# A Feature Discretization Method Based on Fuzzy Rough Sets for High-resolution Remote Sensing Big Data Under Linear Spectral Model

Qiong Chen, Mengxing Huang, *Member, IEEE,* Hao Wang, *Senior Member, IEEE,* and Guangquan Xu, *Member, IEEE*

***Abstract*—As one of the most relevant data preprocessing techniques, discretization has played an important role in data mining, which is widely applied in industrial control. It can transform continuous features to discrete ones, thus improving the efficiency of data processing and adapting to learning algorithms that require discrete data as inputs. However, traditional discretization methods have shortcomings, such as highly complex programs, excessive numbers of intervals obtained, and significant loss of necessary information in the preprocessing of high-resolution remote sensing big data. Moreover, the large number of mixed pixels in the image is a primary reason for the uncertainty of remote sensing information systems, and current discretization methods are based on the assumption that one pixel only corresponds to the spectral information of a single object, without considering the influence of the uncertainty caused by a mixed spectrum, which causes the classification accuracy to drop after discretization. We propose a discretization method for high-resolution remote sensing big data. We determine the membership degree of each pixel in training samples through linear decomposition, and establish the individual fitness function based on a fuzzy rough model. An adaptive genetic algorithm selects discrete breakpoints, and a MapReduce framework calculates the individual fitness of the population in parallel, to obtain the optimal discretization scheme in the minimum time. Our method is compared to the best state-of-the-art discretization algorithms on the authentic remote sensing datasets. Experiments verified the effectiveness of the proposed method, which provides strong support for the subsequent processing of images.**

***Index Terms*—Adaptive genetic algorithm, Discretization, Fuzzy rough model, MapReduce framework, Mixed pixels.**

## I. Introduction

CLASSIFICATION is always a concern in remote sensing image processing [1]. Although high-resolution remote sensing data can provide more detailed spectrally continuous spatial information about the land surface than low- and medium-resolution data [2], this rich information brings huge space-time overhead and greatly increases the complexity of data analysis [3]. In addition, the use of all possible spectral information in the classification process of high-resolution remote sensing images will lead to information redundancy and weaken the generalization ability of the learning model, thus significantly reducing the overall accuracy [4]. Discretization plays an important role in data mining as used in industrial control [5]. It can transform continuous features in spectral information to discrete ones that are closer to the knowledge layer representation, reducing system overhead, removing redundant information, and enhancing the robustness of learning algorithms [6]. Therefore, feature discretization is a key step for the success of high-resolution remote sensing image processing [7]. However, in the feature discretization of high-resolution remote sensing image, due to the large number of different values on each band, the complex correlations among multiple bands, and the unknown distribution of pixel categories, making the selection of breakpoints extremely difficult [7]. On the other hand, due to the addition of more detailed information in high-resolution remote sensing images, gaps and shadows between objects that are neglected and weakened in low- and medium-resolution images are enhanced, boundaries are weakened, and intra-class heterogeneity is enhanced, which greatly increases the proportion of mixed pixels [8]. The existence of a large number of mixed pixels makes high-resolution remote sensing data highly complex and uncertain, which greatly increases the difficulty of feature discretization [9].

Genetic algorithm is a mature global optimization algorithm, which can achieve good results on many complex remote sensing processing problems [10], [11]. A genetic algorithm was used for parameter optimization and feature selection of support vector machines (SVM), which improved the classification accuracy of RADARSAT-2 (RS2) images and Thaichote (THEOS) images [10]. The genetic algorithms have been applied for simultaneous refinement of multiple disparity maps, which effectively solved the problem of dense stereo image matching in remote sensing images [11]. On the other hand, fuzzy sets and rough sets, as mathematical tools to solve uncertainty problems, are also widely used in the analysis of remote sensing data [12], [13]. Deep learning technology and fuzzy clustering method were combined to deal with the uncertainty of high-resolution remote sensing data, which efficiently achieved the broad-area thematic and contextual understanding of the geographic land cover [12]. Rough set

This work was supported by Hainan Provincial Natural Science Foundation of China under Grant 2019CXTD400, and the National Key Research and Development Program of China under Grant 2018YFB1404400. (Corresponding authors: Mengxing Huang and Hao Wang.)

Q. Chen and M. Huang are with the State Key Laboratory of Marine Resource Utilization in South China Sea and School of Information and Communication Engineering, Hainan University, Haikou 570228, China (e-mail: 13907534385@163.com; huangmx09@163.com).

H. Wang is with the Department of Computer Science, Norwegian University of Science and Technology, Norway (e-mail: hawa@ntnu.no).

G. Xu is with the Tianjin Key Laboratory of Advanced Networking (TANK), College of Intelligence and Computing, Tianjin University, Tianjin China; Qingdao Huanghai University, No. 1145, Linghai Road, Huangdao District, Qingdao City, Shandong Province, China (e-mail: losin@tju.edu.cn).

theory was used to encode domain knowledge to reform the structure of granular neural network, which enhanced the performance of progressive granular neural network (PGNN) model for the classification of remote sensing images [13]. The fuzzy rough set is considered to be a more powerful uncertainty analysis model for big data than either fuzzy or rough sets [14]–[19]. This model introduces a fuzzy set on the basis of a rough set, and describes the relationship between the pixels by a similarity relation instead of the equivalent relation of a rough set. Thus, it deals more flexibly with uncertain information introduced by mixed pixels. In our previous research work [20], we used genetic algorithm to optimize the multidimensional data discretization scheme, and determined the crossover segments and mutation points of the discretization scheme to be optimized through Q-learning mechanism, and achieved good discretization results. However, this method uses conventional genetic operations on other individuals in the population, so it is difficult to obtain high-quality individuals to cross with the discretization scheme to be optimized, which reduces search efficiency. In addition, the fitness function of this method is based on rough set model, which is difficult to describe the fuzziness caused by mixed pixels. Therefore, by establishing fitness function based on fuzzy rough model in adaptive genetic algorithm to search discrete breakpoints, more accurate classification results can be obtained.

Discretization can be considered supervised or unsupervised depending on whether the data contain category information [5]. Supervised discretization makes full use of labels and target attribute information because the calculations are based on class information [21]. Therefore, it is easier to find the appropriate locations of breakpoints than in unsupervised discretization [22]–[24]. The method of discretization based on information entropy was used to rank label data, which improved the sensitivity of homogeneity of sample ranking in the set [25]. However, using this standard, only one attribute at a time can be partitioned separately in multi-feature data, and the compatibility of the information system after discretization cannot be guaranteed. Particle swarm optimization (PSO) was used for discretization-based feature selection, which could generate a stronger and more compact representation in high-dimensional datasets to achieve better classification performance [26]. However, this representation is static, and the number of possible solutions increases exponentially with the size of the feature set, making the search space larger. The method can only select one discrete breakpoint at a time for each feature, and its performance on data requiring discretization into multiple intervals has not been verified. A discretization method was proposed to classify remote sensing image features [27]. The method first defines the uncertainty of the decision system based on the equivalence class model of the rough set. Using an information entropy criterion, it then selects breakpoints by controlling the change of the uncertainty under a given threshold. Since this method considers the intrinsic relationship between multiple attributes during discretization, it achieved a high classification accuracy on SPOT-5 (Systeme Probatoire d'Observation de la Terre-5) images. However, it only considers the stability of the interval, and not that of adjacent intervals, so it cannot effectively filter the noise interval, and it will produce more intervals. A discretization method based on Cramers V-test (CVD) has been applied for feature selection during remote sensing image classification [28]. Discrete features generated by this method on QuickBird and PHI (Pushbroom Hyperspectral Imagery) images could improve the performance of J48 decision tree and naive Bayes classifiers. However, as in the method reported in Ref [25], the intrinsic correlation between multiple attributes was not considered, and it was difficult to ensure the compatibility of the system after discretization.

The above methods assume that only one pixel corresponds to the spectral information of a single object. In practice, the ground unit covered by one pixel usually contains more than one object type. Under linear spectral model, the spectral value of each pixel is derived from the linear combination of the reflectance of all objects within the pixel and the coefficient weighted by their proportions of the pixel area [9]. In remote sensing image feature discretization, neglecting the influence of the uncertainty caused by the mixed spectrum will cause errors in the data and reduce the classification accuracy after discretization [29]. With the advantage of rough set in dealing with incomplete information, the discretization method of combining rough set and genetic algorithm [20] can make full use of the powerful search ability and the known knowledge base without any prior information to obtain the minimum number of breakpoints while ensuring that the compatibility of the system is not destroyed. However, the equivalent class of rough set is an ordinary set, which is difficult to describe the fuzzy components in the data, and the accuracy obtained in the discretization of mixed spectrum is low. In addition, using genetic algorithm to determine the discrete breakpoints has the following challenges: (1) due to the large number of different values on each band, it is difficult to generate high-quality initial population which directly affects the performance of the algorithm; (2) the time cost increases sharply with the expansion of population size; (3) the convergence speed is slow in the processing of big data, and the unreasonable parameter setting is easy to make the search space fall into local optimum.

In view of the above problems, we propose a discretization method called the fuzzy-rough-set-based genetic algorithm (FRSGA) for high-resolution satellite remote sensing of big data. Our innovation points to the article are as follows: (1) different from the current discretization methods, which are based on the assumption that one pixel only corresponds to the spectral information of a single object, the influence of uncertainty caused by a mixed spectrum is fully considered in our method, each category of ground object in high-resolution remote sensing image is regarded as a fuzzy set, and the Euclidean distance between pixels is calculated according to the pixel values on each band to obtain the fuzzy relation matrix on the set of pixels; (2) based on the rough set model in the original work, the membership degree of pixels to each category is determined through linear decomposition, and the individual fitness function based on a fuzzy rough model [19] is established; (3) in contrast to the conventional genetic operations in the previous work [20], chaotic mapping is

employed to generate the initial population of adaptive genetic algorithm [30] to select discrete breakpoints, and MapReduce [31] is used to calculate the fitness of all individuals of the population in parallel, hence obtaining the optimal discretization scheme in the minimum time. We compared our method with mainstream discretization methods on the authentic remote sensing datasets. The results showed that our method can effectively mine the data association, reduce the response time of data access, improve the accuracy of classification rules, and provide strong support for subsequent image processing.

The rest of this article is arranged as follows. Section II introduces basic concepts and reviews related work. Section III describes the discretization problem of mixed pixels. Section IV elaborates the discretization algorithm based on a fuzzy rough model applied to high-resolution remote sensing images. The experimental environment and datasets are introduced in Section V. We analyze and discuss the results in Section VI. Section VII summarizes the paper and presents prospects for future research work. The specific implementation of the improved adaptive genetic algorithm and the parallel calculation of the individual fitness of the population in the MapReduce framework provided by MATLAB are detailed in the Appendix.

## II. Preliminaries

We introduce the basic concepts of the decision table and rough set, describe the fuzzy set and fuzzy relation, define the fuzzy rough model, and explain related concepts in this section.

### A. Decision Table and Rough Set

A decision table can be described as a 4-tuple $S = (U, R, V, f)$, in which $U = \{x_1, x_2, \ldots, x_n\}$ is a finite set of objects, i.e., a universe. $R = C \cup D$ refers to the set of attributes, where $C$ and $D$ are nonempty subsets of the condition and decision attributes, respectively, and $C \cap D = \varnothing$. $V = \bigcup_{a \in R} V_a$, where $V_a$ is the range of values of attribute $a$. $f : U \times R \to V$ is an information function that specifies a unique value $f(x, a) \in V_a$ for each object $x \in U$ on each attribute $a \in R$.

