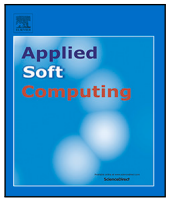
Applied Soft Computing Journal 155 (2024) 111450

Contents lists available at ScienceDirect

Applied Soft Computing

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



A local rough set method for feature selection by variable precision composite measure

Kehua Yuan [a], Weihua Xu [b], Duoqian Miao [a,*]

[a] Department of Computer Science and Technology, Tongji University, Shanghai, 201804, PR China
[b] College of Artificial Intelligence, Southwest University, Chongqing, 400715, PR China

ARTICLE INFO

Keywords:
Granular computing
Feature selection
Neighborhood rough set
Variable precision composite measure

ABSTRACT

Feature selection using variable precision neighborhood rough sets (VPNRS) has garnered considerable attention in data mining and knowledge discovery. Nevertheless, the positive region of VPNRS may not be strictly divided due to the introduction of variable parameters, which could reduce the credibility of feature significance. Meanwhile, the calculation of approximate space is also complex and expensive. Hence, how to improve the computation efficiency is also an investigated issue. As for these issues, we propose a variable precision composite measure and design a novel local method for the feature selection of decision data. Firstly, we introduce the variable precision neighborhood rough set model to process uncertain information from global and local viewpoints. Furthermore, the variable precision composite measure is defined to evaluate the model's accuracy and further used to select the essential features. Finally, a local forward algorithm is provided for feature selection to improve computing efficiency. All experiments on twelve datasets show that the local method is efficient, and the feature selection algorithm based on variable precision composite measure performs well in classification performance. Our work will provide a convenient tool for feature selection methods with uncertainty measures.

1. Introduction

Feature selection is a critical data processing technology to identify the most relevant and informative features in information systems [1–3]. By removing redundant and unimportant features, efficient feature selection can significantly lower the complexity and improve the performance of the machine learning models. Thus, how to evaluate the features is essential in the selection process [4–6]. Currently, feature selection using uncertainty measures allows for a more systematic approach to identifying and selecting the most informative features, which has attracted extensive attention in decision-making [7–9], rule extraction [10,11], and data mining [12–14].

As an emerging computing paradigm of intelligence information process, granular computing emphasizes analyzing and dealing with complex problems from multi-granular, multi-lever, and multiple perspectives [11,15,16]. Hence, it has the natural advantage in uncertain knowledge discovery with complex data. Currently, several granular computing models, including three-way decision [17–20], fuzzy sets [21–23], rough sets [24–26], and concept cognitive learning [27–29] are widely proposed to describe the uncertainty in the information system. Specifically, the rough set (RS) theory proposed by Pawlak [30] provides a convenient and effective mathematical tool for processing imprecision and incomplete information, which has been widely applied to uncertainty reasoning and feature selection [31–33]. Typically, RS approximates the target concept by two definite sets based on the equivalence relation. However, the equivalence relation can only directly process categorical data, and the discretion will lead to losing information. Thus, it is extended to handle numerical data [34], order data [35], interval data [36], and fuzzy data [37,38]. Significantly, the neighborhood rough set model based on similarity relation is more fault-tolerant for dealing with numerical data [39–41]. Yao [19] proposed the relational interpretations of neighborhood operators and rough set approximation operators. Luo et al. [12] designed a novel object dissimilarity measure to construct a psi-neighborhood rough set model in nominal data. Xia et al. [42] presented a granular-ball rough set to represent the Pawlak rough set and neighborhood rough set. The neighborhood thought is also applied to construct a granular rough set model and corresponding feature selection algorithms [43–45]. Wang et al. [46] investigated a neighborhood rough set model to process the uncertain information under semi-supervised data. In addition, the approximate operators need to be more relaxed when approximating the target concept. For example, there are 1000 objects in a similarity class, only one different from the object in the target concept. We will

* Corresponding author.
E-mail addresses: yuan945527632@163.com (K. Yuan), chxuwh@gmail.com (W. Xu), dqmiao@tongji.edu.cn (D. Miao).

https://doi.org/10.1016/j.asoc.2024.111450
Received 4 April 2023; Received in revised form 18 December 2023; Accepted 22 February 2024
Available online 4 March 2024

lose much helpful information if we do not own this object to the lower approximation of the target concept.

Note that the proposition of variable precision rough set(VPRS) overcomes this weakness [47–49]. Mi et al. [48] designed a knowledge reduction approach based on a variable precis rough set model. Liu et al. [50] constructed a VPRS model based on neighborhood geometric similarity and developed a novel decision tree algorithm. A variable precision diversified attribute multi-granulation fuzzy rough set is proposed to investigate the diversified attribute group decision-making problem in papers [51,52]. Moreover, this variable precision thought is also applied to fuzzy rough sets to approximate uncertainty concepts in decision information systems [53,54]. Thus, this also motivates us to use VPRS to study uncertainty measure problems for feature selection. Meanwhile, note that the time complexity of approximation space is $O(n^2)$ due to the computation of similarity class, which is time-consuming in large-scale data [34,55]. To improve the computational efficiency, Xia et al. [56] defined granular ball neighborhood rough sets and proposed a fast adaptive attribute reduction in classification. Qian et al. [55] investigated the local rough set model for knowledge discovery and attribute reduction. In the local rough set, there is only a need to obtain the object classes related to the target concept that save the computational time of objects out of the target concept, which significantly improves the efficiency and can applied to the approximation process in different scenarios. Inspired by this, we investigate a local strategy to approximate vague concepts in decision information systems.

The evaluation function is an essential issue that directly influences the final classification results in feature selection. In rough set theory, it usually designs evaluation functions based on a lower approximation to depict the importance of features [2,6]. Dai [57] proposed the concept of reduced maximal discernibility pairs and developed two attribute selection algorithms. Jensen [58] has proposed attribute reduction algorithms based on the measure of dependency defining upon the lower approximation of the target concept. Wang [59] constructed a fuzzy rough set model based on distance measures with a variable parameter and designed a greedy convergent algorithm for attribute reduction in real datasets. Yao [19] proposed a novel approach to credit scoring features based on the positive measure in a variable rough set. Based on the positive region of a variable precision neighborhood, Chen [60] also gave the significance of an attribute and used it to select feature subsets. Note that the positive region in VPNRS may not be strictly divided due to the introduction of variable parameters, which could reduce the credibility of the evaluation function. Thus, the above measures only focus on the lower approximation is insufficient to describe the data fully.

Inspired by the above analysis, we propose a variable precision composite measure based on the approximation space to describe uncertainty knowledge and design a local heuristic algorithm to determine the optimal feature subset in real-valued decision data. The main contributions of this are as follows:

- We propose a variable precision composite measure based on VPNRS to process uncertain knowledge for real-valued data. Compared with other uncertainty measures based on lower approximation, it provides a comprehensive thought for depicting uncertainty knowledge by combining certain and possible information.
- We define a novel target reduction function based on the proposed composite measure. It could keep the consistency of the approximation structure with the original information system and avoid information loss.
- We design a forward feature selection algorithm from global and local viewpoints. All the experimental results show that the local strategy is efficient, and the method based on the variable precision composite measure is superior to other compared methods in classification performance.

The rest of this paper is organized as follows. Section 2 overviews some basic notions about neighborhood rough set. Section 3 introduces our proposed model, the local variable precision neighborhood rough set model, and defines a composite measure to evaluate its accuracy. Meanwhile, three significance measures and a local heuristic forward algorithm are proposed in Section 4. Moreover, in Section 5, some numerical experiments are carried out to evaluate the efficiency and superiority of our approach on twelve public datasets. Finally, we conclude with a summary of our findings and suggestions for future research in Section 6. The proposed approach is depicted in Fig. 1.

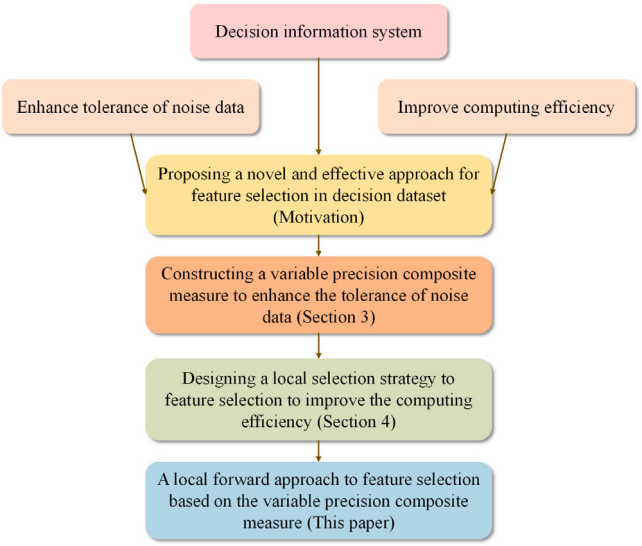


Fig. 1. Block diagram of steps of the proposed approach.

2. Related work

This section briefly reviews some basic notions about neighborhood rough set (NRS) and its extensions, including variable precision neighborhood rough set (VPNRS) and local variable precision neighborhood rough set. The details can be obtained from their corresponding papers[33,46,61].

2.1. Neighborhood rough set

Let $DIS = (U, B \cup D)$ be a decision information system, where $U = \{x_1, x_2, \ldots, x_n\}$ is the universe, $B = \{b_1, b_2, \ldots, b_m\}$ is the conditional attribute set and $D = \{d\}$ is the decision attribute set. The neighborhood class based on distance function is proposed to improve the tolerance of equivalent relations in processing numerical data. For $A \subseteq B$, $x \in U$ and parameter δ, the neighborhood class of x on A is defined as follows.

$$\delta_A(x) = \{y | \Delta(x, y) \leq \delta, y \in U\}, \tag{1}$$

where $\Delta(x, y)$ is the Euclidean distance between x and y.

Based on the above definition of neighborhood class, for $A \subseteq B$ and $X \subseteq U$, the lower and upper neighborhood approximations of X on B can be represented as follows [61].

$$\begin{aligned} \underline{R_A}(X) &= \{x | \delta_A(x) \subseteq X, x \in U\}, \\ \overline{R_A}(X) &= \{x | \delta_A(x) \cap X \neq \emptyset, x \in U\}. \end{aligned} \tag{2}$$

The pair $\langle \underline{R_A}(X), \overline{R_A}(X) \rangle$ is called a neighborhood rough set. The positive region of X on A is $POS_A(X) = \underline{R_A}(X)$, negative region of X on A is $NEG_A(X) = U - \overline{R_A}(X)$, and boundary region of X on A is $BON_A(X) = \overline{R_A}(X) - \underline{R_A}(X)$.

Table 1
A decision information system.

U	b_1	b_2	b_3	b_4	d
x_1	0.18	0.74	0.21	0.32	1
x_2	0.52	0.49	0.35	0.49	1
x_3	0.21	0.76	0.29	0.31	2
x_4	0.23	0.75	0.32	0.40	1
x_5	0.50	0.76	0.26	0.59	2
x_6	0.49	0.73	0.27	0.56	1
x_7	0.62	0.80	0.75	0.82	2
x_8	0.54	0.63	0.49	0.32	2
x_9	0.45	0.55	0.49	0.35	2
x_{10}	0.35	0.85	0.42	0.35	2

2.2. Variable precision neighborhood rough set

According to the analysis in Section 2.1, a sample is considered to be in the lower approximation of the target concept X only if its neighborhood class satisfies an inclusion relation with X. However, this operator may be too strict for dividing the approximation space, especially when dealing with large-scale data. Therefore, it is necessary to introduce a more flexible method for approximating the target concept. The proposal of a variable precision rough set based on the inclusion degree solves this limitation effectively.

Definition 1 ([46]). Let $DIS = (U, B \cup D)$ be a decision information system. For $A \subseteq B$, $X \subseteq U$, and the parameters $0 \le \beta < \alpha \le 1$, the α-lower and β-upper approximations of X on A can be defined as follows.

$$\begin{aligned} \underline{R_A}_{\alpha}(X) &= \{x | \mathscr{D}(X/\delta_A(x)) \ge \alpha, x \in U\}, \\ \overline{R_A}_{\beta}(X) &= \{x | \mathscr{D}(X/\delta_A(x)) > \beta, x \in U\}, \end{aligned} \tag{3}$$

where $\mathscr{D}(X/\delta_A(x)) = \frac{|X \cap \delta_A(x)|}{|\delta_A(x)|}$.

The pair $\langle \underline{R_A}_{\alpha}(X), \overline{R_A}_{\beta}(X) \rangle$ is called a variable precision neighborhood rough set. In addition, the α-positive region of X on A is $POS_{A,\alpha}(X) = \underline{R_A}_{\alpha}(X)$, β-negative region of X on A is $NEG_{A,\beta}(X) = U - \overline{R_A}_{\beta}(X)$, and the boundary region of X on A is $BND_{A,\alpha,\beta}(X) = \overline{R_A}_{\beta}(X) - \underline{R_A}_{\alpha}(X)$.

2.3. Local variable precision neighborhood rough set

From Eq. (3), to obtain the variable neighborhood rough set, we need to compute all dependency degree functions of objects in the universe, thus the $\langle \underline{R_A}_{\alpha}(X), \overline{R_A}_{\beta}(X) \rangle$ is also called global neighborhood rough set. This computing process is time-consuming, especially in big-scale data. A local approximate strategy is introduced to VPNRS to improve computing efficiency further to solve this issue, which can be depicted in the following definition.

Definition 2 ([46]). Let $DIS = (U, B \cup D)$ be a decision information system. For $A \subseteq B$, $X \subseteq U$, and $0 \le \beta < \alpha \le 1$. The local α-lower approximation $\underline{R_A}_{L,\alpha}(X)$ and $\beta-$upper approximation $\overline{R_A}_{L,\beta}(X)$ of X on A are represented as follows.

$$\begin{aligned} \underline{R_A}_{L,\alpha}(X) &= \{x | \mathscr{D}(X/\delta_A(x)) \ge \alpha, x \in X\}, \\ \overline{R_A}_{L,\beta}(X) &= \{x | \mathscr{D}(X/\delta_A(x)) > \beta, x \in X\}. \end{aligned} \tag{4}$$

The $\langle \underline{R_A}_{L,\alpha}(X), \overline{R_A}_{L,\beta}(X) \rangle$ is called a local variable precision neighborhood rough set. Similarly, the local α-positive region of X on A is $POS_{L,A,\alpha}(X) = \underline{R_A}_{L,\alpha}(X)$, the β-negative region of X on A is $NEG_{L,A,\beta}(X) = U - \overline{R_A}_{L,\beta}(X)$ and the boundary region of X on A is $BND_{L,A,\alpha,\beta}(X) = \overline{R_A}_{L,\beta}(X) - \underline{R_A}_{L,\alpha}(X)$.

Example 1. This example shows the approximation process from global and local viewpoints. A decision information system with ten patients is shown in Table 1, where $U = \{x_1, x_2, \ldots, x_{10}\}$ and $B = \{b_1, b_2, b_3, b_4\}$. The four attributes represent cough, fever, weakness, and breathing, while the decision attribute d represents the diagnosis of patients. A diagnosis of "healthy" and "infected" is represented by 1 and 2. Given the target concept D_1, we can obtain the variable precision rough set from global and local perspectives according to the above definitions. We first compute the distance matrix D as follows.

$$D = \begin{bmatrix} 0 & 0.48 & 0.09 & 0.15 & 0.42 & 0.40 & 0.86 & 0.47 & 0.43 & 0.29 \\ 0.48 & 0 & 0.45 & 0.40 & 0.30 & 0.26 & 0.61 & 0.26 & 0.22 & 0.43 \\ 0.09 & 0.45 & 0 & 0.10 & 0.40 & 0.38 & 0.80 & 0.41 & 0.38 & 0.21 \\ 0.15 & 0.40 & 0.10 & 0 & 0.34 & 0.31 & 0.72 & 0.38 & 0.35 & 0.19 \\ 0.42 & 0.30 & 0.40 & 0.34 & 0 & 0.04 & 0.56 & 0.38 & 0.40 & 0.34 \\ 0.40 & 0.26 & 0.38 & 0.31 & 0.04 & 0 & 0.57 & 0.34 & 0.36 & 0.32 \\ 0.86 & 0.61 & 0.80 & 0.72 & 0.56 & 0.57 & 0 & 0.59 & 0.62 & 0.64 \\ 0.47 & 0.26 & 0.41 & 0.38 & 0.38 & 0.34 & 0.59 & 0 & 0.12 & 0.30 \\ 0.43 & 0.22 & 0.38 & 0.35 & 0.40 & 0.36 & 0.62 & 0.12 & 0 & 0.32 \\ 0.29 & 0.43 & 0.21 & 0.19 & 0.34 & 0.32 & 0.64 & 0.30 & 0.32 & 0 \end{bmatrix}.$$

Let $\delta = 0.1$, the neighborhood classes determined by the objects coming U on B are obtained as follows.

