

A Feature Discretization Method for Classification of High-resolution Remote Sensing Images in Coastal Areas

Qiong Chen, Mengxing Huang, *Member, IEEE*, and Hao Wang, *Member, IEEE*

Abstract—Feature discretization is one of the most relevant techniques for data preprocessing in remote sensing research area. Its main goal is to transform the continuous features of images into discrete ones to improve the efficiency of intelligent image processing algorithms, thus helping experts to more easily understand and use the acquired remote sensing data. In this paper, we focus on feature discretization for classification of high-resolution remote sensing images in coastal areas. In these images, 1) interactions among multiple bands exist, 2) noises interfere, and 3) maritime domain-specific prior knowledge is difficult to get. To address these challenges, we propose a hybrid metric method, based on information entropy and chi-square test, to calculate the stability of the discrete interval and the similarity of adjacent intervals. In addition, we use the degree of dependence among knowledge from the rough set theory as the evaluation criterion for discretization schemes, then scan each band in turn with the strategy of first splitting then merging, to obtain the optimal set of discrete features. Our method has been compared with the best state-of-the-art discretization algorithms on the GF-2 and Landsat 8 satellite dataset. Experiments show that the proposed method achieves better classification accuracy for high-resolution remote sensing images in coastal areas. It can not only effectively mine the correlation between features, but also filter the outliers in bands, thus producing as few discrete intervals as possible while ensuring data consistency.

Index Terms—Feature discretization, Image processing, Coastal areas, Hybrid metric, Rough set theory, Classification accuracy.

I. INTRODUCTION

As an advanced technology, high-resolution remote sensing satellites have been widely used in the field of oceans, providing important data support for monitoring and detection of seas and coastal areas [1]. However, with the accumulation of more and more data, the uncertainty in remote sensing information system is becoming more and more significant.

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 2018YFB140400. (Corresponding author: Mengxing Huang.)

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 (email: hawa@ntnu.no).

Compared with the previous low and medium spatial resolution remote sensing data, high-resolution remote sensing data greatly increase the amount of surface information, which can reflect the details of ground objects more abundantly, but also greatly increase the data volume and data complexity. Although the current database system can efficiently implement data entry, query, deletion, modification, and statistical analysis, it is still very difficult for database users to find the relationships and rules among the data, and to effectively predict and analyze future data based on existing data. In order to extract potential and valuable information from a large number of high-resolution remote sensing data with disorder and strong interference, it is necessary to introduce the data mining technique to analyze and process high-resolution remote sensing images [2]. However, most of the data stored in the actual database are continuous numerical attributes, such as spectral features, spatial features, time series features, and polarization features of remote sensing images. Most existing data mining algorithms can only deal with discrete data, and the processing efficiency of continuous features is inefficient. Before using these methods to analyze and process remote sensing data, it is necessary to transform the continuous features into discrete ones to reduce the complexity of program and the overhead of time and space, and enhance the clustering ability and anti-noise ability of the system. Therefore, feature discretization has become an important fundamental building block in remote sensing data processing, which directly affects the accuracy of subsequent image classification [3].

Before feature discretization of a high-resolution remote sensing image, we need to extract features from the original image. As shown in Fig. 1, the position coordinates of the pixels as the basic units and the corresponding values in each band are obtained after the feature information of the original image is extracted [4]-[6]. These pixel values are a set of continuous values ranging from 0 to 1. To discretize image features is to represent continuous pixel values with as few discrete intervals as possible on the premise that the compatibility of information systems is not compromised. As a basic reduction technology, feature discretization has attracted increasing attention from researchers all over the world with the continuous development of data mining and knowledge engineering, and has been extensively studied [7]. However, existing algorithms are far from satisfactory. Moreover, most of the research works on discretization technology mainly come from other areas than remote sensing. There are very few

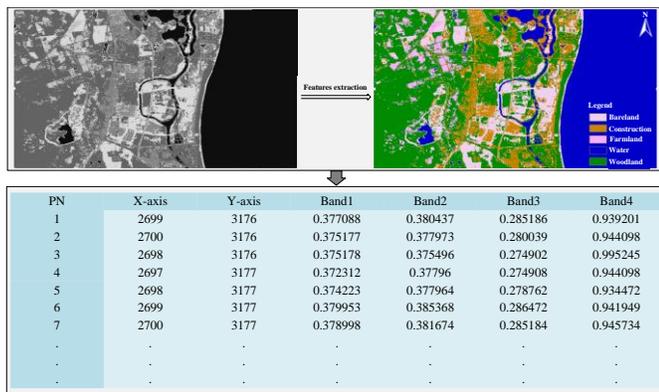


Fig. 1. Feature extraction from a high-resolution remote sensing image.

methods customized based on the characteristics of remote sensing data, therefore they are ineffective and inefficient for remote sensing applications.

Discretization can be divided into supervised discretization and unsupervised discretization according to whether the data contains label or category information [8]. At present, there are few mature unsupervised discretization methods, and it is difficult to obtain qualified results without class information. EqualWidth and EquaFrequency are two commonly used unsupervised discretization methods [9]. They divide continuous features according to given interval length and interval frequency respectively. Although they are simple and convenient, they will lead to uneven distribution of data and loss of some important information. In addition, many researchers adopt the idea of clustering in the process of discretization to derive breakpoints through the necessary projection intervals of the clustering region on each continuous attribute axis [10]. K-means is the earliest simple and effective partition-based clustering algorithm [11]. Its basic idea is to randomly select k objects as centroids of the initial intervals in the continuous feature space. The remaining objects are assigned to the nearest interval according to their Euclidean distances from the centroids, then the position of the centroids are updated. The above operation is repeated until each centroid no longer changes. Its main disadvantage is that the clustering ability is limited. Because the separation of data is based on the Voronoi graph, it cannot find non-convex clusters and is sensitive to isolated points, so that noises cannot be effectively filtered [12].

Supervised discretization has the advantage of making full use of class label and target attribute information because of the calculation based on class information. Therefore, it is easier to find the appropriate locations of breakpoints than unsupervised discretization. The most common supervised discretization method using information gain to divide intervals is based on the Minimum Description Length Principle (MDLP) [13]. After sorting the continuous feature values, the boundary between different target classes is set as candidate partition point, then the partition boundary which maximizes the information gain is found as the two-separation boundary in the candidate partition points. The potential partition points are determined iteratively until the principle of minimum description length principle is satisfied. Because of the information gain criterion, the algorithm can ensure the consistency of sample class in the interval to a large extent, and is suitable for the case of uniform distribution of target attribute

values. However, the sampled data are often affected by noises and other impurities in remote sensing image processing. The distribution of target attribute values is still scattered after sorting, so, it is difficult to filter noises by setting the threshold of partition. CADD [14], CAIM [15] and CACC [16] are discretization methods using class-attribute correlation as performance evaluation indicators. The redundancy of class-attribute correlation obtained by calculating the ratio of class-attribute mutual information and class-attribute joint entropy is used as the measurement standard in CADD. CADD requires users to specify the number of intervals at initialization, while CAIM does not need to preset the number of intervals, and has made some improvements in the calculation of the criterion of class-attribute correlation. Since only considering the most diverse class attribute in the interval and ignoring the distribution of other class attributes, it will result in overfitting to generate too many intervals. CACC uses logarithmic function of interval number to replace interval number in class-attribute correlation redundancy standard, which speeds up the process of discretization and prevents overfitting, thus making up for the deficiency of CAIM [17]. Because CACC adopts the criterion of maximizing the degree of correlation between classes, it only optimizes the result of interval discretization only for a single attribute, lacks the description of the overall data information and does not consider the consistency of data before and after discretization, which will inevitably result in the loss of important information in the original data. ChiMerge [18], Chi2 [19] and Extended Chi2 [20] adopt the method of class-attribute information calculation based on interval similarity to apply Pearson statistics to discriminate and merge the adjacent intervals. Their advantage is that the adjacent intervals have distinct structures, but they are sensitive to parameters.

In [21], a method of discretization based on information entropy is applied to rank label data, which improves the sensitivity of homogeneity of sample ranking in the set. However, adopting this standard, only one attribute can be partitioned separately in multi-feature data at a time, and the compatibility of the whole information system after discretization cannot be guaranteed.

In [22], a discretization method is proposed to classify remote sensing image features. The method first defines the uncertainty of the decision system based on the equivalence class model of the rough set. Then, using information entropy criterion, selects the breakpoints by controlling the change of uncertainty under a given threshold. Since this method takes into account the intrinsic relationship among multiple attributes in the process of discretization, it achieves high classification accuracy on SPOT5 images. However, this method only considers the stability of the interval, and does not consider the similarity of adjacent intervals, so, it cannot effectively filter the noises, and will produce more intervals.

A discretization method based on Cramer's V-Test is presented in [23], which is applied to feature selection of remote sensing image classification. This method uses Cramer's V-test as a discretization criterion to measure the dependence between the target class and the discretization variable in a given feature, and divides χ^2 by $\ln(n)$ to accelerate the discretization process and reduce the enormous

influence of n in the discretization scheme, where n is the number of intervals. Discrete features generated by this method on QuickBird and PHI hyperspectral images can improve the classification performance of J48-DT and NB classifiers. However, just like [21], the intrinsic correlation between multiple attributes is not considered, and it is difficult to ensure that the compatibility of the system is not destroyed after discretization.

In addition, some common discretization algorithms have simple steps and low complexity, but only consider a single continuous feature at a time, and do not solve the problem of joint discretization of multiple continuous features under the premise of ensuring the compatibility of the whole information system, such as PKID [24], AEFD [25], 1R [26], D2 [27] and discretization method based on Hellinger [28], etc.

