

Diversity-Driven Multikernel Collaborative Representation Ensemble for Hyperspectral Image Classification

Hongjun Su¹, Senior Member, IEEE, Yuzheng Hu, Hongliang Lu², Weiwei Sun³, Senior Member, IEEE, and Qian Du⁴, Fellow, IEEE

Abstract—Recently, kernel collaborative representation classification (KCRC) has shown its outstanding performance in dealing with the problem of linear inseparability in hyperspectral remote sensing image classification. Meanwhile, ensemble learning has attracted great attention in improving the performance of a single classifier. Aiming at the limitation of a single classifier, bagging algorithm based on KCRC (KCRC-bagging) is presented in this article. The KCRC-bagging method uses bootstrap to increase the diversity of base classifiers, thus improving the classification accuracy and generalization performance. In order to reduce the scale of ensemble, a diversity-driven multikernel collaborative representation classifier ensemble approach (DIV-KCRC) is proposed. DIV-KCRC verifies the effectiveness of the representation classifier with the pair of diversity measures, and classifiers with high accuracy and diversity are selected to improve the classification performance and efficiency of the ensemble system. Three real hyperspectral data sets were applied to prove the validity of the proposed method. The experimental results demonstrate that both KCRC-bagging and DIV-KCRC can yield better classification performance than their corresponding base classifiers. In particular, DIV-KCRC provides more reliable classification results than KCRC-bagging.

Index Terms—Bagging, diversity measure, ensemble learning, kernel collaborative representation classification (KCRC).

I. INTRODUCTION

HYPERSPECTRAL remote sensing images can obtain information with hundreds of continuous spectral bands of surface objects, and provide rich spectral information to enhance the ability to distinguish ground objects [1], [2]. Hyperspectral

remote sensing plays an important role in the diversity applications, such as national defense, environmental monitoring, agriculture, etc. Among them, supervised classification is a critical process for hyperspectral remote sensing image applications. However, the limited training samples, uneven quality, and high dimensionality results in great challenges for hyperspectral images classification [3].

In view of the above-mentioned problems, support vector machine (SVM) and extreme learning machine (ELM) are applied to hyperspectral image classification and achieved meaningful performance. However, a single classifier is often limited and cannot accomplish the best performance with complex and diverse scenarios of hyperspectral images. Therefore, how to integrate the advantages of different classifiers and achieve the effect of “1+1>2” is an important research direction. Ensemble learning has attracted much attention for its ability to use the advantages of multiple classifiers to complete the final decision and has shown promising performance than individual classifiers. Such as the classic Bagging and Boosting [4], and the improved ensemble method: a scalable end-to-end tree boosting system (XGBoost) [5], a highly efficient gradient boosting decision tree (LightGBM) [6], etc.

At present, ensemble learning is investigated from two aspects: improving the base classifier, and taking advantage of the diversity between classifiers. To improve the performance of the base classifiers, that is, to design a strong classifier suitable for hyperspectral image classification. Some traditional machine learning models have been used for hyperspectral image classification, and the most representative ones of which are SVM and ELM. SVM is one of the representative algorithms of the kernel transformation technique. Its main idea is to transform low-dimensional linear non-separable problems into a high-dimensional space for accurate classification. Melgani and Bruzzone were the first to test the application of SVM in the field of hyperspectral remote sensing image classification [7]. In order to further improve the performance of SVM, SVM-based extensions such as hybrid kernel SVM have been introduced, which have achieved certain improvements, but the selection of kernel function and optimal parameter combination is still a difficulty [8]. ELM has proven to be effective for high-dimensional data [9]. The output weights of the learning network can be obtained only through one-shot calculation,

Manuscript received January 13, 2022; revised February 24, 2022 and March 10, 2022; accepted March 18, 2022. Date of publication March 24, 2022; date of current version April 20, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 42122008 and Grant 41871220. (Corresponding author: Hongjun Su.)

Hongjun Su, Yuzheng Hu, and Hongliang Lu are with the School of Earth Sciences and Engineering, Hohai University, Nanjing 211100, China, and also with the Jiangsu Province Engineering Research Center of Water Resources and Environment Assessment using Remote Sensing, Hohai University, Nanjing 211100, China (e-mail: hjsu@hhu.edu.cn; huyuzheng116@163.com; hllu_hhu@163.com).

Weiwei Sun is with the Department of Geography and Spatial Information Techniques, Ningbo University, Ningbo 315201, China (e-mail: sunweiwei@nbu.edu.cn).

Qian Du is with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS 39762 USA (e-mail: du@ece.msstate.edu).

Digital Object Identifier 10.1109/JSTARS.2022.3162121

which effectively improves the parameter setting and improves computing efficiency in the training process, and has stronger generalization ability than SVM. Convolutional neural network methods is also widely used [10], [11]. All of the above methods need complex parameters. In recent years, the representation model has been widely concerned because of its advantages of simplicity and few parameters [12]. The core idea of this classifier is that the class labels of testing samples can be represented by a dictionary of labeled samples linearly. The representation coefficients are solved by different regularization functions, namely, sparse representation (SR) with l_0 -norm or l_1 -norm minimization constraints [12], [13], and collaborative representation (CR) using l_2 -norm minimization constraints. CR has attracted more attention due to its lower computational complexity and better classification performance [13], [14]. For example, the proposed novel CR-based nearest neighbor algorithms in [15] uses the weight coefficient of CR to find the truly closest training sample for each testing sample, which outperforms the traditional Euclidean distance. Tangent distance-based CR for classifier (TCRC) in [16], using manifold learning to project hyperspectral data into a simplified tangent space to achieve better performance. In addition to linear methods, some nonlinear CRC methods that use nuclear techniques are also proposed. For example, kernel CR with Tikhonov regularization is proposed in [17] incorporates neighborhood information into a kernel space to increase separability between classes, a novel kernel nonlocal joint CR classification method uses the similarity measure between spectral pixels as the mapping feature in [18], [19], and a multiple kernel CRC incorporates multiple kernel functions, enhance CRC performance [20], [21]. Liu *et al.* [22] proposed a probabilistic kernel cooperative representation classifier to solve the problem that the original KCRC could not do. Karaca *et al.* [23] designed a spatial aware probabilistic multikernel CR to overcome the problem of small samples in hyperspectral classification. Ma *et al.* [24] designed a discriminative kernel CR and Tikhonov regularization method to improve the utilization rate of correlation between different categories, which effectively improved the separability between samples. Su *et al.* [25] designed a shape-adaptive neighborhood KCRC model to explore the nonlinear characteristics of spatial features, which effectively improved the performance of the kernel cooperative representation model. Tu *et al.* [26] used the density peak to improve the super pixel model and extract local spatial information of images and designed a structural-kernel CR algorithm, which effectively improved the reliability of extracted spatial information. An optimized single classifier provides more choices of base classifiers for ensemble learning, but there is still room for further improvement.