$\delta_B(x_1) = \{x_1, x_3\}$, $\delta_B(x_2) = \{x_2\}$, $\delta_B(x_3) = \{x_1, x_3, x_4\}$, $\delta_B(x_4) = \{x_3, x_4\}$, $\delta_B(x_5) = \{x_5, x_6\}$,
$\delta_B(x_6) = \{x_5, x_6\}$, $\delta_B(x_7) = \{x_7\}$, $\delta_B(x_8) = \{x_8\}$, $\delta_B(x_9) = \{x_9\}$.
$\delta_B(x_{10}) = \{x_{10}\}$.

Then, we compute the degree of inclusion of all object similarity classes with the target concept D_1 as follows.

$$\mathscr{D}(D_1/\delta_B(x_1)) = \frac{|D_1 \cap \delta_B(x_1)|}{|\delta_B(x_1)|} = \frac{|\{x_1\}|}{|\{x_1, x_3\}|} = 1/2,$$

$$\mathscr{D}(D_1/\delta_B(x_2)) = \frac{|D_1 \cap \delta_B(x_2)|}{|\delta_B(x_2)|} = \frac{|\{x_2\}|}{|\{x_2\}|} = 1,$$

$$\mathscr{D}(D_1/\delta_B(x_3)) = \frac{|D_1 \cap \delta_B(x_3)|}{|\delta_A(x_3)|} = \frac{\{x_4\}}{|\{x_1, x_3, x_4\}|} = 2/3,$$

$$\mathscr{D}(D_1/\delta_B(x_4)) = \frac{|D_1 \cap \delta_B(x_4)|}{|\delta_A(x_4)|} = \frac{|\{x_4\}|}{|\{x_3, x_4\}|} = 1/2,$$

$$\mathscr{D}(D_1/\delta_B(x_5)) = \frac{|D_1 \cap \delta_B(x_5)|}{|\delta_B(x_5)|} = \frac{|\{x_6\}|}{|\{x_5, x_6\}|} = 1/2,$$

$$\mathscr{D}(D_1/\delta_B(x_6)) = \frac{|D_1 \cap \delta_B(x_6)|}{|\delta_B(x_6)|} = \frac{|\{x_6\}|}{|\{x_5, x_6\}|} = 1/2,$$

$$\mathscr{D}(D_1/\delta_B(x_7)) = \frac{|D_1 \cap \delta_B(x_7)|}{|\delta_B(x_7)|} = \frac{|\emptyset|}{|\{x_7\}|} = 0,$$

$$\mathscr{D}(D_1/\delta_B(x_8)) = \frac{|D_1 \cap \delta_B(x_8)|}{|\delta_B(x_8)|} = \frac{|\emptyset|}{|\{x_8\}|} = 0,$$

$$\mathscr{D}(D_1/\delta_B(x_9)) = \frac{|D_1 \cap \delta_B(x_9)|}{|\delta_B(x_9)|} = \frac{|\{x_9\}|}{|\{x_9\}|} = 1,$$

$$\mathscr{D}(D_1/\delta_B(x_{10})) = \frac{|D_1 \cap \delta_B(x_{10})|}{|\delta_B(x_{10})|} = \frac{|\emptyset|}{|\{x_{10}\}|} = 0.$$

When $\alpha = 0.8$ and $\beta = 0.4$, the α-lower and β-upper approximations of D_1 on B are shown as follows.

$\underline{R_B}_{\alpha}(D_1) = \{x_2\}$, $\overline{R_B}_{\beta}(D_1) = \{x_1, x_2, x_3, x_4, x_5, x_6\}$.

Similarly, the lower and upper approximations of target decision D_2 on attribute set B can be obtained as follows.

$\underline{R_B}_{\alpha}(D_2) = \{x_7, x_8, x_9, x_{10}\}$, $\overline{R_B}_{\beta}(D_2) = \{x_1, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\}$.

Unlike the global α-lower and β-upper approximations, the local approximation space only needs to compare the relationship between the target concept and the neighborhoods of objects in $D_i, i = 1, 2$. Subsequently, we display the local approximation process as follows.

Let $\delta = 0.1$, the neighborhood classes determined by the objects coming D_1 on B are obtained as follows.

$\delta_B(x_1) = \{x_1, x_3\}$, $\delta_B(x_2) = \{x_2\}$, $\delta_B(x_4) = \{x_3, x_4\}$, $\delta_B(x_6) = \{x_5, x_6\}$.

The inclusion degree of x_1, x_2, x_4 and x_6 in D_1 are computed as follows.

$\mathscr{D}(D_1/\delta_B(x_1)) = \frac{|D_1 \cap \delta_B(x_1)|}{|\delta_B(x_1)|} = \frac{|\{x_1\}|}{|\{x_1, x_3\}|} = 1/2$, $\mathscr{D}(D_1/\delta_B(x_2)) = \frac{|D_1 \cap \delta_B(x_2)|}{|\delta_B(x_2)|} = \frac{|\{x_2\}|}{|\{x_2\}|} = 1$,
$\mathscr{D}(D_1/\delta_B(x_4)) = \frac{|D_1 \cap \delta_B(x_4)|}{|\delta_A(x_4)|} = \frac{|\{x_4\}|}{|\{x_3, x_4\}|} = 1/2$, $\mathscr{D}(D_1/\delta_B(x_6)) = \frac{|D_1 \cap \delta_B(x_6)|}{|\delta_B(x_6)|} = \frac{|\{x_6\}|}{|\{x_5, x_6\}|} = 1/2$.

Under the same parameters($\alpha = 0.8, \beta = 0.4$), we obtain the local α-lower approximation and β-upper approximation of D_1 on B as follows.

$\underline{R_B}_{L,\alpha}(D_1) = \{x_2\}, \quad \overline{R_B}_{L,\beta}(D_1) = \{x_1, x_2, x_4, x_6\}.$

Similarly, the local α-lower approximation and β-upper approximation of D_2 on B are obtained as follows.

$\underline{R_B}_{L,\alpha}(D_2) = \{x_7, x_8, x_9, x_{10}\}, \quad \overline{R_B}_{L,\beta}(D_2) = \{x_5, x_7, x_8, x_9, x_{10}\}.$

According to the computing process of approximation space in the above example, we could find that the local strategy can significantly reduce time complexity compared to the global VPNRS.

3. Variable precision neighborhood rough set in DIS

Considering the effectiveness of variable precision neighborhood rough set in handling numerical data, in this section, we mainly introduce VPNRS to decision information systems and propose a novel composite measure for uncertainty measure.

3.1. VPNRS from global and local viewpoints

Variable precision neighborhood rough set is an effective tool for uncertainty reasoning in real-valued data, which is usually introduced into decision information system to depict uncertainty. This subsection mainly introduces the construction process of VPNRS in DIS and proposes two representative accuracy measures of models from global and local viewpoints.

Definition 3 ([46]). Let $DIS = (U, B \cup D)$ be a decision information system, where $U/D = \{D_1, D_2, \dots, D_s\}$. For $A \subseteq B$ and $0 \le \beta < \alpha \le 1$, the global $\alpha-$lower approximation and $\beta-$upper approximation of D on A are defined as follows.

$$\begin{aligned} &\underline{\eta}_{A,\alpha}(D) = \{\underline{R_A}_{\alpha}(D_1), \underline{R_A}_{\alpha}(D_2), \dots, \underline{R_A}_{\alpha}(D_s)\}, \\ &\overline{\eta}_{A,\beta}(D) = \{\overline{R_A}_{\beta}(D_1), \overline{R_A}_{\beta}(D_2), \dots, \overline{R_A}_{\beta}(D_s)\}. \end{aligned} \tag{5}$$

The α-positive region of D on A is $POS_{A,\alpha}(D) = \bigcup_{i=1}^{s} \underline{R_A}_{\alpha}(D_i)$, the β-negative region of D on A is denoted by $NEG_{A,\beta}(D) = \bigcup_{i=1}^{s}(U - \overline{R_A}_{\beta}(D_i))$ and the boundary region of D on A is $BND_{A,\alpha,\beta}(D) = \bigcup_{i=1}^{s}(\overline{R_A}_{\beta}(D_i) - \underline{R_A}_{\alpha}(D_i))$.

According to Definition 3, two measures based on global α-lower approximation and β-upper approximation are defined to evaluate the accuracy of models. Let $DIS = (U, B \cup D)$ be a decision information system, for $A \subseteq B$, the accuracy measure based on $\underline{\eta}_{A,\alpha}(D)$ can be defined as follows.

$$\gamma_{1,A}(D) = \frac{\sum_{i=1}^{s} |\underline{R_A}_{\alpha}(D_i)|}{\sum_{i=1}^{s} |D_i|}. \tag{6}$$

The $\gamma_{1,A}(D)$ depicts the accuracy of VPNRS from certain information in lower approximation, which has a positive relationship with the model's accuracy. Next, we define another accuracy measure based on the upper approximation, which can be represented as follows.

$$\gamma_{2,A}(D) = \frac{\sum_{i=1}^{s} |\overline{R_A}_{\beta}(D_i)|}{\sum_{i=1}^{s} |D_i|}. \tag{7}$$

The $\gamma_{2,R_A}(D)$ characterizes the model's accuracy from possibility information in upper approximation. Since $NEG_{A,\beta}(D) = \bigcup_{i=1}^{s}(U - \overline{R_A}_{\beta}(D_i))$, the $\gamma_{2,A}$ has a negative trend with the model's accuracy.

According to the analysis in Section 2.3, the local rough set could significantly reduce the time complexity of the approximation process due to ignoring information out of the target concept. Thus, to improve the efficiency in the concept approximation, we further introduce the local VPNRS model to decision information systems as follows.

Definition 4 ([46]). Let $DIS = (U, B \cup D)$ be a decision information system, where $U/D = \{D_1, D_2, \dots, D_s\}$. For $A \subseteq B$ and $0 \le \beta < \alpha \le 1$, the local $\alpha-$lower approximation and $\beta-$upper approximation of D on A are defined as follows.

$$\begin{aligned} &\underline{\eta}_{L,A,\alpha}(D) = \{\underline{R_A}_{L,\alpha}(D_1), \underline{R_A}_{L,\alpha}(D_2), \dots, \underline{R_A}_{L,\alpha}(D_s)\}, \\ &\overline{\eta}_{L,A,\beta}(D) = \{\overline{R_A}_{L,\beta}(D_1), \overline{R_A}_{L,\beta}(D_2), \dots, \overline{R_A}_{L,\beta}(D_s)\}. \end{aligned} \tag{8}$$

The local variable positive, negative and boundary region of D on A are respectively defined by $POS_{L,A,\alpha}(D) = \bigcup_{i=1}^{s} \underline{R_A}_{L,\alpha}(D_i)$, $NEG_{L,A,\beta}(D) = \bigcup_{i=1}^{s}(U - \overline{R_A}_{L,\beta}(D_i))$ and $BND_{L,A,\alpha,\beta}(D) = \bigcup_{i=1}^{s}(\overline{R_A}_{L,\beta}(D_i) - \underline{R_A}_{L,\alpha}(D_i))$. The local accuracy measures are also obtained as follows. Let $DIS = (U, B \cup D)$ be a decision information system, for $A \subseteq B$, the local accuracy measure based on $\underline{\eta}_{L,A,\alpha}$ can be defined as follows.

$$\gamma_{L,1,A}(D) = \frac{\sum_{i=1}^{s} |\underline{R_A}_{L,\alpha}(D_i)|}{\sum_{i=1}^{s} |D_i|}. \tag{9}$$

Similarly, the accuracy measure $\gamma_{L,2,A}(D)$ of D based on local upper approximation can be obtained as follows.

$$\gamma_{L,2,A}(D) = \frac{\sum_{i=1}^{s} |\overline{R_A}_{L,\beta}(D_i)|}{\sum_{i=1}^{s} |D_i|}. \tag{10}$$

According to the above definitions, we know the two kinds of accuracy measures $\gamma_{L,1,A}$ and $\gamma_{L,2,A}$ are only considered certain information in lower approximation and possibility information in upper approximation. Meanwhile, it should be noted that the positive and negative regions may not be strictly divided due to the introduction of variable parameters. Thus, the accuracy measures based on the single information are unilateral in uncertainty measure. Subsequently, we will propose a composite measure for characterizing the model's accuracy by combining certain and possible information.

3.2. Composite measure based on VPNRS

Variable precision neighborhood rough set improves the tolerance of noise data while making the approximation space not strictly accurate. Therefore, it could not comprehensively describe the model's accuracy, relying on single information in lower or upper approximation. This subsection investigates a variable precision composite measure by combining lower and upper approximations and then proposes a novel target function for feature selection.

Definition 5. Let $DIS = (U, B \cup D)$ be a decision information system. For $A \subseteq B$ and $0 \le \beta < \alpha \le 1$, the variable precision composite measure is defined as follows.

$$\gamma_{3,A}(D) = \sum_{i=1}^{s} C_{A,\alpha,\beta}(D_i), \tag{11}$$

where $C_{A,\alpha,\beta}(D_i) = \frac{|\underline{R_A}_{\alpha}(D_i)|}{|\overline{R_A}_{\beta}(D_i)|}$ is applied to depict the approximate ability of VPNRS. Due to $\beta \in [0, 1]$, the β-upper approximation may be an empty set, and then the ratio value is meaningless, thus $C_{A,\alpha,\beta}(D_i) = 0$ when $\overline{R_A}_{\beta}(D_i) = \emptyset$. The larger the variable precision composite measure, the higher the model's accuracy.

Compared with $\gamma_{1,A}$ and $\gamma_{2,A}$, the defined composite measure provides a comprehensive method to evaluate the model's accuracy. The local variable precision composite measure can also be obtained with the local VPNRS in the decision information system.

Definition 6. Let $DIS = (U, B \cup D)$ be a decision information system. For $A \subseteq B$ and $0 \le \beta < \alpha \le 1$, the local variable precision composite measure is defined as follows.

$$\gamma_{L,3,A}(D) = \sum_{i=1}^{s} C_{L,A,\alpha,\beta}(D_i), \tag{12}$$

where $C_{L,A,\alpha,\beta}(D_i) = \frac{|\underline{R_A}_{L,\alpha}(D_i)|}{|\overline{R_A}_{L,\beta}(D_i)|}$. These composite measures combining with whole approximation space have the following properties. In this part, we only take the $\gamma_{L,3,A}(D)$ as an example to prove its properties.

Property 1. *Let $DIS = (U, B \cup D)$ be a decision information system. For $A, A' \subseteq B$ and $0 \leq \beta < \alpha \leq 1$, we have*

(1) Non-Negative: $\gamma_{L,3,A}(D) \geq 0$;
(2) For $A \subseteq A'$, the $\gamma_{L,3,A}(D)$ and $\gamma_{L,3,A'}(D)$ is nonmonotonic.

Proof.

(1) According to the definition of VPNRS, we know $|\underline{R_A}_{L,\alpha}(D_j)| \geq 0$ and $|\overline{R_A}_{L,\beta}(D_j)| \geq 0$ for $j = 1, 2 \ldots, s$, thus $C_{L,A,\alpha,\beta}(D_1) = \frac{|\underline{R_A}_{L,\alpha}(D_j)|}{|\overline{R_A}_{L,\beta}(D_j)|} \geq 0$ and the sum of them should be non-negative.
(2) For $\forall x \in U$, the $\delta_A(x) \supseteq \delta_{A'}(x)$ if $A \subseteq A'$, and the $\delta_A(x) \cap D_j \supseteq \delta_{A'}(x) \cap D_j$, thus $\underline{R_A}_{\alpha}(D_j) \supseteq \underline{R_{A'}}_{\alpha}(D_j)$ and $\overline{R_A}_{L,\beta}(D_j) \supseteq R_{A'\,L,\beta}(D_j)$ for $j = 1, \ldots, s$. However, the $\frac{|\underline{R_{A'}}_{L,\alpha}(D_j)|}{|\overline{R_{A'}}_{L,\beta}(D_j)|}$ and $\frac{|\underline{R_A}_{L,\alpha}(D_j)|}{|\overline{R_A}_{L,\beta}(D_j)|} (j = 1, \ldots, s)$ are incomparable, thus $\gamma_{L,3,A}$ is not monotonous with the attributes increase.

Note that the global variable precision composite measure also has the above properties. According to the lower and upper variable precision approximations, we can directly obtain the two target selection functions used for selecting necessary properties while keeping them constant. Nevertheless, they are limited as they all consider the importance of attributes from one point of view. Hence, based on the local composite measure, we introduce a novel local target function $\kappa_{L,A,\alpha,\beta}$.

$$\kappa_{L,A,\alpha,\beta}(D) = \left\{C_{L,A,\alpha,\beta}(D_1), C_{L,A,\alpha,\beta}(D_2), \ldots, C_{L,A,\alpha,\beta}(D_s)\right\}. \tag{13}$$

According to Property 1, the local variable precision composite measure is not monotonous with increased attributes. That is to say that the feature subset obtained by the forward algorithm may contain redundant features. Accordingly, it is necessary to remove any unnecessary attributes during the selection process. Below are the definitions of necessary and unnecessary attributes about $\kappa_{L,A,\alpha,\beta}$.