Based on the state of the art, we can see that most of the discretization algorithms used in the analysis and processing of high-resolution remote sensing images in coastal areas encounter the following problems: 1) Since a coastal zone is the natural area with the most abundant phenomena and processes [29]-[31], correspondingly, the high-resolution remote sensing images collected in this area contain not only a large number of pixels but also complex categories, which brings many difficulties to compute the discretization of image features. 2) Influenced by the complexity and variability of the coastal areas, and in every link of high-resolution remote sensing image acquisition, such as periodic deviation of sensors, electromagnetic interference between load components, etc., a certain number of abnormal values and noises will inevitably occur in each band of the image [32], when these abnormal values or noises cannot be effectively filtered, the quality of interval division will be greatly reduced. 3) The prior knowledge of the sea is difficult to obtain or because of the age and the great changes in the marine environment, the existing prior knowledge is no longer applicable [33], which greatly reduces the accuracy of the algorithm and destroys the compatibility of the decision system after discretization.

Although the above three main characteristics may separately exist in remote sensing in other regions [34]-[35], the remote sensing in coastal zone contains all of these three main characteristics. In addition to the influence of noise, the special geographical environment and frequent human activities cause the diversity of underlying surface, which makes the spectral signals collected in this area complicated. For example, in the recognition process of coastal water, due to the reflection of water surface and the transparency of water, the water signal received by the sensor includes spectral signals from the bottom, water body, water surface and atmosphere. Therefore, the signal of coastal water is a superimposed comprehensive signal containing the spectral signals of chlorophyll, suspended sediment, various pollutants, flow field, etc., which is more difficult to identify. In addition, due to the high risk of offshore operation, long time-consuming and high cost of investigation, it is very difficult to conduct field investigation of remote sensing in coastal zone, so that prior knowledge is hard to obtain, which increases the difficulty of remote sensing data analysis in coastal zone.

In response to these problems, based on the feature preprocessing framework built in our previous work [36], we propose a feature discretization method (Hybrid metric of

information entropy and chi-square test using rough set as evaluation criterion for discretization, termed as ECRSD) for classification of high-resolution remote sensing images in coastal areas. Firstly, we use the top-down splitting strategy to calculate the information gain generated by each breakpoint and select the breakpoint splitting interval with the largest value. Then, we use chi-square test to scan all the intervals to judge the similarity of adjacent intervals and merge adjacent intervals satisfying the conditions. Finally, we use the degree of dependence among knowledge from the rough set theory as the criterion for discretization schemes to evaluate the compatibility of the information system, and outputs the results if the conditions are satisfied, otherwise, the corresponding splitting threshold and chi-square significance level are changed, and each band is scanned again until the optimal set of discrete features is obtained. We simulate and analyze the high-resolution remote sensing image collected in the coastal area of the South China Sea, and compare with the current mainstream methods. Experiments show that the proposed method achieves better classification results on SVM and neural network classifier. It can not only effectively mine the correlation between features, but also filter the abnormal values in bands, and obtain fewer discrete intervals while guaranteeing data consistency.

The rest of this paper is arranged as follows: Section II introduces the basic concepts and describes the problem models. Section III elaborates the proposed algorithm flow. The results of experiment are analyzed and discussed in Section IV. Section V summarizes this paper.

II. BASIC CONCEPTS AND PROBLEM MODELS

In this section, we formally define the problem of feature discretization of high-resolution remote sensing images. In addition, we analyze the possible distribution of pixels in intervals partitioned by feature discretization of high-resolution remote sensing images in coastal areas. Finally, we describe the idea of information entropy and Chi-Square test applied to the discretization of remote sensing images, and the evaluation model of rough set.

A. Feature Discretization

Turing Award winner Herbert Simon once pointed out that the discrete features are closer to the knowledge layer than the continuous features, and the data is easier to understand, use and explain for users and experts after discretization [37]. Dougherty et al. also confirmed in a research report that discretization makes machine learning more accurate and efficient, extending the boundaries of many learning algorithms [38]. Some classification methods can only deal with discrete features [39]. In industry, it is seldom to directly input continuous values as the features of LR (logistic regression model), but to convert continuous features into discrete features and then hand these discrete features to LR. Generally, in the LR model, continuous feature A will correspond to a weight W . If A is discretized, then A will be extended to features A_1, A_2, A_3, \dots , etc. Each feature corresponds to a weight. If the feature A_k does not appear in the training samples, then the training model has no weight for A_k . Even if the feature A_k

appears in the test samples, the feature A_k will not affect the model, which is equivalent to invalid. However, if continuous feature A is used directly without discretization, then in the LR model, $y = W \times A$, A is the feature and W is the weight corresponding to A . For example, if A represents age and $A \in [1, \dots, 100]$, even if a test case with a value of "300" on A appears in the test set (obviously this "300" is an outlier), the LR model will still substitute the outlier "300" into $y = W \times A$ to get a value, and this value will be very large. Therefore, outliers will have a very large impact on the final result, and discretization can eliminate this effect. Feature discretization also simplifies the LR model and reduces the risk of overfitting. When using continuous features, a feature corresponds to a weight, then, if the weight of this feature is large, the model will be very dependent on this feature, so that a small change in this feature may cause a big change in the final result. Such a model is very unreliable. When encountering a new sample, it is likely to get the wrong classification result because it is too sensitive to this feature, that is, the generalization ability is poor and overfitting is easy to occur. However, after feature discretization, a feature becomes multiple sub-features, and correspondingly, the number of weights also changes from one to multiple, so the influence of the previous continuous feature on the model is dispersed and weakened, thereby reducing the risk of overfitting. Therefore, feature discretization can improve the generalization ability of classifiers, thus improving the prediction accuracy of classification.

After atmospheric correction, the spectral features of high-resolution remote sensing image can be obtained as continuous brightness DN (Digital Number) in the range of 0-1. The feature discretization of high-resolution remote sensing image simply means dividing the continuous DN values of the image into a limited number of sub-intervals according to a specific rule, and associating these sub-intervals with a set of discrete values. The basic flow of feature discretization of a high-resolution remote sensing image is shown in Fig. 2. Firstly, image features are sorted according to DN values. Secondly, the intervals are merged or split according to the breakpoint selection strategy and the corresponding evaluation criteria. Finally, the final discretization features are generated if the discretization results meet the termination conditions, otherwise, the intervals continue to be merged or split.

Considering a classification problem of the remote sensing image including C types of land cover, N is the number of pixels to be processed and M is the number of bands, so, the discretization of the remote sensing image is defined as follows.

$$D_A = \{[d_0, d_1], (d_1, d_2], \dots, (d_{k_A-1}, d_{k_A}]\} \quad (1)$$

The discretization algorithm divides the continuous DN values on band A into k_A discrete and disjoint intervals. Among them, d_0 and d_{k_A} are the minimum and maximum DN values, and all values are arranged in ascending order in D_A . D_A is called a discretization scheme on band A . $P_A = \{d_0, d_1, d_2, \dots, d_{k_A-1}\}$ is the set of breakpoints on band A .

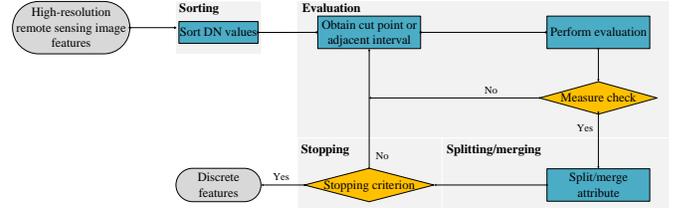


Fig. 2. Feature discretization of a high-resolution remote sensing image.

$P = \{P_1, P_2, \dots, P_A, \dots, P_M\}$ represents the complete set of all breakpoints in all M bands. Therefore, the search space for feature discretization of a high-resolution remote sensing image is formed by all candidate breakpoints in each band, which are from all different DN values in each band of the training set.

B. Distribution of Pixels in Discrete Interval

Due to the huge amount of information contained in each band of the high-resolution remote sensing image, correspondingly, the pixels of each interval in the discretization process are not only numerous but also complex in categories. In particular, the existence of outliers or noises greatly affects the quality of discretization results. We analyze several possible distributions of internal pixels in discrete intervals from high-resolution remote sensing images in coastal areas.

1) When using MDLP [13] or class-attribute correlation redundancy criterion CAIR [14] to split the intervals of the bands from top to bottom, there will be adjacent intervals with the same structure, as shown in Fig. 3. The column attribute PN of the information table is the ordinal number of the pixels, File X and File Y represent the two-dimensional coordinates X and Y of the pixels, the brightness values of the discretized pixels are recorded in B1, and the categories of the pixels are recorded in Class. Under a given segmentation threshold, both interval 1 and interval 2 contain 5 pixels, of which four are type A and one is type B. Obviously, the internal structures of interval 1 and interval 2 are the same, and they are adjacent intervals, so they can be merged to reduce the number of discrete intervals generated.

2) When using ChiMerge [18], Chi2 [19] and Extended Chi2 [20] to merge discrete intervals from bottom to top, there will be discrete intervals with many types, as shown in Fig. 4. Although these methods can ensure that the internal structures of adjacent intervals are different from each other, with the progress of interval merging, there will be a large number of different types of pixels in the resulting discrete intervals due to the lack of judgment on the occurrence frequency of pixels in the interval. N represents noises or outliers. It can be seen that besides the two types of pixels A and B, discrete interval 1 and the discrete interval 2 also have noises or outliers, which makes

PN	File X	File Y	B1	Class
1	4114	996	0.1	A
2	4113	996	0.1	A
3	4113	997	0.1	B
4	4114	997	0.1	A
5	4115	997	0.1	A
6	4115	998	0.2	A
7	4114	998	0.2	A
8	4115	999	0.2	A
9	4114	999	0.2	A
10	4130	1001	0.2	B

Fig. 3. Adjacent intervals with the same internal structure.