The decision table in the classification of a remote sensing image was introduced previously [6], and the elements in $U$ are the pixels of the image. The bands and land cover classes correspond respectively to condition attributes in $C$ and the decision attribute in $D$. $V_{C_i}$ refers to the range of pixel values of the $i$th band, and $V_d$ is the scope of the land cover class.

A rough set is based on a classification mechanism. It interprets a classification problem as an indiscernible relation that is a division of the feature space. Each division corresponds to knowledge in the universe. The idea of rough set theory [32] is to use the known knowledge base to approximate imprecise or uncertain knowledge. The concepts of the equivalent class, lower approximation set, and upper approximation set [33] are needed to describe the uncertainty.

Given a decision table $S = (U, R, V, f)$, for each attribute subset $A \subseteq R$, the indiscernible binary relation $IND(A)$ and equivalent classes for attribute subset $A$ in $U$ are defined as follows:

$$IND(A) = \left\{(x, y) \mid (x, y) \in U^2, \forall a \in A(a(x) = a(y))\right\} \tag{1}$$

$$U \mid IND(A) = \{X \mid X \subseteq U \wedge \forall x \in X \forall y \in X \forall a \in A(a(x) = a(y))\} \tag{2}$$

According to the above decision table $S$, for each subset $X \subseteq U$ and the equivalent classes for attribute subset $A$ in $U$, the lower and upper approximation sets of $X$ are defined as follows:

$$A_{-}(X) = \cup\{Y | Y \in U | IND(A) \wedge Y \subseteq X\} \tag{3}$$

$$A^{-}(X) = \cup\{Y | Y \in U | IND(A) \wedge Y \cap X \neq \varnothing\} \tag{4}$$

### B. Fuzzy Set and Fuzzy Relation

Fuzzy sets are used to describe fuzzy phenomena that are difficult to measure precisely because of the lack of strict boundaries. In the theory proposed by Zadeh [34], fuzzy sets can express both fuzzy and clear concepts. The definition of a fuzzy set is as follows:

$$\forall A \in F(U), A \leftrightarrow \mu_A(x) : \\ U \to [0, 1], A(x) = \begin{cases} 1, & x \in A \\ \mu_A(x), & 0 < \mu_A(x) < 1 \\ 0, & x \notin A \end{cases} \tag{5}$$

Let $A$ be a fuzzy subset determined by the mapping function $\mu_A$ on $U$. $A(x)$ denotes the degree to which element $x$ in $U$ belongs to $A$. $F(U)$ is the set of all fuzzy subsets on $U$, which is called a fuzzy power set. The complementary set, union, intersection, and inclusion operations between fuzzy subsets $A$ and $B$ in $F(U)$ are specified as follows:

$$(-A)(x) = 1 - A(x), \tag{6}$$

$$(A \cup B)(x) = \max\{A(x), B(x)\}, \tag{7}$$

$$(A \cap B)(x) = \min\{A(x), B(x)\}, \tag{8}$$

$$A \subseteq B \Leftrightarrow A(x) \leq B(x). \tag{9}$$

A binary fuzzy relation $\tilde{R}$ from set $X$ to set $Y$ is a fuzzy set on $U$, denoted as $X \xrightarrow{\tilde{R}} Y$. The membership function $\tilde{R}(x, y)$ is a mapping from $X \times Y$ to the real number interval [0,1], indicating the degree of relationship $\tilde{R}$ between $x$ and $y$. Let $\tilde{R}$ be a fuzzy relation from $X$ to $Y$, denoted as $r_{ij} = \tilde{R}(x_i, y_j)$, where $X = \{x_1, x_2, \ldots, x_m\}$, $Y = \{y_1, y_2, \ldots, y_n\}$, $i = 1, 2, \ldots, m$, and $j = 1, 2, \ldots, n$. If $r_{ij} \in [0, 1]$ is satisfied for any $i$ and $j$, then $\tilde{R} = [r_{ij}]_{m \times n}$ is called a fuzzy matrix,

$$\tilde{R}_{(m \times n)} = \begin{bmatrix} r_{11} & r_{12} & \cdot & \cdot & \cdot & r_{1n} \\ r_{21} & r_{22} & \cdot & \cdot & \cdot & r_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{m1} & r_{m2} & \cdot & \cdot & \cdot & r_{mn} \end{bmatrix}. \tag{10}$$

### C. Fuzzy Rough Set

The theory of the fuzzy rough set was proposed by Dubois and Prade to deal with data inconsistencies [35], [36]. They introduced the theory of fuzzy sets into rough sets to solve the problem of information loss caused by discretization of continuous attributes. After many years, the fuzzy rough set has a relatively complete theoretical framework [37]–[39], and has seen wide use in machine learning and data mining [40], [41].

Let $(U, \tilde{R})$ be a fuzzy approximation space and $\tilde{R}$ be a binary fuzzy relation on $U$. The lower and upper approximation sets of the fuzzy set $B \in F(U)$ are still fuzzy sets in $(U, \tilde{R})$. $\forall y \in U$, and the membership functions of the lower and upper approximation sets of $B$ are specified as follows:

$$\underline{B}(y) = \inf_{x \in U} \{(1 - \tilde{R}(x, y)) \vee B(x)\}, \tag{11}$$

$$\overline{B}(y) = \sup_{x \in U} \{\tilde{R}(x, y) \wedge B(x)\}. \tag{12}$$

If $\forall s \in U$ and $\tilde{R}(s, s) = 1$, then the lower and upper approximation sets of $B$ have the property

$$\underline{B} \subseteq B \subseteq \overline{B}. \tag{13}$$

The cardinality of the fuzzy set $B$ is:

$$|B| = \sum_{y \in U} B(y). \tag{14}$$

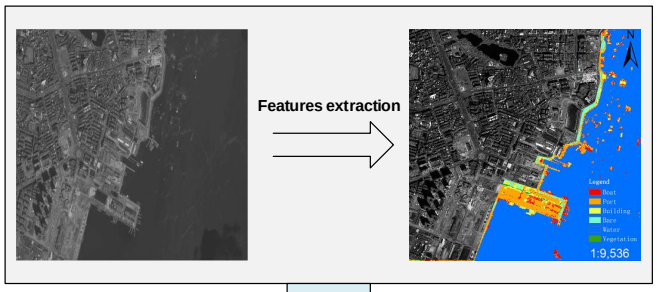
Hence its approximate precision is:

$$\eta(B) = \frac{|\underline{B}|}{|\overline{B}|}, \tag{15}$$

where $|\underline{B}|$ and $|\overline{B}|$ are the respective cardinalities of $\underline{B}$ and $\overline{B}$. The approximate precision is used to measure the overall accuracy of describing fuzzy set $B$ from known knowledge. The closer $\eta(B)$ is to 1 the higher the overall approximate precision is. If $\eta(B) = 1$, then $B$ can be precisely defined by known knowledge. In the feature discretization of high-resolution remote sensing image, the existence of a large number of mixed pixels makes the collected massive remote sensing data highly complex and uncertain. Fuzzy rough set has the advantages of both fuzzy set and rough set. It can make full use of the known knowledge base without any prior information, and use membership function to fuzzify the equivalence relationship to describe the fuzziness and uncertainty between and within the mixed pixels, so as to improve the accuracy of discrete breakpoints on continuous spectral features.

## III. Problem Statement

We formally define the problem of feature discretization of a high-resolution remote sensing image and introduce the concept and decomposition models of mixed pixels. On this basis, we analyze the influence of mixed pixels on the discretization results through a simple example.




| PN | X-axis | Y-axis | Band 1 | Band 2 | Band 3 | Band 4 |
|---|---|---|---|---|---|---|
| 1 | 2244 | 644 | 0.03193 | 0.042641 | 0.048458 | 0.152716 |
| 2 | 2244 | 645 | 0.031914 | 0.043818 | 0.050262 | 0.154377 |
| 3 | 2244 | 646 | 0.032913 | 0.042615 | 0.049649 | 0.158527 |
| 4 | 2243 | 647 | 0.034912 | 0.046159 | 0.053861 | 0.1469 |
| 5 | 2243 | 648 | 0.036924 | 0.050306 | 0.060491 | 0.142741 |
| 6 | 2243 | 649 | 0.038933 | 0.052667 | 0.060482 | 0.179236 |
| 7 | 2243 | 650 | 0.040436 | 0.056213 | 0.06409 | 0.188325 |
| . | . | . | . | . | . | . |
| . | . | . | . | . | . | . |
| . | . | . | . | . | . | . |

Fig. 1. Results of feature extraction from high-resolution remote sensing image.

### A. Definition of Feature Discretization

Before feature discretization, it is necessary to extract the features of the original images [42] to obtain the corresponding digital number (DN) of the pixels in each band. As shown in Fig. 1, these DN values are a set of continuous values from 0 to 1. To discretize the image features is to represent the continuous DN values in a band by as few discrete intervals as possible while not destroying the compatibility of the information system.

Consider the classification problem of a remote sensing image, including $k$ types of land cover and $m$ bands. The discretization algorithm divides the continuous DN values on the $i$th band into $n_i$ discrete and disjoint intervals,

$$\begin{aligned} D_i = \{&[d_0, d_1], (d_1, d_2], \ldots, (d_{n-1}, d_{n_i}]\} \\ &\text{where } \min(DN(i)) = d_0 \leq d_1 < d_2 < \ldots \\ &< d_{n_1 - 1} < d_{n_i} = \max(DN(i)), i \in \{1, 2, \ldots, m\} \end{aligned}, \tag{16}$$

$D_i$ is called a division on the $i$th band, and each interval is represented by a breakpoint (discrete value). $BP = [D_1, D_2, \ldots, D_m]$ is a discretization scheme obtained by the discretization algorithm on a remote sensing image. Therefore, the search space for feature discretization is formed by all candidate breakpoints in each band, which are from all the different DN values in each band of the training set [6].

### B. Decomposition of Mixed Pixels

The pixel is the basic unit that reflects the features of a high-resolution remote sensing image [43]. The phenomenon of similar ground objects being in different states or the coexistence of multiple heterogeneous ground objects generally occurs at the pixel scale, which affects the recognition accuracy of ground objects and restricts the development of quantitative high-resolution remote sensing [8]. A component

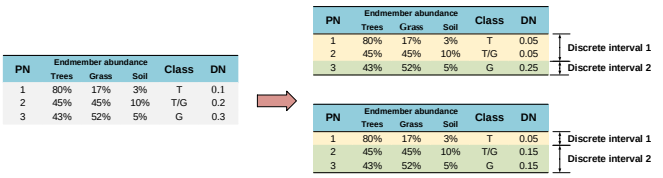



Fig. 2. A simple example of discretization of three mixed pixels. Left: original information table; Top right: the first discretization scheme; Bottom right: the second discretization scheme.

of the spectral signal of a mixed pixel is called an endmember. The percentage of the area of the endmember in the mixed pixel is called its abundance, and this can be obtained from the mixed pixel by decomposition [44]–[46].

Decomposition methods of mixed pixels include linear spectral, probability, random geometric, geometric optical, and fuzzy models [47]. The linear spectral model is simple and intuitive, and it is suitable for most applications [48]. In this model, the radiance or reflectivity of each mixed pixel is regarded as a linear combination of the internal endmembers and their proportions of the area, as follows:

$$R_b = \sum_{i=1}^{N} F_i R_{i,b} + E_b, b = \{1, \ldots, M\}, \tag{17}$$

where $R_b$ is the reflectivity of the mixed pixel on band $b$, $N$ is the number of extracted endmembers, $M$ is the number of bands, and $F_i$ is the area proportion of the endmember of type $i$ in the mixed pixel, $R_{i,b}$ is the reflectivity of the endmember of type $i$ on band $b$, and $E_b$ is the fitting residual of band $b$.