Property 2. *Let $DIS = (U, B \cup D)$ be a decision information system. For $A \subseteq B$, α and β, we have*

- *The $a \in A$ is the necessary feature of subset A about local novel target function when $\kappa_{L,A,\alpha,\beta}(D) \neq \kappa_{L,A-a,\alpha,\beta}(D)$;*
- *The $a \in A$ is the unnecessary feature of subset A about local novel target function when $\kappa_{L,A,\alpha,\beta}(D) = \kappa_{L,A-a,\alpha,\beta}(D)$.*

The necessary and unnecessary attributes of $\kappa_{L,A-a,\alpha,\beta}(D)$ have been defined above. Furthermore, we can also define the necessary and unnecessary attributes of $\underline{\eta}_{L,A,\alpha}(D)$ and $\overline{\eta}_{L,A,\beta}(D)$ from both a global and local perspective.

Definition 7. Let $DIS = (U, B \cup D)$ be a decision information system. For $A \subseteq B$ and $0 \leq \beta < \alpha \leq 1$, the A is called a local feature reduct of B if the following properties hold.

(1) $\kappa_{L,A,\alpha,\beta}(D) = \kappa_{L,B,\alpha,\beta}(D)$;
(2) There are no redundant features in A, i.e. $\kappa_{L,A',\alpha,\beta}(D) \neq \kappa_{L,A,\alpha,\beta}(D)$ for any $A' \subset A$.

Example 2. According to the DIS shown in Table 1, the feature selection criteria based on lower approximation, upper approximation, and variable precision composite measure from a global viewpoint are shown as follows.

$\underline{\eta}_{B,\alpha}(D) = \{\{x_2\}, \{x_7, x_8, x_9, x_{10}\}\}$, $\overline{\eta}_{B,\beta}(D) = \{\{x_1, x_2, x_3, x_4, x_5, x_6\}, \{x_1, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\}\}$, $\kappa_{B,\alpha,\beta}(D) = \{1/6, 1/2\}$.

The feature selection criteria based on lower approximation, upper approximation and variable precision composite measure from local viewpoint are as follows.

$\underline{\eta}_{L,B,\alpha}(D) = \{\{x_2\}, \{x_7, x_8, x_9, x_{10}\}\}$,
$\overline{\eta}_{L,B,\beta}(D) = \{\{x_1, x_2, x_4, x_6\}, \{x_5, x_7, x_8, x_9, x_{10}\}\}$,
$\kappa_{L,B,\alpha,\beta}(D) = \{1/4, 4/5\}$.

From a global viewpoint, we can observe that x_2 is definitely healthy, while x_7, x_8, x_9, and x_{10} are undoubtedly infected. x_3 may be healthy, and x_1, x_4, x_5, and x_6 are located in the boundary region of D_1 and D_2, requiring further examination of their condition. From a local viewpoint, we can ascertain that x_2 is healthy, and x_7, x_8, x_9, and x_{10} are infected. x_1, x_4, and x_6 could be healthy, while x_5 may be infected. Consequently, their condition also requires further investigation. The inclusion degree between the target decisions D_i and the neighborhood classes determined by objects in D_i can be used to approximate the target concepts, which is a time-efficient approach based on the definitions of local variable approximations.

4. Feature selection based on variable precision composite measure

According to the analysis in Section 3.2, in this part, we design a local rough set method for feature selection by variable precision composite measure. Firstly, we define three essential measures for evaluating features, then propose corresponding global and local methods for feature selection.

4.1. Significant measures based on VPNRS

Based on the variable precision neighborhood rough set theory, the approximation space is usually adopted to design a measure for feature selection. In this subsection, we investigate three significant measures for feature evaluation. We only use the local VPNRS as an example to design the corresponding measures for simplification.

Definition 8. Let $DIS = (U, B \cup D)$ be a decision information system. For $A \subseteq B$ and $c \in B - A$, the local significance of c relative A is defined as follow.

(1) $SIG_{L,1}(c, A, D) = \gamma_{L,1,A\cup\{c\}}(D) - \gamma_{L,1,A}(D)$;
(2) $SIG_{L,2}(c, A, D) = \gamma_{L,2,A}(D) - \gamma_{L,2,A\cup\{c\}}(D)$;
(3) $SIG_{L,3}(c, A, D) = \gamma_{L,3,A\cup\{c\}}(D) - \gamma_{L,3,A}(D)$.

To ensure generality, we make the assumption that $\gamma_{L,1,A} = 0$, $\gamma_{L,2,A} = 1$, and $\gamma_{L,3,A} = 0$ when $A = \emptyset$. Similarly, we can also define global significance measures $SIG_1(c, A, D)$, $SIG_2(c, A, D)$, and $SIG_3(c, A, D)$. The larger the value of the significance measure, the more important the feature. Therefore, we select the attribute corresponding to the maximum value $SIG_{L,i}(c, A, D)$ at each iteration, where $i = 1, 2, 3$.

4.2. Algorithm design

Note that there may be multiple reduction sets according to the reduction definition, but in some cases, one feature subset is sufficient. In this subsection, we choose a heuristic forward strategy for feature selection based on three different significance measures of attributes to select an attribute subset with the same approximate ability as the original data. Moreover, since the feature measures are not monotonic when the neighbor radius is flexible, we need to evaluate further whether there are redundant attributes in the selected attribute subset, as per the propositions of measures. To address this issue, we have developed a local feature selection algorithm based on the variable

Algorithm 1: Feature selection algorithm based on local VPCM.

Input : A decision information system $DIS = (U, B \cup D)$, parameters α and β, and radius δ.
Output : Feature selection subset A.

begin
  Initialize $A \leftarrow \emptyset$, $r_0 \leftarrow \emptyset$, $redunt = \emptyset$, and $p = 1$;
  Compute the composite measure $\kappa_{L,B,\alpha,\beta}(D)$;
  while *P=1* **do**
    for *each* $c_k \in B - A$ **do**
      for $i = 1 : s$ **do**
        Compute $\underline{R_{A\cup\{c_k\}}}_{L,\alpha}(D_i)$, $\overline{R_{A\cup\{c_k\}}}_{L,\alpha}(D_i)$, and $C_{L,A\cup\{c_k\},\alpha,\beta}(D_i)$;
      end
      Compute the local significance measure $SIG_{L,3}(c_k, A, D)$;
    end
    $A \leftarrow c_k$ satisfying $c_k = \underset{c_k \in B-A}{argmax}\, SIG_{L,3}(c_k, A, D)$;
    if $\kappa_{L,A,\alpha,\beta}(D) = \kappa_{L,B,\alpha,\beta}(D)$ **then**
      $P = 0$;
    else
      $P = 1$;
    end
  end
  for $a \in A$ **do**
    Compute $\kappa_{L,A-\{a\},\alpha,\beta}(D)$;
    if $\kappa_{L,A-\{a\},\alpha,\beta}(D) = \kappa_{L,A,\alpha,\beta}(D)$ **then**
      $redunt = redunt \cup \{a\}$;
    end
  end
  $A = A - redunt$;
  return : A.
end

precision composite measure (LVPCM). This algorithm's detailed selection process and pseudo code are presented in Algorithm 1 and Fig. 2, respectively.

In the LVPCM method, the reduction set contains l properties finally. In this algorithm, we first compute the local lower and upper approximations, and the time complexity is $O(|B|\sum_{j=1}^{s}(|D_j \parallel U| + |D_j|^2))$, then we obtain the composite measure $\kappa_{L,B,\alpha,\beta}$ whose time complexity is $O(s)$. Thus the complexity in pseudo-code 3 is $O(|B|\sum_{j=1}^{s}(|D_j \parallel U| + |D_j|^2)) + O(s)$ while the global rough set is $O(|B|\sum_{j=1}^{s}(|U|^2 + |D_j \parallel U|) + O(s)$. Suppose the size of original feature subset is l, the time complexity for selecting the ith feature is $O((|B|-i)i\sum_{j=1}^{s}(|D_j \parallel U| + |D_j|^2) + s)$, thus the total time complexity of original feature selection is $O(\sum_{i=1}^{l}((|B| - i)i\sum_{j=1}^{s}(|D_j \parallel U| + |D_j|^2) + s))$. When the selected features based on local algorithm is l, the corresponding time complexity is $O(\sum_{i=1}^{l}((|B| - i)i\sum_{j=1}^{s}(|U|^2 + |D_j \parallel U|) + s))$. Moreover, we compute the $\kappa_{L,A-\{a\},\alpha,\beta}(D)$ for $a \in A$ to check whether there exists redundant attribute, the time complexity of them is $l(l-1)\sum_{j=1}^{s}(|D_j \parallel U| + |D_j|^2 + s)$. Similarly, the computational complexity of deleting redundant attributes on the global algorithm is $l(l-1)\sum_{j=1}^{s}(|U|^2 + |D_j \parallel U| + s)$. Also, each of the other steps of the LVPCM method is constant. Due to $|D_j| \le |U|$, the local algorithm can significantly improve the selection efficiency.

Algorithm 1 is based on the composite measure to select attributes from the point of local, which contains more comprehensive information compared with lower and upper approximations measures. In addition, it saves time relative to global rough sets. Similarly to Algorithm 1, the local feature selection methods based on lower and upper approximations, namely LMLA and LMUA, can be proposed, respectively. Moreover, to illustrate the effectiveness of the local strategy, the global feature selection methods based on lower and upper approximations and the composite measure are correspondingly designed, called GMLA, GMUA, and GVPCM.

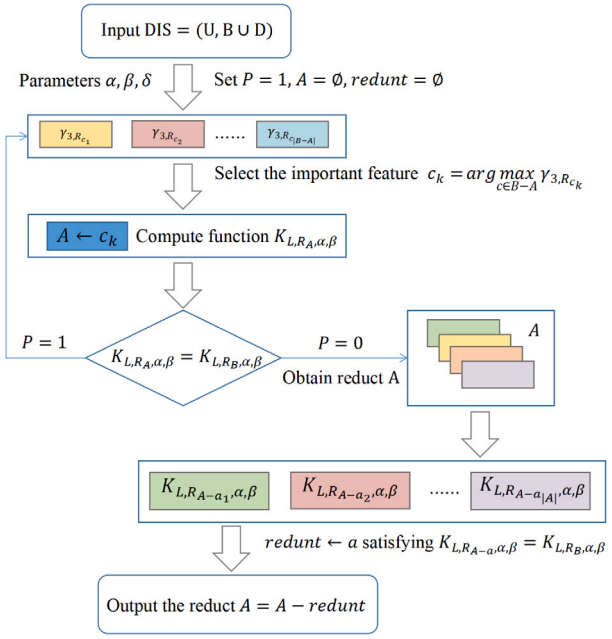


Fig. 2. Illustration of the overall procedure for LVPCM.

Table 2
The selected features by different methods.

R	$\underline{\eta}_{R,\alpha}(D)$	$\overline{\eta}_{R,\beta}(D)$	$\kappa_{R,\alpha,\beta}(D)$
Global	$\{b_2, b_3\}$	$\{b_1, b_3, b_4\}$	$\{b_1, b_2, b_3\}$
Local	$\{b_2, b_3\}$	$\{b_1, b_2, b_3\}$	$\{b_3, b_4\}$

Table 3
The average accuracy of different algorithms on KNN classifier.

Featuresubset	$\{b_2, b_3\}$	$\{b_3, b_4\}$	$\{b_1, b_2, b_3\}$
$k = 3$	0.50	0.70	0.60
$k = 5$	0.50	0.70	0.60

Example 3 (*Continue to the Example 2*). According to the Algorithm 1, we could obtain the attribute reduction results about designed feature selection methods, shown in Table 2.

We evaluate the different feature subset according to their classify performance. Considering the size of individuals, we adopt leave-one-out method to classify the individual in Table 1 on KNN classifier. The average accuracy of different feature subset is shown in the following Table 3.

According to the results shown in Table 3, we know the $\{b_1, b_2, b_3\}$ is superior to other subsets $\{b_2, b_3\}$ and $\{b_1, b_3, b_4\}$, its average accuracy achieves the highest 0.6 from a global viewpoint. The accuracy of $\{b_3, b_4\}$ gets the highest value 0.7 higher than that of $\{b_2, b_3\}$ and $\{b_1, b_2, b_3\}$ when $k = 3$ and $k = 5$. These results show that the feature selection based on the composite measure could select the relatively better subset with better classification performance. Meanwhile, the local algorithm could select an equally excellent subset of attributes while improving efficiency. Thus, the local algorithm can be further applied to feature selection.

5. Experiment and evaluation

In this section, variable numerical experiments are carried out on the public to verify the effectiveness of the proposed variable precision composite measure and target function for feature selection. To illustrate the effectiveness of the proposed method, the local methods, including LMLA, LMUA, and LNRS [46], and global methods, including

Table 4
Detailed information of 12 datasets.

No. s	Datasets	Abbreviations	Objects	Features
1	Seeds	Seed	210	7
2	Page-blocks	Page	5472	10
3	Hcv	Hcv	615	12
4	Speaker Accent Recognition	Spea	329	12
5	Vowel	Vowe	990	13
6	Wine	Wine	178	13
7	Climate Model Simulation Crashes	Clim	540	17
8	Segmentation	Segm	2310	19
9	Absenteeismatwork	Absen	740	20
10	South German Credit	Sout	1000	21
11	Wdbc	Wdbc	569	30
12	Ionosphere	Iono	351	33

GVPCM, GMLA, GMUA, and GNRS [46]. are selected to make comparisons in feature selection. All the methods are implemented in Matlab 2016b and carried out on a personal computer with Intel(R) Core(TM) i5-1135G7 CPU@2.40GH 2.42GH, and 16 GB memory.

5.1. Experimental design

In order to verify the effectiveness of the proposed LVPCM method, we select twelve datasets to make comparisons from the UCI and KEEL Repository. Table 4 shows detailed information about datasets. Each dataset is first normalized by the following equation.

$$\widehat{f}(x_i, b_j) = \frac{f(x_i, b_j) - Min(b_j)}{Max(b_j) - Min(b_j)}, \tag{14}$$

where $f(x_i, b_j)$ is the value of object x_i under feature $b_j \in B$, $Max(b_j)$ and $Min(b_j)$ are the maximum and minimum value of all objects in b_j.

Note that δ is an important parameter influencing neighborhood rough set model construction. Considering the number of attributes m is varied in the different datasets, we set the parameter $\delta = 0.01 * m$. Without loss of generality, we only let $\alpha = 0.7, \beta = 0.3$, $\alpha = 0.8, \beta = 0.2$, and $\alpha = 0.9, \beta = 0.1$ for the experiments. Moreover, to further verify the advantages of the LVPCM method, two classical classifiers Decision Tree(Gini index) and KNN(K = 3), are selected to estimate the classification accuracies of the feature selected by these four methods, respectively.

5.2. Classification performance evaluation of LVPCM

This subsection mainly compares the LVPCM method with other feature selection methods from three parts, including (1) the number of selected attributes, (2) the time comparison of various methods between global and local viewpoints, and (3) the classification performances of various algorithms under Decision Tree and KNN classifiers.

5.2.1. The number of selected features

The number of selected features is an essential index for comparing various feature selection methods, but it does not mean that the fewer, the better. Feature selection aims to select as few attributes as possible based on high classification accuracy. This subsection mainly analyzes the number of selected features on different methods, shown in Tables 5–7, where the "RAW" represents the information of the original dataset. From these tables, we find that the size of the selected attribute subset is all smaller than the original data, which illustrates the imperative of feature selection. From a global viewpoint, the average size of selected attribute subsets by the GNRS method is the smallest compared with those of the other feature selection methods. The average size of the attribute subset obtained by the LMUA is also the smallest from the local viewpoint. Moreover, in most situations, we know that the number of selected features by local algorithms is greater than that of global algorithms, except for the GMUA method. These results about selected features are further shown in Fig. 3.

5.2.2. Time consumption of different feature selection methods

According to the analysis in Section 3, the local method could significantly reduce the time complexity in the approximation process. In this subsection, we mainly verify the effectiveness of local methods compared with corresponding global methods by analyzing the running time of different feature selection methods.

Tables 8–10 record the running time of different methods on 12 datasets, where the unit of time is seconds(s). From these tables, we obtain that the local method takes much less time than the corresponding global method when the size of the local attribute subset is smaller than or equal to that of the global algorithm. When the number of selected features by the local method is greater than that of global ones, the local method could show superiority, such as the LNRS on Clim when $\alpha = 0.7, \beta = 0.3$ and $\alpha = 0.8, \beta = 0.2$.

