATISTICAL TEST OF FOUR METHODS UNDER CLASSIFIERS KNN AND CART

Classifiers	Mean rankings				χ_F^2	F_F
	FNCE	IFPR	FSNTDJE	FNPME-FS		
KNN	2.4	3	3.4	1.2	8.28	4.93
CART	2.6	3	3.2	1.2	7.32	3.81



(a) KNN classifier (b) CART classifier

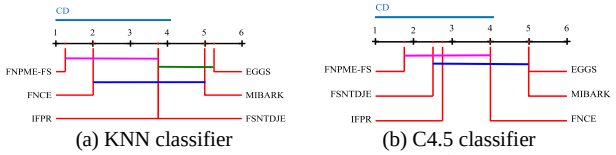
Fig. 4. Accuracy comparison with four methods on classifiers KNN and CART.

In this final portion of this subsection, the statistical results and analyses focus on the accuracy of classification of the six compared algorithms on the C4.5 and KNN classifiers for four gene datasets, namely, Colon, DLBCL, Leukemia, and Prostate in Tables XIV and XV. The Friedman test is wielded to indicate these differences in classification among six algorithms. Then the mean rankings of six algorithms under classifiers KNN and C4.5 can be achieved and illustrated in Table XVIII.

TABLE XVIII
STATISTICAL TEST OF SIX METHODS ON CLASSIFIERS KNN AND C4.5

Classifiers	Mean rankings						χ_F^2	F_F
	FNCE	IFPR	FSNTDJE	MIBARK	EGGS	FNPME-FS		
KNN	2	3.75	3.75	5	5.25	1.25	14.57	8.05
C4.5	4	2.75	2.5	5	5	1.75	10.71	3.46

From Table XVIII, one can calculate $\chi_F^2 = 14.57$ and $F_F = 8.05$ under classifier KNN and $\chi_F^2 = 10.71$ and $F_F = 3.46$ under classifier C4.5. It follows that $\chi_F^2(5) = 9.24$, $F(5, 30) = 2.05$ and $CD = 3.08$. Then χ_F^2 and F_F is larger. All null hypotheses are rejected, and six algorithms are different under classifiers KNN and C4.5. Similarly, Fig. 5, respectively, illustrates the grouping of six algorithms under the KNN and C4.5 classifiers, and the significant differences can be exhibited easily.



(a) KNN classifier (b) C4.5 classifier

Fig. 5. Accuracy comparison with six methods on classifiers KNN and C4.5.

In Fig. 5(a) under classifier KNN, FNPME-FS gets the lowest mean ranking, and its performance is significantly better than MIBARK and EGGS because the different value is larger than 3.08. In terms of the classifier C4.5 in Fig. 5(b), FNPME-FS has the same group as FNCE, IFPR, and FSNTDJE, which means the differences among the four algorithms are not obvious although FNPME-FS performs better than FNCE,

IFPR, and FSNTDJE. In conclusion, FNPME-FS indeed surpasses the other five compared methods under the results of the Friedman statistic test.

F. Discussion

From the above experimental results, this part summarizes the main objectives of our experiments as follows.

(1) As far as the five gene expression datasets with high dimensions in Table I, the Fisher score model is first employed for gene extraction and preliminary dimension reduction and improves classification efficiency in the process of data classification. Then, for setting the fuzzy neighborhood decision system for each dataset in Table I, the classification accuracy and the reduction rate under three classifiers KNN, CART, and C4.5 are attained on seven UCI datasets and five gene datasets, their suitable fuzzy neighborhood radius parameter values are found and shown in Tables II and III.