Another idea is to increase and utilize the diversity among classifiers for ensemble learning [27], [28]. Bagging and boosting, the most typical ones, create diversity by changing the training sample subset. SVM and ELM are often integrated as base classifiers. Among them, rotation-based SVM integration uses data transformation and random features to generate training results of multiple SVMs to participate in the final decision [29]. Bagging-based ELM and Adaboost-based ELM use ELM as the base classifier for bagging and boosting respectively,

which overcomes the problem in the original ELM caused by randomness of input weights [9], [21]. Artificial neural networks are also used in ensemble learning as weak classifiers. SR has been used as a base classifier in random subspace (RS) and bagging ensemble methods in the area of signal processing [32], [33]. CRC-bagging subsequently proved to be effective in hyperspectral classification [34]. Tangent space corepresentation projects hyperspectral data to a simplified tangent space to achieve better performance. Besides, TCRC is sensitive to training data and parameters, so TCRC combined with bagging and boosting can show better performance than CRC-bagging. It also effectively solves the problems of limited accuracy and weak generalization performance in a single TCRC [35]. In order to effectively utilize the diversity information of features, Chen *et al.* combined multifeature and Adaboost algorithm to extract Gabor features, gray level co-occurrence matrix, and extended morphological profile (EMP) features for stack, and each can increase the diversity between base classifiers by using different spatial features [36], [37]. Chen *et al.* [38] also proposed an SVM algorithm based on RS, which effectively improved the classification accuracy and obtained feature sequences by feature selection. Then, feature sequences were divided into sections and randomly sampled according to a sampling ratio, and the performance and diversity of base classifiers were also improved. In order to improve the computational efficiency, an adaptive shape neighborhood RS-based k-nearest class CR with Tikhonov algorithm was proposed by combining CR and RS [39], and the spatial structure information of training and testing samples was further explored.

It is worth noting that in addition to creating diversity, it is also important to quantify the diversity between classifiers [40]–[42]. Bi *et al.* studied the impact of diversity on the accuracy of classifier ensemble, and proved that effective diversity indicators can improve ensemble accuracy while reducing the integration scale [41], [42]; Kuncheva *et al.* [43] used nine measures to test the performance of two classic algorithms, bagging and boosting, and experiments show that boosting can produce more diversity even for stable classifiers. Nan Li *et al.* used voting and paired difference evaluation indicators to measure the difference between classifiers and ensemble pruning, reducing the integration scale, and improving the calculation efficiency [26], [28], [44]. Zhao *et al.* [45] proposed a classifier ensemble difference measure based on complementary information entropy under fuzzy relation. It measures the uncertainty contained in the classification data space, and the difference between base classifiers is gauged according to the information, which solves the problem that fuzzy data cannot be directly processed in the diversity assessment of multiclassifier systems. A heuristic classifier ensemble algorithm considering sparsity and diversity is proposed in [46]. In addition, Tan *et al.* [47] combined the three diversity metrics of different (D), double fault (DF), and correlation coefficient (ρ) to evaluate the difference between SVM, k-nearest neighbor, multinomial logistic regression, and ELM, and select the three most diverse classifiers for integration, which effectively improves the classification accuracy.

From the above-mentioned analysis, the first problem is the existing differentiation-guided classifier ensemble methods are

mostly verified on UCI data sets, and the effect of hyperspectral feature classification is unknown. Moreover, traditional machine learning models are mainly used. In other words, the base classifier involved in the ensemble is relatively simple. Another problem is that although the ensemble learning based on the representation model classifier can increase the diversity of the ensemble system, but without the evaluation on diversity. There is no relevant research on the diversity of representing model classifier. Therefore, how to solve the problem of diversity and precision balance in ensemble learning is very important, and how to choose a strong classifier with diversity as the basic classifier is also the key.

In this article, a bagging ensemble learning algorithm based on KCRC (KRC-bagging) and a diversity-driven multikernel CR classifier ensemble (DIV-KCRC) method for hyperspectral image classification are proposed. KCRC-bagging selects the KCRC as the base classifier and uses the Bootstrap sampling method to generate training subsets with differences, and then creates a series of classifier combinations with diversity to participate in the ensemble to improve the generalization performance of KCRC. The DIV-KCRC explores the diversity among CR classifiers guided by different kernel functions and effectively utilizes the advantages of KCRC. In this method, the CR classifiers of different kernel functions are combined as base classifiers, and the differences among base classifiers are evaluated by four diversity measures. With the verification accuracy and diversity evaluation results, the optimal combination is selected to participate in the integration for the final classification result. Through the final classification results, the effect of several diversity indicators is analyzed, and then the diversity indicators with good effects are selected, and the validity of the selected indicators is validated on other data sets. The main contributions of this article are as follows.

- 1) The novel nonlinear ensemble learning method combining KCRC and bagging ensemble strategy is presented for hyperspectral image classification. The KCRC-bagging improves the CRC-bagging by combining the kernel cooperative representation classifier with the bagging ensemble method and can solve the problem of the linear inseparability of low-dimensional space more effectively.
- 2) The differences between different KCRC is evaluated by the diversity measures for the first time, and multikernel CRs with more diversity were used in ensemble learning, making full use of the advantages of different KCRC. Meanwhile, the difference in the ensemble system is ensured through the guidance of diversity.
- 3) The prior knowledge of accuracy and diversity is added before the ensemble to avoid base classifiers with poor performance in the ensemble, which reduces the ensemble scale and improves the generalization performance.

The rest of this article is organized as follows. In Section II, a detailed explanation of CRC, KCRC, pairwise diversity measures, and diversity in ensemble learning are presented. Section III is dedicated to introducing the proposed KCRC-bagging and introduced the DIV-KCRC algorithm to reduce ensemble complexity. Section IV introduces the experiment and analysis of three real hyperspectral data. Section V discussed

the parameters selection of our method. Finally, Section VI concludes this article.

II. THEORETICAL BASIS

A. CRC and KCRC

Let training data be denoted as $X \in R^{F \times N}$ (which contains F bands and N samples for M classes). Let the dictionary D be constructed using M different subdictionaries as $[D_1, D_2, \dots, D_M]$. The subdictionary $D_n = \{x_{ni}\}_{i=1}^{N_n}$, $n \in \{1, 2, \dots, M\}$, and $\sum_{n=1}^M N_n$. In CRC, an approximation of a testing sample $y \in R^{F \times 1}$ can be represented via the linear combination of atoms. The best approximation in the subspace is regarded as the class of sample y , which can be expressed as

$$\text{class}(y) = \arg \min_{m=1,2,\dots,M} \{\|y - D_m \alpha_m\|_2 / \|\alpha_m\|_2\} \quad (1)$$

where the coefficient vector α can be estimated by solving the following optimization problem with the l2-norm regularization constraint:

$$\alpha = \arg \min_{\alpha^*} \left\{ \|y - D \alpha^*\|_2^2 + \lambda \|\alpha^*\|_2^2 \right\} \quad (2)$$

here λ is the regularization parameter as the tradeoff between the residual term and regularization. The analytical solution to (2) is

$$\alpha = (D^T D + \lambda I)^{-1} D^T y \quad (3)$$

where I is an identity matrix. After obtaining α , the class label of y can be determined according to (1). However, samples of HSI may be linearly indivisible in the original data space due to the complex scene environment and mixed pixels. A better approach is to use the kernel trick to map data from a lower dimensional space to a higher one to increase class separability. Define the nonlinear mapping $\Phi: R^F \mapsto H$ corresponding to a kernel function $K(d_i, d_j) = \Phi(d_i)^T \Phi(d_j)$. In the KCRC, the optimization problem in the kernel-induced space is rewritten as

$$\alpha = \arg \min_{\alpha^*} \left\{ \|\Phi(y) - \Phi(D \alpha^*)\|_2^2 + \lambda \|\alpha^*\|_2^2 \right\} \quad (4)$$

where $\Phi(y)$ is the mapped testing sample y , and training atoms X mapping into $\Phi(D)$. The closed-form solution to α according to (3) is reformulated as

$$\alpha = \left(\Phi(D)^T \Phi(D) + \lambda I \right)^{-1} \Phi(D)^T \Phi(y) \quad (5)$$

using the kernel function $K(d_i, d_j)$, the coefficient α can be denoted as

$$\alpha = (K(D, D) + \lambda I)^{-1} K(D, y) \quad (6)$$

where $K(D, D)$ is an $N \times N$ Gram matrix and $K(D, y)$ is an $N \times 1$ vector for the inner products between $\Phi(y)$ and $\Phi(D)$. Similar to (1), the label of sample y is determined by

$$\begin{aligned} \text{class}(y) &= \arg \min_{m=1,2,\dots,M} \{\|\Phi(y) - \Phi(D_m) \alpha_m\|_2 / \|\alpha_m\|_2\} \\ &= \arg \min_{m=1,2,\dots,M} \{(K(y, y) + \alpha_m^T K(D_m, D_m) \alpha_m \\ &\quad - 2\alpha_m^T K(D_m, y)) / \|\alpha_m\|_2\}. \end{aligned} \quad (7)$$

Algorithm 1: KCRC-Bagging.