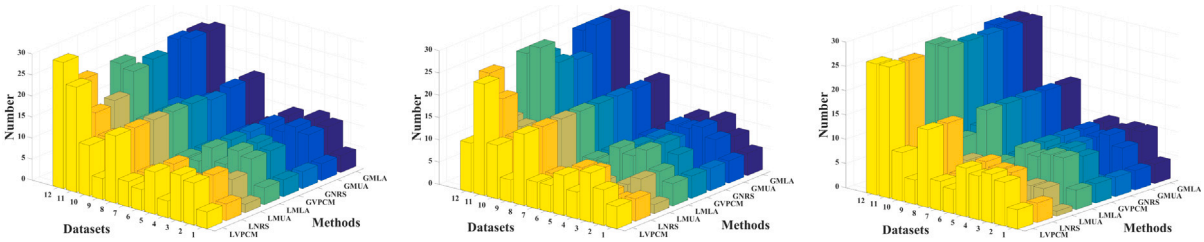
The sharper contrast between the global and local methods is shown in Fig. 4, where eight bars represent the GLMA, LMLA, GMUA, LMUA, GNRS, LNRS, GVPCM, and LVPCM, respectively. As can be seen from this figure, the bars of the local method are significantly lower than that of global ones in most cases at three precision. When $\alpha = 0.7, \beta = 0.3$, the local LMLA takes more time compared with that of the global method on Sout, and the LNRS method takes more time than that of the global method on Iono. When $\alpha = 0.8, \beta = 0.2$, the LNRS takes longer than the global method on Page and Hcv datasets. Moreover, when $\alpha = 0.9, \beta = 0.1$, the LNRS spends more time than global ones on Hcv. The above issue is because the local subset is larger than the global subset. The above analysis illustrates that the local method can significantly improve the efficiency of feature selection compared with the global method.

5.2.3. Classification performance of different methods

Different feature subsets could achieve different classification accuracy, and the higher the accuracy, the stronger the ability of the subset to classify. In this section, we adopt 10-fold cross-validation to evaluate the selected feature subset by four methods through their classification performances on Decision Tree and KNN classifiers. The classification accuracy and error of twelve datasets under different precisions are recorded in Tables 11–16, where the excellent results are in bold.

Tables 11–13 record the classification performance on the Decision Tree classifier. From the global viewpoint, the average accuracy of GMLA methods is higher than the original data with $\alpha = 0.7, \beta = 0.3$ and $\alpha = 0.8, \beta = 0.2$, the performance of the GNRS method is not excellent compared with the original data for all situations, while the GMUA and GVPCM are superior to the original data. In addition, the global GVPCM method obtains the highest accuracy 8($\alpha = 0.7, \beta = 0.3$), 9($\alpha = 0.8, \beta = 0.2$), 9($\alpha = 0.9, \beta = 0.1$) times, and minimum error 8($\alpha = 0.7, \beta = 0.3$), 7($\alpha = 0.8, \beta = 0.2$), 6($\alpha = 0.9, \beta = 0.1$) times in 12 datasets on Decision Tree classifier. Moreover, the average accuracy and error of 12 datasets in the GVPCM method are the highest and smallest among the four global methods, respectively. Therefore, we could obtain the GVPCM as an excellent and robust feature selection method compared with the other three methods. From the local viewpoint, the local LVPCM method achieves the highest accuracy 10($\alpha = 0.7, \beta = 0.3$), 9($\alpha = 0.8, \beta = 0.2$), 9($\alpha = 0.9, \beta = 0.1$) times and most minor error 7($\alpha = 0.7, \beta = 0.3$), 7($\alpha = 0.8, \beta = 0.2$), 4($\alpha = 0.9, \beta = 0.1$) times in 12 datasets. Meanwhile, it obtains the highest average accuracy of these compared methods, illustrating feature selection's effectiveness based on the proposed composite measure. It is noted that the classification performance of LVPCM is also higher than or equal to that of the global method. Thus, the LVPCM method performs well on the Decision Tree classifier.

Tables 14–16 record the classification performance on the KNN classifier. All the classified performances of GVPCM and LVPCM are better than the original data with various precisions. From global viewpoint, GVPCM method achieves 8($\alpha = 0.7, \beta = 0.3$), 10($\alpha = 0.8, \beta = 0.2$), 9($\alpha = 0.9, \beta = 0.1$) times get the highest accuracy and 6($\alpha = 0.7, \beta = 0.3$), 5($\alpha = 0.8, \beta = 0.2$), 7($\alpha = 0.9, \beta = 0.1$) times get the



(a) $\alpha = 0.7, \beta = 0.3$ (b) $\alpha = 0.8, \beta = 0.2$ (c) $\alpha = 0.9, \beta = 0.1$

Fig. 3. Number of selected features induced by different methods on 12 datasets.

Table 5
Number of selected features by different methods when $\alpha = 0.7, \beta = 0.3$.

Dataset	RAW	Global				Local			
		GMLA	GMUA	GNRS	GVPCM	LMLA	LMUA	LNRS	LVPCM
Seed	7	4	4	4	4	4	2	4	4
Page	10	10	10	1	10	10	7	10	10
Hcv	12	11	11	1	11	11	6	5	11
Spea	12	7	9	7	4	7	2	7	4
Vowe	13	10	10	10	10	10	4	10	10
Wine	13	6	6	5	4	6	1	6	5
Clim	17	6	6	1	6	6	1	6	6
Segm	19	16	16	15	16	16	16	16	16
Absen	20	7	7	4	5	7	1	7	5
Sout	21	11	11	11	12	12	3	12	12
Wdbc	30	25	25	11	24	23	18	17	25
Iono	33	24	24	1	18	24	3	24	30
Average	17.25	11.42	11.58	5.92	10.33	11.33	5.33	10.33	11.50

Table 6
Number of selected features by different methods when $\alpha = 0.8, \beta = 0.2$.

Dataset	RAW	Global				Local			
		GMLA	GMUA	GNRS	GVPCM	LMLA	LMUA	LNRS	LVPCM
Seed	7	5	5	5	5	5	2	5	5
Page	10	8	9	1	9	7	7	4	8
Hcv	12	11	11	1	11	11	5	5	11
Spea	12	9	10	9	6	9	2	9	6
Vowe	13	10	10	9	9	10	4	9	9
Wine	13	6	6	6	6	6	1	6	6
Clim	17	6	6	1	6	6	3	6	6
Segm	19	16	16	16	16	16	16	16	16
Absen	20	7	6	4	5	7	1	7	5
Sout	21	12	12	12	12	12	1	12	12
Wdbc	30	28	28	22	23	28	12	20	25
Iono	33	26	26	15	24	26	13	25	11
Average	17.25	12.00	12.08	8.42	11.00	11.92	5.58	10.33	10.00

Table 7
Number of selected features by different methods when $\alpha = 0.9, \beta = 0.1$.

Dataset	RAW	Global				Local			
		GMLA	GMUA	GNRS	GVPCM	LMLA	LMUA	LNRS	LVPCM
Seed	7	4	4	4	4	4	1	4	4
Page	10	10	8	1	9	10	5	1	9
Hcv	12	10	10	2	10	10	5	9	10
Spea	12	9	9	9	9	9	1	9	9
Vowe	13	10	10	10	10	10	1	10	10
Wine	13	5	5	5	5	5	1	5	5
Clim	17	6	6	6	6	6	4	6	6
Segm	19	16	16	16	16	16	7	16	16
Absen	20	2	6	5	3	3	1	3	5
Sout	21	9	9	9	10	9	1	9	10
Wdbc	30	27	27	27	27	27	5	27	27
Iono	33	27	27	14	13	27	1	12	27
Average	17.25	11.25	11.42	9.00	10.17	11.33	2.75	9.25	11.50

Table 8
Running time of different methods when $\alpha = 0.7, \beta = 0.3$.

Dataset	Global				Local			
	GMLA	GMUA	GNRS	GVPCM	LMLA	LMUA	LNRS	LVPCM
Seed	0.67	0.58	0.58	0.59	0.23	0.23	0.23	0.27
Page	189.75	162.72	35.55	219.30	82.52	95.69	78.30	89.17
Hcv	7.65	7.61	1.32	8.04	2.84	2.97	1.93	3.08
Spea	4.80	4.56	4.61	4.57	1.08	1.07	1.13	1.19
Vowe	33.88	38.93	35.03	34.90	4.82	5.47	4.97	4.57
Wine	1.31	1.23	1.03	1.40	0.46	0.47	0.48	0.55
Clim	3.95	3.94	2.39	3.95	2.32	2.40	1.00	2.47
Segm	129.24	144.33	119.75	135.77	32.56	35.09	32.31	35.53
Abse	142.71	135.28	57.66	71.58	5.39	7.25	5.32	3.30
Sout	18.28	28.74	47.56	20.59	28.00	24.65	27.89	12.47
Wdbc	23.93	23.63	14.78	23.89	14.39	15.27	12.12	15.10
Iono	15.31	15.17	1.05	13.73	8.56	8.38	8.47	9.66
Average	47.62	46.89	26.66	44.86	15.26	16.92	14.63	14.78

Table 9
Running time of different methods when $\alpha = 0.8, \beta = 0.2$.

Dataset	Global				Local			
	GMLA	GMUA	GNRS	GVPCM	LMLA	LMUA	LNRS	LVPCM
Seed	0.61	0.62	0.60	0.62	0.25	0.26	0.24	0.27
Page	607.86	435.73	89.96	412.88	217.34	213.73	143.92	221.06
Hcv	7.47	7.71	1.28	7.52	2.92	2.93	2.01	2.89
Spea	4.44	4.74	4.41	4.81	1.02	0.99	1.00	1.08
Vowe	37.37	38.76	33.07	32.91	5.31	5.30	4.67	4.58
Wine	1.16	1.14	1.07	1.23	0.45	0.45	0.45	0.50
Clim	3.57	3.75	0.95	3.86	2.37	2.24	2.40	2.44
Segm	111.19	145.82	117.44	152.62	34.70	40.66	34.61	39.88
Abse	146.79	312.69	98.75	71.01	12.08	7.25	5.52	3.29
Sout	36.26	36.96	35.82	15.13	21.71	27.89	21.92	9.81
Wdbc	23.53	23.79	22.73	23.96	14.77	15.30	13.56	15.01
Iono	14.95	15.05	11.84	16.15	8.92	8.39	8.48	9.75
Average	82.93	85.56	34.83	61.89	26.82	27.12	19.90	25.88

Table 10
Running time of different methods when $\alpha = 0.9, \beta = 0.1$.

Dataset	Global				Local			
	GMLA	GMUA	GNRS	GVPCM	LMLA	LMUA	LNRS	LVPCM
Seed	0.52	0.54	0.52	0.56	0.21	0.18	0.21	0.24
Page	452.99	430.66	90.04	440.92	210.70	215.34	42.48	218.96
Hcv	7.78	7.77	2.46	7.59	3.06	2.82	2.93	2.78
Spea	4.65	4.49	4.35	4.51	0.96	1.00	1.02	1.07
Vowe	34.44	38.90	33.75	34.73	4.84	4.66	4.75	4.88
Wine	0.97	1.04	1.02	1.07	0.38	0.39	0.38	0.44
Clim	3.78	3.78	3.84	4.89	2.35	1.74	2.35	2.44
Segm	122.57	109.22	115.55	129.35	35.95	38.19	35.83	36.60
Abse	157.65	199.51	68.39	91.85	8.61	7.02	7.51	3.67
Sout	43.16	44.08	42.80	16.58	26.13	3.58	26.20	10.33
Wdbc	23.21	23.57	23.33	23.69	14.16	15.02	14.30	14.83
Iono	14.79	14.97	10.48	10.75	8.41	0.60	6.61	9.95
Average	72.21	73.21	33.04	63.87	26.31	24.21	12.05	25.52

smallest error in 12 datasets. Also, for the LVPCM method, there are $10(\alpha = 0.7, \beta = 0.3)$, $10(\alpha = 0.8, \beta = 0.2)$, $9(\alpha = 0.9, \beta = 0.1)$ times to reach the highest accuracy and $6(\alpha = 0.7, \beta = 0.3)$, $7(\alpha = 0.8, \beta = 0.2)$, $7(\alpha = 0.9, \beta = 0.1)$ times to obtain the smallest error in 12 datasets from the local viewpoint. In addition, the average accuracy of the LVPCM method is higher than the other three methods, and the average error is smaller than the others from global and local viewpoints, respectively. Thus, the feature selection method based on composite measures could achieve excellent classification results compared with lower and upper measures. Moreover, the overall performance of LVPCM is superior to GVPCM when $\alpha = 0.8, \beta = 0.2$ and $\alpha = 0.9, \beta = 0.1$. Therefore, the LVPCM is efficient and robust among the four feature selection methods. Figs. 5 and 6 further show the advantages of the proposed method in classification performance.

In order to test whether there exists a significant difference between the compared methods in classification performance, the Wilcoxon pairwise test is adopted to make comparisons on 12 datasets. The test results are recorded in Tables 17. These tables show that when $\alpha = 0.9, \beta = 0.1$, there is no significant difference between the GVPCM method and GMUA and GNRS methods on the Decision Tree classifier. The classification performance of LVPCM on two classifiers differs from the other three methods. Also, considering that the average accuracy of the GVPCM and LVPCM is higher than that of the other three methods, we could obtain that the feature selection method based on the composite measure is better and more robust than others. The above analysis shows that LVPCM performs excellently in classification compared to other methods.

5.3. Robustness analysis

This subsection mainly illustrates the robustness of different methods in noise environments. To obtain noise data, on each dataset, we

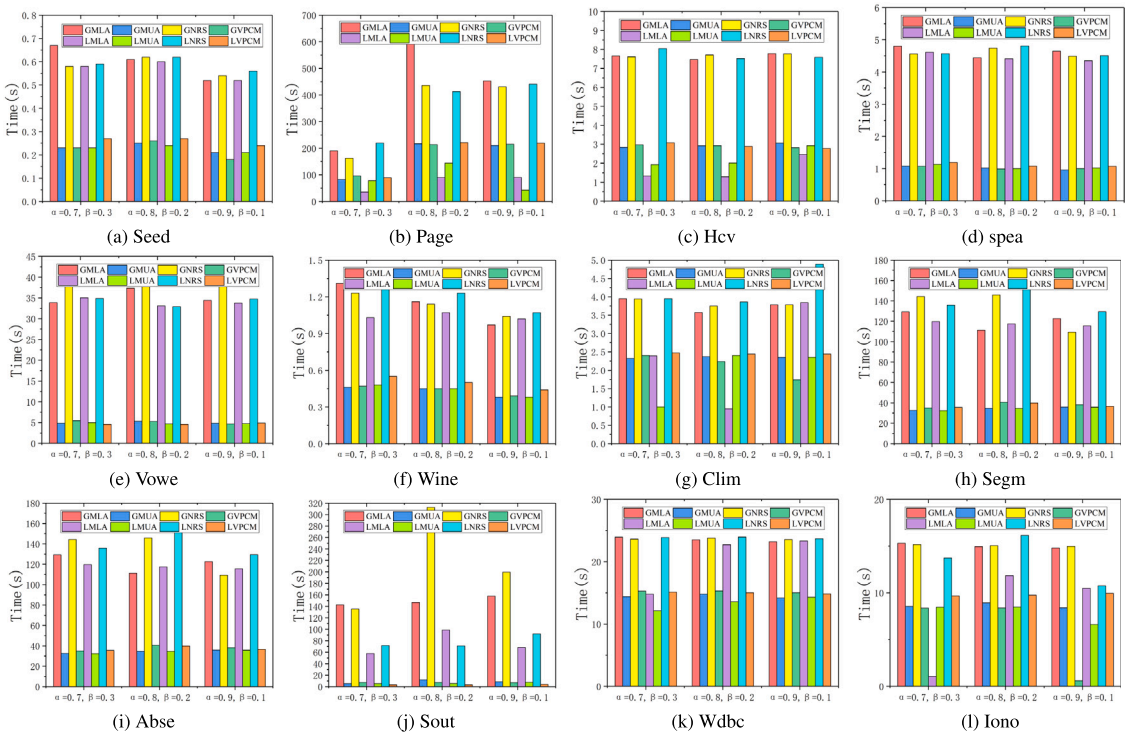


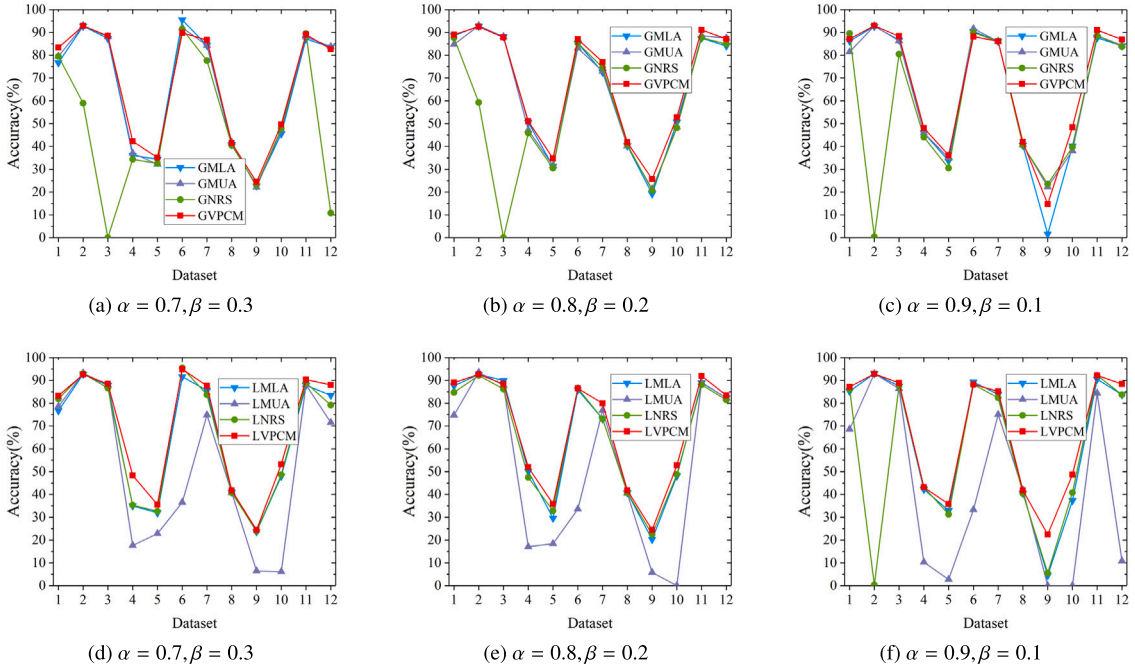
Fig. 4. Time consumption of different methods on 12 datasets.