(2) As for the seven UCI datasets with low dimensions, in metrics of the sizes of the selected feature subsets in Table IV, our FNPME-FS method chooses fewer features than three algorithms FSNTDJE, FNCE, and IFPR on almost all datasets. From Table VI, under classifier KNN, FNPME-FS gains the better accuracy of classification on almost UCI datasets; nevertheless, the classification accuracy of FNPME-FS is 5.46% less than FNCE for the Wdbc dataset and 1.52% less than FSNTDJE for the Wpbc dataset, respectively. On classifier CART, FNPME-FS manifests the best result in Table VII, except for the Heart and Wdbc datasets by using OMGE; however, the mean accuracy of FNPME-FS is 5.27% higher than OMGE. In general, from Tables IV to VII, FNPME-FS exhibits the best values. In metrics of three metrics *Acc*, *TPR* and *FPR* in Tables VIII and IX, FNPME-FS performs better than the other methods on almost the datasets. What is more, FNPME-FS achieves higher values than the other methods in terms of *TPR*, except for the Wdbc datasets under the classifier KNN. Under the classifier CART in Table IX, the *TPR* values of FNPME-FS are superior to the other six methods for almost all datasets. In general, FNPME-FS can exhibit the great classification performance on all UCI datasets under two different classifiers KNN and CART.

(3) In terms of gene expression datasets with high dimensions, as shown in Table X, the size of the selected gene subset by FNPME-FS is a little bit larger than the other four methods for three datasets Breast, Colon, and Prostate. However, the classification accuracy of FNPME-FS under classifiers KNN and CART in Tables XII and XIII are higher than the other four methods, except for the Colon and Prostate datasets under two classifiers KNN and CART. From Tables XIV and XV, on four datasets Colon, DLBCL, Leukemia, and Prostate, although the accuracy of FNPME-FS for dataset Colon is 6.83% lower than FNCE under classifier KNN and 10.54% lower than FSNTDJE under classifier C4.5, respectively, our results perform the best on three datasets DLBCL, Leukemia, and Prostate under two different classifiers. In summary, from Tables X to XV, FNPME-FS indicates the better classification performance for these selected gene expression datasets.

(4) The statistical analysis is performed to further demonstrate the statistical significance of our algorithm. By analyzing all the classification results from Tables VI and VII for all UCI datasets, the statistical differences between FNPME-FS and the other algorithms under classifiers KNN and CART are indicated in Fig. 3. As shown in Fig. 3, FNPME-FS outperforms the other five algorithms. As for these gene expression datasets, the comparison results of FNPME-FS with the other algorithms are illustrated in Fig. 5. It follows that FNPME-FS excels at all the statistic tests against the other algorithms. Therefore, in neighborhood decision systems, our FNPME-FS method can produce worthy classification results and exhibit great statistical performance for these datasets.

VI. Conclusion and Future Work

In this article, a new FNMRS-based feature selection approach is brought forward to ameliorate the performance of classification for neighborhood decision systems. The definitions of the FNRS and NMRS models are first investigated, and some efficient uncertainty measures based on fuzzy neighborhood entropy are expanded in neighborhood decision systems. Then by integrating algebra with information views in the FNMRS model, the fuzzy neighborhood pessimistic multigranulation entropy is put forward to analyze the noises, uncertainty, and fuzziness of neighborhood decision systems. Furthermore, some properties are found. Accordingly, a novel forward search algorithm for feature selection is devised to make the performance of classification better for the selected feature subset. All the designed experiments exhibit that our FNPME-FS model is able to select a feasible and stable feature subset owning the optimal performance of classification for some low-dimensional UCI datasets and several gene datasets with high dimensions. Nevertheless, the developed feature selection method cannot fully balance the size of the selected feature subset and the classification accuracy for all heterogeneous datasets. Therefore, in our future work, more effective search methods and uncertainty measures for the FNMRS will be further studied to achieve an optimal balance and improve the computational efficiency of our method on many large-scale and high-dimensional datasets.

References

[1] X. Zhang, C. L. Mei, D. G. Chen, Y. Y. Yang, and J. H. Li, "Active incremental feature selection using a fuzzy rough set-based information entropy," *IEEE Transactions on Fuzzy Systems*, 2019, doi: 10.1109/TFUZZ.2019.2959995.

[2] W. P. Ding, C. T. Lin, M. Prasad, Z. H. Cao, and J. D. Wang, "A layered-coevolution-based attribute-boosted reduction using adaptive quantum behavior PSO and its consistent segmentation for neonates brain tissue," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 3, pp. 1177–1191, 2018.