PN	File X	File Y	B1	Class
1	4134	1005	0.3	A
2	4135	1005	0.3	A
3	4123	1021	0.3	B
4	4123	1022	0.3	N
5	4122	1022	0.3	N
6	4123	1023	0.4	A
7	4122	1023	0.4	A
8	4123	1024	0.4	B
9	4122	1024	0.4	N
10	4122	1025	0.4	N

Fig. 4. Discrete intervals with many types.

PN	File X	File Y	B1	Class
1	4941	1246	0.1	N
2	4945	1246	0.1	N
3	4938	1247	0.2	N
4	4943	1247	0.2	N
5	4945	1247	0.3	N
6	4937	1247	0.3	N
7	4942	1248	0.4	N
8	4943	1248	0.4	N
9	4944	1248	0.5	N
10	4940	1248	0.5	N

Fig. 5. Continuous noise blocks.

the stabilities of the two intervals very poor. In addition, because of the influence of noises or outliers, when the significance level of these methods is set improperly, discrete interval 1 and discrete interval 2, which are similar in structure, cannot be merged.

3) When a large number of noises and outliers exist in the image and the parameters of the algorithms are unreasonable, continuous noise blocks will appear after the band is discretized, as shown in Fig. 5. The five adjacent discrete intervals are composed of noises, because of the lack of sufficient consideration of the stability of the frequency of pixel classes within the interval or the similarity of adjacent interval structures in the process of selecting candidate breakpoints. As the discretization proceeds, continuous noise blocks are gradually formed, and the adjacent noise intervals cannot be merged and filtered due to the inability to identify the types of pixels in each noise block, thus increasing the number of discrete intervals.

Obviously, the above three special cases show that some of the intervals that should be merged in the discretization process cannot be merged eventually, resulting in too many discrete intervals. It can be seen that in the process of feature discretization of high-resolution remote sensing images, not only the distribution of pixels within the interval but also the similarity of adjacent interval structures should be considered, so that the optimal discretization scheme can be obtained by adjusting the relevant parameters of the algorithm.

C. Information Entropy Metric of Discrete Interval

Information entropy is a famous mathematical theory put forward by Shannon in order to solve the problem of quantitative measurement of information in the field of communication [40]. Catlette, Fayyad and Irani respectively introduce information entropy into discretization algorithms of continuous features [27], [41]. According to Fayyad's and Irani's expositions, the formulas of information entropy and breakpoint information entropy are given respectively.

$$E(S) = -\sum_{i=1}^k P(C_i, S) \log_2(P(C_i, S)) \quad (2)$$

$$E(A, T, S) = \frac{|S_1|}{|S|} E(S_1) + \frac{|S_2|}{|S|} E(S_2) \quad (3)$$

Information entropy is a good metric, which can reflect the stability of frequency of all categories within the discrete interval [21], thus ensuring the validity of the discrete interval partition. For this reason, when information entropy is applied to the feature discretization of high-resolution remote sensing images, S represents the set of image pixels, K is the number of land cover types, C_i is the number of instances where category is i in S , A and T denote band A and the breakpoint T in it, S_1 and S_2 represent the sets of pixels of two discrete intervals divided by breakpoint T on band A , and $|S|$ represents the cardinality of S , that is, the number of pixels contained in S .

D. Chi-square Metric of Discrete Interval

Chi-square test is a widely used hypothesis test method. It judges the degree of consistency between theoretical deduced values and actual observed values by the deviation between the actual observed values of statistical samples and theoretical deduced values [42]. It uses the method of class-attribute information calculation based on interval similarity to discriminate and merge adjacent intervals [43], as shown in (4).

$$\chi^2 = \sum_{i=1}^m \sum_{j=1}^k \frac{(A_{ij} - E_{ij})^2}{E_{ij}} \quad (4)$$

Chi-square test formula uses Pearson statistics to determine whether the current breakpoint should be removed, that is, whether the two discrete intervals adjacent to the breakpoint should be merged. According to Pearson's theorem in mathematical statistics, the asymptotic distribution of statistic is the χ^2 distribution with degree of freedom $k-1$, and k is the total number of categories of the system, m is the number of variables to be tested for correlation. Obviously, when applied to the discretization problem of judging whether two discrete intervals should be merged, $m=2$. When significance level α is given, the corresponding critical value χ_α can be determined. To this end, chi-square test formula is applied to the determination of the similarity of adjacent discrete intervals in high-resolution remote sensing image features. Among them, χ^2 represents the degree of deviation between observed values and theoretical values, A_{ij} is the number of pixels belonging to class j in the i th discrete interval, and E_{ij} is the expected frequency of class j in the i th discrete interval.

E. Rough Set Evaluation Model

Rough set theory is an important mathematical tool for dealing with inaccurate data [44]. In rough set theory, knowledge is regarded as the division of universe, that is, knowledge is considered to be granular, and the inaccuracy of knowledge is caused by the granularity of knowledge. Different from DS evidence theory [45], fuzzy set theory [46], etc., the membership function values of the objects in the rough set theory depend on the knowledge base, which can be obtained

directly from the required data without any prior knowledge or additional information about the data. Therefore, it is more objective to use it to reflect the uncertainty of marine knowledge in the case that the prior knowledge of the ocean is not easy to acquire [47].

In rough sets, data table is called information system. Information system is a quaternion $S = (U, A, V, f)$, where U is a non-empty set of finite objects, A is a non-empty set of finite attributes, $V = U(V_a)$ is a set of attribute values, V_a is the range of attribute a , $f: U \times A \rightarrow V$ is a mapping function indicating that each attribute of each object maps to a value of the set of attribute values. If an attribute of the attribute set is considered as a decision attribute, then the information system S defined above is called a decision table, where $A = C \cup D$, including the condition attribute set C and the decision attribute set D .

Since the high-resolution remote sensing image contains multiple bands, i.e., multiple feature variables, if only one feature is discretized at a time, the results will largely destroy the compatibility of the original system, thus affecting the subsequent classification accuracy. So, in the analysis and processing of remote sensing images, based on the above-mentioned rough set theory, we establish a multi-feature model of high-resolution remote sensing images, then U represents the set of image pixels, attributes in condition attribute set C represents the bands, D contains only one decision attribute, which corresponds to the land cover class in the remote sensing image, and V_a represents the value range of band a [3]. The model is represented by a matrix, as shown below.

$$DT = \begin{bmatrix} u_1 & c_{11} & c_{12} & \cdot & \cdot & \cdot & c_{1m} & d_1 \\ u_2 & c_{21} & c_{22} & \cdot & \cdot & \cdot & c_{2m} & d_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ u_n & c_{n1} & c_{n2} & \cdot & \cdot & \cdot & c_{nm} & d_n \end{bmatrix} \quad (5)$$

Each row represents a pixel instance. $U = \{u_1, u_2, \dots, u_n\}$ is a set of pixels, condition attribute column $C = \{c_1, c_2, \dots, c_m\}$ represents the brightness values of the pixel in m bands, and the last column is the decision attribute column, which identifies the class information of the sample. Each pixel is composed of pixel number, band attribute and category attribute. The range of brightness value is $0 \leq c_{ij} \leq 1$, where c_{ij} is the brightness value of the i th sample in the j th band, d_i is the category of the i th sample. The category is expressed by natural values, and the range of category values is determined by the number of categories defined in the system. For example, assuming that the number of categories defined is 6, the range of category values is $D = \{1, 2, 3, 4, 5, 6\}$.

III. HYBRID METRIC METHOD OF FEATURE DISCRETIZATION

The essence of feature discretization method is to determine the number of breakpoints and the location of breakpoints, then to split or merge intervals according to certain criteria for the continuous features. Its ultimate goal is to maximize the interdependence between categories and discrete intervals. In other words, it must minimize the loss of important information caused by discretization. Therefore, we should balance the simplicity and accuracy, and try to get as few discrete intervals as possible while ensuring data consistency.

Here, we describe the flow of ECRSD algorithm applied to the classification of high-resolution remote sensing images in coastal areas. ECRSD calculates the stability of pixel categories in discrete intervals and the similarity of adjacent interval structures with the idea of hybrid metric of information entropy and chi-square test. The degree of dependence among knowledge from the rough set is used as the evaluation criterion of discretization scheme. The optimal discrete feature set is obtained by scanning each band in turn with the strategy of splitting and merging.

A. Interval Entropy Table of Discrete Features

In order to quickly find the discrete subinterval with the largest entropy value in the band, it is necessary to establish a table to record the entropy value of all the discrete feature intervals in the current band, that is, the discrete feature interval entropy table (DFIET). DFIET consists of three columns, the lower bound of the interval in the first column, the upper bound of the interval in the second column and the corresponding entropy value in the third column, as shown in Table I.

Each row in the table corresponds to a discrete feature subinterval, which is arranged from small to large according to the value of entropy. Each time the separable discrete feature subinterval with the largest entropy value is searched from the last item in DFIET. The separable discrete feature subinterval refers to the interval that contains at least two breakpoints (that is, the lower bound of the interval is not equal to the upper bound the interval) and the entropy value is greater than the given threshold value. In this paper, the formulas for calculating the entropy values of intervals and the stability of the overall interval under given discretization scheme are introduced as follows.

$$EV_r = - \sum_{i=1}^k \frac{p_{ir}}{p_{+r}} \log_2 \frac{p_{ir}}{p_{+r}} \quad (6)$$

$$IS_D(C, F) = \sum_{r \in D} p_{+r} EV_r \quad (7)$$

Among them, r denotes the r th subinterval under the discretization scheme D and k denotes the number of land cover categories in the image. p_{ir} represents the ratio of the number of pixels of category i to the total number of instances in the r th interval. p_{+r} represents the ratio of the number of pixels in the r th interval to the total number of instances. C and F represent the category variable and the discrete feature interval variable respectively.