### C. Discretization Problem of Mixed Pixels

The pixel is usually regarded as the basic unit when dividing the continuous features of a remote sensing image. A pixel is treated as a single nominal entity, i.e., the category of the pixel is determined and unique [27], [28]. However, each pixel contains a variety of objects. The endmember, as the smallest unit of a pixel, represents a single feature of information. The reflectivity and abundance of the endmembers in each spectral band determine the DN value of a pixel [9], [46]. Therefore, methods that regard a pixel as a single nominal entity are challenged by fuzzy logic and fuzzy datasets that consider blur, haziness, and uncertainty in image processing [49].

Fig. 2 shows an example of the discretization of three mixed pixels. Assuming that under linear spectral model, these pixels contain only the terrain information of trees (T), grass (G), and soil (S), the abundance of each endmember satisfies $\mathrm{T} + \mathrm{G} + \mathrm{S} = 1$, and the candidate breakpoint set is $\{0.05,0.15,0.25,0.35\}$. The abundances of these three endmembers in pixel 1 are 80%, 17%, and 3%, respectively, represented by vector $\mathrm{PN}(1) = [0.8, 0.17, 0.03]$, so its label shows type T with the largest proportion, and $\mathrm{DN}(1) = 0.1$. Similarly, $\mathrm{PN}(2) = [0.45, 0.45, 0.1]$, $\mathrm{PN}(3) = [0.43, 0.52, 0.05]$, because of the same proportions of trees and grass, the label of pixel 2 is uncertain, and the label of pixel 3 is G, $\mathrm{DN}(2) = 0.2$, $\mathrm{DN}(3) = 0.3$. $|\mathrm{PN(i)} - \mathrm{PN(j)}|$ is the Euclidean distance between PN(i) and PN(j). We should divide the continuous features of the three pixels into two discrete intervals. There are two feasible schemes. Under the first, the system regards pixel 2 as the T type and chooses $\{0.05,0.25,0.35\}$ as the discrete breakpoint set. $0.05 < \mathrm{DN}(1) < \mathrm{DN}(2) < 0.25$, so pixel 1 and pixel 2 of the same type T are placed in the same interval of value 0.05. However, we can see that the abundance of each endmember of $\mathrm{PN}(1)$ is quite different from that of $\mathrm{PN}(2)$, and the abundance of each endmember of $\mathrm{PN}(2)$ is closer to that of $\mathrm{PN}(3)(|\mathrm{PN}(2) - \mathrm{PN}(3)| < |\mathrm{PN}(1) - \mathrm{PN}(2)|)$. When the label type of pixel 2 is uncertain, to classify it with pixel 1 in the same interval will inevitably cause data errors. In the second scheme, $\{0.05,0.15,0.35\}$ is chosen as the discrete breakpoint set. Since $0.15 < \mathrm{DN}(2) < \mathrm{DN}(3) < 0.35$, pixel 2 and pixel 3 are placed in the same interval, with a value of 0.15. This division is more reasonable.

## IV. Discretization Algorithm Based on Fuzzy Rough Model

We first construct the individual fitness function based on the fuzzy rough model. We then analyze the models feasibility and describe the binary coding scheme of the genetic algorithm. Finally, we explain the flow of the algorithm.

### A. Fitness Function Based on Fuzzy Rough Model

As mentioned above, failure to consider the impact of the uncertainty caused by the mixed spectrum will significantly increase the data error rate. Blur, haziness, and uncertainty between and within mixed pixels in the discretization can be described by a fuzzy rough model. Based on the decision table $S = (U, R, V, f)$, as applied to an image in Section II-A, we introduce the distance matrix in the pixel space,

$$D_{(n\times n)} = \begin{bmatrix} d_{11} & d_{12} & . & . & . & d_{1n} \\ d_{21} & d_{22} & . & . & . & d_{2n} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ d_{n1} & d_{n2} & . & . & . & d_{nn} \end{bmatrix}, \tag{18}$$

where $n$ is the number of pixels, and $d_{ij}$ is the Euclidean distance between the $i$th and $j$th pixels, $1 \le i, j \le n$, $d_{ij}$ is defined as

$$d_{ij} = \sqrt{\sum_{s=1}^{m} \left(DN_i^s - DN_j^s\right)^2}, \tag{19}$$

where $m$ is the number of bands, and $DN_i^s$ and $DN_j^s$ respectively represent the DN values of the $i$th and $j$th pixels in the $s$th band. The element $r_{ij}$ in the fuzzy relationship $\tilde{R}$ represents the degree of similarity between the $i$th and $j$th pixel,

$$r_{ij} = \begin{cases} 1 - \frac{d_{ij}}{\max(D)}, & \max(D) > 0 \\ 1, & \max(D) = 0 \end{cases}, \tag{20}$$

where $\max(D)$ is the maximum Euclidean distance in matrix $D$. Suppose the image contains $k$ types of ground objects. For any mixed pixel $x \in U$, let $p_1(x), \ldots, p_k(x)$ be the abundance of each endmember in the mixed pixel $x$, i.e., the proportion of $k$ types of ground objects in the mixed pixel $x$, where $0 \le p_1(x), \ldots, p_k(x) \le 1$ and $\sum_{s=1}^{k} p_s(x) = 1$. $C_t$ is defined

as the fuzzy set corresponding to the $t$th type of land cover, $C_t(x) = p_t(x)$, $t \in [1, \ldots, k]$. Since the principle of selecting the optimal breakpoint set is to obtain the minimum number of breakpoints under the premise of satisfying the data accuracy within a given range, the fitness function of chromosome $c$ should be determined by the obtained number of breakpoints and the average accuracy of the fuzzy sets corresponding to all land cover types,

$$\text{Fitness } (c) = u \times \left(1 - \frac{N_c}{N_I}\right) + (1-u) \times \frac{1}{k} \times \sum_{t=1}^{k} \frac{|\underline{C}_t|}{|\overline{C}_t|}, \quad (21)$$

where $N_I$ is the number of candidate breakpoints, and $N_c$ is the number of breakpoints obtained by decoding chromosome $c$. $\underline{C}_t$ and $\overline{C}_t$ are respectively the lower and upper approximation sets of fuzzy set $C_t$ corresponding to the $t$th type of land cover, and $u$ is the weight coefficient, $0 \leq u \leq 1$. The selection of the weight coefficient is an open problem, as no specific weight coefficient can adapt to all datasets. The rationality of the weight selection is usually judged by the characteristics of the dataset and experimental observations [50]. The purpose of this paper is to improve the classification accuracy of data after discretization, and classification accuracy is directly related to data accuracy. Therefore, we set the weight of $\frac{1}{k} \times \sum_{t=1}^{k} \frac{|\underline{C}_t|}{|\overline{C}_t|}$ to be larger than that of $\left(1 - \frac{N_c}{N_I}\right)$, set $u = 0.1$, and verified it using the experimental results.

We use this fitness function to verify the results of the two discretization schemes in Fig. 2. Suppose $U = \{x_1, x_2, x_3\}$ represents the pixel set composed of pixels 1, 2, and 3, and $C_T$, $C_G$, and $C_S$ represent the fuzzy sets corresponding to types T, G, and S, respectively. $D_1$, $\tilde{R}_1$, and $\eta_1$ respectively represent the distance matrix, fuzzy relation matrix, and average approximation precision of all the fuzzy sets in the first discretization scheme, and $D_2$, $\tilde{R}_2$, and $\eta_2$ are the corresponding variables in the second discretization scheme. For any pixel $x \in U$, the membership function values in fuzzy sets $C_T$, $C_G$, and $C_S$ correspond to the abundance of internal endmembers of types T, G, and S, respectively. Since these two schemes have the same number of intervals, we can judge the results by comparing $\eta_1$ and $\eta_2$. From (19) and (20), $D_1 = \begin{bmatrix} 0 & 0 & 0.2 \\ 0 & 0 & 0.2 \\ 0.2 & 0.2 & 0 \end{bmatrix}$, $D_2 = \begin{bmatrix} 0 & 0.1 & 0.1 \\ 0.1 & 0 & 0 \\ 0.1 & 0 & 0 \end{bmatrix}$, $\tilde{R}_1 = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, and $\tilde{R}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$. From (11) and (14), $\underline{C}_T(x_1) = 0.45$, $\underline{C}_T(x_2) = 0.45$, $\underline{C}_T(x_3) = 0.43$; thus $|\underline{C}_T| = 1.33$. Similarly, $|\underline{C}_G| = 0.79$ and $|\underline{C}_S| = 0.11$. From (12) and (14), $\overline{C}_T(x_1) = 0.8$, $\overline{C}_T(x_2) = 0.8$, $\overline{C}_T(x_3) = 0.43$; thus $|\overline{C}_T| = 2.03$. Similarly, $|\overline{C}_G| = 1.42$ and $|\overline{C}_S| = 0.25$. Therefore, from (15), $\eta_1 \approx 0.5505$ (reserving four decimal places) can be found. In the same way, $\eta_2 \approx 0.8087$. $\eta_2 > \eta_1$, i.e., the average approximation precision of all the fuzzy sets obtained by the second discretization scheme was higher than that of the first scheme, which is consistent with the conclusion of Section III-C. Furthermore, in remote sensing image feature discretization, due to a large number of mixed pixels, even if the numbers of intervals of the discretization schemes are the same, the obtained data accuracy will differ greatly.

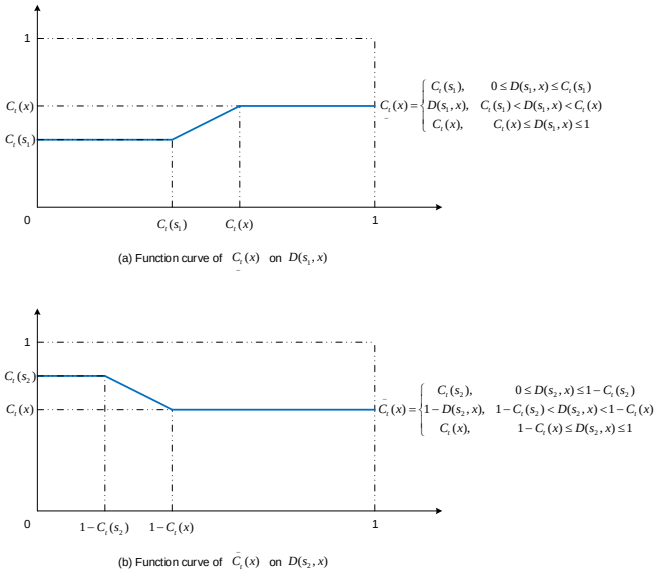



Fig. 3. Function curves of $\underline{C}_t(x)$ and $\overline{C}_t(x)$ on $D(s_1, x)$ and $D(s_2, x)$.

### B. Feasibility Analysis and Genetic Coding

The discretization problem can be seen as a classification of the data itself [5]. Its purpose is to categorize data of the same type into the same interval as much as possible. A greater distance between the different types of data is better. Since $\underline{C}_t \subseteq C_t \subseteq \overline{C}_t$, $|\underline{C}_t| \leq |C_t| \leq |\overline{C}_t|$. Because the approximate precision of fuzzy set $C_t$ is $\eta_t = |\underline{C}_t| / |\overline{C}_t|$, $0 \leq \eta_t \leq 1$. The larger the value of $\eta_t$ is, the more accurate the data will be after discretization. To maximize $\eta_t$ requires $|\underline{C}_t|$ and $|\overline{C}_t|$ to be as close as possible to $|C_t|$.