Input: KCRC: base classifier, K : Ensemble times
 $X \in \mathbf{R}^{F \times N}$ Input training sets, λ : regularized parameter
 y : testing samples.
For $k = 1$ to K
 Get new samples $X_k \in \mathbf{R}^{F \times N}$ from X using Bootstrap
 Get a new dictionary $D_K = \{D_{1k}, D_{2k}, \dots, D_{Tk}\}$
 by new differentiated training set X_k .
 Calculating the class label class (y) for the testing sample
 y according to Eq. (4-7) on the new samples
End for
 Return label matrix: $l = [l_1, l_2, \dots, l_K]$
 Received the final classification result using the majority
 voting rule
Output: class (y) = mode (l)

TABLE I
 CLASSIFICATION RESULTS OF CLASSIFIER C_i AND CLASSIFIER C_j

	C_j correct (1)	C_j false (0)
C_i correct (0)	a	b
C_i false (0)	c	d

B. Pairwise Diversity Measures

Generally speaking, diversity measures can be divided into two categories: pairwise and nonpairwise [48].

Suppose the classifier set $C = \{C_1, C_2, \dots, C_P\}$, where P is the number of classifiers in the classifier set. Let a, b, c, and d respectively represent the number of correct and incorrect sample classification by classifiers, and the details are shown in Table I. The total number is $T = a + b + c + d$.

1) Q statistics: For any two classifiers $C_i, C_j \in C$, Q is calculated as

$$Q = \frac{ad - bc}{ad + bc} \quad (8)$$

and $Q \in [-1, 1]$. For statistically independent classifiers $Q = 0$.

The overall diversity calculation method Q_{av} of classifier set C is

$$Q_{av} = \frac{2}{P(P-1)} \sum_{i=1}^{P-1} \sum_{j=i+1}^P Q_{ij}. \quad (9)$$

The diversity of classifier sets decreases with the increasing of Q [49].

2) *Correlation Coefficient*: For classifiers $C_i, C_j \in C$, Cor is defined as

$$Cor_{ij} = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}. \quad (10)$$

Correlation Coefficient (Cor) and Q are the same as positive and negative, and it can be proved that $|Cor| \leq |Q|$.

Algorithm 2 DIV-KCRC

Input: KCRC-RBF, KCRC-LA, KCRC-Linear, KCRC-Pairwise, KCRC-Poly: base classifiers.
 $D = [D_1, D_2, \dots, D_K]$ λ : regularized parameter,
 y : testing sample, k : the number of pre-defined class
For $p = 1$ to P
 1. Calculate the distance between y and training samples per class
 2. Sort **dis** to find the k -nearest classes of y to obtain V
 3. Computing the validation label class (v) for the verification sample v according to Eq. (4-7).
 4. Calculating the **Diversity** between classifiers separately according to Eq. (8-15).
 5. Sort **Diversity** to selected the most different classifier combination.
End for
 Return label matrix: $l = [l_1, l_2, \dots, l_C]$
 Get the final classification result by majority voting rules
Output: class (y) = mode (l)

The diversity calculation method Cor_{av} of classifier set C is

$$Cor_{av} = \frac{2}{P(P-1)} \sum_{i=1}^{P-1} \sum_{j=i+1}^P Cor_{ij}. \quad (11)$$

The diversity of classifier sets decreases with the increasing of Cor [40].

3) *Disagreement Measurement*: The disagreement measurement (Dis) between C_i and C_j can be computed as

$$Dis_{ij} = \frac{b+c}{a+b+c+d} \quad (12)$$

and the Dis [40] of the entire classifier combination can be obtained as

$$Dis_{av} = \frac{2}{P(P-1)} \sum_{i=1}^{P-1} \sum_{j=i+1}^P Dis_{ij}. \quad (13)$$

4) *Double Fault Measurement*: The misclassification probability of both classifiers can be measured by double fault measurement (DF) as

$$DF_{ij} = \frac{d}{a+b+c+d}. \quad (14)$$

The overall diversity DF of the classifier system is expressed as

$$DF_{av} = \frac{2}{P(P-1)} \sum_{i=1}^{P-1} \sum_{j=i+1}^P DF_{ij} \quad (15)$$

The diversity of classifier sets increases with the decrease of DF [49].

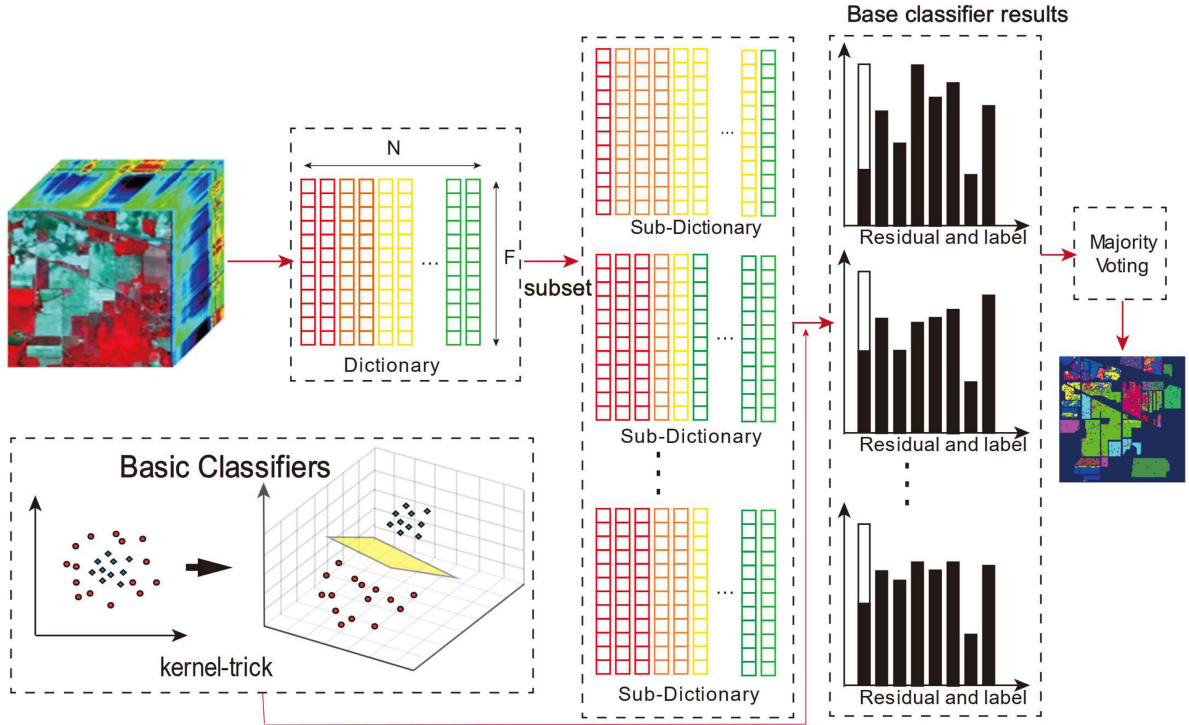


Fig. 1. Illustration of KCRC-bagging for hyperspectral image classification.