[3] C. Z. Wang, Y. Huang, M. W. Shao, Q. H. Hu, and D. G. Chen, "Feature selection based on neighborhood self-information," *IEEE Transactions on Cybernetics*, 2019, doi: 10.1109/TCYB.2019.2923430.

[4] W. P. Ding, C. T. Lin, and Z. H. Cao, "Shared nearest neighbor quantum game-based attribute reduction with hierarchical co-evolutionary Spark and its consistent segmentation application in neonatal cerebral cortical surfaces," *IEEE Transactions on Neural Network and Learning System*, vol. 30, no. 7, pp. 2013-2027, 2019.

[5] 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," *Information Sciences*, vol. 502, pp.18–41, 2019.

[6] N. L. Giang, L. H. Son, T. T. Ngan, T. M. Tuan, H. T. Phuong, M. Abdel-Basset, A. R. L. De Macêdo, and V. H. C. De Albuquerque, "Novel incremental algorithms for attribute reduction from dynamic decision tables

using hybrid filter–wrapper with fuzzy partition distance," *IEEE Transactions on Fuzzy Systems*, 2019, doi: 10.1109/TFUZZ.2019.2948586.

[7] L. Sun, L. Y. Wang, W. P. Ding, Y. H. Qian, and J. C. Xu, "Neighborhood multi-granulation rough sets-based attribute reduction using Lebesgue and entropy measures in incomplete neighborhood decision systems," *Knowledge-Based Systems*, vol. 192, article ID. 105373, 2020.

[8] C. Z. Wang, Y. Wang, M. W. Shao, Y. H. Qian, and D. G. Chen, "Fuzzy rough attribute reduction for categorical data," *IEEE Transactions on Fuzzy Systems*, 2019, doi: 10.1109/TFUZZ.2019.2949765.

[9] Z. Pawlak, "Rough sets," *International Journal of Computer & Information Sciences*, vol. 11, pp. 341–356, 1982.

[10] C. Z. Wang, M. W. Shao, Q. He, Y. H. Qian, and Y. L. Qi, "Feature subset selection based on fuzzy neighborhood rough sets," *Knowledge-Based Systems*, vol. 111, pp. 173–179, 2016.

[11] L. L. Chen, D. G. Chen, and H. Wang, "Fuzzy kernel alignment with application to attribute reduction of heterogeneous data," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 7, pp. 1469–1478, 2019.

[12] L. A. Zadeh, "Probability measures of fuzzy events," *Journal of Mathematical Analysis and Applications*, vol. 23, no. 2, pp. 421–427, 1968.

[13] A. H. Tan, W. Z. Wu, Y. H. Qian, J. Y. Liang, J. K. Chen, and J. J. Li, "Intuitionistic fuzzy rough set-based granular structures and attribute subset selection," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 3, pp. 527–539, 2019.

[14] J. C. Xu, Y. Wang, H. Y. Mu, and F. Z. Huang, "Feature genes selection based on fuzzy neighborhood conditional entropy," *Journal of Intelligent & Fuzzy Systems*, vol. 36, no. 1, pp. 117–126, 2019.

[15] M. Zabihimayvan and D. Doran, "Fuzzy rough set feature selection to enhance phishing attack detection," 2019 *IEEE International Conference on Fuzzy Systems*, New Orleans, LA, USA, pp. 1–6, 2019.

[16] Q. H. Hu, D. R. Yu, J. F. Liu, and C. X. Wu, "Neighborhood rough set based heterogeneous feature subset selection," *Information Sciences*, vol. 178, pp. 3577–3594, 2008.

[17] G. Selvachandran, S. G. Quek, L. T. H. Lan, L. H. Son, N. L. Giang, W. P. Ding, M. Abdel-Basset, and V. H. C. De Albuquerque, "A new design of mamdani complex fuzzy inference system for multi-attribute decision making problems," *IEEE Transactions on Fuzzy Systems*, 2019, doi: 10.1109/TFUZZ.2019.2961350.

[18] D. Dubois and H. Prade, "Rough fuzzy sets and fuzzy rough sets," *International Journal of General Systems*, vol. 17, pp. 191–208, 1990.