We divide $U$ into two parts, $U_1$ and $U_2$. $\forall s_1 \in U_1$, $C_t(s_1) < C_t(x)$, while $\forall s_2 \in U_2$, $C_t(s_2) \geq C_t(x)$; then $\underline{C}_t(x) = \inf\left\{C_t(x), \inf_{s_1 \in U_1}\left\{\left(1 - \tilde{R}(s_1, x)\right) \vee C_t(s_1)\right\}\right\}$. Let $d$ be the absolute distance between $s_1$ and $x$. According to (18) and (20), $\left(1 - \tilde{R}(s_1, x)\right) = \frac{d}{\max(D)}$, where $\max(D)$ is the maximum distance between pixels in the distance matrix, which is a constant, and $\frac{d}{\max(D)}$ is called the relative distance between $s_1$ and $x$. $D(s_1, x) = 1 - \tilde{R}(s_1, x)$ represents the relative distance between $s_1$ and $x$, where $0 \leq D(s_1, x) \leq 1$, as shown in Fig. 3(a). The value of $\underline{C}_t(x)$ is determined by $D(s_1, x)$. When $\underline{C}_t(x)$ takes the minimum value $\min_{s_1 \in U_1}(C_t(s_1))$, $0 \leq D(s_1, x) \leq C_t(s_1)$, and when it takes the maximum value $C_t(x)$, $C_t(x) \leq D(s_1, x) \leq 1$. For $\underline{C}_t(x)$ to approach $C_t(x)$, it is necessary to increase the relative distance $D(s_1, x)$ between $s_1$ and $x$ as much as possible. In general, for two mixed pixels, the greater the difference of the membership degree is, the greater the distance between them becomes. We assume that $\exists u_1 \in U_1$, where $\min_{s_1 \in U_1}(C_t(s_1)) = C_t(u_1)$ and $\max_{s_1 \in U_1}(D(s_1, x)) = D(u_1, x)$. To prevent some pixels with smaller membership values from playing a decisive role in $\underline{C}_t(x)$, it is necessary to shorten the distance between $u_1$ and pixels with similar membership values to increase the distance

between them and $x$ and to maximize $\underline{C}_t(x)$. As shown in Fig. 3(b), the value of $\overline{C}_t(x)$ is also determined by $D(s_2, x)$. We assume that $\exists u_2 \in U_2$, where $\max_{s_2 \in U_2}(C_t(s_2)) = C_t(u_2)$ and $\max_{s_2 \in U_2}(D(s_2, x)) = D(u_2, x)$. To prevent some elements with larger membership values from playing a decisive role in $\overline{C}_t(x)$, it is necessary to shorten the distance between $u_2$ and pixels with similar membership values as much as possible to increase the distance between these pixels and $x$, thereby minimizing $\overline{C}_t(x)$.

Determining the distance matrix $D$ of the pixel set is equivalent to determining $\eta_t$ and $1-\frac{N_c}{N_I}$. Therefore, solving the optimal discretization scheme is equivalent to determining an optimal distance matrix $D$ such that the interval division result can maximize the average distance between the heterogeneous pixels and minimize the average distance between similar pixels when the number of intervals is as small as possible. According to the definition of $D$, each selected set of breakpoints $B$ uniquely determines $D$, and the problem becomes to select the optimal set of breakpoints. We let $g(B) = 1-\left(\frac{N_c}{N_I}\right)$, $h(B) = \left(\sum_{t=1}^{k} \eta_t\right)/k$. Furthermore, we let $P$ be the set of all of the candidate breakpoints in all of the bands, and $u$ be the weight coefficient, where $0 \le u \le 1$. The problem of choosing the best set of breakpoints can be reduced to:

$$\min \mathrm{f(B)} = \frac{1}{u \times g(B) + (1-u) \times h(B)}, B \subseteq P \quad \text{s.t.} \begin{cases} 0 \le D(x,y) \le 1, & \forall x, y \in U \\ \sum_{t=1}^{k} C_t(z) = 1, & \forall z \in U \\ 0 \le C_t(z) \le 1, & t \in \{1, \ldots, k\} \end{cases} . \tag{22}$$

Thus, the determination of the optimal discretization scheme is a constrained optimization problem, which has been proven to be NP-complete [51]. It is difficult to solve this kind of problem by traditional methods, and global optimization algorithms are more effective than traditional methods because their group search strategies and calculation methods do not depend on gradient information [52]. Global optimization algorithms for discretization mainly include the genetic algorithm (GA) [53], PSO [26], and ant colony optimization (ACO) [54]. Compared to the GA, the PSO has no crossover and mutation, which makes the operation principle simpler. Furthermore, the PSO has fewer parameters, and is more easily implemented. However, the theoretical research on convergence analysis is still weak, and we cannot use mature analysis methods to estimate the convergence speed like for the GA [55], [56]. The ACO algorithm can obtain good results when the number of intervals is known, but the number of intervals for discretization must be given in advance. The best number of intervals usually cannot be determined, especially with datasets including a large number of samples, where the method fails because it cannot provide the optimal number of intervals [54], [57]. Therefore, given the characteristics of a high-resolution remote sensing image, we use the improved adaptive genetic algorithm to solve the above constraint optimization problem. By introducing the adaptive mechanism in each stage of the genetic operation, we can dynamically adjust the search range, thus improving the search efficiency and accelerating the convergence speed.

**Algorithm 1** Initial Breakpoints Acquisition

**Input:** DN value set of all bands $VC = [Vc_1, Vc_2, \ldots, Vc_m]$
**Output:** Initial breakpoint set $B = [B_1, B_2, \ldots, B_m]$
1: $B_i = \varnothing, 1 \le i \le m$.
2: **for all** $Vc_i$ **do**
3: Sort the DN values in $Vc_i$ by ascending order;
4: Remove duplicate DN values in $Vc_i$;
5: $B_i \leftarrow Vc_i$;
6: **end for**

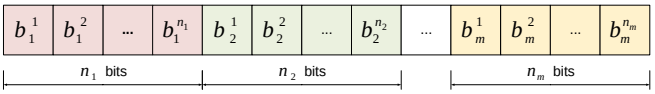



Fig. 4. Chromosome structure.

Applying $S = (U, R, V, f)$ to high-resolution remote sensing image, $U = \{x_1, x_2, \ldots, x_n\}$ is a finite set of objects representing the set of pixels in the image; $R = C \cup D$ is the set of all attributes; and $C = \{c_1, c_2, \ldots, c_m\}$ and $D = \{d\}$ are respectively the band and class attribute sets, where is the number of bands. The steps to obtain the initial set of candidate breakpoints are shown in Algorithm 1.

The feature discretization of a high-resolution remote sensing image is the selection of some of the DN values as discrete breakpoints to divide the whole feature interval. Therefore, the binary coding scheme can be used to directly encode the selection state of the initial breakpoint. Each bit in the binary code corresponds to a breakpoint, and the values 1 and 0 respectively represent reserving or abandoning a breakpoint. The set of initial breakpoints obtained by Algorithm 1 in the $i$th band is $B_i = \left\{b_i^1, b_i^2, \ldots, b_i^{n_i}\right\} (i = 1, 2, \ldots, m)$, where $n_i$ is the number of different values of pixels in the $i$th band, and the structure of each chromosome in the population is shown in Fig. 4. The colors in Fig. 4 represent bands in the image. Each chromosome with length $\sum_{i=1}^{m} n_i$ represents a feature discretization scheme. With this coding, all the bands in the image can be discretized simultaneously, laying the foundation for multivariable discretization.

### C. Flow of Proposed Discretization Method

The proposed discretization method is shown in Algorithm 2. The initial breakpoint set is obtained based on the features of the image. Based on the size of the initial breakpoint set, the initial population is generated by chaotic mapping. We introduce an individual replacement strategy based on roulette selection to solve the problem of population diversity. A dynamic multi-point crossing strategy is adopted to ensure that each band of a high-resolution remote sensing image can participate in the crossover operation, and the cross individuals are combined according to the Hamming distance between them to expand the search scope, thus avoiding premature convergence. Finally, drawing on the idea of particle evolution in a bare bones particle swarm optimization (BBPSO) algorithm, the number and locations of mutation points are determined according to the absolute dissimilarity of the local and global optimal individuals of a given chromosome in

**Algorithm 2** FRSGA Algorithm Process

**Input:** Decision table $S = (U, R, V, f)$, initial breakpoint set $B = [B_1, B_2, \ldots, B_m]$, number of categories $k$, reflectivity of the mixed pixels $R = \{R_1^1, \ldots, R_m^1, \ldots, R_1^n, \ldots, R_m^n\}$, reflectivity of the endmembers $r = \{r_{1,1}, \ldots, r_{1,m}, \ldots, r_{k,1}, \ldots, r_{k,m}\}$

**Output:** Optimal discretization scheme *gbest*

Initialize the fuzzy sets $C_1, \ldots, C_k$;
**for all** $x \in U$ **do**
    Calculate the abundance of each endmember $p_1(x), \ldots, p_k(x)$ by (17);
    $C_1(x) = p_1(x), \ldots, C_k(x) = p_k(x)$;
**end for**
Binary encoding of $B$;
t=0;
Initialize population P(t) by chaotic mapping;
Calculate the fitness of each individual in P(t) by (21) in MapReduce;
Update *gbest* with the maximum fitness value;
**while** t is less than the user's termination iterations **do**
    Perform selection operation of adaptive genetic algorithm;
    Perform crossover operation of adaptive genetic algorithm;
    Calculate the fitness of each individual in P(t) by (21) in MapReduce;
    Update *gbest* with the maximum fitness value;
    Perform mutation operation of adaptive genetic algorithm;
    Calculate the fitness of each individual in P(t) by (21) in MapReduce;
    Update *gbest* with the maximum fitness value;
    P(t+1)=P(t);
    t = t + 1;
**end while**

the form of Gaussian sampling, which makes the mutation operator directional and improves the search efficiency of the population. The adaptive genetic algorithm designed above is used to select, cross, and mutate the current population to generate the next generation population, and the above process is repeated until the termination condition is met. The fitness function based on a fuzzy rough model of each genetic operation is computed in parallel under the MapReduce framework. The list of local and global optimal variables is updated after each genetic operation. In the end, the optimal chromosome preserved by the global optimal variable is the optimal discretization scheme.

## V. Experimental Design

We briefly introduce the experimental data source, environment configuration, and dataset used in the experiment in this section.

### A. Data Source

The experimental data used in this paper are from a GF-2 satellite image of the coastal area of the South China Sea on September 16, 2016, as shown in Fig. 5. The image contains four bands. The objects in the image are divided into the categories of bare land, construction, farmland, water, and woodland.

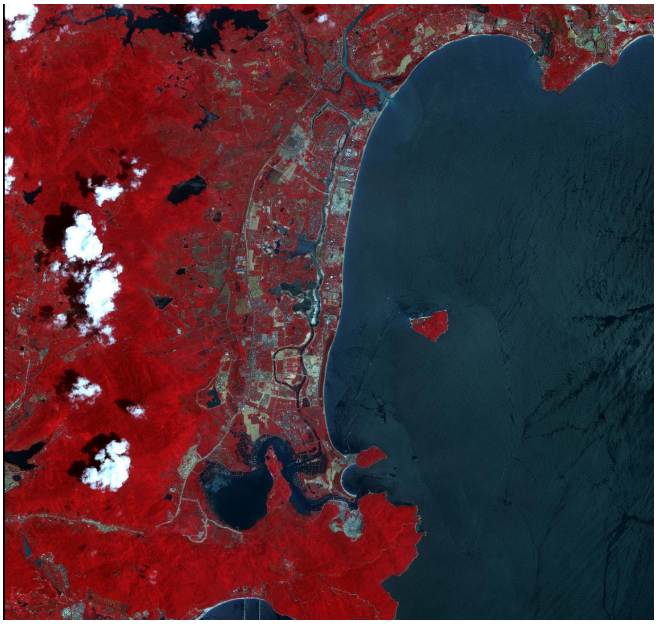


Fig. 5. Area used for study.