Table 11
Classification accuracy of different methods on Decision Tree classifier when $\alpha = 0.7, \beta = 0.3$.

$\alpha = 0.7, \beta = 0.3$		Global				Local			
Dataset	RAW	GMLA	GMUA	GNRS	GVPCM	LMLA	LMUA	LNRS	LVPCM
Seed	85.24(9.84)	76.67(9.90)	79.52(9.27)	79.52(**5.04**)	**83.33**(7.86)	76.67(9.38)	78.57(**5.14**)	81.90(7.71)	**83.33**(7.86)
Page	92.36(1.52)	92.75(1.37)	**92.95**(1.11)	58.90(2.06)	92.91(**0.90**)	92.69(1.15)	**93.20**(1.32)	92.96(1.16)	92.56(**0.90**)
Hcv	86.50(2.25)	87.31(3.93)	88.12(**2.61**)	0.00(0.00)	**88.46**(5.16)	87.80(**3.27**)	87.96(3.60)	86.67(4.85)	**88.46**(4.73)
Spea	20.35(11.52)	36.16(8.02)	37.05(8.42)	34.38(8.84)	**42.23(6.88)**	34.98(8.50)	17.62(6.34)	35.29(9.99)	**48.32(6.25)**
Vowe	32.02(4.74)	34.34(4.12)	32.22(7.29)	32.53(7.25)	**35.15(3.95)**	31.92(6.37)	22.83(3.96)	32.63(5.15)	**35.56(3.95)**
Wine	91.57(6.02)	95.52(**5.75**)	92.09(8.48)	91.47(8.25)	**89.84**(9.18)	91.57(11.50)	36.54(9.59)	**95.52(4.11)**	94.97(4.39)
Clim	84.81(5.00)	84.26(4.80)	84.26(4.80)	77.59(6.38)	**86.67(3.75)**	85.56(5.58)	74.81(6.37)	83.70(**3.79**)	**87.59**(3.85)
Segm	40.52(4.92)	40.77(3.71)	**41.81**(3.01)	40.56(3.01)	41.47(**2.39**)	41.04(**2.52**)	40.82(3.33)	40.77(3.24)	**41.77**(3.10)
Abse	22.03(5.87)	22.70(5.51)	22.16(5.26)	23.11(8.81)	**24.46(3.95)**	23.65(7.19)	6.49(2.91)	23.92(5.74)	**24.46(3.95)**
Sout	51.20(5.43)	45.50(6.11)	47.10(6.94)	47.80(6.30)	**49.60(3.89)**	48.00(3.80)	6.10(2.42)	48.60(5.19)	**53.20(2.57)**
Wdbc	88.22(6.19)	87.17(4.53)	88.06(4.58)	**89.46**(5.60)	89.10(**2.73**)	87.86(5.25)	88.05(4.96)	88.92(4.49)	**90.33(2.38)**
Iono	85.74(6.48)	83.49(7.02)	**83.77(4.22)**	10.81(5.46)	82.62(5.79)	83.48(6.38)	71.51(10.65)	79.17(5.22)	**88.03(5.19)**
Average	65.05(5.82)	65.55(5.30)	65.76(5.50)	48.84(5.59)	**67.15(4.08)**	65.43(5.91)	52.04(5.05)	65.84(5.08)	**69.05(4.07)**

Table 12
Classification accuracy of different methods on Decision Tree classifier when $\alpha = 0.8, \beta = 0.2$.

$\alpha = 0.8, \beta = 0.2$		Global				Local			
Dataset	RAW	GMLA	GMUA	GNRS	GVPCM	LMLA	LMUA	LNRS	LVPCM
Seed	85.24(9.84)	88.57(8.16)	84.76(9.73)	87.62(9.31)	**89.05(7.12)**	87.62(8.16)	74.76(10.05)	84.76(10.48)	**89.05(6.75)**
Page	92.93(1.10)	92.76(1.30)	**92.98**(1.48)	59.25(2.06)	92.49(**0.83**)	92.60(1.23)	93.51(1.12)	92.12(**0.68**)	**92.60**(1.34)
Hcv	86.50(4.05)	**88.13(3.48)**	**88.13(3.48)**	0.00(0.00)	87.81(3.84)	**89.92**(3.95)	87.82(**2.33**)	86.18(4.81)	88.44(5.59)
Spea	20.35(3.94)	50.50(7.18)	47.13(8.83)	45.86(9.91)	**51.02(5.91)**	50.14(9.54)	17.03(**7.49**)	47.48(11.23)	**51.93(7.49)**
Vowe	32.02(6.06)	31.31(6.00)	31.62(**3.30**)	30.51(4.41)	**34.75**(3.90)	29.60(6.04)	18.38(4.41)	32.83(7.50)	**35.86(2.34)**
Wine	88.10(9.15)	85.26(8.80)	83.10(**7.97**)	85.29(7.81)	**86.99**(8.12)	85.88(7.74)	33.56(15.33)	**86.57(5.71)**	86.50(6.56)
Clim	84.81(5.00)	72.59(**4.35**)	72.96(6.77)	74.63(4.79)	**77.04**(5.47)	73.15(7.68)	76.85(4.21)	73.15(7.42)	**80.00(2.87)**
Segm	40.52(4.92)	40.30(4.85)	40.17(4.21)	40.82(5.12)	**41.77(3.10)**	40.78(3.93)	40.48(4.21)	40.74(3.16)	**41.77(3.10)**
Absen	21.89(5.05)	19.19(5.58)	21.35(4.60)	20.81(5.28)	**25.68(4.46)**	20.27(4.85)	5.81(3.37)	22.43(**3.32**)	**24.46**(3.95)
Sout	51.20(5.43)	50.60(5.08)	48.40(5.99)	48.10(4.48)	**52.70(3.23)**	48.10(4.63)	0.00(0.00)	48.80(5.92)	**52.70(3.23)**
Wdbc	88.22(6.19)	87.52(**4.17**)	88.57(6.40)	87.70(5.02)	**91.05**(5.69)	88.93(4.31)	89.10(5.96)	88.05(4.68)	**91.92(3.32)**
Iono	85.47(6.38)	84.02(6.17)	**87.48**(5.98)	84.89(7.78)	87.18(**3.61**)	82.33(6.64)	82.34(**5.16**)	81.45(5.47)	**83.44**(7.85)
Average	64.77(5.59)	65.90(5.43)	65.55(5.73)	55.46(5.50)	**68.13(4.61)**	65.78(5.73)	51.64(5.30)	65.38(5.87)	**68.22(4.53)**



(a) $\alpha = 0.7, \beta = 0.3$ (b) $\alpha = 0.8, \beta = 0.2$ (c) $\alpha = 0.9, \beta = 0.1$

(d) $\alpha = 0.7, \beta = 0.3$ (e) $\alpha = 0.8, \beta = 0.2$ (f) $\alpha = 0.9, \beta = 0.1$

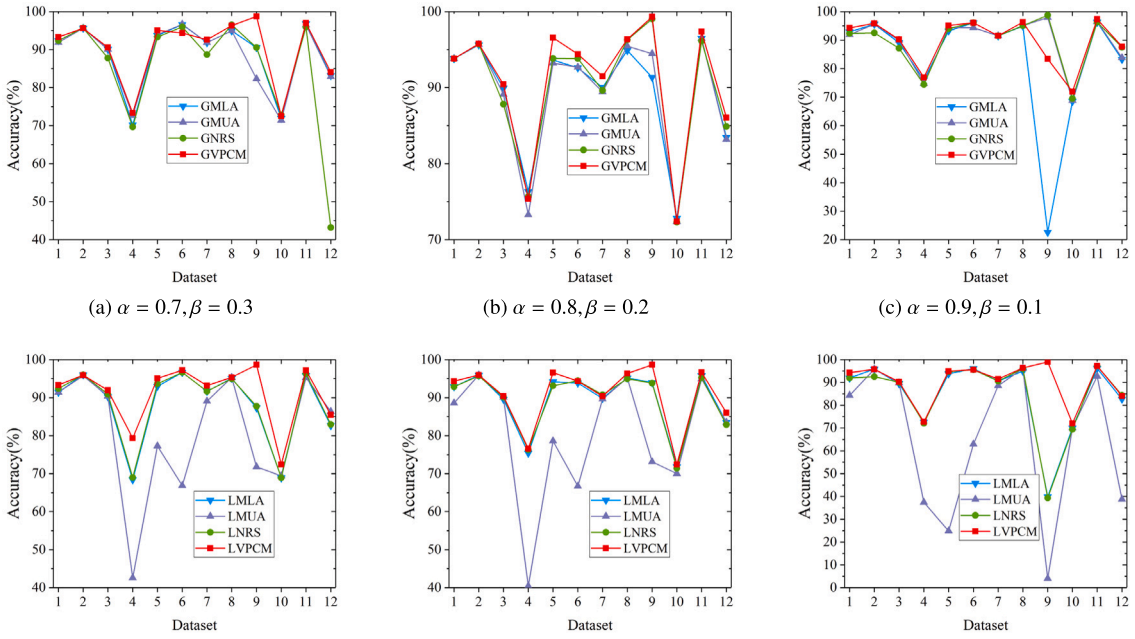
Fig. 5. Classification accuracy of different methods on Decision Tree classifier.

Table 13
Classification accuracy of different methods on Decision Tree classifier when $\alpha = 0.9, \beta = 0.1$.

$\alpha = 0.9, \beta = 0.1$		Global				Local			
Dataset	RAW	GMLA	GMUA	GNRS	GVPCM	LMLA	LMUA	LNRS	LVPCM
Seed	85.24(9.84)	86.19(7.60)	81.43(7.60)	**89.52**(7.71)	87.14(**4.52**)	85.24(9.10)	68.57(8.46)	86.19(6.13)	**87.14(4.52)**
Page	93.00(1.10)	92.64(1.28)	92.86(1.02)	0.42(**0.37**)	**92.96**(0.72)	**92.95**(1.03)	**92.95**(0.54)	0.35(**0.37**)	92.85(0.82)
Hcv	86.50(5.02)	86.68(4.18)	86.34(5.89)	80.50(4.74)	**88.28(3.35)**	87.49(4.81)	86.36(**2.74**)	87.65(2.99)	**88.95**(6.23)
Spea	20.35(5.70)	45.89(11.78)	45.89(**8.46**)	44.07(9.27)	**48.02**(10.74)	42.23(7.95)	10.34(**6.27**)	43.13(7.80)	**43.14**(9.66)
Vowe	32.02(3.67)	33.74(8.78)	35.05(3.60)	30.51(6.00)	**36.16(2.93)**	32.94(7.84)	2.73(1.58)	31.31(5.15)	**35.76(4.62)**
Wine	90.92(8.77)	90.46(7.43)	**91.63**(9.16)	89.93(5.04)	88.20(**4.94**)	**89.28**(6.24)	33.27(10.81)	88.20(**6.11**)	88.27(10.58)
Clim	84.81(5.00)	**86.11(5.03)**	**86.11(5.03)**	**86.11(5.03)**	**86.11(5.03)**	83.89(5.87)	75.00(**4.72**)	82.41(5.87)	**85.19**(5.98)
Segm	40.52(4.92)	40.48(3.14)	40.69(**2.87**)	40.48(3.36)	**41.90**(3.28)	40.48(3.14)	41.95(2.17)	40.35(**1.83**)	**41.77**(3.10)
Absen	21.89(5.05)	1.62(**1.66**)	22.30(7.73)	**23.51**(7.91)	14.73(4.74)	4.46(2.63)	0.00(0.00)	5.54(**2.25**)	**22.43**(5.85)
Sout	51.20(5.43)	40.2(**3.68**)	38.1(6.01)	39.70(4.60)	**48.40**(4.45)	37.30(4.95)	0.00(0.00)	40.80(5.09)	**48.70(4.11)**
Wdbc	88.22(6.19)	87.70(**3.15**)	88.58(4.15)	88.57(5.24)	**91.03**(3.28)	90.51(4.62)	84.35(4.67)	**92.23(3.43)**	**92.23(3.43)**
Iono	85.74(5.58)	84.03(6.02)	84.35(5.89)	83.78(5.63)	**86.89(4.52)**	83.75(8.54)	10.85(5.53)	83.74(**4.91**)	**88.33**(5.24)
Average	65.44(5.98)	64.64(5.31)	66.11(5.62)	58.09(5.41)	**67.49(4.38)**	64.21(5.56)	42.20(**3.96**)	56.82(4.33)	**67.90**(5.35)

Table 14
Classification accuracy of different methods on KNN classifier when $\alpha = 0.7, \beta = 0.3$.

$\alpha = 0.7, \beta = 0.3$		Global				Local			
Dataset	RAW	GMLA	GMUA	GNRS	GVPCM	LMLA	LMUA	LNRS	LVPCM
Seed	90.48(5.96)	92.38(**3.33**)	91.90(4.52)	92.38(7.84)	**93.33**(4.60)	91.43(5.41)	91.43(8.64)	92.38(**3.33**)	**93.33**(4.60)
Page	94.80(0.73)	95.65(1.07)	95.63(0.72)	**95.69**(1.45)	95.60(**0.67**)	**95.98**(1.10)	95.74(1.03)	**95.98(0.79)**	95.91(0.67)
Hcv	89.93(2.58)	90.08(4.04)	90.07(**3.75**)	87.80(5.55)	**90.56**(4.35)	90.24(2.99)	90.41(**2.37**)	90.90(3.26)	**91.94**(4.54)
Spea	43.15(9.27)	70.21(9.78)	72.64(**7.15**)	69.64(7.86)	**73.27**(8.20)	68.41(9.75)	42.58(9.70)	69.00(9.56)	**79.34(6.49)**
Vowe	93.74(3.44)	93.74(2.51)	93.94(2.69)	93.33(3.51)	**95.05(2.35)**	92.93(2.52)	77.27(5.29)	93.64(2.78)	**95.05(2.35)**
Wine	95.46(3.67)	96.60(**4.03**)	**96.67**(4.68)	95.98(4.81)	94.38(5.93)	96.63(5.38)	66.86(9.28)	96.63(**2.90**)	**97.22**(3.93)
Clim	92.59(2.62)	91.85(4.47)	91.85(4.47)	88.70(3.75)	**92.59(3.15)**	91.67(4.12)	89.07(3.85)	91.67(5.18)	**93.15(3.15)**
Segm	94.97(1.45)	94.94(1.33)	94.94(**0.97**)	**96.54**(1.10)	96.32(1.08)	95.06(1.45)	95.28(1.23)	94.94(2.03)	**95.32(0.86)**
Absen	89.59(6.37)	90.54(2.50)	82.30(4.10)	90.54(1.57)	**98.78(1.34)**	87.30(5.41)	71.76(4.20)	87.70(4.15)	**98.65(1.42)**
Sout	73.00(5.25)	71.90(5.00)	71.40(5.66)	**72.80(4.92)**	72.50(5.13)	68.90(4.20)	69.40(5.32)	68.90(**2.75**)	**72.40**(3.92)
Wdbc	96.31(2.26)	96.84(1.69)	96.48(2.21)	95.95(1.81)	**97.01(1.38)**	95.78(2.65)	95.25(2.63)	96.31(2.40)	**97.19(2.19)**
Iono	85.18(5.85)	83.17(6.00)	82.91(7.67)	43.21(17.83)	**84.05(2.39)**	82.61(**5.85**)	**86.33**(7.64)	82.93(7.79)	85.47(3.43)
Average	86.60(4.12)	88.99(3.81)	88.39(4.05)	85.21(5.16)	**90.29(3.38)**	88.08(4.24)	80.95(5.10)	88.42(3.91)	**91.25(3.13)**



(a) $\alpha = 0.7, \beta = 0.3$ (b) $\alpha = 0.8, \beta = 0.2$ (c) $\alpha = 0.9, \beta = 0.1$

(d) $\alpha = 0.7, \beta = 0.3$ (e) $\alpha = 0.8, \beta = 0.2$ (f) $\alpha = 0.9, \beta = 0.1$

Fig. 6. Classification accuracy of different methods on KNN classifier.