[19] R. Jensen and Q. Shen, "Fuzzy-rough attributes reduction with application to web categorization," *Fuzzy sets and systems*, vol. 141, no. 3, pp. 469–485, 2004.

[20] W. P. Ding, C. T. Lin, and Z. H. Cao, "Deep neuro-cognitive co-evolution for fuzzy attribute reduction by quantum leaping PSO with nearest-neighbor memeplexes," *IEEE Transactions on Cybernetics*, vol. 49, no. 7, pp. 2744-2757, 2019.

[21] Y. H. Qian, S. Y. Li, J. Y. Liang, Z. Z. Shi, and F. Wang, "Pessimistic rough set-based decisions: A multi-granulation fusion strategy," *Information Sciences*, vol. 264, no. 20, pp. 196–210, 2014.

[22] Y. Y. Yao and Y. H. She, "Rough set models in multigranulation spaces," *Information Sciences*, vol. 327, pp. 40-56, 2016.

[23] C. X. Hu, L. Zhang, B. J. Wang, Z. Zhang, and F. Z. Li, "Incremental updating knowledge in neighborhood multigranulation rough sets under dynamic granular structures," *Knowledge-Based Systems*, vol. 163, pp. 811–829, 2019.

[24] G. P. Lin, Y. H. Qian, and J. J. Li, "NMGRS: Neighborhood-based multi-granulation rough sets," *International Journal of Approximate Reasoning*, vol. 53, no. 7, pp. 1080–1093, 2012.

[25] P. Liu, A. Ali, and N. Rehman, "Multi-granulation fuzzy rough sets based on fuzzy preference relations and their applications," *IEEE Access*, vol. 7, pp. 147825–147848, 2019.

[26] H. D. Zhang, J. M. Zhan, and Y. P. He, "Multi-granulation hesitant fuzzy rough sets and corresponding applications," *Soft Computing*, vol. 23, no. 24, pp. 13085–13103, 2019.

[27] L. Sun, L. Y. Wang, J. C. Xu, and S. G. Zhang, "A neighborhood rough sets-based attribute reduction method using Lebesgue and entropy measures," *Entropy*, vol. 21, article ID. 138, 2019.

[28] X. D. Fan, W. D. Zhao, C. Z. Wang, and Y. Huang, "Attribute reduction based on max-decision neighborhood rough set model," *Knowledge–Based Systems*, vol. 151, pp. 16–23, 2018.

[29] C. Z. Wang, Y. Huang, M. W. Shao, and X. D. Fan, "Fuzzy rough set-based attribute reduction using distance measures," *Knowledge-Based Systems*, vol. 164, pp. 205–212, 2019.

[30] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[31] K. Zeng, K. She, and X. Z. Niu, "Multi-granulation entropy and its applications," *Entropy*, vol. 15, no. 6, pp. 2288–2302, 2013.

[32] G. Y. Wang, "Rough reduction in algebra view and information view," *International Journal of Intelligent Systems*, vol. 18, pp. 679–688, 2003.

[33] P. Jain, A. K. Tiwari, and T. Som, "A fitting model based intuitionistic fuzzy rough feature selection," *Engineering Applications of Artificial Intelligence*, vol. 89, article ID. 103421, 2020.

[34] C. Z. Wang, H. Yang, M. W. Shao, and D. G. Chen, "Uncertainty measures for general fuzzy relations," *Fuzzy Sets and Systems*, vol. 360, pp. 82–96, 2019.

[35] K. Zhang, J. M. Zhan, and W. Z. Wu, "Novel fuzzy rough set models and corresponding applications to multi-criteria decision-making," *Fuzzy Sets and Systems*, vol. 383, pp. 92–126, 2020.

[36] X. D. Yue, Y. F. Chen, D. Q. Miao, and H. Fujita, "Fuzzy neighborhood covering for three-way classification," *Information Sciences*, vol. 507, pp. 795–808, 2020.

[37] R. W. R. De Souza, J. V. C. De Oliveira, L. A. Passos, W. P. Ding, J. P. Papa, and V. H. C. De Albuquerque, "A novel approach for optimum-path forest classification using fuzzy logic," *IEEE Transactions on Fuzzy Systems*, 2019, doi: 10.1109/TFUZZ.2019.2949771.