### B. Configuration of Experimental Environment

To verify the effectiveness of the proposed algorithm, comparative experiments were carried out on an Intel Core i5-5200U CPU at 2.2 GHz with 12 GB memory and a 512-GB hard disk. The visualization, programming, simulation, testing, and numerical processing of the experiments were implemented in MATLAB R2016a. The radiometric calibration and atmospheric correction of the image, decomposition of mixed pixels, training of the neural network classifier for discrete results, and comparison of the classification prediction accuracy were completed in an ENVI5.3 environment. BP neural network with three hidden layers was selected as the classifier. Each hidden layer had 20 nodes. Learning rate $\eta = 0.01$. Sigmoid function was selected as the activation function of the hidden layer. The number of nodes in the input layer was the same as the number of bands, and the number of nodes in the output layer was the same as the number of land cover types.

### C. Preparation of Experimental Datasets

High resolution makes the number of pixels corresponding to the same size area more than that of low- and medium-resolution, correspondingly, the number of pixel values is more, resulting in a large number of candidate breakpoints. In addition, each mixed pixel records the comprehensive spectral information of a variety of surface features [9]. These all reflect the characteristics of big data [4]. We randomly selected several areas covering the categories of bare land, construction, farmland, water, and woodland from the image, and used

the pixels of these areas as the experimental data, including training and test samples. The training set contained 10,000 samples, including 1853 bare land samples, 1014 construction samples, 1324 farmland samples, 2084 water samples, and 3725 woodland samples. The testing set contained 1016 samples, including 207 bare land samples, 200 construction samples, 201 farmland samples, 202 water samples, and 206 woodland samples. We used the linear spectral decomposition model [48] to extract the abundances of each endmember in each mixed pixel in the training set, then we sorted the pixel values and deleted the duplicate values in each band to obtain the initial breakpoints of the four bands, which were 6738, 7114, 8039, and 8367, respectively, totaling 30,258 and constituting a candidate set of breakpoints. We used the proposed method to select breakpoints from the candidate set to discretize the pixel values of the training set, and we compared this with the results of other algorithms.

## VI. Results and Discussion

First, we compare our method to four discretization algorithms using a heuristic search strategy on running time and search efficiency. These are the classical GA based on the consistency principle of the decision system in the early stage [58], multivariate discretization based on evolutionary cut points selection (EMD) [53], potential particle swarm optimization (PPSO) algorithm [26], and ACO based on information distance criterion [54]. We then compare the optimal set of breakpoints obtained by our method with the discretization results of the current mainstream supervised discretization algorithms, mainly based on the evaluation of the number of intervals and the consistency of the data. These are entropy-based discretization for ranking (EDiRa) [25], ChiMerge [59], 1R [21], NCAIC (Novel Class-Attribute Interdependency Discretization Algorithm) [60], FUDC (Feature Discretization Method Accommodating Uncertainty in Classification Systems) [27], CVD [28], and Chi2 [61]. Finally, we train the neural network classifier with the discretization results of all the methods, and the effectiveness of the proposed method is verified by comparison to the classification accuracy obtained by each method.

### A. Time Efficiency

The essence of the FRSGA, GA, EMD, PPSO, and ACO algorithms is to find the best individual by simulating their evolution process. The core operations are based on the evaluation of the individual fitness. Therefore, the time overhead of these algorithms mainly comes from calculating the individual fitness function.

To compare the time efficiencies of the five algorithms, they all had the same population size of 30. We divided the population file generated by FRSGA into five slices, each with six chromosomes. Table I compares the running times of the algorithms in one iteration. FRSGA ran the fastest because it used MapReduce to parallelize the fitness function. The time overhead of FRSGA was less than twice that of GA, was slightly less than that of ACO, and had the lowest time complexity. GA and EMD must perform three genetic operations (selection, crossover and, mutation) in an iteration, and the crossover and mutation both must calculate the fitness function. However, due to the chromosome reduction mechanism [53], the running time of EMD was slightly reduced. PPSO and ACO each had just one evolutionary operation, so GA generates the highest time overhead.

TABLE I
Comparison of Running Time in One Iteration

| Method | Population size | Iterations | Running time |
|---|---|---|---|
| FRSGA | 30 | 1 | 1530 s |
| GA | 30 | 1 | 3622 s |
| EMD | 30 | 1 | 2335 s |
| PPSO | 30 | 1 | 2255 s |
| ACO | 30 | 1 | 1779 s |

ACO = ant colony optimization; EMD = multivariate discretization based on evolutionary cut points selection; FRSGA = fuzzy rough sets-based genetic algorithm (ours); GA = genetic algorithm; PPSO = potential particle swarm optimization.

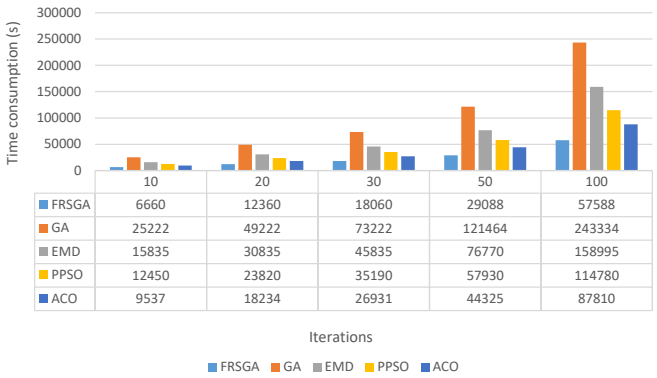



Fig. 6. Time overheads of five algorithms at different iterations.

Fig. 6 shows the time overhead values of the five algorithms at different iterations. FRSGA showed a significant improvement in the running speed due to the parallel computation of the fitness function by MapReduce. Although its time complexity was similar to that of PPSO [62], and it involved one more fitness calculation than PPSO and ACO per iteration, the time overhead was always smallest. EMD, PPSO, and ACO initially maintained similar time overhead values, but with the number of iterations increasing, the gap between EMD, PPSO, and ACO increased, and the time overhead of GA was always largest.

### B. Search Efficiency

We kept the population size of the five algorithms at 30 and set the number of iterations to 50, running them independently 10 times. The simulation results showed that the five algorithms could obtain the optimal discretization, but the fitness values differed due to the different fitness functions. Thus, we used the average relative deviation as an evaluation index, calculated as:

$$Avg_{dev} = \frac{\mathrm{Max}_{fv} - E_{fv}}{\mathrm{Max}_{fv} - \mathrm{Min}_{fv}}, \tag{23}$$

TABLE II
OPTIMAL SOLUTIONS OF FIVE ALGORITHMS

| Method | Number of simulations | Times to obtain maximum of fitness values | Maximum of fitness values |
|---|---|---|---|
| FRSGA | 10 | 9 | 0.9686 |
| GA | 10 | 6 | 3.4686 |
| EMD | 10 | 6 | 0.963 |
| PPSO | 10 | 7 | 0.9090 |
| ACO | 10 | 5 | 67.5919 |

TABLE III
AVERAGE RELATIVE DEVIATION OF THE FIVE ALGORITHMS

| Method | Maximum of fitness values | Minimum of fitness values | Mathematical expectation of fitness value | Average relative deviation |
|---|---|---|---|---|
| FRSGA | 0.9686 | 0.9594 | 0.9677 | 0.0978 |
| GA | 3.4686 | 3.4594 | 3.4662 | 0.2609 |
| EMD | 0.963 | 0.9591 | 0.9618 | 0.3077 |
| PPSO | 0.9090 | 0.8589 | 0.8946 | 0.2874 |
| ACO | 67.5919 | 50.7487 | 62.5216 | 0.3010 |

where $\mathrm{Max}_{fv}$ and $\mathrm{Min}_{fv}$ are the maximum and minimum of the fitness values of all optimal solutions, respectively, and $E_{fv}$ is the mathematical expectation of the optimal solution fitness value. A smaller $Avg_{dev}$ means the fitness value independently obtained by an algorithm was closer to the maximum, i.e., the solution performance was better.

Table II compares the optimal solutions, and Table III shows the average relative deviations of the algorithms. In the experiment, we set the weight coefficient $u$ of FRSGA at 0.1; the weight factors $\alpha$ and $\beta$ of GA to 1 and 2.5, respectively; the weight factor of EMD to 0.7; the crossover rate $p_c$ and mutation rate $p_m$ of FRSGA, GA, and EMD to 0.6 and 0.1, respectively; the weight factor $\mu$ of PPSO to 0.5; the control variables of the pheromone and heuristic value of ACO to 1 and 2, respectively; and the volatility $\rho$ of the pheromone trace to 0.1. Since the number of classes of a given training samples was five, the uniform class conditional probability vector $p_e = \left\{\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}\right\}$ in the ACO algorithm [54]. Because the smaller the fitness function ($Fitness \in [0, 1]$) of EMD is, the higher the discretization quality is. In the other four methods, the higher the fitness is, the higher the discretization quality is. In order to compare the five algorithms according to (23), we used $Fitness' = 1 - Fitness$ as the new fitness function of EMD. The larger $Fitness'$ is, the higher the quality of discretization is. In ten independent simulations, ACO obtained the maximum fitness value of the optimal solutions five times, the least of the five algorithms. FRSGA obtained the maximum fitness value of the optimal solution nine times, most among the algorithms. Its mathematical expectation of the optimal solution was 0.9677, and its average relative deviation was the smallest at 0.0978. PPSO obtained the maximum fitness value of the optimal solution seven times, which was second only to FRSGA. EMD accelerated the convergence rate through chromosome reduction mechanism, but the negative effect of chromosome reduction was that it was easy to fall into local optimum. Although EMD created population diversity by means of reseeding in mutation operation [53], the number of times to obtain the maximum fitness value of the optimal solution was consistent with that of GA, which was six times. Fig. 7 shows the number of iterations for the five algorithms to reach the optimal solution for the first time in 10 independent experiments. Table IV compares the convergence rates of the algorithms. The search efficiency is

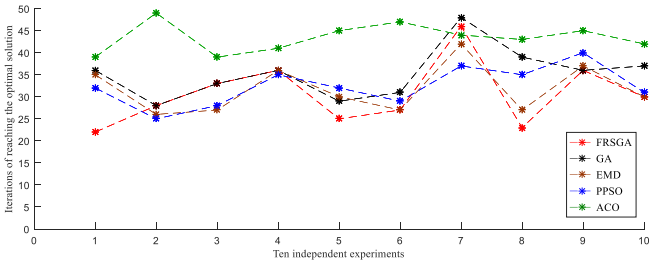



Fig. 7. Iterations of five algorithms needed to obtain optimal solution.