Table 15
Classification accuracy of different methods on KNN classifier when $\alpha = 0.8, \beta = 0.2$.

$\alpha = 0.8, \beta = 0.2$		Global				Local			
Dataset	RAW	GMLA	GMUA	GNRS	GVPCM	LMLA	LMUA	LNRS	LVPCM
Seed	90.48(5.96)	**93.81(4.52)**	**93.81**(5.04)	**93.81**(6.37)	**93.81**(5.04)	92.86(5.61)	88.57(7.51)	92.86(5.14)	**94.29(3.01)**
Page	94.80(0.41)	95.63(**0.80**)	95.67(1.08)	95.72(1.45)	**95.76**(0.55)	95.89(1.14)	**95.96**(1.27)	95.67(1.00)	**95.96(0.60)**
Hcv	89.93(2.95)	89.93(**2.35**)	89.10(4.58)	87.81(2.43)	**90.41**(3.38)	89.58(5.25)	90.09(**3.48**)	90.16(3.58)	**90.309**(3.83)
Spea	43.15(7.06)	**76.33**(9.18)	73.28(9.51)	75.67(**6.44**)	75.37(8.59)	75.39(7.32)	40.43(8.45)	76.28(**5.38**)	**76.59**(6.41)
Vowe	93.74(1.95)	93.64(3.20)	93.23(2.70)	93.84(3.91)	**96.57(1.66)**	94.14(1.77)	78.59(4.31)	93.13(4.17)	**96.57(1.66)**
Wine	95.49(5.20)	92.61(6.65)	92.68(5.32)	93.82(**4.09**)	**94.41**(6.42)	93.82(6.65)	66.70(15.51)	**94.41**(5.86)	94.35(**4.63**)
Clim	92.59(2.62)	90.00(4.88)	89.44(5.09)	89.63(4.47)	**91.48(2.92)**	89.81(3.52)	89.63(**3.40**)	**90.74**(4.78)	90.37(4.60)
Segm	94.97(1.45)	94.89(1.83)	95.45(2.02)	96.28(**1.06**)	**96.36**(1.62)	95.15(1.93)	95.15(1.81)	94.94(1.86)	**96.36(1.62)**
Absen	90.27(4.02)	91.35(2.93)	94.46(2.72)	99.05(0.96)	**99.32(0.91)**	93.92(2.23)	73.11(5.27)	93.78(1.82)	**98.65(1.42)**
Sout	73.00(5.25)	**72.80**(6.71)	72.40(3.37)	72.30(5.21)	72.40(**2.59**)	71.80(5.61)	70.00(4.59)	71.40(6.55)	**72.40(2.59)**
Wdbc	96.31(2.26)	96.49(2.04)	96.48(2.75)	96.14(2.15)	**97.37(1.71)**	95.79(**1.50**)	94.90(2.54)	95.25(2.97)	**96.67**(2.92)
Iono	84.90(5.89)	83.44(7.14)	83.19(5.41)	84.87(**5.29**)	**86.04**(5.78)	83.48(**5.06**)	83.48(6.29)	82.87(5.63)	**86.02**(5.49)
Average	86.64(3.75)	89.24(4.35)	89.10(4.13)	89.91(3.65)	**90.78(3.43)**	89.30(3.97)	80.55(5.37)	89.29(4.06)	**90.72(3.23)**

Table 16
Classification accuracy of different methods on KNN classifier when $\alpha = 0.9, \beta = 0.1$.

$\alpha = 0.9, \beta = 0.1$		Global				Local			
Dataset	RAW	GMLA	GMUA	GNRS	GVPCM	LMLA	LMUA	LNRS	LVPCM
Seed	90.48(5.96)	92.86(**4.05**)	91.9(6.37)	92.38(6.02)	**94.29**(5.04)	91.90(5.96)	84.29(7.46)	91.90(**3.92**)	**94.29**(5.04)
Page	94.80(0.81)	95.61(1.20)	95.58(0.96)	92.53(1.09)	**95.91(0.76)**	**95.83**(0.84)	95.81(0.97)	92.51(1.09)	95.80(**0.65**)
Hcv	89.93(3.95)	89.10(4.49)	89.93(4.80)	87.17(5.35)	**90.24(3.75)**	89.42(5.08)	89.75(**2.07**)	90.07(2.48)	**90.25**(4.97)
Spea	43.15(5.09)	76.00(6.72)	74.79(**6.33**)	74.45(8.54)	**76.90**(6.43)	72.61(**4.70**)	37.42(7.96)	72.03(7.43)	**72.68**(7.67)
Vowe	93.74(2.48)	93.23(2.29)	94.44(2.62)	94.04(2.15)	**95.15(1.83)**	93.74(2.73)	24.85(3.73)	94.75(2.17)	**94.95(1.90)**
Wine	96.08(3.77)	96.08(3.77)	94.38(3.71)	**96.11(2.68)**	96.08(3.77)	**96.08**(5.90)	62.94(6.35)	95.56(9.00)	95.56(**4.38**)
Clim	92.59(2.62)	**91.48(2.92)**	**91.48(2.92)**	**91.48(2.92)**	**91.48(2.92)**	90.56(3.08)	88.52(3.47)	90.56(3.08)	**91.48(2.92)**
Segm	94.97(1.45)	94.94(1.29)	95.11(1.71)	95.06(1.54)	**96.32(0.94)**	94.94(1.29)	96.06(1.75)	96.02(**1.08**)	**96.36**(1.19)
Absen	90.81(2.28)	22.57(4.42)	97.97(1.72)	**98.78(1.49)**	83.38(3.66)	39.86(4.94)	4.05(3.06)	39.32(6.54)	**99.05(1.11)**
Sout	71.90(5.25)	68.50(4.67)	68.80(3.79)	69.50(**2.72**)	**71.90**(3.51)	70.00(5.44)	70.00(5.35)	69.40(6.33)	**72.00(4.32)**
Wdbc	96.31(2.26)	96.31(2.67)	96.49(2.16)	96.31(2.34)	**97.36(1.91)**	96.14(3.18)	92.63(3.77)	**97.19(1.69)**	**97.19(1.69)**
Iono	85.78(8.34)	83.21(5.78)	83.79(6.02)	87.46(4.16)	**87.74(4.08)**	82.65(4.56)	38.75(14.28)	**84.31(4.96)**	84.06(5.33)
Average	86.71(3.69)	83.32(3.69)	89.56(3.59)	89.61(3.42)	**89.73(3.22)**	84.48(3.97)	65.42(5.02)	84.47(4.15)	**90.31(3.43)**

Table 17
The P-value of Wilcoxon pairwise signed rank test between compared methods on 12 datasets.

Parameters	Global						Local					
	Decision Tree			KNN			Decision Tree			KNN		
	GMLA	GMUA	GNRS	GMLA	GMUA	GNRS	LMLA	LMUA	LNRS	LMLA	LMUA	LNRS
$\alpha = 0.7, \beta = 0.3$	0.052	0.055	0.005	0.027	0.034	0.034	0.001	0.002	0.003	0.001	0.002	0.001
$\alpha = 0.8, \beta = 0.2$	0.002	0.007	0.001	0.019	0.002	0.010	0.010	0.002	0.001	0.001	0.001	0.007
$\alpha = 0.9, \beta = 0.1$	0.014	0.206	0.123	0.002	0.054	0.067	0.003	0.002	0.001	0.003	0.001	0.010

Table 18
Average classification accuracy of compared methods on Decision Tree classifier under different noise levels.

$\alpha = 0.7, \beta = 0.3$		Global				Local			
Dataset	RAW	GMLA	GMUA	GNRS	GVPCM	LMLA	LMUA	LNRS	LVPCM
seed	54.21(10.67)	53.57(10.62)	54.05(10.01)	53.41(9.70)	**54.37(9.50)**	54.92(10.66)	57.94(8.96)	54.37(11.36)	**55.00(10.11)**
page	90.45(1.12)	90.48(1.55)	90.82(1.14)	75.00(1.88)	**90.97(1.26)**	90.99(1.14)	49.21(1.43)	91.25(1.04)	**91.83(1.12)**
hcv	84.55(3.96)	**82.48(4.63)**	78.75(5.88)	22.11(5.01)	81.98(5.28)	82.09(5.77)	75.73(5.29)	81.69(4.72)	**81.75(5.01)**
Spea	12.32(6.80)	8.77(5.03)	8.76(4.99)	9.27(5.45)	**9.38(5.89)**	9.33(4.58)	34.74(8.16)	10.58(6.96)	**11.11(7.66)**
vowe	24.81(5.41)	24.21(5.10)	21.73(4.81)	**24.24(4.63)**	23.60(4.62)	24.70(4.77)	7.98(2.98)	24.75(4.35)	**25.42(4.11)**
wine	62.12(13.90)	60.60(12.61)	63.20(11.12)	53.38(11.15)	**64.26(11.59)**	58.78(11.14)	37.32(10.02)	58.92(11.66)	**62.09(11.58)**
Clim	79.26(5.12)	79.51(5.64)	80.59(4.66)	77.22(5.72)	**81.42(4.87)**	80.46(4.90)	79.20(4.93)	79.66(4.84)	**82.19(5.69)**
Segm	42.37(3.30)	42.11(3.59)	32.34(2.94)	41.89(3.84)	**42.49(3.51)**	**43.83(3.41)**	1.79(0.83)	42.34(3.83)	41.84(3.43)
Abse	21.73(7.03)	20.97(4.96)	21.01(5.54)	21.24(5.55)	**21.78(5.57)**	23.09(5.72)	0.56(0.85)	**24.12(6.10)**	22.84(6.48)
Sout	39.23(5.05)	**37.12(4.19)**	36.93(5.04)	36.40(4.31)	36.75(5.08)	36.77(4.72)	2.28(1.43)	37.12(5.32)	38.08(5.78)
wdbc	84.11(4.83)	85.47(4.79)	**84.50(4.64)**	85.03(5.28)	85.42(4.52)	85.62(5.08)	59.52(5.72)	85.29(4.46)	**85.80(3.96)**
Iono	74.48(6.94)	75.12(8.77)	75.69(8.90)	74.77(8.01)	**76.45(7.26)**	78.31(7.55)	51.37(9.60)	**79.20(6.27)**	76.31(8.31)
Average	55.80(6.18)	55.03(5.96)	54.03(5.81)	47.83(5.88)	**55.74(5.75)**	55.74(5.79)	38.14(5.02)	55.77(5.91)	**56.19(6.10)**

Table 19
Average classification accuracy of compared methods on Decision Tree classifier under different noise levels.

$\alpha = 0.8, \beta = 0.2$		Global				Local			
Dataset	RAW	GMLA	GMUA	GNRS	GVPCM	LMLA	LMUA	LNRS	LVPCM
seed	53.97(9.91)	53.97(10.16)	53.57(11.68)	53.57(11.68)	**54.37(11.3)**	55.00(11.17)	53.73(9.48)	54.29(11.66)	**55.48(9.83)**
page	91.62(1.15)	90.63(1.21)	**90.76(1.01)**	67.76(1.91)	88.74(4.02)	91.03(1.38)	72.29(1.64)	**91.23(1.14)**	90.49(1.28)
hcv	80.09(4.85)	81.75(5.10)	81.60(5.31)	54.99(6.34)	**82.28(4.89)**	81.43(4.66)	75.24(5.11)	82.03(4.37)	**82.40(4.81)**
Spea	11.29(7.62)	9.17(5.29)	8.36(5.29)	6.08(4.03)	**11.67(5.99)**	9.89(5.87)	10.13(5.26)	11.04(5.23)	**12.77(6.32)**
vowe	12.71(3.21)	13.87(3.64)	10.32(3.39)	**14.11(3.56)**	13.40(4.01)	**14.53(4.12)**	3.45(1.99)	13.99(3.68)	13.18(3.49)
wine	61.49(11.98)	60.67(12.40)	60.03(12.96)	59.93(13.26)	**61.72(13.71)**	60.19(11.87)	36.23(12.33)	60.89(11.03)	**62.42(12.08)**
Clim	75.44(6.4)	75.43(6.19)	75.00(5.46)	72.41(5.77)	**78.52(6.03)**	74.41(5.02)	75.65(6.65)	74.72(5.65)	**79.91(5.56)**
Segm	42.80(3.42)	42.14(3.50)	36.85(4.13)	42.37(3.15)	**42.88(2.82)**	41.79(3.19)	1.81(0.82)	42.43(3.81)	**42.90(2.64)**
Abse	22.55(6.05)	21.58(5.86)	23.54(6.24)	**23.87(5.94)**	21.58(6.54)	**24.08(6.22)**	1.08(1.12)	24.05(7.05)	22.57(6.28)
Sout	41.53(4.60)	36.32(4.40)	36.27(4.76)	37.75(4.26)	**37.77(4.82)**	36.57(5.18)	12.4(3.57)	36.77(4.75)	**37.82(4.57)**
wdbc	84.13(4.05)	84.98(5.30)	85.23(5.29)	85.00(4.77)	**85.38(4.97)**	85.27(3.83)	69.19(6.65)	84.42(4.78)	**85.85(4.72)**
Iono	71.67(7.30)	62.58(9.89)	**64.91(6.98)**	64.26(8.93)	62.96(9.12)	63.58(8.15)	40.98(7.47)	**64.38(8.47)**	63.29(7.51)
Average	54.11(5.88)	52.76(6.08)	52.20(6.04)	48.51(6.13)	**53.44(6.52)**	53.15(5.89)	37.68(5.17)	53.35(5.97)	**54.09(5.76)**

add random noise to different proportions of samples from 10% to 60% with a step of 10%. For each feature $b_j \in B$, the noise data are obtained as follows:

$$\hat{v}(x_i, b_j) = \begin{cases} \hat{f}(x_i, b_j) + r_{ij}, 0 \le \hat{f}(x_i, b_j) + r_{ij} \le 1 \\ \hat{f}(x_i, b_j) \quad , \quad otherwise \end{cases} \tag{15}$$

where $\hat{f}(x_i, b_j)$ is normalized data, and $0 \le r_{ij} \le 1$ is the added random noise.

To observe the classification performance of compared methods in noise environments, we run these methods on different noise levels and record the corresponding classification accuracy on the Decision Tree and KNN classifiers. Tables 18–20 record the average classification accuracy of different methods on six noise levels. It can be seen from these tables that the proposed GVPCM and LVPCM methods are better than other methods from the whole average viewpoint on three parameters. On the Decision Tree classifier, the GVPCM and LVPCM achieve the maximum value 8 times in 12 datasets when $\alpha = 0.7, \beta = 0.3$ and $\alpha = 0.8, \beta = 0.2$, and also performs excellent in 7 and 10 datasets when $\alpha = 0.9, \beta = 0.1$. Meanwhile, the average accuracy of LVPCM is higher than GVPCM, demonstrating the excellent of local strategy compared with global strategy. Moreover, the detailed classification results of compared methods with different noise levels in four datasets are shown in Fig. 7. From this figure, we find that the classification accuracy of these compared methods all present a downward trend with the noise adding, and the proposed GVPCM and LVPCM perform better than other methods under different noise levels in most cases.