[38] X. Y. You, J. J. Li, and H. K. Wang, "Relative reduction of neighborhood-covering pessimistic multigranulation rough set based on evidence theory," *Information*, vol. 10, no. 11, article ID. 334, 2019.

[39] L. Sun, X. Y. Zhang, Y. H. Qian, J. C. Xu, S. G. Zhang, and Y. Tian, "Joint neighborhood entropy-based gene selection method with fisher score for tumor classification," *Applied Intelligence*, vol. 49, no. 4, pp. 1245–1259, 2019.

[40] L. Sun, L. Y. Wang, Y. H. Qian, J. C. Xu, and S. G. Zhang, "Feature selection using Lebesgue and entropy measures for incomplete neighborhood decision systems," *Knowledge-Based Systems*, vol.186, article ID. 104942, 2019.

[41] L. Sun, X. Y. Zhang, J. C. Xu, and S. G. Zhang, "An attribute reduction method using neighborhood entropy measures in neighborhood rough sets," *Entropy*, vol. 21, no. 2, article ID. 155, 2019.

[42] Q. H. Hu, D. R. Yu, and Z. X. Xie, "Information-preserving hybrid data reduction based on fuzzy-rough techniques," *Pattern Recognition Letters*, vol. 27, no. 5, pp. 414–423, 2006.

[43] F. F. Xu, D. Q. Miao, and L. Wei, "Fuzzy-rough attribute reduction via mutual information with an application to cancer classification," *Computers & Mathematics with Applications*, vol. 57, no. 6, pp. 1010–1017, 2009.

[44] Y. M. Chen, Z. J. Zhang, J. Z. Zheng, Y. Ma, and Y. Xue, "Gene selection for tumor classification using neighborhood rough sets and entropy measures," *Journal of Biomedical Informatics*, vol. 67, pp. 59–68, 2017.

[45] M. Friedman, "A comparison of alternative tests of significance for the problem of *m* rankings," *The Annals of Mathematical Statistics*, vol. 11, no. 1, pp. 86–92, 1940.

[46] J. Demsar and D. Schuurmans, "Statistical comparison of classifiers over multiple data sets," *Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.

APPENDIX

TABLE A
SUMMARIZED ALL ACRONYMS AND FIRST DEFINED REFERENCE

Full Name	Abbr.	Ref.
Neighborhood Rough Sets	NRS	[16]
Fuzzy Rough Sets	FRS	[18]
Fuzzy Neighborhood Rough Sets	FNRS	[10]
Multigranulation Rough Sets	MRS	[21]
Neighborhood Multigranulation Rough Sets	NMRS	[24]
Fuzzy Multigranulation Rough Sets	FMRS	[25]
Original Data Processing	ODP	
Neighborhood Tolerance Dependency Joint Entropy-Based Feature Selection	FSNTDJE	[40]
Fuzzy Neighborhood Conditional Entropy-Based Feature Selection	FNCE	[14]
Intuitionistic Fuzzy Positive Region-Based Attribute Subset Selection	IFPR	[13]
Optimistic Multi-Granulation Entropy-Based Feature Selection for Classification	OMGE	[31]
Pessimistic Multi-Granulation Entropy-Based Feature Selection for Classification	PMGE	[31]
Fuzzy Neighborhood Rough Sets-Based Feature Subset Selection	FNRS	[10]
Fuzzy Entropy-Based Information-Preserving Hybrid Data Reduction	FISEN	[42]
Mutual Information-Based Attribute Reduction for Knowledge Reduction	MIBARK	[43]
Entropy Gain-Based Gene Selection	EGGS	[44]
Accuracy	*Acc*	
True Positive Rate	*TPR*	
False Positive Rate	*FPR*	
k-Nearest Neighbor	KNN	
Classification and Regression Tree	CART	
Version of the C4 Induction Tree Methodology Known as ID3	C4.5	