C. Diversity in Ensemble Learning

The diversity between classifiers is a necessary condition to improve the generalization performance of ensemble learning.

Obviously, if the error rates of the individual classifiers participating in the ensemble are highly correlated, then the ensemble is meaningless. Taking binary classification as an example, assuming that the error rates of the individual classifiers participating in the ensemble are independent of each other, it can be known from the Hoeffding [49] inequality that with the increase of the ensemble times K , the error rate of ensemble eventually tends to zero, namely

$$\begin{aligned}
 P(EL(x) \neq Y(x)) &= \sum_{t=0}^{\lfloor K/2 \rfloor} \binom{T}{t} (1 - \varepsilon)^t \varepsilon^{K-t} \\
 &\leq \exp\left(-\frac{1}{2}K(1 - 2\varepsilon)^2\right) \quad (16)
 \end{aligned}$$

where ε is the error rate of the base classifier, $EL(x)$ represents the decision result of the ensemble system, and $Y(x)$ represents the true value. However, (16) holds on the premise that the error rates of the base classifiers are independent of each other, which does not exist in real-world tasks. In the actual learning process, the base classifiers participating in the ensemble often have a unified learning objective, which determines the inevitable relationship between them. In fact, there is a certain conflict between the accuracy and diversity of base classifiers in ensemble learning. Generally, after the variance is high, increasing the accuracy will sacrifice the variance; on the contrary, in order to ensure the accuracy, a part of the

diversity is often sacrificed. This phenomenon can be explained more intuitively from the perspective of error-ambiguity decomposition, which was proposed by Krogh and Vedelsby [28], [48], [49].

Assuming the classifier set $C = \{c_1, c_2, \dots, c_K\}$, in the regression learning task, for x , the error-ambiguity decomposition of the ensemble system that completes the decision through the weighted average method can be defined as

$$Err(EL) = \overline{Err} - \overline{AM} \quad (17)$$

among them, $\overline{Err} = \sum_{i=1}^K \omega_i Err_i$ and $\overline{AM} = \sum_{i=1}^K \omega_i AM_i$ represent the generalization error and weighted divergence value of c_i .

Since the definition of $Err(EL)$ is based on the entire sample space and \overline{AM} cannot be obtained a priori, $Err(EL)$ is difficult to optimize directly. It should also be noted that the above derivation is only applicable to regression learning, and it is difficult to directly generalize to classification learning tasks.

III. PROPOSED METHOD

A. KCRC Bagging

In this work, KCRC-bagging is proposed using the KCRC as the based classifier within the bagging framework. Benefited from the Bootstrap sampling method, a series of training sample sets $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K\}$ is generated randomly from the raw training set \mathbf{X} , where K represents the number of ensemble size. Then, a subtraining set constructs a discrepant dictionary $\mathbf{D}_K = \{\mathbf{D}_{1k}, \mathbf{D}_{2k}, \dots, \mathbf{D}_{Tk}\}$ in a new subdata \mathbf{X}_k . Therefore, each training subset trains a KCRC model, and the testing samples are classified by K representation coefficients α gained from

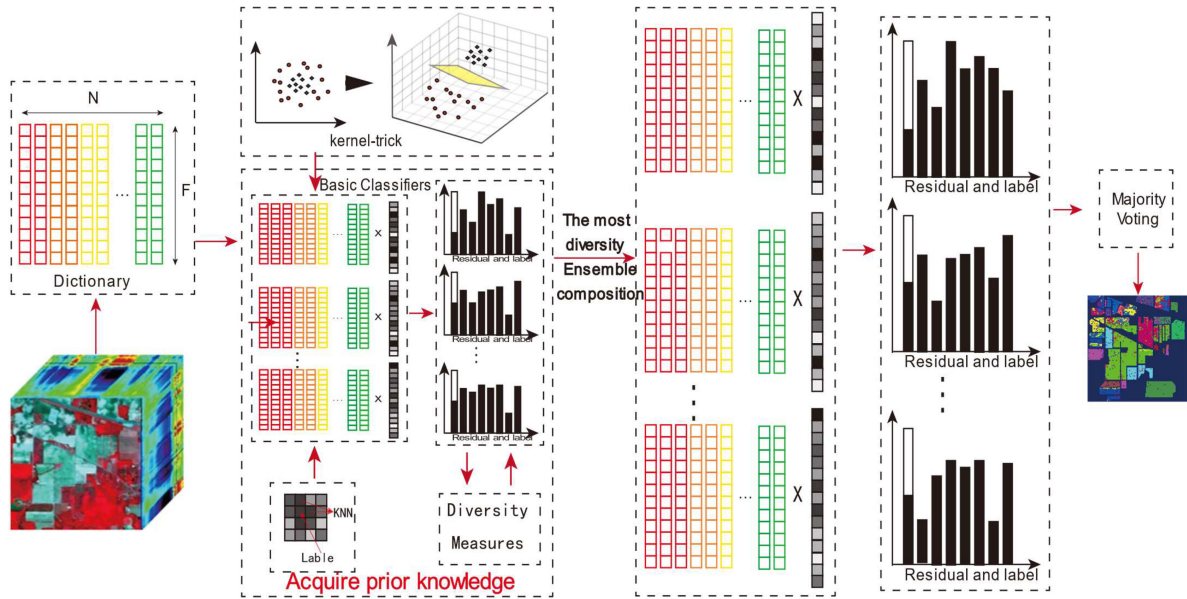


Fig. 2. Illustration of DIV-KCRC for hyperspectral image classification.

each bag. Finally, a set of different base classifier combinations are obtained. Each classifier may vary in the same category. The final classification results are generated by combining the results of all bags using the majority voting rules. The main steps of KCRC-bagging are listed in Algorithm 1 and its process chart is shown in Fig. 1.

B. Kernel CR Classifiers Ensemble Based on Diversity

Although bagging increases the overall diversity of the integrated system by generating different training samples, it lacks the diversity evaluation between the base computational efficiency and poor integration effects.

In this section, a heterogeneous integration method for different kernel cooperative representation classifiers with diversity evaluation is proposed. The classifier pool $C = \{C_1, C_2, \dots, C_P\}$ consists of CRC models guided by different kernel functions. Meanwhile, the difference index is used to evaluate the classification similarity among different classifier models.

The higher the similarity, the smaller the diversity of the ensemble. The combination of accuracy and diversity is combined to determine the combination of base classifiers. First, the EMP feature of spectral data is extracted, and the spectral feature and EMP feature are divided into training and testing sets. Then, a fixed number of training sets were randomly selected from each class. Assume the k -nearest pixels similar to the testing sample belong to the same category and are taken as the verification set. A priori knowledge of the accuracy and diversity of basic classifiers is obtained through validation sets. Finally, using the diversity measures and the overall classification accuracy based on the verification set, a best group can be found. Finally, obtaining the classification result by the majority voting rule. It is worth noting that the most varied combinations were determined by a combination of four indicators. If the results of the four diversity

TABLE II
NINE GROUND-TRUTH CLASSES OF THE ROSIS DATA SET

Class	Name	Samples
1	Asphalt (Asp)	6631
2	Meadows (Mea)	18649
3	Gravel	2099
4	Trees	3064
5	Painted metal sheets (PMS)	1345
6	Bare Soil (Bas)	5029
7	Bitumen (Bit)	1330
8	Self-Blocking Bricks (SBB)	3682
9	Shadows (Sha)	947
Total		9345

indexes are inconsistent, the combination with high precision is taken as the final output result. Algorithm 2 shows the key steps of DIV-KCRC and its flow diagram is shown in Fig. 2.