Moreover, Tables 21–23 also show the classification accuracy of compared methods on the KNN classifier. When $\alpha = 0.7, \beta = 0.3$, GVPCM and LVPCM achieve maximum accuracy values 9 and 11 times, and their average accuracy values on 12 datasets are also the highest from global and local viewpoints. Meanwhile, the GVPCM and LVPCM perform better on 8 datasets when $\alpha = 0.8, \beta = 0.2$, and obtain the excellent classification 5 and 7 times when $\alpha = 0.9, \beta = 0.1$. The average accuracy of LVPCM is better than that of GVPCM, which further illustrates the excellent performance of local strategy in noise data. In particular, the detailed results of classification accuracy varying with noise data are shown in Fig. 8. This figure shows that the classification results of compared methods vary with noise adding, and the Hcv dataset presents an obvious downward trend. Meanwhile, the GVPCM and LVPCM are also higher than other methods. All the above analyses

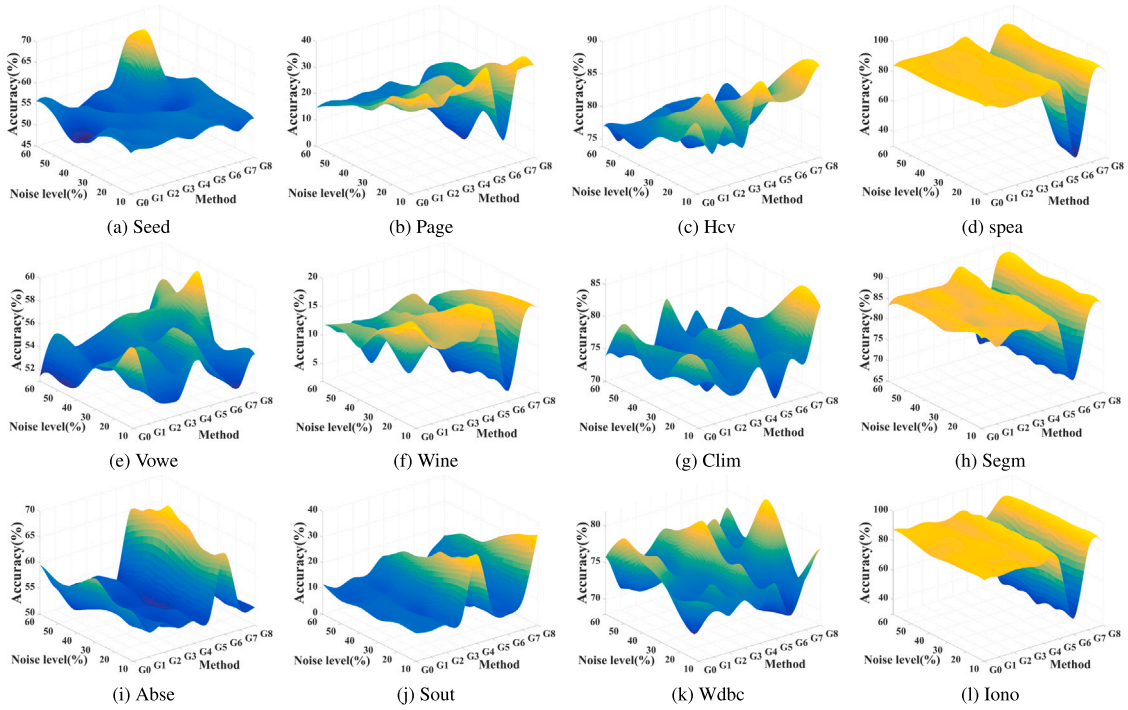


Fig. 7. Classification accuracy of different methods on the Decision Tree classifier when adding noise data. (a)–(d), (e)–(h), and (i)–(l) show the classification results of compared methods when $\alpha = 0.7, \beta = 0.3$, $\alpha = 0.8, \beta = 0.2$, and $\alpha = 0.9, \beta = 0.1$.

Table 20
Average classification accuracy of compared methods on Decision Tree classifier under different noise levels.

$\alpha = 0.9, \beta = 0.1$		Global				Local			
Dataset	RAW	GMLA	GMUA	GNRS	GVPCM	LMLA	LMUA	LNRS	LVPCM
seed	55.76(10.34)	54.05(11.09)	**54.84(9.52)**	53.57(11.23)	54.05(10.43)	53.81(8.05)	**66.51(10.45)**	53.97(12.25)	53.49(12.19)
page	90.27(0.99)	91.09(1.23)	90.42(1.08)	91.09(1.14)	**91.43(0.96)**	90.23(1.45)	80.53(1.43)	90.15(1.22)	**91.58(1.07)**
hcv	82.15(4.45)	**82.34(4.34)**	81.12(6.11)	82.06(4.59)	**82.34(4.62)**	81.72(4.45)	59.12(6.50)	81.98(4.92)	**82.00(5.01)**
Spea	15.10(6.21)	9.29(5.29)	8.77(6.34)	9.18(5.29)	**9.39(5.97)**	9.14(5.35)	8.20(4.38)	9.73(5.27)	**10.08(5.33)**
vowe	10.20(3.33)	9.53(2.69)	9.85(3.37)	10.47(3.16)	**25.34(5.02)**	10.61(3.72)	5.20(2.28)	10.64(3.43)	**25.57(4.39)**
wine	63.45(14.07)	60.74(12.89)	60.08(12.64)	60.76(12.27)	**60.80(11.77)**	60.00(10.87)	50.01(12.92)	60.65(12.15)	**60.87(10.47)**
Clim	71.67(5.71)	73.12(5.26)	73.67(4.93)	74.41(5.72)	**77.25(5.84)**	73.49(5.63)	73.43(6.27)	74.48(5.72)	**77.10(5.31)**
Segm	42.66(3.40)	45.02(3.78)	41.62(3.49)	44.51(3.15)	**45.11(3.82)**	43.86(3.69)	9.78(1.80)	44.65(3.96)	**45.02(3.50)**
Abse	23.15(5.79)	12.77(3.83)	11.80(3.53)	**24.55(5.82)**	22.50(6.01)	12.77(4.16)	1.13(1.33)	12.50(4.87)	**22.21(5.19)**
Sout	42.50(4.59)	**37.80(5.12)**	37.57(5.00)	36.70(5.39)	37.23(4.09)	37.22(4.72)	24.17(4.16)	36.87(5.06)	**37.52(5.43)**
wdbc	83.15(4.37)	85.24(3.3)	**85.35(5.07)**	84.77(5.13)	84.36(4.95)	84.74(5.29)	45.90(7.42)	84.62(4.8)	**85.41(4.56)**
Iono	67.73(7.92)	**58.63(8.91)**	57.83(8.12)	58.03(9.00)	58.02(7.38)	**58.80(7.58)**	31.31(8.22)	57.92(8.51)	58.74(8.88)
Average	53.98(5.93)	51.63(5.64)	51.08(5.77)	52.51(5.99)	**53.98(5.91)**	51.37(5.41)	37.94(5.60)	51.51(6.01)	**54.13(5.95)**

Table 21
Average classification accuracy of compared methods on KNN classifier under different noise levels.

$\alpha = 0.7, \beta = 0.3$		Global				Local			
Dataset	RAW	GMLA	GMUA	GNRS	GVPCM	LMLA	LMUA	LNRS	LVPCM
seed	87.06(8.34)	88.41(7.40)	87.38(6.57)	88.33(6.70)	**88.49(6.87)**	87.62(6.83)	86.27(8.47)	87.62(6.88)	**88.81(6.24)**
page	95.78(0.90)	94.84(0.98)	94.73(0.81)	90.65(1.09)	**95.26(0.71)**	95.09(0.84)	90.48(1.33)	95.05(1.06)	**95.38(0.90)**
hcv	94.34(3.13)	**90.72(3.48)**	89.99(4.01)	90.13(4.42)	**90.72(3.90)**	90.12(4.39)	90.24(3.80)	90.38(3.57)	**90.70(3.65)**
Spea	80.96(6.70)	53.08(8.13)	48.32(7.07)	52.47(9.44)	**55.90(8.32)**	53.34(7.57)	34.65(6.76)	53.70(8.71)	**56.54(6.94)**
vowe	75.88(4.00)	75.98(3.88)	68.86(4.46)	76.11(3.86)	**76.18(3.93)**	76.55(4.24)	35.47(4.18)	76.87(4.23)	**77.36(3.89)**
wine	71.23(9.94)	**80.28(7.94)**	81.93(7.32)	76.32(7.72)	78.46(8.83)	72.02(10.50)	67.02(10.48)	71.26(10.31)	**78.28(8.26)**
Clim	90.49(3.67)	90.43(3.92)	90.56(3.47)	89.75(3.95)	**91.05(3.52)**	89.72(4.03)	89.85(3.95)	89.35(4.01)	**91.48(3.73)**
Segm	95.56(1.19)	81.28(2.41)	68.28(3.32)	82.28(2.25)	**83.12(2.44)**	81.56(2.43)	17.03(2.17)	81.50(2.34)	**82.26(2.04)**
Abse	47.75(5.02)	94.12(2.44)	83.45(3.63)	98.00(1.73)	**98.92(1.29)**	81.26(4.75)	12.43(3.75)	81.89(5.06)	**98.83(1.33)**
Sout	62.57(3.89)	62.93(5.12)	62.62(5.45)	63.17(5.04)	**63.38(4.29)**	62.48(4.32)	**70.00(4.73)**	62.75(4.97)	62.77(4.06)
wdbc	92.50(3.26)	91.01(3.65)	**91.39(3.43)**	91.10(4.42)	91.36(3.79)	91.18(3.81)	79.61(5.32)	91.18(3.74)	**91.54(4.04)**
Iono	79.57(6.52)	82.39(6.98)	81.77(5.26)	**82.95(5.79)**	82.89(6.08)	82.96(5.31)	70.27(6.32)	83.95(5.93)	**84.37(5.57)**
Average	81.14(4.71)	82.12(4.69)	79.10(4.57)	81.77(4.70)	**82.98(4.50)**	80.33(4.92)	61.94(5.11)	80.46(5.07)	**83.19(4.22)**

Table 22
Average classification accuracy of compared methods on KNN classifier under different noise levels.

$\alpha = 0.8, \beta = 0.2$		Global				Local			
Dataset	RAW	GMLA	GMUA	GNRS	GVPCM	LMLA	LMUA	LNRS	LVPCM
seed	87.30(6.47)	88.57(6.85)	88.25(6.07)	88.25(6.07)	**88.73(5.63)**	88.81(6.57)	88.17(5.97)	88.73(6.42)	**89.52(5.99)**
page	95.81(0.80)	94.91(0.86)	**94.98(0.85)**	88.44(1.29)	94.66(0.83)	94.92(0.94)	90.80(1.19)	**94.93(1.01)**	94.76(1.03)
hcv	90.77(3.63)	90.46(3.71)	90.43(3.40)	89.75(3.54)	**90.80(3.22)**	90.30(3.52)	**90.69(4.11)**	90.29(3.46)	**90.69(3.74)**
Spea	81.21(6.35)	57.14(8.93)	47.72(9.89)	35.71(8.00)	**63.63(8.17)**	56.82(8.39)	33.72(6.81)	56.78(8.55)	**64.39(7.44)**
vowe	48.57(4.74)	51.03(4.66)	43.30(4.57)	**51.65(4.64)**	49.09(4.81)	**51.77(5.28)**	20.12(3.28)	50.94(4.27)	51.08(4.61)
wine	71.15(10.91)	71.66(9.88)	71.68(12.03)	72.24(11.87)	**72.29(9.89)**	72.00(11.15)	67.16(13.12)	72.24(9.04)	**72.34(9.01)**
Clim	90.03(3.82)	89.48(4.02)	89.48(4.28)	89.78(4.07)	**90.28(3.81)**	89.85(3.41)	90.06(3.68)	90.06(3.86)	**90.77(3.03)**
Segm	95.48(1.37)	76.86(2.64)	**78.55(2.44)**	76.99(2.51)	77.11(2.71)	76.85(2.85)	17.23(2.81)	76.78(2.7)	**77.55(2.63)**
Abse	47.50(6.26)	95.16(2.44)	80.63(3.55)	93.20(2.88)	**98.87(1.30)**	90.63(3.20)	15.09(3.88)	90.23(3.11)	**98.69(1.54)**
Sout	62.50(4.16)	62.68(4.51)	62.88(4.42)	63.07(5.53)	**63.08(4.76)**	62.92(4.17)	**68.50(3.49)**	62.35(4.71)	63.38(4.91)
wdbc	91.41(3.91)	91.36(3.40)	91.33(3.88)	91.01(3.66)	**91.45(3.95)**	91.39(3.04)	84.82(4.67)	91.48(3.19)	**91.62(3.53)**
Iono	70.26(9.12)	77.26(6.68)	76.44(7.58)	**76.54(7.50)**	74.88(9.12)	74.73(6.63)	67.91(8.07)	**74.93(7.50)**	74.74(5.69)
Average	77.67(5.13)	78.88(4.88)	76.31(5.25)	76.39(5.13)	**79.57(4.85)**	78.41(4.93)	61.19(5.09)	78.31(4.82)	**79.96(4.43)**

Table 23
Average classification accuracy of compared methods on KNN classifier under different noise levels.

$\alpha = 0.9, \beta = 0.1$		Global				Local			
Dataset	RAW	GMLA	GMUA	GNRS	GVPCM	LMLA	LMUA	LNRS	LVPCM
seed	87.70(7.96)	87.70(6.61)	**87.78(8.07)**	87.54(5.81)	87.70(6.88)	87.94(6.05)	83.25(8.26)	88.02(7.74)	**88.65(7.21)**
page	95.05(0.86)	95.41(1.01)	94.51(0.82)	**95.42(0.79)**	94.58(0.87)	**94.93(1.01)**	92.25(1.05)	94.91(0.82)	94.58(0.87)
hcv	94.14(3.32)	90.49(3.81)	90.32(4.20)	90.72(3.52)	**90.80(3.63)**	90.29(3.46)	90.47(3.47)	90.49(3.94)	**90.58(4.04)**
Spea	80.13(7.07)	**56.52(7.49)**	46.89(8.72)	55.37(8.99)	55.37(6.69)	55.11(7.77)	43.27(9.60)	**56.29(8.22)**	55.35(8.19)
vowe	40.10(4.41)	43.10(4.12)	43.94(5.23)	43.65(5.01)	**76.16(3.73)**	43.28(5.05)	27.73(3.51)	44.09(4.43)	**76.33(3.59)**
wine	70.80(9.74)	72.76(11.21)	**73.87(10.81)**	73.18(10.97)	73.77(10.12)	72.30(9.87)	**74.62(8.46)**	74.05(8.21)	72.99(9.19)
Clim	89.94(3.65)	**90.12(3.68)**	90.28(3.94)	89.17(4.20)	89.94(3.95)	89.97(4.01)	89.63(4.33)	89.97(4.10)	**90.46(3.37)**
Segm	95.48(1.31)	80.80(2.53)	**88.15(2.01)**	80.69(2.66)	80.86(2.44)	80.65(2.12)	30.42(2.49)	80.83(2.64)	**81.03(2.42)**
Abse	47.75(6.27)	55.70(3.81)	55.54(3.31)	93.27(2.91)	**98.83(1.25)**	55.54(3.62)	13.60(4.95)	55.61(4.63)	**98.83(1.58)**
Sout	62.63(3.91)	62.85(4.65)	62.93(4.72)	62.70(4.30)	**63.02(5.12)**	62.20(5.24)	**68.10(4.50)**	63.20(3.71)	62.92(4.70)
wdbc	91.53(3.21)	91.36(3.83)	91.33(3.99)	91.33(3.81)	**91.42(3.37)**	91.42(3.22)	69.10(6.56)	91.33(3.35)	**91.44(4.05)**
Iono	66.61(7.36)	70.74(6.64)	**70.99(8.06)**	70.32(7.17)	69.70(8.04)	**70.70(6.94)**	60.03(7.60)	70.80(6.37)	70.04(7.51)
Average	76.82(4.92)	74.80(4.95)	74.71(5.32)	77.78(5.01)	**81.01(4.67)**	74.53(4.86)	61.87(5.40)	74.96(4.85)	**81.10(4.73)**

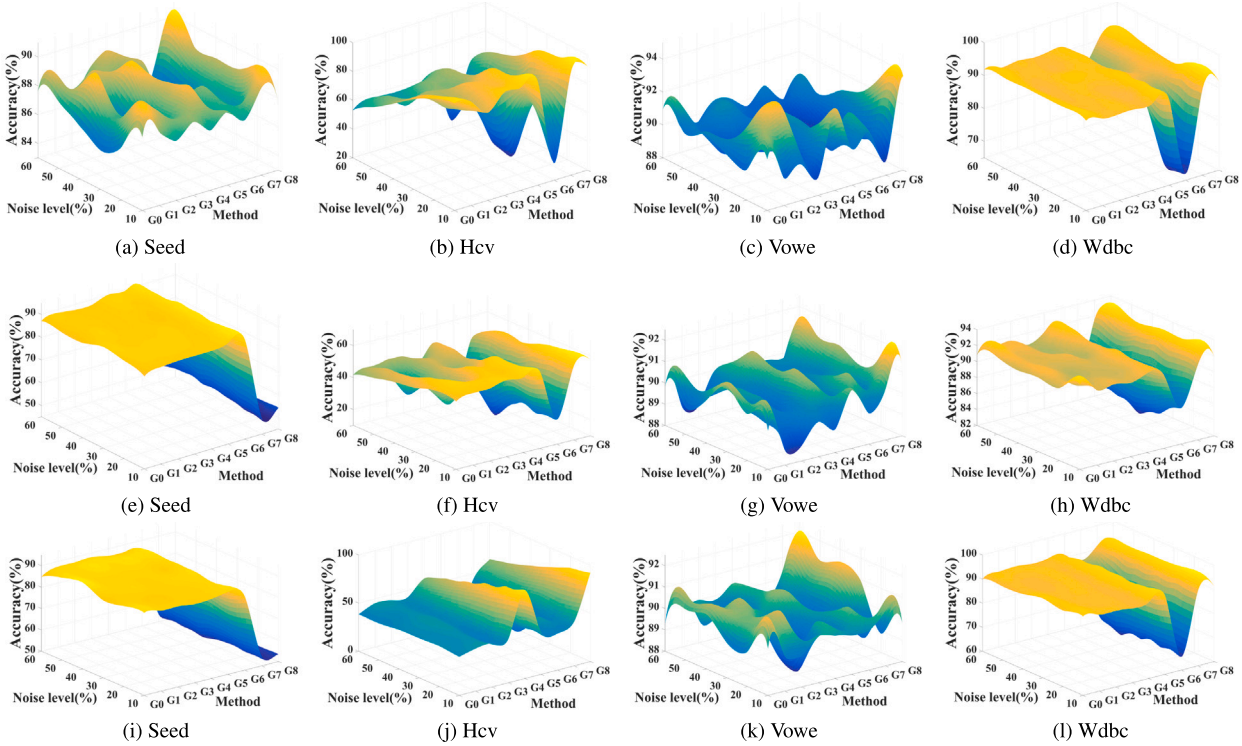


Fig. 8. Classification accuracy of different methods on the KNN classifier when adding noise data. (a)–(d), (e)–(h), and (i)–(l) show the classification results of compared methods when $\alpha = 0.7, \beta = 0.3$, $\alpha = 0.8, \beta = 0.2$, and $\alpha = 0.9, \beta = 0.1$.

show that the proposed GVPCM and LVPCM perform better than other compared methods, illustrating their robustness with noise adding.