TABLE I
DFIET Structure

Lower bound	Upper bound	Entropy value
L1	U1	EV1
L2	U2	EV2
⋮	⋮	⋮
Ln	Un	EVn

After the separable discrete feature subinterval with the largest entropy value is found, it is necessary to determine the location of the breakpoint in this subinterval, that is, to select the optimal segmentation breakpoint among all the candidate breakpoints in this subinterval. To this end, a formula is introduced to measure the mutual exclusion information between C and F , as shown below.

$$IM_D(C, F) = \sum_{r \in D1} EV_r - IS_{D2}(C, F) \quad (8)$$

Where, $D1$ and $D2$ are the overall intervals before and after discretization respectively. From (8) and the description of the selection of breakpoints, it can be seen that the change of mutual exclusion information is actually caused by the position of breakpoints in the separable discrete feature subinterval. Therefore, we only need to consider the change of information in the separable discrete feature subinterval with the largest entropy value. We can narrow the calculation scope of mutual exclusion information to a separable discrete feature subinterval, as shown in (9).

$$IM_m(C, F) = EV_m - IS_m(C, F) \quad (9)$$

The mutual exclusion information is computed for each candidate breakpoint in the separable discrete feature subinterval m with the largest entropy value, and the largest $IM_m(C, F)$ of candidate breakpoint is chosen as the segmentation breakpoint. In the beginning, DFIET only contains one row, that is, the entire continuous feature interval. As the algorithm runs, it starts to split. By adding a new row at the current split subinterval position and updating the boundary values and entropy values of the two discrete feature subintervals separately, DFIET can preserve all discrete feature subintervals.

B. Interval Chi-square Table with Discrete Features

Because it is necessary to use chi-square formula to evaluate the similarity between adjacent discrete feature subintervals, in order to quickly find the smallest chi-square value of discrete feature subintervals, it is also necessary to establish a table to record the chi-square values of all discrete feature subintervals in the current band, namely the discrete feature interval chi-square table (DFICT). DFICT consists of three columns, the lower bound of the interval in the first column, the upper bound of the interval in the second column and the corresponding chi-square value in the third column, as shown in Table II.

Each row in the table corresponds to a discrete feature subinterval, which is arranged from small to large according to the value of boundary, so as to ensure that the adjacent rows are adjacent intervals. The third column of the current row records

TABLE II
DFICT Structure

Lower bound	Upper bound	Chi-square value
L1	U1	CV1
L2	U2	CV2
⋮	⋮	⋮
Ln	Un	---

the chi-square value between the current row and the next row, and the third column of the last row is assigned invalid value because there is no discrete feature subinterval represented by the next row. Each time the combinable discrete feature subinterval with the smallest chi-square value is searched from DFICT (the smaller the chi-square value is, the higher the similarity between two adjacent subintervals). The combinable discrete feature subinterval refers to the interval where the chi-square value is valid (that is, not the last discrete feature subinterval in DFICT) and the chi-square value is less than the given threshold value. The threshold value is set by the chi-square critical probability corresponding to the degree of freedom, and the calculation of the degree of freedom is shown in (10).

$$v = k - 1 \quad (10)$$

Among them, k represents the number of land cover categories in the two adjacent intervals. Initially, the discrete feature subintervals in DFIET are mapped to DFICT. As the algorithm runs, it starts to merge. By deleting the location of the current merging subinterval and updating the upper bound and chi-square value of the adjacent discrete feature subinterval, DFICT can preserve all discrete feature subintervals.

C. Measurement of System Compatibility

We use the degree of dependence among knowledge from the rough set as the evaluation criterion of system compatibility after discretization [22]. In order to calculate the degree of dependence among knowledge, we need to introduce the concepts of indiscernible relationship, lower approximation set and upper approximation set.

Given an information system $S = (U, R, V, f)$, where U is a finite set of objects, i.e., the set of pixels in the image, R is a set of attributes, including condition attribute set C and decision attribute set D . For each attribute subset $A \subseteq R$, the indiscernible relationship $IND(A)$ is defined below.

$$IND(A) = \{ \langle x, y \rangle \mid \langle x, y \rangle \in U^2, \forall a \in A (a(x) = a(y)) \} \quad (11)$$

The equivalence class of A in U , also known as knowledge A defined on U , is given by (12).

$$U \mid IND(A) = \{ X \mid X \subseteq U \wedge (\forall x \in X \forall y \in X \Rightarrow \forall a \in A (a(x) = a(y))) \} \quad (12)$$

For each subset $X \subseteq U$ and the equivalence class of the attribute subset A in U , the lower and upper approximation sets of X are respectively defined as follows.

$$A_-(X) = \cup \{ Y \mid Y \in U \mid IND(A) \wedge Y \subseteq X \} \quad (13)$$

$$A^-(X) = \cup \{ Y \mid Y \in U \mid IND(A) \wedge Y \cap X \neq \emptyset \} \quad (14)$$

$$A_-(X) \subseteq A^-(X) \quad (15)$$

The equivalence class generated on U by the set of condition attributes consisting of all bands in the image is recorded as knowledge C , and the equivalence class generated on U by the set of decision attribute consisting of land cover categories is recorded as knowledge D . Generally, in D , the lower approximation set is included in the upper approximation set, and only related to the equivalence class of a land cover category, that is, the lower approximation set of D has uniqueness. Therefore, when evaluating the system compatibility, the lower approximation set can more effectively reflect the change of equivalence class after discretization. (16) is the positive domain of decision attribute set D under knowledge system $U|C$. The dependence of knowledge D on knowledge C is given by (17).

$$POS_C(D) = C_-(D) \quad (16)$$

$$\gamma_C(D) = \frac{card(POS_C(D))}{card(U)} \quad (17)$$

Where $card(\bullet)$ is the cardinality of the set, that is, the number of elements contained in the set. Obviously, $0 \leq \gamma_C(D) \leq 1$. When $\gamma_C(D) \rightarrow 1$, knowledge D is highly dependent on knowledge C . When $\gamma_C(D) = 1$, knowledge D is completely dependent on knowledge C , and the compatibility of information system is not destroyed. The size of $\gamma_C(D)$ reflects the degree to which system compatibility is destroyed after discretization.

D. Flow of ECRSD Algorithm

ECRSD algorithm calculates the stability of pixel categories in discrete intervals and the similarity of adjacent interval structures by using the hybrid metric of information entropy and chi-square test, as shown in Algorithm 1. Then, we take the degree of dependence among knowledge from the rough set as the evaluation criterion of discretization scheme, and scan each band in turn with the strategy of splitting and merging, so as to obtain the optimal discrete feature set. The basic process is shown in Algorithm 2.

At the beginning of the program, a high-resolution remote sensing image is input, after the preprocessing of the image, the brightness values of all pixels in each band are obtained, and the decision information table of continuous feature space is generated. Discretization is performed in order from the first band to establish DFIET. Each time, the separable interval with the largest entropy value is found from DFIET for segmentation. If all the separable intervals are completed, DFIET is mapped to DFICT for interval merging. After all the bands are discretized, the compatibility of the decision information system is evaluated by the degree of dependence among knowledge from the rough set. If the termination conditions specified by users are not satisfied, the splitting and merging thresholds are reset and the relevant parameters are adjusted. Then, the continuous features are discretized again until the termination condition are finally met, and the discrete

Algorithm 1 Hybrid Metric Method of Information Entropy and Chi-square Test

```

begin
  for each band i do
    Initialize DFIET;
    while Splittable discrete feature interval with the largest entropy
    value in DFIET exists do
      Select interval of which entropy value is the largest;
      obp <- null;
      lmv <- min;
      for each candidate breakpoint j in this interval do
        Calculate the mutual information generated by candidate
        breakpoint j using Eq. (9);
        if lmv < mutual information j do
          lmv <- mutual information j;
          obp <- candidate breakpoint j;
        end
      end
      Split this interval by obp;
      Update DFIET;
    end
  Map DFIET to DFICT in ascending order of boundary value;
  Initialize DFICT by using Eq. (4);
  Set the merge threshold according to Eq. (10);
  while Mergeable discrete feature interval with the smallest
  chi-square value in DFICT exists do
    Select interval of which chi-square value is the smallest;
    Merge this interval with its next adjacent interval;
    Update DFICT;
  end
end

```

Algorithm 2 ECRSD Algorithm Basic Scheme

```

Input: Remote sensing image features
Output: Discretized features
begin
  Initialize discretization thresholds and related parameters;
  while System compatibility does not satisfy the user's termination
  condition do
    Adjust discretization parameters;
    for each band i do
      Get discrete intervals by the hybrid metric method; end for
    end
    Evaluate system compatibility using Eq. (17);
  end
  Return discretization scheme D;
end

```

feature set of this high-resolution remote sensing image is output, and the program ends.

IV. EXPERIMENTS AND ANALYSIS

In this section, we briefly introduce the experimental data source, the experimental environment configuration and the data set used in the experiment. Then, we compare the optimal set of breakpoints obtained by ECRSD algorithm with the discretization results of the current mainstream methods, mainly from the aspects of the number of intervals and the consistency of data. Finally, we train the SVM and neural network classifiers separately from the discretization results of these methods, and verify the effectiveness of the proposed method by comparing the classification accuracy obtained by each method.