IV. EXPERIMENTS

A. Hyperspectral Data Sets

1) *Pavia University*: The data set of University of Pavia provides 103 bands in the 0.43–0.86 μm wavelength range, excluding noise and water absorption bands. This data set is located in Pavia, Italy, and was acquired in 2003 by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor, and the size of the data is 610×340 pixels, which includes classes. The detailed information of each class is described in Table II, and images of this data set are shown in Fig. 3.

2) *Purdue Campus*: The second data set, Purdue Campus, was obtained by an airborne hyperspectral mapper (HYMAP) sensor. This data provides 126 bands in the spectrum of 0.45–2.48 μm , after bad band removal. The data size is 377×512 pixels under the 3.6 m spatial resolution. Six classes are contained in the HYMAP data set shown in Table III, and the false color image and ground-truth image are shown in Fig. 4(a) and (b).

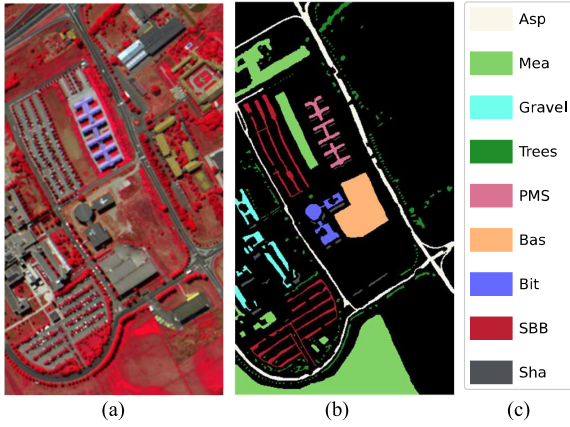


Fig. 3. (a) False-color image (b) Ground truth of the Pavia University data set. (c) Class legend.

TABLE III
NINE GROUND-TRUTH CLASSES OF THE ROSIS DATA SET

Class	Name	Samples
1	Road	1287
2	Grass	1114
3	Shadow	219
4	Soil	379
5	Tree	1351
6	Roof	1285
Total		5635

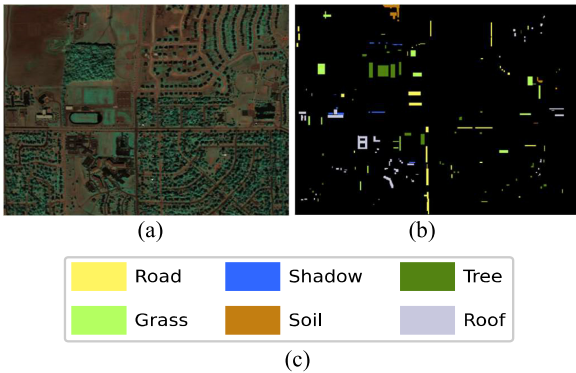


Fig. 4. (a) False-color image. (b) Ground truth of the Purdue data set. (c) Class legend.

3) *Yellow River Delta*: The third data set is Yellow River Delta, which contains 330 spectral bands from visible light to shortwave infrared. The data of the remaining 285 bands after removing the 45 water absorption bands participated in the experiment. The Yellow River Delta hyperspectral data was collected on January 7, 2019, by the Gaofen-5 with a spatial resolution of 30 m, and the image size of the experimental area is 1185×1324 . It is a typical coastal wetland area in Dongying City, Shandong Province, China ($36^{\circ}55' - 38^{\circ}16'N$, $117^{\circ}31' - 119^{\circ}18'E$) [50], [51]. The data features complex types, including 21 types of features mainly composed of artificial vegetation, wet vegetation, halophyte vegetation, and complex water bodies, and the spatial distribution of various features presents a block structure, which is of great representative significance [52], [53]. The detailed information of each class is described in Table IV, and images of this data set are shown in Fig. 5.

TABLE IV
TWENTY-ONE CLASSES OF THE YELLOW RIVER DELTA

Class	Name	Samples
1	Acquaculture (Acq)	393
2	Seep sea (Ses)	796
3	Soybean (Soy)	110
4	Rice	190
5	Building (Bui)	83
6	Maize	96
7	Broomcorn (Bro)	95
8	Locust	211
9	Spartina (Spa)	200
10	Shallow (Sha)	936
11	Mud flat (Muf)	553
12	River	469
13	Suaeda salsa (Sus)	361
14	Reed	240
15	Salt marsh (Sam)	595
16	Intertidal saltwater (Ins)	454
17	Tamarix(Tam)	133
18	Pond	377
19	Flood plain(Flp)	68
20	Freshwater herbaceous marsh (FHM)	72
21	Aquatic vegetation (Aqv)	39
Total		6471

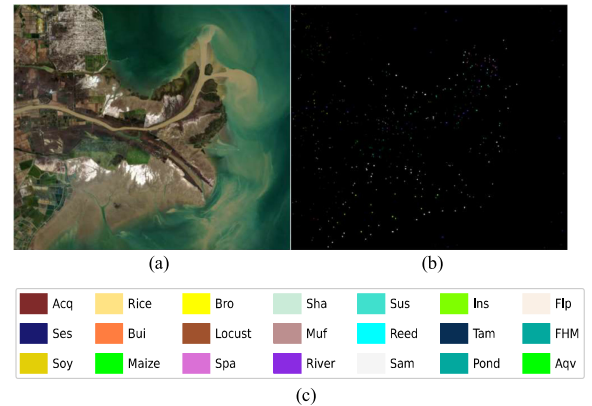


Fig. 5. (a) False-color image. (b) Ground truth of the Yellow River Delta. (c) Class legend.

B. Experimental Setup

In order to evaluate the performance of the two proposed methods, several related classifiers are chosen for comparison, including the classical single classifier SVM, typical integration methods RF [4], XGBOOST and LingtGBM, single classifier CRC and KCRC based on presentation model framework, the CRC-bagging algorithm combined with CR and bagging, and the ensemble method KCRC-All which combines ALL classifiers without difference evaluation. The CRC-bagging and KCRC-bagging are running for 20 iterations.

C. Classification Performance

In this paper, the commonly used indices such as overall accuracy (OA), average accuracy (AA), and Kappa coefficient (KAPPA) are selected to measure the performance of classifier.

1) Overall Accuracy:

$$P(C_i)_{OA} = \frac{\sum_{m=1}^M P_{mm}}{N} \times 100\% \quad (18)$$

TABLE V
PARAMETERS SETTING

Data		Pavia University	Purdue	Yellow River Delta
KCRC-Bagging	λ	1e-2	1e-2	1e-5
	K	20	20	20
DIV-KCRC	λ_1	1e-5	1e-5	1e-5
	λ_2	1e-4	1e-2	1e-1
	λ_3	1e-3	1e-3	1e-5

where $P(C_i)_{OA}$ represents the overall classification accuracy of classifier C_i , M represents the total number of ground object classes, and N represents the total number of test samples. P represents a confusion matrix of size $M \times M$. P_{mm} represents the number of correctly classified ground objects of class m , and this element is located on the main diagonal of matrix P .