TABLE IV
CONVERGENCE SPEED OF FIVE ALGORITHMS

| Method | Iterations | Average iterations to reach optimal solution | Search efficiency |
|---|---|---|---|
| FRSGA | 50 | 30.6 | 0.388 |
| GA | 50 | 35.3 | 0.294 |
| EMD | 50 | 31.7 | 0.366 |
| PPSO | 50 | 32.4 | 0.352 |
| ACO | 50 | 43.4 | 0.132 |

$$Eff_{\text{search}} = 1 - \frac{A_I}{T_I}, \tag{24}$$

where $A_I$ is the average number of iterations to obtain the optimal solution, $T_I$ is the total number of iterations, and $Eff_{\text{search}}$ is the search efficiency. The larger the value of $Eff_{\text{search}}$ is, the faster the convergence speed is. Table IV shows that the average number of iterations for ACO to obtain the optimal solution was the largest at 43.4, and its search efficiency was the lowest at 0.132. The average number of iterations for FRSGA to obtain the optimal solution was the smallest at 30.6, and its search efficiency was the highest at 0.388.

### C. Interval Quality

We compared the optimal set of discrete breakpoints with the discretization results of state-of-the-art discretization methods EDiRa, ChiMerge, 1R, NCAIC, FUDC, CVD, and Chi2. We evaluated the interval quality from the number of intervals and data consistency [21]. The interval quality of the discretization schemes is calculated as follows:

$$E_{diq} = \frac{\omega_1 \times (N_I - N_D)}{N_I} + \frac{\omega_2 \times (N_S - N_F)}{N_S}, \tag{25}$$

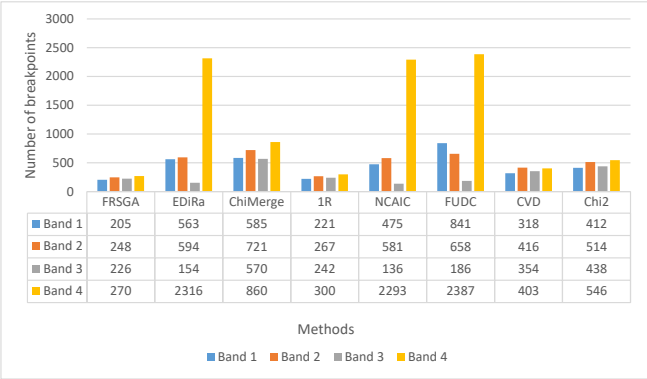



Fig. 8. Number of breakpoints of eight algorithms in different bands.

TABLE V
COMPARISON OF EIGHT ALGORITHMS ON INTERVAL QUALITY

| Method | Intervals | Inconsistencies | Quality |
|---|---|---|---|
| FRSGA | 949 | 0 | 0.9508 |
| EDiRa | 3627 | 36 | 0.7320 |
| ChiMerge | 2736 | 57 | 0.7242 |
| 1R | 1030 | 78 | 0.7443 |
| NCAIC | 3485 | 68 | 0.6668 |
| FUDC | 4072 | 4 | 0.7804 |
| CVD | 1491 | 63 | 0.7635 |
| Chi2 | 1910 | 59 | 0.7554 |

1R = one rule; Chi2 = second generation ChiMerge; ChiMerge = chi square-based discretization; CVD = Cramer's V-test discretization; EDiRa = entropy-based discretization for ranking; FUDC = feature discretization method accommodating uncertainty in classification systems; NCAIC = novel class-attribute interdependency discretization algorithm.

where $E_{diq}$ is the interval quality of the discretization scheme, $N_I$ is the number of initial breakpoints, $N_D$ is the number of discrete intervals, $N_S$ is the number of all instances, $N_F$ is the number of data errors, and $\omega_1$ and $\omega_2$ are weight coefficients. As shown in (25), the ideal discretization result is the best balance between the minimum number of breakpoints and minimum number of errors [26], [50], [58]. A larger $E_{diq}$ means a higher interval quality of the discretization scheme. However, the ratios of $\frac{N_D}{N_I}$ and $\frac{N_F}{N_S}$ are usually quite different, and it is difficult to determine the weight coefficients $\omega_1$ and $\omega_2$. Therefore, we evaluate the interval quality as follows:

$$E'_{diq} = \left(1 - \frac{N_D}{Sum_D}\right) \times \left(1 - \frac{N_F}{Sum_F}\right). \quad (26)$$

The weight coefficients are omitted from (25), and the interval quality is determined only by the number $N_D$ of discrete intervals and the number $N_F$ of the data errors. $Sum_D$ is the sum of the numbers of discrete intervals obtained by all the algorithms in the experiment, and $Sum_F$ is the sum of the numbers of data errors.

Fig. 8 shows the number of discrete intervals obtained by eight algorithms in each band. Table V compares the algorithms in terms of the interval quality. The number of discrete intervals obtained by FRSGA was 949, which was smallest of the algorithms, and there were no data errors.

TABLE VI
NUMBER OF CORRECT RECOGNITIONS OF FIVE CATEGORIES BY EIGHT ALGORITHMS

| Method | Bare land | Construction | Farmland | Water | Woodland |
|---|---|---|---|---|---|
| FRSGA | 207 | 160 | 105 | 198 | 202 |
| EDiRa | 197 | 152 | 95 | 198 | 173 |
| ChiMerge | 200 | 155 | 72 | 198 | 182 |
| 1R | 177 | 138 | 65 | 180 | 162 |
| NCAIC | 201 | 148 | 65 | 198 | 178 |
| FUDC | 200 | 153 | 100 | 198 | 192 |
| CVD | 198 | 150 | 72 | 198 | 180 |
| Chi2 | 200 | 155 | 72 | 198 | 182 |

Therefore, the interval quality of FRSGA, at 0.9508, was highest. The number of discrete intervals obtained by 1R was 1030, which was smallest of all of the algorithms except for that of FRSGA. However, the number of data errors was largest at 78, and the interval quality was 0.7443. The number of discrete intervals obtained by NCAIC was 3485, the number of data errors was 68, and the interval quality was the lowest at 0.6668. The number of discrete intervals obtained by FUDC was the largest at 4072, but there were only four data errors, and the interval quality was 0.7804, second only to FRSGA.

### D. Classification Accuracy

We trained the neural network classifier [63] by using the discrete samples obtained by the eight algorithms, and compared the output for the test samples in the model with the actual value of the label to obtain the classification accuracy. Through the confusion matrix [64], we obtained the overall average prediction accuracy and corresponding kappa coefficient [65]. Tables VI and VII respectively show the number of correct recognitions of the five categories and the overall classification accuracy of the neural network classifier by the analysis of the test samples after the neural network classifier was trained by using the discrete samples of the eight algorithms. FRSGA had the best comprehensive recognition results in bare land, construction, farmland, water, and woodland. The overall accuracy and kappa coefficient of FRSGA were respectively about 3 and 4 percentage points higher than those of FUDC, closely following FRSGA. FRSGA used the individual fitness function based on the fuzzy rough model, so the mixed pixels in the same interval had similar distributions of the endmember abundance after discretization, which greatly improved the recognition rate of objects. Data inconsistency also had a great impact on the classification accuracy. FRSGA and FUDC had zero and four data errors, respectively, with better classification accuracy, while NCAIC and 1R had more data errors, with the classification accuracies ranking last and second to last, respectively.

### E. Scalability

We used FRSGA and reinforcement learning-based genetic algorithm (RLGA) [20] to optimize the discretization results

TABLE VII
COMPARISON OF CLASSIFICATION ACCURACY OF EIGHT ALGORITHMS

| Method | Overall accuracy | Kappa coefficient |
|---|---|---|
| FRSGA | 0.8583 | 0.8227 |
| EDiRa | 0.8022 | 0.7525 |
| ChiMerge | 0.7943 | 0.7426 |
| 1R | 0.7106 | 0.6379 |
| NCAIC | 0.7776 | 0.7217 |
| FUDC | 0.8297 | 0.7870 |
| CVD | 0.7854 | 0.7315 |
| Chi2 | 0.7943 | 0.7426 |

TABLE VIII
NUMBER OF BREAKPOINTS OPTIMIZED BY FRSGA IN GF-2 IMAGE

| Method | Band 1 | Band 2 | Band 3 | Band 4 |
|---|---|---|---|---|
| FRSGA | 310 | 301 | 222 | 198 |
| RLGA | 310 | 301 | 222 | 198 |
| MFD-mvtR | 350 | 350 | 346 | 299 |

MFD-mvtR = multivariable optical remote sensing image feature discretization applied to marine vessel target recognition; RLGA = reinforcement learning-based genetic algorithm.

of multivariable optical remote sensing image feature discretization applied to marine vessel target recognition (MFD-mvtR) [7] on a GF-2 image. There were 2607 training samples in this image, including 676 boats, 742 ports, 143 buildings, 116 bare land shoals, 807 water bodies, and 123 vegetation. We used the linear spectral decomposition model [48] to extract the abundances of each endmember in each mixed pixel in the training set, sorted the pixel values of the training set, and deleted the duplicate values in each band to obtain the initial breakpoints of the four bands, which were 502, 493, 358, and 359, respectively, totaling 1712 breakpoints. The number of breakpoints after optimization is shown in Tables VIII and IX. MFD-mvtR obtained 1345 breakpoints and five data errors on the GF-2 image, which was unsatisfactory compared to other mainstream algorithms. We took the discretization scheme of MFD-mvtR as the individual of the initial population, and on this basis, used FRSGA to optimize it. By further optimizing the results of MFD-mvtR, the numbers of breakpoints in bands 1C4 were reduced to 310, 301, 222, and 198, respectively. The total number of breakpoints was 314 less than that before optimization, and the number of data errors was reduced to zero, similar to RLGA.

TABLE IX
QUALITY OF INTERVALS OPTIMIZED BY FRSGA IN GF-2 IMAGE

| Method | Intervals | Inconsistencies |
|---|---|---|
| FRSGA | 1031 | 0 |
| RLGA | 1031 | 0 |
| MFD-mvtR | 1345 | 5 |

TABLE X
CLASSIFICATION ACCURACY OF DISCRETIZATION SCHEME OPTIMIZED BY FRSGA IN GF-2 IMAGE

| Method | Overall accuracy | Kappa coefficient |
|---|---|---|
| FRSGA | 95.6250% | 0.9475 |
| RLGA | 91.0417% | 0.8925 |
| MFD-mvtR | 85.2083% | 0.8225 |

TABLE XI
QUALITY OF INTERVALS OPTIMIZED BY FRSGA IN LANDSAT 8 IMAGE

| Method | Intervals | Inconsistencies |
|---|---|---|
| FRSGA | 2247 | 0 |
| RLGA | 2247 | 0 |

Table X shows the classification accuracy of the neural network classifier after being trained by the discretization results of the above algorithms. After optimizing the discretization scheme of MFD-mvtR, the classification accuracy obtained by the confusion matrix was about 10 percentage points higher than the original, and the kappa coefficient was 0.9475. In the case of the same number of intervals and the same data inconsistencies, the classification accuracy was improved by about 5 percentage points compared to that of RLGA, which showed that the classification accuracy decreased when the influence of mixed pixels was not considered in the discretization. Fig. 9(a)-(j) correspond to FRSGA, RLGA, MFD-mvtR, EDiRa, ChiMerge, 1R, NCAIC, FUDC, CVD, and Chi2, respectively. FRSGA, RLGA, and MFD-mvtR yielded clearer textures of the ground object information in their classification effect maps, and more effectively identified boats in the image; in particular, the junction between berthing ships and the port could be well identified. Compared to other algorithms, there were fewer patches and stripes, and the boundaries of each category were clear. The classification effect map of FRSGA even accurately distinguished ports, buildings, bare land shoals, and vegetation in some local areas with complex surface features. However, in the middle of Fig. 9(d)-(j), there are stripes of varying degrees. There are also unrecognizable spots in the water area, especially in Fig. 9(f), where the boundaries between the berthing ships and the port are vague, and there are many spots in the water area. FRSGA yielded the highest quality classification effect map.