A. Data Source

The experimental data used in this paper are from two GF-2 satellite images and one Landsat 8 satellite image collected in

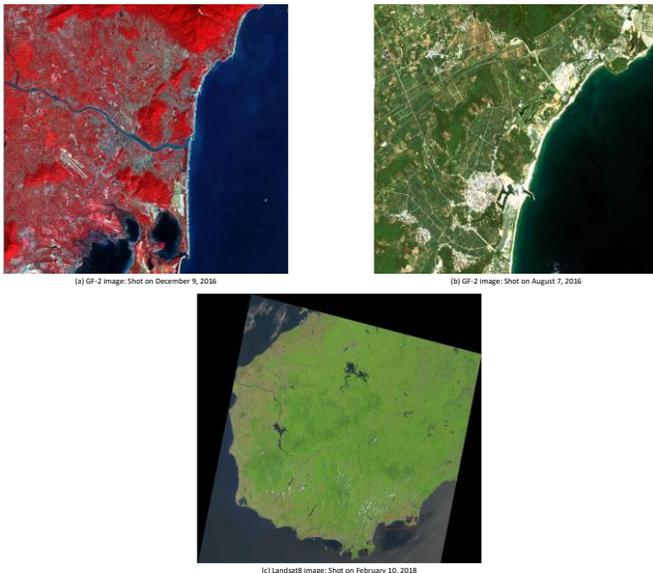


Fig. 6. Area used for research.

coastal areas of the South China Sea, as shown in Fig. 6. Among them, GF-2 satellite data contains four bands, and Landsat 8 satellite data contains seven bands. In the experiment, the objects in the first GF-2 image are divided into eight categories: construction, coniferous forest, broadleaf forest, fallow land, arable land, bare land, water and cloud. The objects on the second GF-2 image are divided into seven categories: coniferous forest, broadleaf forest, arable land, fallow land, bare land, construction and water. The objects on Landsat 8 image are divided into seven categories: coniferous forest, fallow land, bare land, construction, broadleaf forest, arable land and water.

B. Configuration of Experimental Environment

To verify the effectiveness of the proposed algorithm, the comparative experiments are carried out under the hardware conditions of Intel (R) Core (TM) i5-5200U CPU@2.20GHZ processor, 12G memory and 512G hard disk. The visualization, programming, simulation, testing and numerical processing of the experiments are implemented in MATLAB (R2016a version). The radiometric calibration, atmospheric correction, and the comparison of classification of discrete feature sets of the high-resolution remote sensing image generated by algorithms in the experiment are completed in ENVI5.3 environment.

We use Radial Basis Function as the kernel function of SVM.

We set the parameter $\gamma = \frac{1}{2\sigma^2} = \frac{1}{k}$ of the RBF kernel, where

k is the number of categories, so, γ of the first GF-2 image is $1/8$, γ of the second GF-2 image is $1/7$, and γ of Landsat 8 image is $1/7$. In addition, we choose BP neural network with three hidden layers as another classifier. Each hidden layer has 20 nodes, and Sigmoid function is selected as the activation function of the hidden layer. We set the number of input and output nodes of BP neural network model based on three different images in the experiment. The first GF-2 image contains four bands and is divided into eight categories. Therefore, the number of nodes in the input layer is 4 and the

number of nodes in the output layer is 8. The second GF-2 image contains four bands and is divided into seven categories. Therefore, the number of nodes in the input layer is 4 and the number of nodes in the output layer is 7. Similarly, Landsat 8 image contains seven bands and is divided into seven categories. Therefore, the number of nodes in the input layer is 7 and the number of nodes in the output layer is 7. The activation function of the output node is Softmax function.

C. Data preprocessing

We randomly select and tag several regions covering the eight categories from the first GF-2 image. After integration, 6862 samples are used as training samples to be discretized. Among them, 360 instances are construction, 198 instances are coniferous forest, 654 instances are broadleaf forest, 247 instances are fallow land, 229 instances are arable land, 389 instances are bare land, 3012 instances are water and 1773 instances are cloud. We then sort all the pixels in each band according to the brightness value and delete the duplicate values. The initial breakpoints of four bands are respectively 4551, 3882, 3668 and 2256, totaling 14357 breakpoints. In the second GF-2 image, there are 5855 samples used as training samples to be discretized. Among them, 729 instances are coniferous forest, 1552 instances are broadleaf forest, 968 instances are arable land, 517 instances are fallow land, 626 instances are bare land, 945 instances are construction and 518 instances are water. The initial breakpoints of four bands are respectively 4336, 4108, 4090 and 1209, totaling 13743 breakpoints. Similarly, in Landsat 8 image, there are 2621 samples used as training samples to be discretized. Among them, 308 instances are coniferous forest, 245 instances are fallow land, 322 instances are bare land, 675 instances are construction, 296 instances are broadleaf forest, 262 instances are arable land and 513 instances are water. The initial breakpoints of seven bands are respectively 1204, 1276, 1424, 1491, 1786, 1883 and 1755, totaling 10819 breakpoints.

D. Quality Assessment of Discretization Scheme

The quality of discrete feature set depends on the number of intervals and the inconsistency of data in the decision information table. Data inconsistency is given by (18).

$$Inconsistencies = \sum_{k=1}^N (T_{CEC}^k - \text{Max}(C_1^k, C_2^k, \dots, C_m^k)) \quad (18)$$

Where, N is the number of equivalent classes on the set of condition attributes under the current discretization scheme, and m is the number of categories in the decision information table of the high-resolution remote sensing image. T_{CEC}^k denotes the number of instances contained in the k th equivalent class. C_i^k denotes the number of instances whose category is i in the k th equivalent class, while $\text{Max}(C_1^k, C_2^k, \dots, C_m^k)$ is the largest number of instances of all categories in the k th equivalent class. The proposed method is used to discretize decision information table of the high-resolution remote sensing image, and the results are compared with those of EDiRa [21], ChiMerge [48], 1R [49], NCAIC [17], FUDC [22], Cramer's V-Test [23], and Chi 2 [50], these seven best state-of-the-art discretization methods. In the

TABLE III

Comparison of the Number of Discrete Intervals in Each Band of the First GF-2 Image

Method	Band 1	Band 2	Band 3	Band 4
ECRSD	255	134	74	63
EDiRa	199	286	157	323
ChiMerge	287	152	91	63
1R	142	74	60	27
NCAIC	384	274	159	120
FUDC	340	221	129	111
CV-Test	288	153	92	64
Chi2	288	153	92	66

1R = one rule; Chi2 = second generation ChiMerge; ChiMerge = chi square-based discretization; CV-Test = Cramer's V-test discretization; ECRSD = entropy and chi-square test using rough set as evaluation criterion for discretization; EDiRa = entropy-based discretization for ranking; FUDC = feature discretization method accommodating uncertainty in classification systems; NCAIC = novel class-attribute interdependency discretization algorithm.

TABLE IV

Comparison of the Number of Data Errors in the First GF-2 Image

Method	Number of intervals	Inconsistency
ECRSD	526	0
EDiRa	965	14
ChiMerge	593	24
1R	303	39
NCAIC	937	23
FUDC	801	4
CV-Test	597	20
Chi2	599	27

experiment, we set the initial threshold value of information entropy to 1.0. In each cycle of the algorithm, the step size is decreased by 0.01, and the confidence level of chi square is set to three levels: 0.9, 0.95 and 0.99. In the discretization process of the first GF-2 image, when the final DFIET is reached, the threshold of information entropy is 0.92, and the confidence level of chi square is 0.99. The number of intervals in each band and data inconsistency obtained by the eight algorithms are shown in Table III and Table IV.

It can be seen that the number of intervals in ECRSD algorithm is 526, only more than 1R algorithm, and no data errors occur. Although 1R algorithm achieves the fewest number of intervals, the data inconsistency is the highest with 39 errors. EDiRa algorithm has the largest number of intervals, reaching 965, followed by NCAIC algorithm, which is 937. The data errors obtained by these two algorithms are 14 and 23. The number of intervals in FUDC algorithm is 136 less than NCAIC algorithm, but the data inconsistency can be controlled at a lower level with 4 errors, which is second only to our method. ChiMerge, Chi2 and Cramer's V-Test algorithms have similar intervals, with data errors of 24, 27 and 20, respectively. In the discretization process of the second GF-2 image, when the final DFIET is reached, the threshold of information entropy is 0.85, and the confidence level of chi square is 0.9. The number of intervals in each band and data inconsistency obtained by the eight algorithms are shown in Table V and VI.

TABLE V

Comparison of the Number of Discrete Intervals in Each Band of the Second GF-2 Image

Method	Band 1	Band 2	Band 3	Band 4
ECRSD	1686	1717	1631	286
EDiRa	2519	2505	2411	380
ChiMerge	1730	1738	1663	310
1R	449	450	414	69
NCAIC	2493	2440	2337	380
FUDC	2527	2516	2415	380
CV-Test	1729	1737	1662	310
Chi2	1711	1715	1650	307

TABLE VI

Comparison of the Number of Data Errors in the Second GF-2 Image

Method	Number of intervals	Inconsistency
ECRSD	5320	0
EDiRa	7815	46
ChiMerge	5441	85
1R	1382	281
NCAIC	7650	76
FUDC	7838	39
CV-Test	5438	61

TABLE VII

Comparison of the Number of Discrete Intervals in Each Band of Landsat 8 Image

Method	Band 1	Band 2	Band 3	Band 4	Band 5	Band 6	Band 7
ECRSD	39	42	44	44	53	55	51
EDiRa	63	62	66	71	84	83	81
ChiMerge	49	51	56	67	99	88	63
1R	32	33	35	35	42	42	41
NCAIC	66	66	71	73	89	86	83
FUDC	69	70	75	77	95	92	88
CV-Test	49	45	55	60	93	81	57
Chi2	49	49	55	62	96	86	57

TABLE VIII

Comparison of the Number of Data Errors in Landsat 8 Image

Method	Number of intervals	Inconsistency
ECRSD	328	0
EDiRa	510	4
ChiMerge	473	5
1R	260	17
NCAIC	534	5
FUDC	566	4
CV-Test	440	5
Chi2	454	5

It can be seen that the number of intervals in ECRSD algorithm is 5320, only more than 1R algorithm, and no data errors occur. Although 1R algorithm achieves the fewest number of intervals, the data inconsistency is the highest with 281 errors. FUDC algorithm has the largest number of intervals, reaching 7838, followed by EDiRa algorithm, which is 7815. The data errors obtained by these two algorithms are 39 and 46. In the discretization process of Landsat 8 image, when the final DFIET is reached, the threshold of information entropy is 0.93, and the confidence level of chi square is 0.95. The number of

intervals in each band and data inconsistency obtained by the eight algorithms are shown in Table VII and VIII.