6. Conclusions

With the rapid development of information science, data size is increasing, which brings challenges for data mining. However, making decisions based on all attributes may not be wise due to the existence of redundant attributes. Therefore, selecting important attributes is necessary for decision-making datasets. This paper proposes a local rough set method for feature selection based on the designed composite measure. The proposed composite measure considers the inaccurate region in variable precision neighborhood rough set to define a composite measure, combining the information presented at lower and upper approximations, which could comprehensively depict the uncertainty knowledge and provide a conventional tool to the research of decision-making and intelligence information processing. Moreover, the designed local method significantly improves the computational efficiency in the rough set. All the numerical experimental results illustrate that the proposed method is efficient and excellent in feature selection compared with some existing methods. These research results provide a new thought for researching uncertainty measures and feature selection. It is worth noting that the calculation of composite measures combining the lower and upper approximation is time-consuming compared with other indicators based on lower approximation, especially in large-scale datasets. Considering the effectiveness of the granular ball, a more efficient mechanism is necessary to be investigated for approximation computation in further work.

CRediT authorship contribution statement

Kehua Yuan: Conceptualization, Investigation, Methodology, Writing – original draft. **Weihua Xu:** Investigation, Methodology, Writing – review & editing. **Duoqian Miao:** Methodology, Writing – review & editing.

Declaration of competing interest

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

Data availability

No data was used for the research described in the article.

Acknowledgments

This paper was supported in part by the National Key Research and Development Program of China “Key Special Project on Cyberspace Security Governance” under Grant 2022YFB3104700, the National Natural Science Foundation of China (Nos. 62376198, 61976245, 61976158, 62376229.

References

[1] M. Akram, G. Ali, J.C.R. Alcantud, Parameter reduction analysis under interval-valued m-polar fuzzy soft information, Artif. Intell. Rev. 54 (2021) 5541–5582.
[2] W.P. Ding, C.T. Lin, W. Pedrycz, Multiple relevant feature ensemble selection based on multilayer co-evolutionary consensus mapreduce, IEEE Trans. Cybern. 50 (2) (2020) 425–439.
[3] W.P. Ding, W. Pedrycz, I. Triguero, et al., Multigranulation super-trust model for attribute reduction, IEEE Trans. Fuzzy Syst. 29 (6) (2021) 1395–1408.
[4] D.D. Guo, W.H. Xu, Fuzzy-based concept-cognitive learning: An investigation of novel approach to tumor diagnosis analysis, Inform. Sci. 639 (2023) 118998.
[5] W.P. Ding, I. Triguero, C.T. Lin, Coevolutionary fuzzy attribute order reduction with complete attribute-value space tree, IEEE Trans. Emerg. Top. Comput. Intell. 5 (1) (2021) 130–142.
[6] H.C. Peng, F.H. Long, C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min redundancy, IEEE Trans. Pattern Anal. Mach. Intell. 27 (8) (2005) 1226–1238.
[7] W.P. Ding, J. Nayak, B. Naik, et al., Fuzzy and real-coded chemical reaction optimization for intrusion detection in industrial big data environment, IEEE Trans. Ind. Inform. 17 (6) (2021) 4298–4307.
[8] D.D. Guo, C.M. Jiang, R.X. Sheng, et al., A novel outcome evaluation model of three-way decision: A change viewpoint, Inform. Sci. 607 (2022) 1089–1110.
[9] W.H. Xu, D.D. Guo, J. Mi, et al., Two-way concept-cognitive learning via concept movement viewpoint, IEEE Trans. Neural Netw. Learn. Syst. (2023) http://dx.doi.org/10.1109/TNNLS.2023.3235800.
[10] J.C.R. Alcantud, The semantics of N-soft sets, their applications, and a coda about three-way decision, Inform. Sci. 606 (2022) 837–852.
[11] Y.Y. Yao, Three-way conflict analysis: reformulations and extensions of the pawlak model, Knowl.-Based Syst. 180 (2019) 26–37.
[12] S. Luo, D.Q. Miao, Z.F. Zhang, et al., A neighborhood rough set model with nominal metric embedding, Inform. Sci. 520 (2020) 373–388.
[13] Y.Y. Yang, D.G. Chen, H. Wang, et al., Incremental perspective for feature selection based on fuzzy rough sets, IEEE Trans. Fuzzy Syst. 26 (3) (2018) 1257–1273.
[14] E.A.K. Zaman, A. Mohamed, A. Ahmad, Feature selection for online streaming high-dimensional data: A state-of-the-art review, Appl. Soft Comput. 127 (2022) 109355.
[15] W.P. Ding, M.A. Basset, H. Hawash, et al., Explainability of artificial intelligence methods, applications and challenges: A comprehensive survey, Inform. Sci. 615 (2022) 238–292.
[16] C.M. Jiang, D.D. Guo, R.Y. Xu, Measuring the outcome of movement-based three-way decision using proportional utility functions, Appl. Intell. 51 (2021) 8598–8612.
[17] D.D. Guo, C.M. Jiang, P. Wu, Three-way decision based on confidence level change in rough set, Internat. J. Approx. Reason. 143 (2022) 57–77.
[18] Y.Y. Yao, L.Q. Zhao, A measurement theory view on the granularity of partitions, Inform. Sci. 213 (2012) 1–13.
[19] Y.Y. Yao, Relational interpretations of neighborhood operators and rough set approximation operators, Inform. Sci. 111 (1) (1998) 239–259.
[20] J.M. Zhan, W.J. Wang, J.C.R. Alcantud, et al., A three-way decision approach with prospect-regret theory via fuzzy set pair dominance degrees for incomplete information systems, Inform. Sci. 617 (2022) 310–330.
[21] O. Castillo, J.R. Castro, P. Melin, A methodology for building interval type-3 fuzzy systems based on the principle of justifiable granularity, Int. J. Intell. Syst. 37 (10) (2022) 7909–7943.
[22] W.P. Ding, M.A. Basset, H. Hawash, et al., Multimodal infant brain segmentation by fuzzy-informed deep learning, IEEE Trans. Fuzzy Syst. 30 (4) (2022) 1088–1101.
[23] D.D. Guo, W.H. Xu, Y.H. Qian, et al., Fuzzy-granular concept-cognitive learning via three-way decision: performance evaluation on dynamic knowledge discovery, IEEE Trans. Fuzzy Syst. (2023) http://dx.doi.org/10.1109/TFUZZ.2023.3325952.
[24] W.T. Li, S.C. Zhai, W.H. Xu, et al., Feature selection approach based on improved fuzzy C-means with principle of refined justifiable granularity, IEEE Trans. Fuzzy Syst. (2022) http://dx.doi.org/10.1109/TFUZZ.2022.3217377.
[25] B.Z. Sun, W.M. Ma, X.T. Chen, et al., Multigranulation vague rough set over two universes and its application to group decision making, Soft Comput. 23 (2019) 8927–8956.
[26] W.H. Xu, K.H. Yuan, W.T. Li, et al., An emerging fuzzy feature selection method using composite entropy-based uncertainty measure and data distribution, IEEE Trans. Emerg. Top. Comput. Intell. 7 (1) (2023) 76–88.
[27] D.D. Guo, W.H. Xu, Y.H. Qian, et al., M-FCCL: Memory-based concept-cognitive learning for dynamic fuzzy data classification and knowledge fusion, Inf. Fusion 100 (2023) 101962.
[28] W.H. Xu, D.D. Guo, Y.H. Qian, et al., Two-way concept-cognitive learning method: A fuzzy-based progressive learning, IEEE Trans. Fuzzy Syst. (2022) http://dx.doi.org/10.1109/TFUZZ.2022.3216110.
[29] X.Y. Zhang, D.D. Guo, W.H. Xu, Two-way concept-cognitive learning with multi-source fuzzy context, Cogn. Comput. (2023) http://dx.doi.org/10.1007/s12559-023-10107-w.
[30] Z. Pawlak, Rough sets, Int. J. Comput. Inf. Sci. 11 (1982) 341–356.
[31] W.P. Ding, C.T. Lin, Z.H. Cao, Shared nearest-neighbor quantum game-based attribute reduction with hierarchical coevolutionary spark and its application in consistent segmentation of neonatal cerebral cortical surfaces, IEEE Trans. Neural Netw. Learn. Syst. 30 (7) (2019) 2013–2027.
[32] L. Sun, T.Y. Yin, W.P. Ding, et al., Feature selection with missing labels using multilabel fuzzy neighborhood rough sets and maximum relevance minimum redundancy, IEEE Trans. Fuzzy Syst. 30 (5) (2022) 1197–1211.
[33] W.H. Xu, K.H. Yuan, W.T. Li, Dynamic updating approximations of local generalized multigranulation neighborhood rough set, Appl. Intell. 52 (2022) 9148–9173.

[34] D.G. Chen, S.Y. Zhao, Local reduction of decision system with fuzzy rough sets, Fuzzy Sets and Systems 161 (13) (2010) 1871–1883.
[35] W.T. Li, X.P. Xue, W.H. Xu, T. Zhan, B.J. Fan, Double-quantitative variable consistence dominance-based rough set approach, Internat. J. Approx. Reason. 124 (2020) 1–26.
[36] M.A. Sanchez, O. Castillo, J.R. Casto, Information granule formation via the concept of uncertainty-based information with interval type-2 fuzzy sets representation and takagi-sugeno-kang consequents optimized with cuckoo search, Appl. Soft Comput. 27 (2015) 602–609.
[37] X.F. Huang, J.M. Zhan, Z.S. Xu, et al., A prospect-regret theory-based three-way decision model with intuitionistic fuzzy numbers under incomplete multi-scale decision information systems, Expert Syst. Appl. 214 (2023) 119144.
[38] K.H. Yuan, W.H. Xu, W.T. Li, et al., An incremental learning mechanism for object classification based on progressive fuzzy three-way concept, Inform. Sci. 584 (2021) 127–147.
[39] H.R. Ju, W.P. Ding, X.B. Yang, et al., Robust supervised rough granular description model with the principle of justifiable granularity, Appl. Soft Comput. 110 (2021) 107612.
[40] K.Y. Liu, T.R. Li, X. Yang, et al., SemiFREE: semi-supervised feature selection with fuzzy relevance and redundancy, IEEE Trans. Fuzzy Syst. 31 (10) (2023) 3384–3396.
[41] S.T. Wang, Y.Y. Wang, D.J. Wang, et al., An improved random forest-based rule extraction method for breast cancer diagnosis, Appl. Soft Comput. 86 (2022) 105941.
[42] S.Y. Xia, C. Wang, G.Y. Wang, et al., GBRS: An unified model of pawlak rough set and neighborhood rough set, 2022, arXiv:2201.03349.
[43] K.Y. Liu, T.R. Li, X.B. Yang, et al., Feature selection in threes: neighborhood relevancy, redundancy, and granularity interactivity, Appl. Soft Comput. 146 (2023) 110679.
[44] S.Y. Xia, X.C. Dai, G.Y. Wang, et al., An efficient and adaptive granular-ball generation method in classification problem, IEEE Trans. Neural Netw. Learn. Syst. (2022) http://dx.doi.org/10.1109/TNNLS.2022.3203381.
[45] S.Y. Xia, S.L. Wu, X.X. Chen, et al., GRRS: Accurate and efficient neighborhood rough set for feature selection, IEEE Trans. Knowl. Data Eng. 35 (9) (2023) 9281–9294.
[46] Q. Wang, Y.H. Qian, X.Y. Liang, et al., Local neighborhood rough set, Knowl.-Based Syst. 153 (2018) 53–64.
[47] H.B. Jiang, J.M. Zhan, D.G. Chen, Covering-based variable precision(i, t)-fuzzy rough sets with applications to multiattribute decision-making, IEEE Trans. Fuzzy Syst. 27 (8) (2019) 1558–1572.
[48] J.S. Mi, W.Z. Wu, W.X. Zhang, Approaches to knowledge reduction based on variable precision rough set model, Inform. Sci. 159 (3–4) (2004) 255–272.
[49] S.Y. Zhao, E.C.C. Tsang, D.G. Chen, The model of fuzzy variable precision rough sets, IEEE Trans. Fuzzy Syst. 17 (2) (2009) 451–467.
[50] C.H. Liu, B.W. Lin, J.Y. Lai, et al., An improved decision tree algorithm based on variable precision neighborhood similarity, Inform. Sci. 615 (2022) 152–166.
[51] B.Z. Sun, W.M. Ma, X.T. Chen, Variable precision multigranulation rough fuzzy set approach to multiple attribute group decision-making based on λ–similarity relation, Comput. Ind. Eng. 127 (2019) 326–343.
[52] B.Z. Sun, C. Qi, W.M. Ma, et al., Variable precision diversified attribute multigranulation fuzzy rough set-based multi-attribute group decision making problems, Comput. Ind. Eng. 142 (2020) 106331.
[53] C.Y. Wan, L.J. Wan, New results on granular variable precision fuzzy rough sets based on fuzzy (co)implications, Fuzzy Sets and Systems 423 (2021) 149–169.
[54] D.D. Zou, X.L. Xu, L.Q. Li, et al., Novel variable precision fuzzy rough sets and three-way decision model with three strategies, Inform. Sci. 629 (2023) 222–248.
[55] Y.H. Qian, X.Y. Liang, Q. Wang, et al., Local rough set: a solution to rough data analysis in big data, Internat. J. Approx. Reason. 97 (2018) 38–63.
[56] S.Y. Xia, H. Zhang, W.H. Li, et al., GBNRS: A novel rough set algorithm for fast adaptive attribute reduction in classification, IEEE Trans. Knowl. Data Eng. 34 (3) (2022) 1231–1242.
[57] J.H. Dai, H. Hu, W.Z. Wu, Others maximal-discernibility-pair-based approach to attribute reduction in fuzzy rough sets, IEEE Trans. Fuzzy Syst. 26 (4) (2017) 2174–2187.
[58] R. Jensen, Q. Shen, Fuzzy-rough attributes reduction with application to web categorization, Fuzzy Sets and Systems 141 (2004) 469–485.
[59] C.Z. Wang, Y. Huang, M.W. Shao, et al., Fuzzy rough set-based attribute reduction using distance measures, Knowl.-Based Syst. 164 (2019) 205–212.
[60] Y.Y. Chen, Y.M. Chen, Feature subset selection based on variable precision neighborhood rough sets, Int. J. Comput. Intell. Syst. 14 (1) (2021) 572–581.
[61] Q.H. Hu, D.R. Yu, J.F. Liu, et al., Neighborhood rough set based heterogeneous feature subset selection, Inform. Sci. 178 (18) (2008) 3577–3594.