2) *Average Accuracy*: AA is obtained by dividing the sum of the precision of all classes by the percentage of the total number of classes

$$P(C_i)_{AA} = \frac{\sum_m^M P(C_i)_{CA,m}}{K} \times 100\%. \quad (19)$$

3) *Kappa Coefficient*: KAPPA is different from OA and AA, which integrates all the information of confusion matrix P . As a quantitative evaluation index for the consistency between the classification results of the analysis algorithm and the distribution of real ground object classes, KAPPA can reflect the overall classification accuracy of the classifier more comprehensively and accurately.

The optimum parameters for the proposed methods are listed in Table V. Table VI shows the corresponding numbers of base classifiers. The value of n denotes the number of neighboring pixels. The classification accuracies, such as OA, AA, and KAPPA, are included in Tables VII–XII. Boldface represents the best result.

For the Pavia University data set, Fig. 6(a)–(j) visually displays the ten classification maps. By comparing the classification maps of several algorithms, it can be found that KCRC-bagging and DIV-KCRC algorithms have better classification effects on the 4th, 6th, 7th, 8th, and 9th class, especially the DIV-KCRC algorithm has a better processing effect on the building boundary. From Table VII, classification accuracy performance of classifier combinations selected by four diversity measures (Cor, Q statistics, DF and Dis) are shown. The classifier combinations selected by the four diversity measures were all 134, which was 3.3% higher than the OA of all classifiers combined together. It is 0.28%, 3.36%, 3.81%, 3.36%, 3.54%, 3.81%, 3.81% higher than that of 123, 124, 125, 145, 234, 235, 245, 345, respectively, and only 0.10% lower than that of 135, and the DF value difference between combination 134 and combination 135 is only 0.0005. The experimental results validate the four difference evaluation indexes. Table VIII records the specific classification accuracy

performance of eight algorithms in ROSIS data. The OA (%) values of the classifiers (RF, SVM, XGBOOST, LingtGBM, CRC, CRC-bagging, KCRC, KCRA-All, KCRC-bagging, and DIV-KCRC) were 77.57, 76.90, 77.00, 85.57, 87.57, 89.31, 87.57, 88.08, 89.01, 91.38, respectively. As can be seen from Table VII, compared with single classifier KCRC and KCRC-all, the OA of KCRC-bagging is improved by 1.53% and 1.02% respectively. And the OA of DIV-KCRC algorithm is 13.81%, 14.48%, 14.38%, 5.81%, 3.81%, 2.07%, 3.81%, 3.3% higher than that of base classifier RF, SVM, XGBOOST, LingtGBM, CRC, CRC-bagging, KCRC, KCRA-All algorithm. The OA of the KCRC-ALL that integrates all base classifiers is increased by 3.3%, and the OA of the classic ensemble algorithm KCRC-bagging is increased by 2.28%. In addition to the overall classification accuracy, the AA and KAPPA of DIV-KCRC ensemble method are the best among the 10 methods, followed by KCRC-bagging.

The experimental results demonstrate the accuracy and effectiveness of the ensemble learning method based on difference evaluation for multikernel classifier, while reducing the ensemble scale.

In order to verify the reliability of the proposed method, the Purdue data sets with grassland, roads, and buildings as main ground object types were used for the experiment. For this data, 7 or 8 training samples were randomly selected for each class by the 10 algorithms. Experimental results of Purdue data show that both KCRC-bagging algorithm and the DIV-KCRC algorithm achieve good results. By comparing the classification results of 10 algorithms for this data as shown in Fig. 7(a)–(j), it is found that the two proposed algorithms have significantly improved their performance in distinguishing buildings and roads compared with the other eight comparison algorithms, especially DIV-KCRC algorithm, which has better fitting effect than KCRC-bagging algorithm.

As illustrated in Table IX, the classifier combinations selected by Cor, Q, DF, Dis were all 134. The accuracy of the selected combination is highest. From Table X, the OA (%) values for RF, SVM, XGBOOST, LingtGBM, CRC, CRC-bagging, KCRC, KCRA-All, KCRC-bagging and DIV-KCRC are 86.83, 83.41, 78.36, 86.08, 88.17, 87.50, 90.96, 90.23, 90.98, and 93.81. Similar to the ROSIS data set, DIV-KCRC can also provide the best performance by producing 93.81% of OA, which increases 3.58% and 2.83% over KCRA-All and the KCRC-bagging respectively.

However, the CRC-bagging and KCRC-bagging have interesting performance, not even as good as a single classifier. It can be seen that ensemble learning without considering accuracy and difference may not achieve ideal results.

The above results show that the four difference indexes can be used to select the suitable classifier combinations for different data types, but the ground object types of the first two data sets

$$KAPPA = \frac{N \sum_m^M P_{m,m} - \sum_{m=1}^M \left(\sum_{m=1}^M P_{m,n} \times \sum_{m=1}^M P_{n,m} \right)}{N^2 - \sum_{m=1}^M \left(\sum_{m=1}^M P_{m,n} \times \sum_{m=1}^M P_{n,m} \right)} \quad (20)$$

TABLE VI
BASE CLASSIFIER SHORTHAND TABLE

Base Classifier	KCRC-LA	KCRC-Linear	KCRC-RBF	KCRC-POLY	KCRC-Pairwise
Shorthand	1	2	3	4	5

TABLE VII
FOUE DIVERSITY MEASURES AND TEN GROUPS CLASSIFICATION ACCURACY(%) FOR THE PAVIA UNIVERSITY DATA SET

Group	Cor	Q	DF	Dis	OA (%)
123	0.59254	0.94814	0.06769	0.08000	91.10
124	0.55567	0.93059	0.07684	0.10208	88.02
125	0.70132	0.96109	0.08342	0.05833	87.57
134	0.51487	0.90623	0.06705	0.11186	91.38
135	0.59254	0.94814	0.06769	0.08000	91.48
145	0.55567	0.93059	0.07684	0.10208	88.02
234	0.55800	0.92300	0.08223	0.10478	87.84
235	0.69178	0.95012	0.08740	0.06385	87.57
245	0.79824	0.98429	0.11311	0.05281	87.57
345	0.55800	0.92300	0.08223	0.10478	87.46

TABLE VIII
CLASSIFICATION ACCURACY (%) FOR THE PAVIA UNIVERSITY DATA SET

Class	Train	RF	SVM	XG BOOST	Light GBM	CRC	CRC-Bagging	KCRC	KCRC-ALL	KCRC-Bagging	DIV-KCRC
1	7	88.98	91.05	94.54	90.84	94.6	94.59	91.89	94.28	93.67	93.48
2	7	65.36	69.42	62.78	80.61	78.96	84.36	86.12	80.15	85.69	89.05
3	8	57.92	50.38	69.72	70.78	86.08	81.81	93.23	86.94	78.88	88.24
4	7	67.66	65.63	62.32	73.43	88.94	81.01	69.10	85.96	75.73	79.59
5	8	99.77	100.00	99.92	99.92	100	100	100	100	100	100
6	7	98.22	81.23	99.86	99.44	99.96	99.64	99.64	99.96	99.60	99.96
7	7	98.71	99.92	97.95	96.27	99.85	99.84	99.69	100	100	100
8	8	88.85	85.55	86.72	91.96	89.53	91.81	88.02	91.81	90.31	90.91
9	8	100.00	99.68	77.68	84.55	99.35	99.46	99.46	99.46	99.46	99.46
OA(%)		77.57	76.90	77.00	85.57	87.57	89.31	87.57	88.08	89.10	91.38
AA(%)		85.05	82.54	83.50	87.53	93.03	92.50	91.9	93.18	91.48	93.64
KAPPA		0.7183	0.7044	0.7131	0.8135	0.841	0.8617	0.8573	0.8469	0.8584	0.8887
Time(s)		0.40	0.48	2.92	2.43	0.65	50.88	2.01	2.13	55.03	12.88