It can be seen that the number of intervals in ECRSD algorithm is 328, only more than 1R algorithm, and no data errors occur. Although 1R algorithm achieves the fewest number of intervals, the data inconsistency is the highest with 17 errors. FUDC algorithm has the largest number of intervals, reaching 566, followed by NCAIC algorithm, which is 534. The data errors obtained by these two algorithms are 4 and 5. ChiMerge, Chi2 and Cramer's V-Test algorithms have similar intervals, and the data errors are all 5. The quality assessment of discretization scheme is usually calculated by the following formula.

$$Q_{dfs} = \omega_1 \times \frac{N_o - N_d}{N_o} + \omega_2 \times \frac{N_s - N_e}{N_s} \quad (19)$$

Where, Q_{dfs} denotes the quality of discretization scheme, N_o is the number of initial breakpoints, N_d is the number of discrete intervals, N_s is the number of all instances, N_e is the number of data errors, ω_1 and ω_2 are weight coefficients. As shown in (19), the ideal discretization result is actually to find the best balance between the minimum number of breakpoints and the minimum number of data errors. The larger the value of Q_{dfs} , the higher the quality of the discretization scheme. We set ω_1 and ω_2 to 0.1 and 0.9, respectively, to obtain Q_{dfs} of all algorithms, as shown in Table IX, X and XI.

We can see that our discretization scheme has the highest quality. Our method introduces the idea of hybrid metric of information entropy and chi-square test. We consider both internal stability and external similarity of intervals, adopt the discretization strategy of first splitting then merging, and evaluate the compatibility of discrete feature set. Therefore, the minimum number of intervals is obtained while ensuring the minimum data errors. In addition, DFIET and DFICT are established respectively to improve the efficiency of data retrieval thus greatly reducing the run time. Although EDiRa algorithm uses information entropy to measure the stability of the interval, it needs to consider the overall similarity between the label rankings in the training set while adopting the top-down splitting strategy. Therefore, when the number of samples increases, the time overhead will increase significantly. Moreover, since it only discretizes one band at a time, the result will destroy the compatibility of the system to a certain extent. So, we can see that while obtaining the maximum number of intervals, it has more data errors than FUDC algorithm which also use information entropy to split the intervals. Different from EDiRa, FUDC algorithm utilizes rough set to evaluate the uncertainty of decision system. Therefore, the number of data errors generated by FUDC is much less than that of EDiRa. NCAIC uses class-attribute correlation as the criterion of interval division, and takes into account the upper approximation of each class and the distribution information of data. However, considering only the upper approximation cannot completely characterize the whole equivalent class. Discrete discriminant still has a certain probability to incline to the class attribute containing the most samples in the interval, resulting in excessive number of intervals. So, we can see that the discrete feature set obtained by NCAIC has mediocre

TABLE IX
Comparison of the Quality of Discretization Schemes in the First GF-2 Image

Method	Number of interval decreases	Number of correct instances	Quality of discretization scheme
ECRSD	13831	6862	0.9963
EDiRa	13392	6848	0.9914
ChiMerge	13764	6838	0.9927
1R	14054	6823	0.9928
NCAIC	13420	6839	0.9905
FUDC	13556	6858	0.9939
CV-Test	13760	6842	0.9932
Chi2	13758	6835	0.9923

TABLE X
Comparison of the Quality of Discretization Schemes in the Second GF-2 Image

Method	Number of interval decreases	Number of correct instances	Quality of discretization scheme
ECRSD	8423	5855	0.9613
EDiRa	5928	5809	0.9361
ChiMerge	8302	5770	0.9473
1R	12361	5574	0.9468
NCAIC	6093	5779	0.9327
FUDC	5905	5816	0.9370
CV-Test	8305	5794	0.9511
Chi2	8360	5768	0.9475

TABLE XI
Comparison of the Quality of Discretization Schemes in Landsat 8 Image

Method	Number of interval decreases	Number of correct instances	Quality of discretization scheme
ECRSD	10491	2621	0.9970
EDiRa	10309	2617	0.9939
ChiMerge	10346	2616	0.9939
1R	10559	2604	0.9918
NCAIC	10285	2616	0.9933
FUDC	10253	2617	0.9934
CV-Test	10379	2616	0.9942
Chi2	10365	2616	0.9941

performance in the number of intervals and the number of data errors. ChiMerge algorithm uses class-attribute information calculation method based on interval similarity to discriminate and merge adjacent intervals. Due to the bottom-up merging strategy, the time overhead is relatively low. Although ChiMerge takes into account the mutual exclusion in the process of merging adjacent intervals, it cannot guarantee the stability of classes within the interval. In order to make the stability of classes within the interval as good as possible, it must be at the cost of increasing the number of intervals. Based on ChiMerge, Cramer's V-Test algorithm divides χ^2 by $\ln(n)$ to reduce the enormous influence of n in the discretization scheme, where n is the number of intervals. Although it can accelerate the process of discretization in some cases, as with ChiMerge, because only the mutual exclusion of adjacent intervals is considered, the number of intervals is larger. Chi2 algorithm has improved on the criterion of the importance of breakpoints, but it lacks the corresponding theoretical basis, the above problems still exist. It can also be

seen from the actual results that the performances of ChiMerge, Cramer's V-Test and Chi2 are not much different. The number of intervals in 1R algorithm is given by the user, but the division criterion of the interval is too simple and lacks flexibility. Although it can obtain a small number of discrete feature intervals in a large-scale data set, it cannot guarantee the mutual exclusion of adjacent intervals and the stability of classes within the interval, which causes great damage to the compatibility of the system. Therefore, we can see that it gets the largest number of data errors. Based on the above analysis, the overall performance of ECRSD is the best in the eight algorithms.

E. Evaluation of Classification Accuracy

The classification accuracy of the remote sensing image is often evaluated using pixel-level calculation [51]. The pixel-level calculation method is to randomly select samples on the classification effect map, and evaluate the classification result by statistical comparison with actual measurement result. The confusion matrix is often used to represent the classification accuracy at the pixel level [52]. The definition of confusion matrix is as follows.

$$CM = \begin{bmatrix} cm_{11} & cm_{12} & \cdot & \cdot & \cdot & cm_{1n} \\ cm_{21} & cm_{22} & \cdot & \cdot & \cdot & cm_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ cm_{n1} & cm_{n2} & \cdot & \cdot & \cdot & cm_{nn} \end{bmatrix} \quad (20)$$

Where, n is the total number of categories in the remote sensing image, and cm_{ij} represents the number of pixels belonging to category i in the test set that are assigned to category j . Obviously, the larger the values of the elements on the diagonal of the confusion matrix, the higher the classification accuracy. Conversely, the lower the classification accuracy. Therefore, we can get the overall average prediction accuracy through the confusion matrix, as shown in (21).

$$P_{Accuracy} = \frac{\sum_{i=1}^n cm_{ii}}{\sum_{i=1}^n \sum_{j=1}^n cm_{ij}} \quad (21)$$

It is actually the ratio of the number of correctly categorized samples to the total number of samples. The overall average prediction accuracy is simple to calculate and has a clear statistical significance. Besides, Kappa coefficient is also widely used in the classification accuracy evaluation of the remote sensing image [53]. It quantifies the overall effectiveness of the classifier on the basis of confusion matrix. The formula for calculating Kappa coefficient is as follows.

$$Kappa = \frac{T \sum_{i=1}^n cm_{ii} - \sum_{i=1}^n (cm_{i+} cm_{+i})}{T^2 - \sum_{i=1}^n (cm_{i+} cm_{+i})} \quad (22)$$

TABLE XII

Comparison of Classification Accuracy on SVM in the First GF-2 Image		
Method	Average accuracy	Kappa coefficient
ECRSD	95.7836%	0.9413
EDiRa	88.3506%	0.8378
ChiMerge	87.1846%	0.8216
1R	84.2581%	0.7809
NCAIC	87.4761%	0.8257
FUDC	91.7027%	0.8845
CV-Test	87.7676%	0.8297
Chi2	87.1730%	0.8215
Original data	87.4178%	0.8249

Original data are image features without discretization.