Lin Sun received the M.S. degree in Computer Science and Technology from Henan Normal University in 2007 and the Ph.D. degree in Pattern Recognition and Intelligent Systems from Beijing University of Technology in 2015. He is currently an Associate Professor at the College of Computer and Information Engineering with Henan Normal University, China. He was a Visiting Scholar at University of Regina, Canada, in 2019. He has become a Postdoctor with Henan Normal University, China, in 2019. He has received funding from ten grants from the National Natural Science Foundation of China, the China Postdoctoral Science Foundation, etc. His main research interests include rough sets, granular computing and big data mining. He has received the title of Henan's Distinguished Young Scholars for Science and Technology Innovation Talents, and has served as a reviewer for several prestigious peer-reviewed international journals.

Lanying Wang is currently a postgraduate student in Computer Science and Technology at the College of Computer and Information Engineering with Henan Normal University. She received the B.Sc. degree in Computer Science and Technology from Henan Institute of Science and Technology in 2017. Her main interests include rough sets and data mining.

Weiping Ding (M'16-SM'19) received the Ph.D. degree in Computation Application, Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China, in 2013. He was a Visiting Scholar at University of Lethbridge (UL), Alberta, Canada, in 2011. From 2014 to 2015, he is a Postdoctoral Researcher at the Brain Research Center, National Chiao Tung University (NCTU), Hsinchu, Taiwan. In 2016, he was a Visiting Scholar at National University of Singapore (NUS), Singapore. From 2017 to 2018, he was a Visiting Professor at University of Technology Sydney (UTS), Ultimo, NSW, Australia. He is a member of Senior IEEE, IEEE-CIS, and Senior CCF. Dr. Ding is the Chair of IEEE CIS Task Force on Granular Data Mining for Big Data. He is a member of Technical Committee on Soft Computing of IEEE SMCS, a member of Technical Committee on Granular Computing of IEEE SMCS, a member of Technical Committee on Data Mining and Big Data Analytics Technical Committee of IEEE CIS. His main research directions involve deep learning, data mining, evolutionary computing, granular computing and big data analytics. He has co-authored more than 70 peer-reviewed journal and conference papers in these fields. Dr. Ding currently serves on the Editorial Advisory Board of Knowledge-Based Systems, and Editorial Board of Information Fusion and Applied Soft Computing. He serves/served as an Associate Editor of several prestigious journals, including IEEE Transactions on Fuzzy Systems, Information Sciences, Swarm and Evolutionary Computation, IEEE Access, and Journal of Intelligent & Fuzzy Systems, as well as the leading guest editor in several international journals. He has delivered more than 10 keynote speeches at international conferences and has co-chaired several international conferences and workshops in the area of fuzzy decision-making, data mining, and knowledge engineering.

Yuhua Qian received the M.S. and Ph.D. degrees in computers with applications from Shanxi University, Taiyuan, China, in 2005 and 2011, respectively. He is currently a Professor with the Key Laboratory of Computational Intelligence and Chinese Information Processing, Ministry of Education, Shanxi University. He is best known for multi-granulation rough sets in learning from categorical data and granular computing. He is involved in research on pattern recognition, feature selection, rough set theory, granular computing, and artificial intelligence. He has published over 80 articles on these topics in international journals. He served on the Editorial Board of the International Journal of Knowledge-Based Organizations and Artificial Intelligence Research. He has served as the Program Chair or Special Issue Chair of the Conference on Rough Sets and Knowledge Technology, the Joint Rough Set Symposium, and the Conference on Industrial Instrumentation and Control, and a PC member of many machine learning, data mining, and granular computing conferences.

Jiucheng Xu is currently a Professor at the College of Computer and Information Engineering, Henan Normal University. He received the M.S. degree and the Ph.D. degree in Computer Science and Technology from Xi'an Jiaotong University in 1995 and 2004, respectively. He has received funding from grants from the National Natural Science Foundation of China, the Key Scientific Research Project of Higher Education of Henan Province, and the Key Scientific and Technological Project of Henan Province. He has published over 100 articles. His research interests include granular computing, data mining, intelligent information processing, and pattern recognition. He has received the title of Henan's Distinguished High Profile Professional, and has served as a reviewer in several prestigious peer-reviewed international journals.