TABLE IX
FOUE DIVERSITY MEASURES AND TEN GROUPS CLASSIFICATION ACCURACY (%) FOR THE PURDUE CAMPUS DATA SET

Group	Cor	Q	DF	Dis	OA (%)
123	0.45	0.90516	0.04917	0.1034	0.92065
124	0.45207	0.90123	0.04983	0.09943	0.91253
125	0.58507	0.91861	0.07154	0.07887	0.86474
134	0.42119	0.89049	0.04142	0.10606	0.93814
135	0.45	0.90516	0.04917	0.1034	0.92696
145	0.45207	0.90123	0.04983	0.09943	0.91253
234	0.50672	0.90982	0.06384	0.09979	0.89847
235	0.65939	0.93527	0.08873	0.07286	0.86474
245	0.74314	0.97183	0.09642	0.05482	0.86474
345	0.50672	0.90982	0.06384	0.09979	0.89702

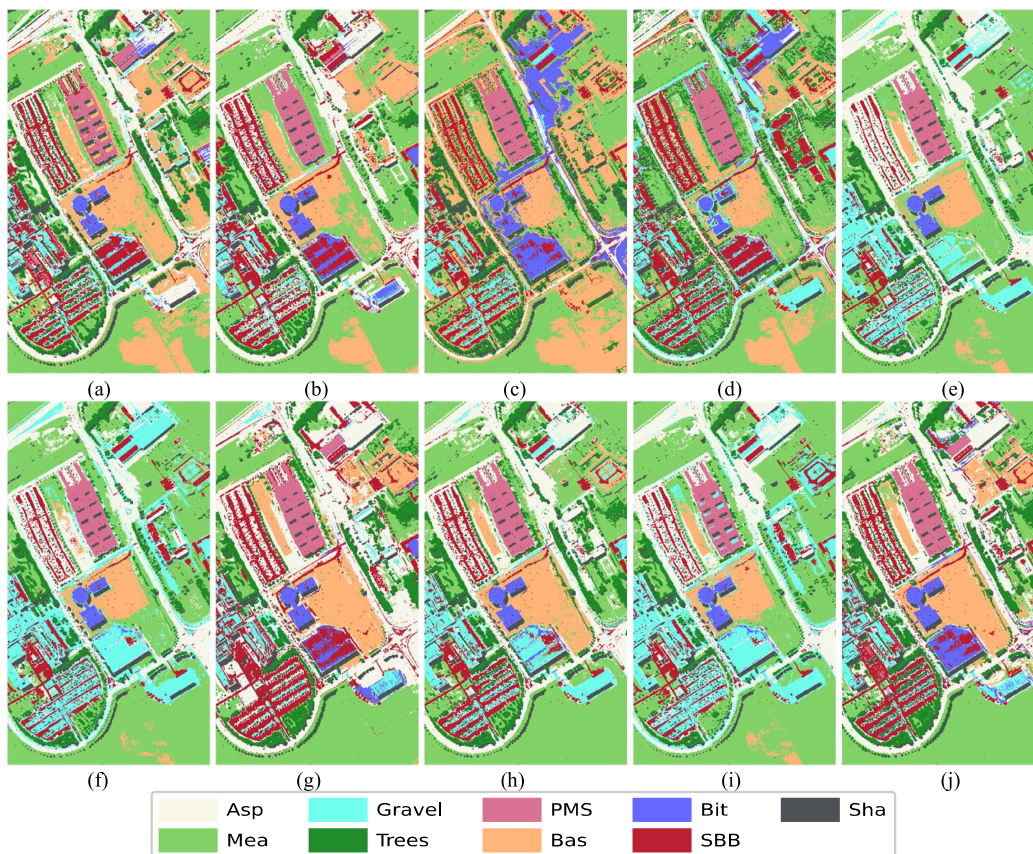


Fig. 6. Classification maps resulting from the classification for the Pavia University data set using eight labeled samples per class. (a) RF (b) SVM. (c) XGBOOST (d) LightGBM. (e) CRC. (f) CRC-Bagging. (g) KCRC. (h) KCRC-ALL. (i) KCRC-Bagging. (j) DIV-KCRC.

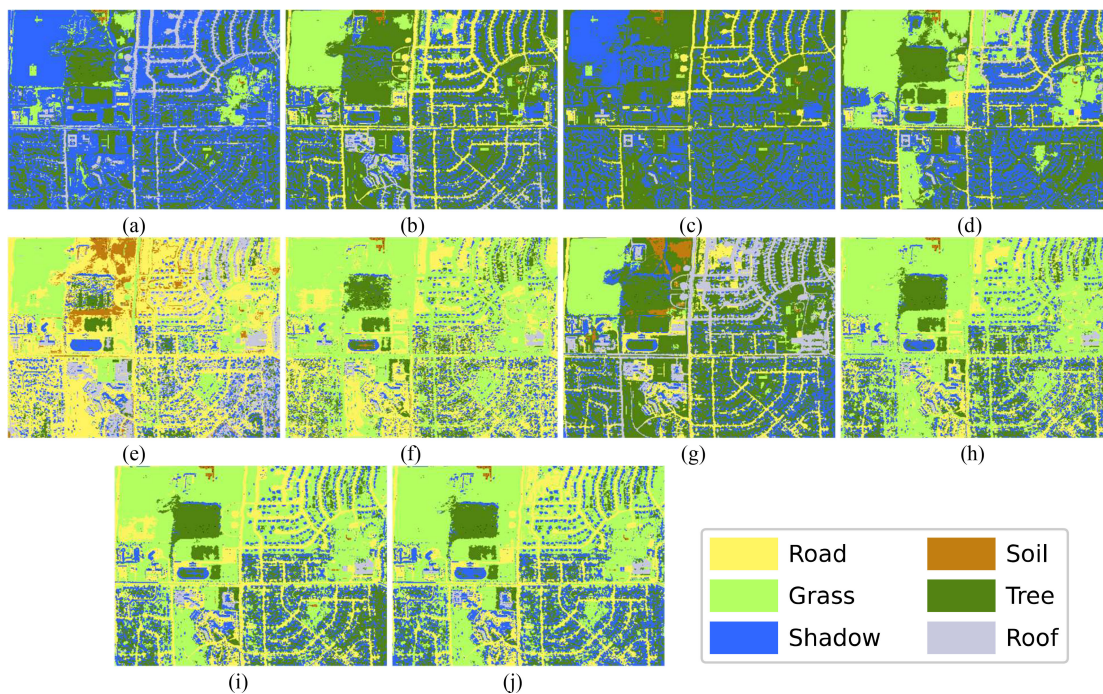


Fig. 7. Classification maps resulting from the classification for the Purdue data set using eight labeled samples per class. (a) RF (b) SVM. (c) XGBOOST. (d) LightGBM. (e) CRC. (f) CRC-Bagging. (g) KCRC. (h) KCRC-ALL. (i) KCRC-Bagging. (j) DIV-KCRC.