We further used FRSGA to optimize the discretization result of RLGA on a Landsat 8 image. The training set in this image included 6331 samples, consisting of 935 impervious surface

TABLE XII
CLASSIFICATION ACCURACY OF DISCRETIZATION SCHEME OPTIMIZED BY FRSGA IN LANDSAT 8 IMAGE

| Method | Overall accuracy | Kappa coefficient |
|---|---|---|
| FRSGA | 88.8254% | 0.8334 |
| RLGA | 83.3680% | 0.7503 |

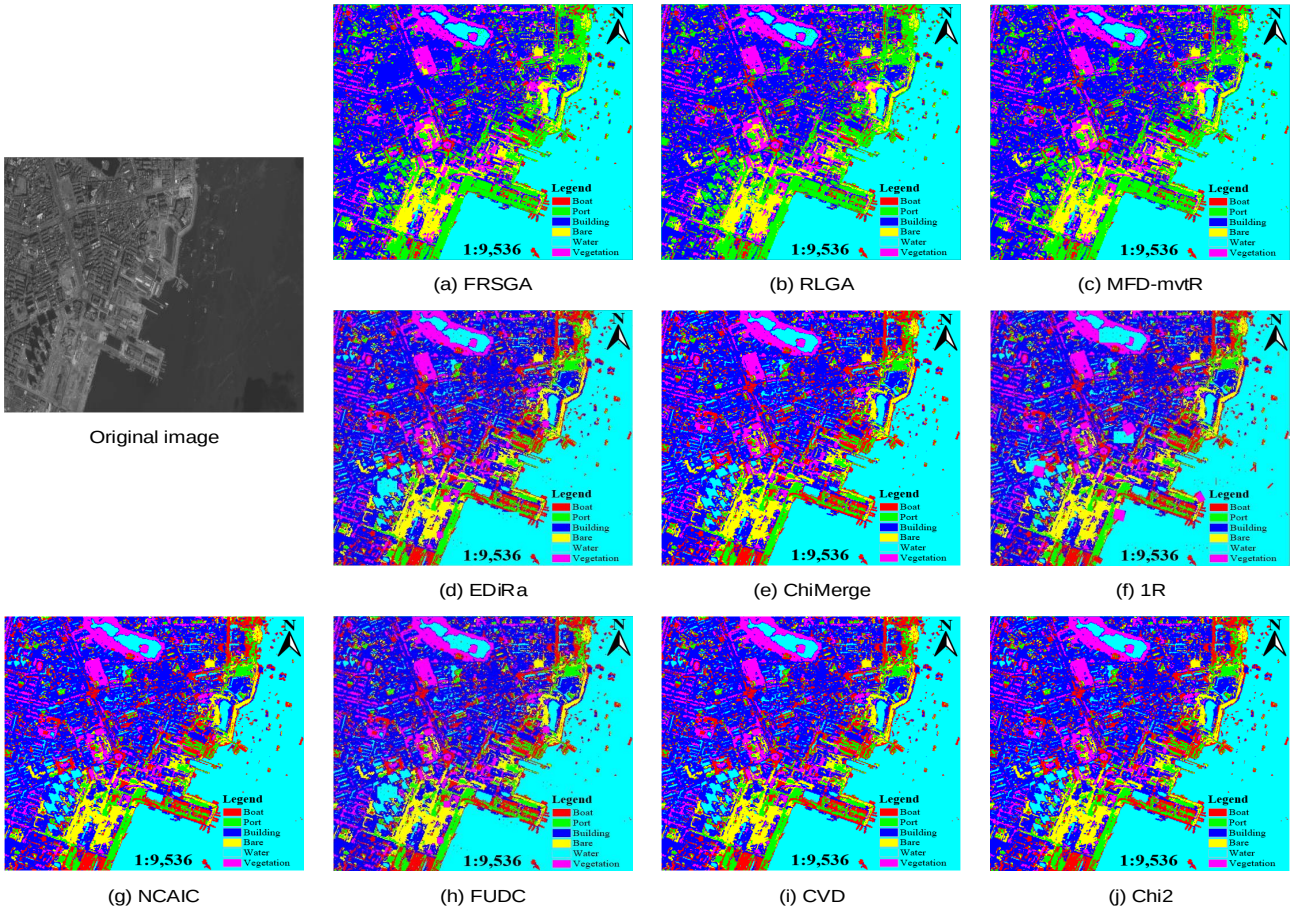



Fig. 9. Classification effect maps of the ten algorithms.

samples, 936 construction samples, 958 bare land samples, 2324 water samples, and 1178 vegetation samples. We used the linear spectral decomposition model [48] to extract the abundances of each endmember in each mixed pixel in the training set, after which we sorted the pixel values of the training set and deleted the duplicate values in each of the seven bands to obtain their initial breakpoints, which were 1403, 1429, 1680, 1869, 2402, 2530, and 2240, totaling 13553 breakpoints. The corresponding numbers of breakpoints after optimization by FRSGA were 193, 214, 148, 77, 703, 475, and 437, totaling 2247. As shown in Table XI, FRSGA and RLGA yielded the same optimization results. Table XII shows the classification accuracy of the neural network classifier after being trained by the discretization results of FRSGA and RLGA. For the same number of intervals and the same data inconsistencies, the classification accuracy was improved by about 5 percentage points compared to that of RLGA, which further showed that in the process of discretization, if the influence of mixed pixels was not considered, the classification accuracy was reduced.

## VII. Conclusions and Future Work

We proposed a discretization algorithm based on a fuzzy rough model to analyze and process high-resolution remote sensing big data. The main contributions of this work were as follows. (1) According to the theory of the fuzzy set and rough set, a decision table was introduced for the classification of a high-resolution remote sensing image, and each category of the ground object was regarded as a fuzzy set. (2) The distance matrix and fuzzy relation matrix on the set of pixels were obtained by calculating the Euclidean distance between pixels according to their DN values on each band. (3) The membership function of each fuzzy set was determined under the linear decomposition model, and the individual fitness function based on the fuzzy rough model was further established. (4) We adopted a chaotic mapping to initialize the population, and dynamically adjusted the search range by introducing the adaptive mechanism in each stage of the genetic operation, thus improving the search efficiency and accelerating the convergence speed. We used MapReduce to calculate the fuzzy-rough-based individual fitness in parallel, which improved the efficiency of data processing. (5) We verified the proposed method by comparing it to state-of-the-art discretization algorithms in terms of the running time, search efficiency, interval quality, and classification accuracy on the authentic GF-2 and Landsat 8 images.

Although our method can effectively solve the problem of feature discretization under the linear decomposition model of mixed pixels, the linear decomposition model must be based on the assumption that photons arrive at the sensor interact with a unique spectral endmember. In the ideal situation, each pixel in the high-resolution remote sensing image can be

approximately regarded as a linear mixture of each endmember in the pixel. However, there is interference from noise and outliers, and intimate mixtures formed by particulate media in close proximity appear in the scene. There are multiple interactions between scatterers of the different layers, which causes the nonlinear mixing of each endmember in a pixel. The fuzzy rough set has great flexibility in dealing with uncertain information. Therefore, our future research will continue to explore how to build more accurate and practical remote sensing image feature discretization models in the process of the nonlinear decomposition of mixed pixels using a type II fuzzy rough set with a stronger anti-noise performance.

## REFERENCES


[1] G. Cheng, X. Xie, J. Han, L. Guo, and G. S. Xia, "Remote sensing image scene classification meets deep learning: challenges, methods, benchmarks, and opportunities," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3735–3756, Jun. 2020.

[2] D. A. Kroodsma, J. Mayorga, T. Hochberg, N. A. Miller, K. Boerder, F. Ferretti, A. Wilson, B. Bergman, T. D. White, B. A. Block, P. Woods, B. Sullivan, C. Costello, and B. Worm, "Tracking the global footprint of fisheries," *Science*, vol. 359, no. 6378, pp. 904–908, Feb. 2018.

[3] X. Li, X. Yao, and Y. Fang, "Building-a-nets: robust building extraction from high-resolution remote sensing images with adversarial networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 10, pp. 3680–3687, Oct. 2018.

[4] G. Aydin, I. R. Hallac, and B. Karakus, "Architecture and implementation of a scalable sensor data storage and analysis system using cloud computing and big data technologies," *J. Sens.*, vol. 2015, pp. 1–11, Mar. 2015.

[5] S. Ramírez-Gallego, S. García, H. Mouriño Talín, D. Martínez-Rego, V. Bolón-Canedo, A. Alonso-Betanzos, J. M. Benítez, and F. Herrera, "Data discretization: taxonomy and big data challenge," *Wiley Interdiscip. Rev: Data Min. Knowl. Discov.*, vol. 6, no. 1, pp. 5–21, Jan./Feb. 2016.

[6] L. Xie, G. Li, M. Xiao, and L. Peng, "Novel classification method for remote sensing images based on information entropy discretization algorithm and vector space model," *Comput. Geosci.*, vol. 89, no. C, pp. 252–259, Apr. 2016.

[7] M. Huang, Q. Chen, and H. Wang, "A multivariable optical remote sensing image feature discretization method applied to marine vessel targets recognition," *Multimed. Tools Appl.*, vol. 79, no. 7-8, pp. 4597–4618, Aug. 2019.

[8] C. Wu, "Quantifying high-resolution impervious surfaces using spectral mixture analysis," *Intl. J. Remote Sens.*, vol. 30, no. 11, pp. 2915–2932, Jun. 2009.

[9] C. Quintano, A. Fernández-Manso, Y. E. Shimabukuro, and G. Pereira, "Spectral unmixing," *Intl. J. Remote Sens.*, vol. 33, no. 17, pp. 5307–5340, Sept. 2012.

[10] C. Sukawattanavijit, J. Chen, and H. Zhang, "Ga-svm algorithm for improving land-cover classification using sar and optical remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 3, pp. 284–288, Mar. 2017.

[11] M. Mahato, S. Gedam, J. Joglekar, and K. M. Buddhiraju, "Dense stereo matching based on multiobjective fitness functionła genetic algorithm optimization approach for stereo correspondence," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3341–3353, Jun. 2019.

[12] R. S. Gargees and G. J. Scott, "Deep feature clustering for remote sensing imagery land cover analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 8, pp. 1386–1390, Aug. 2020.

[13] D. A. Kumar, S. K. Meher, and K. P. Kumari, "Knowledge-based progressive granular neural networks for remote sensing image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 12, pp. 5201–5212, Dec. 2017.

[14] J. Dai, H. Hu, W. Z. Wu, Y. Qian, and D. Huang, "Maximal-discernibility-pair-based approach to attribute reduction in fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 4, pp. 2174–2187, Aug. 2018.

[15] K. Gong, Y. Wang, M. Xu, and Z. Xiao, "Bssreduce an $O(|U|)$ incremental feature selection approach for large-scale and high-dimensional data," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 6, pp. 3356–3367, Dec. 2018.

[16] W. Qu, L. Kong, K. Wu, F. Tang, and G. Chen, "Distributed fuzzy rough set for big data analysis in cloud computing," in *2019 IEEE 25th Intl. Conf. Parall. Distr. Syst.*, Tianjin, China, 2019, pp. 109–116.

[17] L. Kong, W. Qu, J. Yu, H. Zuo, G. Chen, F. Xiong, S. Pan, S. Lin, and M. Qiu, "Distributed feature selection for big data using fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 5, pp. 846–857, May 2020.

[18] X. Zhang, C. Mei, D. Chen, Y. Yang, and J. Li, "Active incremental feature selection using a fuzzy-rough-set-based information entropy," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 5, pp. 901–915, May 2020.