TABLE XIII

Comparison of Classification Accuracy on Neural Network in the First GF-2 Image		
Method	Average accuracy	Kappa coefficient
ECRSD	94.3261%	0.9210
EDiRa	86.8932%	0.8176
ChiMerge	85.7272%	0.8013
1R	84.2872%	0.7813
NCAIC	86.0187%	0.8054
FUDC	90.2453%	0.8642
CV-Test	86.3102%	0.8094
Chi2	85.7155%	0.8012
Original data	86.8349%	0.8168

TABLE XIV

Comparison of Classification Accuracy on SVM in the Second GF-2 Image		
Method	Average accuracy	Kappa coefficient
ECRSD	89.7865%	0.8773
EDiRa	84.9018%	0.8186
ChiMerge	81.4176%	0.7767
1R	71.8190%	0.6613
NCAIC	81.4860%	0.7775
FUDC	86.3706%	0.8362
CV-Test	83.1939%	0.7980
Chi2	81.4176%	0.7767
Original data	85.3459%	0.8239

TABLE XV

Comparison of Classification Accuracy on Neural Network in the Second GF-2 Image		
Method	Average accuracy	Kappa coefficient
ECRSD	90.0598%	0.8805
EDiRa	85.1751%	0.8218
ChiMerge	81.6909%	0.7800
1R	71.1358%	0.6531
NCAIC	81.7592%	0.7808
FUDC	86.6439%	0.8395
CV-Test	83.4671%	0.8013
Chi2	81.6909%	0.7800
Original data	85.6191%	0.8272

TABLE XVI

Comparison of Classification Accuracy on SVM in Landsat 8 Image		
Method	Average accuracy	Kappa coefficient
ECRSD	93.4376%	0.9214
EDiRa	92.2167%	0.9067
ChiMerge	90.3090%	0.8839
1R	86.8752%	0.8427
NCAIC	91.6063%	0.8994
FUDC	92.7509%	0.9131
CV-Test	89.9275%	0.8793
Chi2	90.1564%	0.8820
Original data	91.0721%	0.8930

Where, T is the total number of pixels used for accuracy evaluation, and n is the number of categories. cm_{ii} is the number of pixels in the i th row and the i th column of the confusion matrix, that is, the number of correctly categorized samples. cm_{i+} and cm_{+i} are the number of pixels in the i th

TABLE XVII

Comparison of Classification Accuracy on Neural Network in Landsat 8 Image

Method	Average accuracy	Kappa coefficient
ECRSD	94.2770%	0.9314
EDiRa	92.5982%	0.9113
ChiMerge	90.6906%	0.8884
1R	87.2568%	0.8473
NCAIC	91.9878%	0.9040
FUDC	93.2850%	0.9195
CV-Test	90.3090%	0.8839
Chi2	90.5380%	0.8866
Original data	91.4536%	0.8976

row and the i th column, respectively. Compared with confusion matrix, Kappa coefficients take into account not only the correctly categorized pixels on the diagonal, but also the errors of omission and commission outside the diagonal. Therefore, the results calculated by the two methods of evaluation, confusion matrix and Kappa coefficient, are usually not equal.

At present, the application of SVM [54] and neural network technology [55] in remote sensing image processing is becoming more and more mature and extensive. They have become an efficient and reliable method for remote sensing image classification. Table XII to XVII show the overall evaluation indicators produced by the analysis of the results of the eight algorithms and the original continuous data on SVM and neural network classifiers for the three images, respectively.

It can be seen that the classification accuracy of ECRSD is the best among the eight algorithms. On the other hand, we can also see that data inconsistency has a greater impact on classification accuracy. The smaller the number of data errors, the higher the classification accuracy on the classifiers. ECRSD and FUDC have fewer data errors than other algorithms, and correspondingly, their classification accuracy is higher than other algorithms. The number of data errors in 1R algorithm is the largest, accordingly, it achieves the lowest accuracy on both SVM and neural network classifiers. Fig. 7 is a classification effect map of the first GF-2 image obtained by ECRSD. It can be seen that the texture of the feature information in the figure is clear, the boundaries of different types of features are more distinct, and there are almost no noise speckles. The eight categories of construction, coniferous forest, broadleaf forest, fallow land, arable land, bare land, water and cloud can be effectively identified. Fig. 8 is a classification effect map of the second GF-2 image obtained by ECRSD. The texture of the feature information in the figure is clear, and the boundaries of different types of features are more obvious. The islands on the sea surface can be effectively identified, and even the hard-to-recognize boundaries between construction and bare land can be separated. Fig. 9 is a classification effect map of Landsat 8 image obtained by ECRSD. The texture of the feature information in the figure is clear, and the boundaries of different types of features are more obvious. Construction and bare land, coniferous forest and broadleaf forest, fallow land and arable land, which are easy to be confused, can be identified effectively. Therefore, our method can achieve excellent results in the classification accuracy of high-resolution remote sensing image in coastal areas.

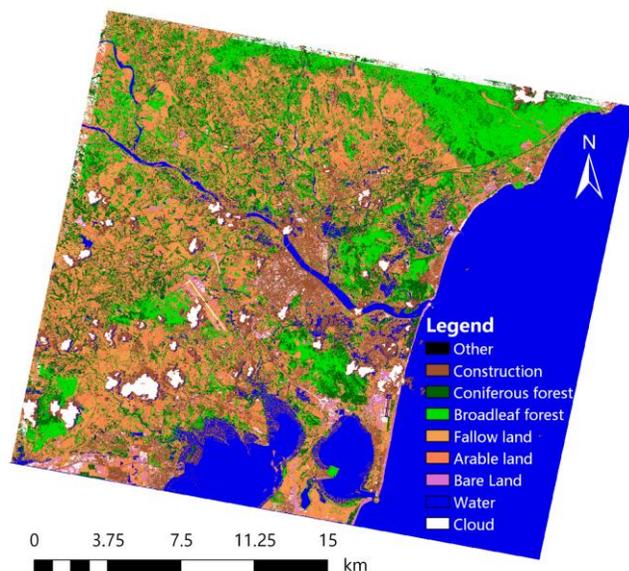


Fig. 7. Classification effect map of the first GF-2 image obtained by ECRSD.

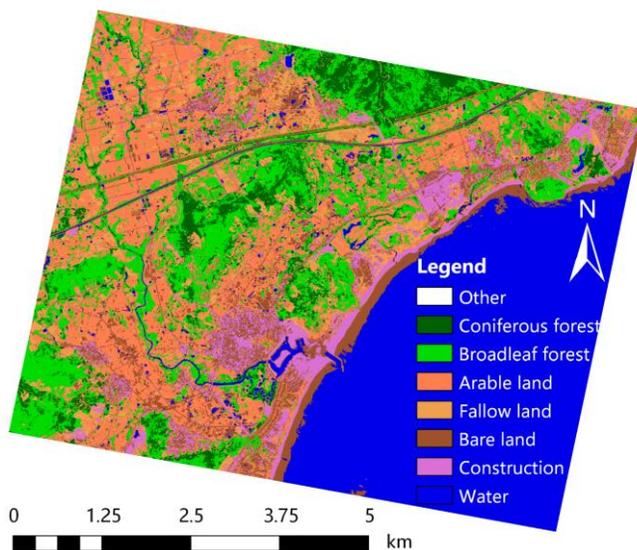


Fig. 8. Classification effect map of the second GF-2 image obtained by ECRSD.

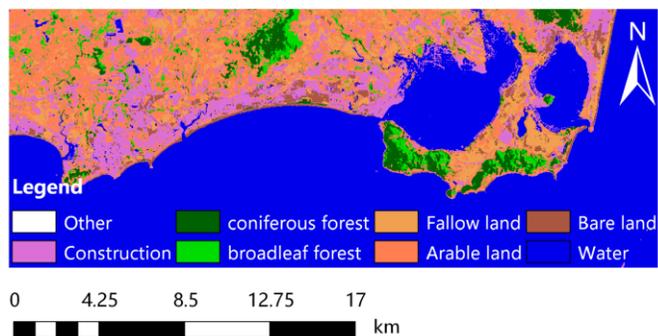


Fig. 9. Classification effect map of Landsat 8 image obtained by ECRSD.

V. CONCLUSION

This paper presents a feature discretization method for classification of high-resolution remote sensing images in coastal areas based on the idea of hybrid metric of information

entropy and chi-square test. We scan each band in turn with the strategy of first splitting then merging by evaluating the stability of the pixel classes within the interval and the similarity of adjacent intervals. Then, we use the degree of dependence among knowledge from the rough set theory as the criterion to evaluate the compatibility of the information system, so as to obtain the optimal discretization scheme. Compared to the statistical methods of multi-channel DN values, ECRSD does not need to rely on the assumption that each class follows a normal distribution in the feature space, and can obtain the knowledge base directly from the data without any prior knowledge or additional information. Moreover, ECRSD can effectively mine the correlation between features, and filter outliers. It has the advantages of strong anti-noise ability and easily generating correct classification rules. So, it is suitable for the discretization of most multi-dimensional data. The simulation results verify the effectiveness of the proposed method. Compared with other best state-of-the-art discretization algorithms, it can get fewer discrete feature intervals and data errors, and achieve better classification results on SVM and neural network classifiers. It not only provides a new idea for data preprocessing of marine high-resolution remote sensing images, but also brings some guidance to the analysis and design of discretization method for high-resolution remote sensing images. Our next step is to test and improve the proposed method in other high-dimensional feature data sets to expand its scope of use, and compare its performance on other classifiers to further optimize the algorithm model.

ACKNOWLEDGMENTS

The authors would like to thank all the anonymous reviewers for their invaluable suggestions. The authors would also like to thank Prof. Hongchao Fan from NTNU Norway for his constructive comments.