[19] N. M. Parthalain, R. Jensen, and R. Diao, "Fuzzy-rough set bireducts for data reduction," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 8, pp. 1840–1850, Aug. 2020.

[20] Q. Chen, M. Huang, Q. Xu, H. Wang, and J. Wang, "Reinforcement learning-based genetic algorithm in optimizing multidimensional data discretization scheme," *Math. Probl. Eng.*, vol. 2020, pp. 1–13, Mar. 2020.

[21] S. García, J. Luengo, J. A. Sáez, V. López, and F. Herrera, "A survey of discretization techniques: taxonomy and empirical analysis in supervised learning," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 4, pp. 734–750, Apr. 2013.

[22] C. J. Tsai, C. I. Lee, and W. P. Yang, "A discretization algorithm based on class-attribute contingency coefficient," *Inf. Sci.*, vol. 178, no. 3, pp. 714–731, Feb. 2008.

[23] X. F. Li, "The research on discretization algorithm based on dynamic hierarchical clustering," in *2012 Fourth Intl. Conf. Comput. Inf. Sci.*, Chongqing, China, 2012, pp. 349–352.

[24] K. Shehzad, "Edisc: a class-tailored discretization technique for rule-based classification," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 8, pp. 1435–1447, Aug. 2012.

[25] C. R. d. Sá, C. Soares, and A. Knobbe, "Entropy-based discretization methods for ranking data," *Inf. Sci.*, vol. 329, pp. 921–936, Feb. 2016.

[26] B. Tran, B. Xue, and M. Zhang, "A new representation in pso for discretization-based feature selection," *IEEE Trans. Cybern.*, vol. 48, no. 6, pp. 1733–1746, Jun. 2018.

[27] G. Zhang, Z. Wu, and L. Yi, "A remote sensing feature discretization method accommodating uncertainty in classification systems," in *Proc. 8th Intl. Symp. Spat. Accuracy Assess. Nat. Res. Environ. Sci.*, Shanghai, China, 2008, pp. 195–202.

[28] B. Wu, L. Zhang, and Y. Zhao, "Feature selection via cramer's v-test discretization for remote-sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2593–2606, May 2014.

[29] Z. Ali and W. Shahzad, "Comparative study of discretization methods on the performance of associative classifiers," in *2016 Intl. Conf. Front. Inf. Technol. (FIT)*, Islamabad, Pakistan, Dec. 2016, pp. 87–92.

[30] J. Gao, L. Dai, and W. Zhang, "Improved genetic optimization algorithm with subdomain model for multi-objective optimal design of spmsm," *CES Trans. Electr. Mach. Syst.*, vol. 2, no. 1, pp. 160–165, Mar. 2018.

[31] S. Zhao and D. Medhi, "Application-aware network design for hadoop mapreduce optimization using software-defined networking," *IEEE Trans. Netw. Service Manag.*, vol. 14, no. 4, pp. 804–816, Dec. 2017.

[32] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning about Data*. Norwell, MA, USA: Kluwer Academic Publishers, 1992.

[33] ——, "Rough set theory and its applications to data analysis," *Cybern. Syst.*, vol. 29, no. 7, pp. 661–688, Oct. 1998.

[34] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, Jun. 1965.

[35] D. Dubois and H. Prade, "Rough fuzzy sets and fuzzy rough sets," *Int. J. Gen. Syst.*, vol. 17, no. 2-3, pp. 191–209, 1990.

[36] ——, "Putting rough sets and fuzzy sets together," *Intell. Decis. Support*, vol. 11, pp. 203–232, 1992.

[37] C. Y. Wang and B. Q. Hu, "Fuzzy rough sets based on generalized residuated lattices," *Inf. Sci.*, vol. 248, pp. 31–49, Nov. 2013.

[38] B. Q. Hu and C. Y. Wang, "On type-2 fuzzy relations and interval-valued type-2 fuzzy sets," *Fuzzy Sets Syst.*, vol. 236, pp. 1–32, Feb. 2014.

[39] C. Y. Wang and B. Q. Hu, "Generalized extended fuzzy implications," *Fuzzy Sets Syst.*, vol. 268, pp. 93–109, Jun. 2015.

[40] H. Jiang, J. Zhan, and D. Chen, "Covering-based variable precision ($I$,$T$)-fuzzy rough sets with applications to multiattribute decision-making," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 8, pp. 1558–1572, Aug. 2019.

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

[42] A. Romero, C. Gatta, and G. Camps-Valls, "Unsupervised deep feature extraction for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1349–1362, Mar. 2016.

[43] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.

[44] N. Keshava, "A survey of spectral unmixing algorithms," *Lincoln Lab. J.*, vol. 14, no. 1, pp. 55–78, 2003.

[45] A. Plaza, P. Martinez, R. Perez, and J. Plaza, "A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 3, pp. 650–663, Mar. 2004.

[46] N. Yokoya, J. Chanussot, and A. Iwasaki, "Nonlinear unmixing of hyperspectral data using semi-nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1430–1437, Feb. 2014.

[47] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[48] Y. Altmann, S. Mclaughlin, and A. Hero, "Robust linear spectral unmixing using anomaly detection," *IEEE Trans. Comput. Imag.*, vol. 1, no. 2, pp. 74–85, Jun. 2015.

[49] M. Ji, "Fuzzy modelling of african ecoregions and ecotones using avhrr ndvi temporal imagery," *Geocarto Int.*, vol. 17, no. 1, pp. 23–32, Mar. 2002.

[50] R. Jin, Y. Breitbart, and C. Muoh, "Data discretization unification," *Knowl. Inf. Syst.*, vol. 19, no. 1, pp. 1–29, May 2008.

[51] B. S. Chlebus and S. H. Nguyen, "On finding optimal discretizations for two attributes," in *Intl. Conf. Rough Sets Curr. Trends Comput.*, Springer, Berlin, Heidelberg, 1998, pp. 537–544.

[52] Y. Ning, X. R. Bai, F. Zhou, and L. Liu, "Method for inverse synthetic aperture radar imaging of space debris using improved genetic algorithm," *IET Radar Sonar Nav.*, vol. 11, no. 5, pp. 812–821, May 2017.

[53] S. Ramirez-Gallego, S. Garcia, J. M. Benitez, and F. Herrera, "Multivariate discretization based on evolutionary cut points selection for classification," *IEEE Trans. Cybern.*, vol. 46, no. 3, pp. 595–608, Mar. 2016.

[54] W. Zhu, J. Wang, Y. Zhang, and L. Jia, "A discretization algorithm based on information distance criterion and ant colony optimization algorithm for knowledge extracting on industrial database," in *2010 IEEE Intl. Conf. Mech. Autom.*, Xi'an, China, 2010, pp. 1477–1482.

[55] P. J. Angeline, "Evolutionary optimization versus particle swarm optimization: philosophy and performance differences," in *Proc. 7th Intl. Conf. Evol. Program. VII*, Springer, Berlin, Heidelberg, 1998, pp. 601–610.

[56] A. A. Naeini, M. Babadi, S. M. J. Mirzadeh, and S. Amini, "Particle swarm optimization for object-based feature selection of vhsr satellite images," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 379–383, Mar. 2018.

[57] D. Yin, S. Du, S. Wang, and Z. Guo, "A direction-guided ant colony optimization method for extraction of urban road information from very-high-resolution images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 10, pp. 4785–4794, Oct. 2015.

[58] J. H. Dai, "A genetic algorithm for discretization of decision systems," in *Proc. 2004 Intl. Conf. Mach. Learn. Cybern.*, Shanghai, China, 2004, pp. 1319–1323.

[59] S. Rosati, G. Balestra, V. Giannini, S. Mazzetti, F. Russo, and D. Regge, "Chimerge discretization method: impact on a computer aided diagnosis system for prostate cancer in mri," in *Intl. Symp. Med. Meas. Appl. (MeMeA) Proc.*, Turin, Italy, 2015, pp. 297–302.

[60] D. Yan, D. Liu, and Y. Sang, "A new approach for discretizing continuous attributes in learning systems," *Neurocomputing*, vol. 133, pp. 507–511, Jun. 2014.

[61] W. Qu, D. Yan, Y. Sang, H. Liang, M. Kitsuregawa, and K. Li, "A novel chi2 algorithm for discretization of continuous attributes," in *Asia-Pacific Web Conf.*, Shenyang, China, 2008, pp. 560–571.

[62] T. Makkar, Y. Kumar, A. K. Dubey, A. Rocha, and A. Goyal, "Analogizing time complexity of knn and cnn in recognizing handwritten digits," in *2017 Fourth Intl. Conf. Image Inf. Process. (ICIIP)*, Shimla, India, 2017, pp. 1–6.

[63] L. Ren, Y. Sun, H. Wang, and L. Zhang, "Prediction of bearing remaining useful life with deep convolution neural network," *IEEE Access*, vol. 6, pp. 13 041–13 049, 2018.

[64] M. Ohsaki, P. Wang, K. Matsuda, S. Katagiri, H. Watanabe, and A. Ralescu, "Confusion-matrix-based kernel logistic regression for imbalanced data classification," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 9, pp. 1806–1819, Sep. 2017.

[65] J. Xiao, H. Wu, C. Wang, and H. Xia, "Land cover classification using features generated from annual time-series landsat data," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 739–743, May 2018.


**Qiong Chen** received his B.Eng. degree at Beijing University of Posts and Telecommunications, P.R. China, 2007, M.Eng. degree at Politecnico di Torino, Italy, 2012, and Ph.D. degree at College of Information Science and Technology, Hainan University, P.R. China, 2020. Now, he is a Postdoctoral Researcher at department of earth system science, Tsinghua University. His research interests include remote sensing image processing, evolutionary computing, granular computing, fuzzy decision-making, rough sets, big data analytics, and multi-source data fusion.

**Mengxing Huang** received the Ph.D. degree from Northwestern Polytechnical University, Xi'an, China, in 2007. He then joined staff with the Research Institute of Information Technology, Tsinghua University as a Postdoctoral Researcher. In 2009, he joined Hainan University. He is currently a Professor and a Ph.D. Supervisor of computer science and technology, and the Dean of School of Information and Communication Engineering. He is also the Executive Vice-Presedent of Hainan Province Institute of Smart City, and the Leader of the Service Science and Technology Team with Hainan University. He has authored or coauthored more than 60 academic papers as the first or corresponding author. He has reported 12 patents of invention, owns 3 software copyright, and published has 2 monographs and 2 translations. He has been awarded Second Class and Third Class Prizes of The Hainan Provincial Scientific and Technological Progress. His current research interests include signal processing for sensor system, big data, and intelligent information processing.

**Hao Wang** is an Associate Professor in the Department of Computer Science in Norwegian University of Science and Technology, Norway. He has a Ph.D. degree (2006) and a B.Eng. degree (2000), both in computer science and engineering, from South China University of Technology, China. His research interests include big data analytics, industrial internet of things, high performance computing, and safety-critical systems. He served as a TPC co-chair for IEEE DataCom 2015, IEEE CIT 2017, ES 2017, and IEEE CPSCom 2020, and a senior TPC member for CIKM 2019. He is the Chair for Sub-TC on Healthcare of IEEE Industrial Electronics Society Technical Committee on Industrial Informatics. He is a senior member of IEEE.

**Guangquan Xu** is a Ph.D. and professor at the Tianjin Key Laboratory of Advanced Networking (TANK), College of Intelligence and Computing, Tianjin University, China. He received his Ph.D. degree from Tianjin University in March 2008. He is a member of the CCF and IEEE. His research interests include cyber security and trust management.