REFERENCES

- [1] David A. Kroodsma et al., "Tracking the global footprint of fisheries," *Science*, vol. 359, no. 6378, pp. 904-908, Feb. 2018.
- [2] Hao Wang et al., "Big data and industrial internet of things for the maritime industry in northwestern norway," in *IEEE Region 10 Conf. (TENCON 2015)*, Macao, China, 2015, pp. 1-5.
- [3] Li Xie et al., "Novel classification method for remote sensing images based on information entropy discretization algorithm and vector space model," *Comput. Geosci.*, vol. 89, pp. 252-259, Apr. 2016.
- [4] Jose M. Pena and Roland Nilsson, "On the complexity of discrete feature selection for optimal classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 8, pp. 1517-1522, Aug. 2010.
- [5] Adriana Romero et al., "Unsupervised deep feature extraction for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1349-1362, Mar. 2016.
- [6] Xiang Li et al., "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.
- [7] Sergio Ramírez-Gallego et al., "Data discretization: taxonomy and big data challenge," *Wil. Int. Rev. Data Min. Knowl. Disc.*, vol. 6, no. 1, pp. 5-21, Jan. 2016.
- [8] Salvador Garcia et al., "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.
- [9] Andrew K. C. Wong and David K. Y. Chiu, "Synthesizing statistical knowledge from incomplete mixed-mode data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, no. 6, pp. 796-805, Nov. 1987.
- [10] Xiao Fei Li, "The research on discretization algorithm based on dynamic hierarchical clustering," in *4th Int. Conf. Com. Inf. Sci.*, Chongqing, China, 2012, pp. 349-352.
- [11] Hu Min, "A global discretization and attribute reduction algorithm based on k-means clustering and rough sets theory," in *Sec. Int. Symp. Knowl. Ac. Mod.*, Wuhan, China, 2009, pp. 92-95.
- [12] Mahito Sugiyama and Akihiro Yamamoto, "A fast and flexible clustering algorithm using binary discretization," in *IEEE 11th Int. Conf. Data Min.*, Vancouver, BC, Canada, 2011, pp. 1212-1217.
- [13] Khurram 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.
- [14] John Y. Ching et al., "Class-dependent discretization for inductive learning from continuous and mixed-mode data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 7, pp. 641-651, Jul. 1995.
- [15] L.A. Kurgan and K.J. Cios, "CAIM discretization algorithm," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 2, pp. 145-153, Feb. 2004.
- [16] Cheng-Jung Tsai et al., "A discretization algorithm based on Class-Attribute Contingency Coefficient," *Inf. Sci.*, vol. 178, no. 3, pp. 714-731, Feb. 2008.
- [17] Deqin Yan et al., "A new approach for discretizing continuous attributes in learning systems," *Neurocomputing*, vol. 133, no. 10, pp. 507-511, Jun. 2014.
- [18] Randy Kerber et al., "ChiMerge: discretization of numeric attributes," in *Proc. 10th nat. conf. Artif. int.*, San Jose, California, USA, 1992, pp. 123-128.
- [19] Huan Liu and Rudy Setiono, "Feature selection via discretization," *IEEE Trans. Knowl. Data Eng.*, vol. 9, no. 4, pp. 642-645, Jul. 1997.
- [20] Chaoton Su and Jyhwa Hsu, "An extended Chi2 algorithm for discretization of real value attributes," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 3, pp. 437-441, Mar. 2005.
- [21] Cláudio Rebelo de Sá et al., "Entropy-based discretization methods for ranking data," *Inf. Sci.*, vol. 329, no. C, pp. 921-936, Feb. 2016.
- [22] Guifeng Zhang et al., "A remote sensing feature discretization method accommodating uncertainty in classification systems," in *Proc. 8th Int. Symp. Spat. Acc. Ass. Nat. Res. Env. Sci.*, Shanghai, P. R. China, 2008, pp. 195-202.
- [23] Bo Wu et al., "Feature selection via Cramer's V-test discretization for remote-sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, May. 2014.
- [24] Ying Yang and Geoffrey I. Webb, "Discretization for naive-Bayes learning: managing discretization bias and variance," *Mach. Learn.*, vol. 74, no. 1, pp. 39-74, Jan. 2009.
- [25] Shengyi Jiang and Wen Yu, "A local density approach for unsupervised feature discretization," in *Proc. 5th Int. Conf. Adv. Data Min. Appl. (ADMA '09)*, Beijing, China, 2009, pp. 512-519.
- [26] Robert C. Holte, "Very simple classification rules perform well on most commonly used datasets," *Mach. Learn.*, vol. 11, no. 1, pp. 63-90, Apr. 1993.
- [27] J. Catlett, "On changing continuous attributes into ordered discrete attributes," in *Proc. Eur. work. sess. Learn. Mach. Learn. (EWSL-91)*, Porto, Portugal, 1991, pp. 164-178.
- [28] Chang-Hwan Lee, "A Hellinger-based discretization method for numeric attributes in classification learning," *Knowledge-Based Syst.*, vol. 20, no. 4, pp. 419-425, May. 2007.
- [29] Di Wu et al., "Study on the assessment method of typhoon regional disaster based on the change of chlorophyll-a concentration in seawater," in *OCEANS*, Aberdeen, UK, 2017, pp. 1-7.
- [30] Autun Purser et al., "Ocean floor observation and bathymetry system (OFOBS): a new towed camera/sonar system for deep-sea habitat surveys," *IEEE J. Ocean. Eng.*, vol. 44, no. 1, Jan. 2019.
- [31] Akira Nagano et al., "Ocean-atmosphere observations in philippine sea by moored buoy," in *OCEANS*, Kobe, Japan, 2018, pp. 1-6.
- [32] Gou S et al., "Coastal zone classification with fully polarimetric sar imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 11, pp. 1616-1620, 2016.
- [33] Di Wu et al., "Strategy for assessment of disaster risk using typhoon hazards modeling based on chlorophyll-a content of seawater," *EURASIP J. Wirel. Commun. Netw.*, vol. 2018, no. 1, pp. 293, Dec. 2018.
- [34] Liu Y et al., "Roadnet: learning to comprehensively analyze road networks in complex urban scenes from high-resolution remotely sensed

- images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2043-2056, 2019.
- [35] Sanches I D et al., “Campo verde database: seeking to improve agricultural remote sensing of tropical areas,” *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 369-373, 2018.
- [36] Qiong Chen et al., “A feature preprocessing framework of remote sensing image for marine targets recognition,” in *OCEANS*, Kobe, Japan, 2018, pp. 1-5.
- [37] Simon H.A., *The Sciences of the Artificial*, 2nd ed., Cambridge, MA: MIT Press, 1981.
- [38] Dougherty J et al., “Supervised and unsupervised discretization of continuous features,” in *Proc. 12th Int. Conf. Mach. Learn.*, Tahoe City, California, 1995, pp. 194-202.
- [39] Hsu C et al., “Why discretization works for naive bayesian classifiers,” in *Proc. 17th Int. Conf. Mach. Learn. (ICML 2000)*, Stanford University, USA, 2000, pp. 399-406.
- [40] C. E. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379-423, Jul. 1948.
- [41] Usama M. Fayyad and Keki B. Irani, “On the handling of continuous-valued attributes in decision tree generation,” *Mach. Learn.*, vol. 8, no. 1, pp. 87-102, Jan. 1992.
- [42] Rong Wang et al., “Chi-square and SPRT combined fault detection for multisensor navigation,” *IEEE Trans. Aerosp. Elec. Syst.*, vol. 52, no. 3, pp. 1352-1365, Jun. 2016.
- [43] K. Lavangananda and S. Chattanachot, “Study of discretization methods in classification,” in *9th Int. Conf. Knowl. Smart Tech. (KST)*, Chonburi, Thailand, 2017, pp. 50-55.
- [44] Yuanyuan Ma et al., “Selection of rich model steganalysis features based on decision rough set α -positive region reduction,” *IEEE Trans. Circ. Syst. Video Tech.*, vol. 29, no. 2, pp. 336-350, Feb. 2019.
- [45] Yanhua Sun et al., “Full diversity reception based on Dempster-Shafer theory for network coding with multiple-antennas relay,” *China Commun.*, vol. 12, no. 10, pp. 76-90, Oct. 2015.
- [46] G. A. Vijayalakshmi Pai, “Fuzzy decision theory based metaheuristic portfolio optimization and active rebalancing using interval type-2 fuzzy sets,” *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 2, pp. 377-391, Apr. 2017.
- [47] Swarnajyoti Patra et al., “Hyperspectral band selection based on rough set,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5495-5503, Oct. 2015.
- [48] S. Rosati et al., “ChiMerge discretization method: impact on a computer aided diagnosis system for prostate cancer in MRI,” in *IEEE Int. Symp. Med. Meas. Appl. Proc. (MeMeA)*, Turin, Italy, 2015, pp. 297-302.
- [49] Zulfiqar Ali and Waseem Shahzad, “Comparative study of discretization methods on the performance of associative classifiers,” in *Int. Conf. Front. Inf. Tech. (FIT)*, Islamabad, Pakistan, 2016, pp. 87-92.
- [50] Wenyu Qu et al., “A novel Chi2 algorithm for discretization of continuous attributes,” in *Proc. 10th Asia-Pacific web conf. Progr. WWW res. dev. (APWeb'08)*, Shenyang, China, 2008, pp. 560-571.
- [51] Ammar Mahmood et al., “Deep image representations for coral image classification,” *IEEE J. Ocean. Eng.*, vol. 44, no. 1, pp. 121-131, Jan. 2019.
- [52] Miho Ohsaki et al., “Confusion-matrix-based kernel logistic regression for imbalanced data classification,” *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 9, pp. 1806-1819, Sept. 2017.
- [53] Jingge Xiao et al., “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.
- [54] Yiqing Guo et al., “Effective sequential classifier training for SVM-based multitemporal remote sensing image classification,” *IEEE Trans. Image Proc.*, vol. 27, no. 6, pp. 3036-3048, Jun. 2018.
- [55] Carlos Bentes et al., “Ship classification in TerraSAR-X images with convolutional neural networks,” *IEEE J. Ocean. Eng.*, vol. 43, no. 1, pp. 258-266, Jan. 2018.



Qiong Chen received his B.Eng. degree at Beijing University of Posts and Telecommunications, P.R. China, 2007 and M.Eng. degree at Politecnico di Torino, Italy, 2012. Now he is a Ph.D. student at College of Information Science and Technology, Hainan University, P.R. China. 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-President 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.