TABLE X
CLASSIFICATION ACCURACY (%) FOR THE PURDUE CAMPUS DATA SET

Class	Train	RF	SVM	XG BOOST	Light GBM	CRC	CRC- Bagging	KCRC	KCRC- ALL	KCRC- Bagging	DIV- KCRC
1	7	82.70	63.44	84.83	84.67	94.81	96.38	97.17	97.96	96.78	98.19
2	7	88.08	97.63	54.32	65.06	68.33	68.24	95.00	82.98	95.18	97.18
3	8	98.04	98.04	98.04	98.53	98.04	98.04	98.04	98.04	98.04	98.04
4	7	98.90	96.98	97.80	97.80	95.88	97.25	95.33	96.98	97.80	96.70
5	8	97.53	96.48	78.44	99.93	96.78	97.23	99.18	97.46	98.35	98.88
6	8	73.39	71.10	83.86	85.75	85.83	80.55	70.24	77.95	70.71	79.69
OA(%)		86.83	83.41	78.36	86.08	88.17	87.50	90.96	90.23	90.98	93.81
AA(%)		89.77	87.28	82.88	88.62	89.95	89.62	92.49	91.89	92.81	94.78
KAPPA		0.8352	0.7928	0.7361	0.8252	0.8321	0.8446	0.8870	0.8780	0.8873	0.9225
Time(s)		0.10	0.12	2.80	2.41	0.68	50.11	2.78	3.01	58.45	11.07

TABLE XI
FOUR DIVERSITY MEASURES AND TEN GROUPS CLASSIFICATION ACCURACY (%) FOR THE YELLOW RIVER DELTA

Group	Cor	Q	DF	Dis	OA (%)
123	0.41711	0.91936	0.02416	0.06291	96.48
124	0.48866	0.94099	0.02832	0.05393	95.56
125	0.56803	0.92937	0.03902	0.04864	92.67
134	0.47126	0.94859	0.02356	0.04799	96.56
135	0.41681	0.91836	0.02316	0.06221	96.61
145	0.48866	0.94099	0.02832	0.05393	95.56
234	0.5043	0.9441	0.03032	0.05437	95.06
235	0.58878	0.9349	0.0414	0.04831	92.67
245	0.782	0.99159	0.0521	0.02626	92.67
345	0.5043	0.9441	0.03032	0.05437	95.33

are relatively simple. In order to further verify the validity of the results, the Yellow River Delta with complex ground object types were selected to verify the reliability of the diversity index again.

For the Yellow River Delta, Fig. 8(a)–(j) shows the classification renderings. The detailed classification results of the diversity values are recorded in Table XI, and the best choice is 135. In addition, the selected results of the four difference indicators are consistent, and all have higher performance accuracy than other combinations with small diversity. Table XII shows the detailed classification of comparison algorithms, where the OA (%) values for RF, SVM, XGBOOST, LingtGBM, CRC, CRC-bagging, KCRC, KCRA-All, KCRC-bagging, and DIV-KCRC are 93.99, 86.75, 82.86, 90.61, 93.69, 93.87, 94.99, 95.50, 95.93, and 96.61, respectively. It is worth noting that KCRC-bagging has a less than 1% improvement over KCRC and KCRC-All, while DIV-KCRC improved by about 2%. There are nearly 1.2% improvements from our methods compared with KCRC-ALL, and the accuracy of the original KCRC is almost up to 1.62%. As with the first two data sets, it is observed that DIV-KCRC achieved better results than KCRC-bagging. Especially for classes 3, 5, 7,

and 18, DIV-KCRC's discrimination effect is significantly improved compared with the KCRC-bagging algorithm and other comparison algorithms. According to the experimental results of the Yellow River Delta, the DIV-KCRC ensemble method with diversity assessment can still achieve a good classification effect on complex data types, that is, prior difference assessment can improve the generalization performance of the ensemble system.

V. DISCUSSIONS

A. Parameters Analysis for KCRC-Bagging

For the KCRC-bagging algorithm, the parameter λ and the ensemble size K significantly impact the algorithm performance. In the experiments, λ is set in the range of $1e-9$ – $1e-1$. The relationship between parameter λ and the OA is shown in Fig. 9.

In the three data, CRC algorithm performs the best at $1e-2$, rapidly declines between $[1e-2, 1e-4]$, and then tends to be stable. Although CRC-bagging also shows a decreasing trend in Pavia University data between $[1e-1, 1e-4]$, its OA is significantly higher than that of CRC after it becomes stable. In the Yellow

TABLE XII
CLASSIFICATION ACCURACY (%) FOR THE YELLOW RIVER DELTA

Class	Train	RF	SVM	XG BOOST	Light GBM	CRC	CRC- Bagging	KCRC	KCRC- ALL	KCRC- Bagging	DIV- KCRC
1	7	89.68	95.50	69.58	80.95	88.89	88.36	86.51	88.62	91.80	86.24
2	7	100.00	100.00	98.08	99.87	100.00	100.00	100.00	100.00	100.00	100.00
3	7	87.63	98.97	63.92	70.10	94.85	90.72	98.97	97.94	98.97	98.97
4	7	88.00	88.00	82.29	86.86	88.00	88.00	88.00	88.00	88.00	88.00
5	7	76.47	57.35	67.65	80.88	86.76	83.82	88.24	92.65	82.35	94.12
6	8	100.00	100.00	100.00	98.77	98.77	98.77	100.00	100.00	100.00	100.00
7	8	98.75	93.75	87.50	86.25	88.75	88.75	87.50	97.50	91.25	98.75
8	8	98.47	100.00	92.35	92.35	100.00	100.00	100.00	100.00	100.00	100.00
9	8	94.65	98.40	93.58	93.58	100.00	100.00	100.00	100.00	100.00	100.00
10	8	83.93	46.36	57.44	82.74	97.94	95.87	90.88	98.37	96.20	97.94
11	8	98.33	97.03	87.57	92.39	100.00	100.00	100.00	100.00	100.00	100.00
12	7	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
13	7	98.84	99.71	91.62	99.71	97.98	98.84	100.00	100.00	100.00	100.00
14	8	85.78	76.00	82.67	87.56	57.33	61.78	81.33	63.11	81.33	78.67
15	7	100.00	100.00	98.28	98.28	96.03	98.28	100.00	99.83	100.00	100.00
16	7	98.63	100.00	88.61	98.18	90.21	90.21	99.77	93.17	99.77	99.54
17	7	91.60	82.35	67.23	70.59	55.46	57.98	78.99	74.79	81.51	76.47
18	7	89.23	60.50	60.50	74.59	87.57	88.95	85.08	89.50	82.04	94.20
19	8	77.78	90.74	68.52	64.81	98.15	98.15	90.74	98.15	88.89	90.74
20	8	84.48	87.93	75.86	84.48	74.14	81.03	87.93	89.66	86.21	87.93
21	8	100.00	100.00	92.86	100.00	100.00	100.00	100.00	100.00	100.00	100.00
OA(%)		93.99	86.75	82.86	90.61	93.69	93.87	94.99	95.50	95.93	96.61
AA(%)		92.49	89.17	82.19	87.76	90.51	90.92	93.52	93.87	93.73	94.84
KAPPA		0.9346	0.8569	0.8147	0.8980	0.9202	0.9333	0.9455	0.9511	0.9557	0.9631
Time(s)		0.30	0.38	12.66	11.89	2.21	168.59	2.13	10.10	235.88	15.96

River Delta, the regularization parameters increase between [1e-1, 1e-3] and show a slight decreasing trend between [1e-6, 1e-9]. KCRC was less affected by parameters, and the results of the three sets of data showed that KCRC achieved the best effect at 1e-2, and then tended to be stable. On the other hand, KCRC-bagging performs better than CRC and CRC-bagging in Pavia University as a whole, but its accuracy is not as good as that of KCRC except for $\lambda=1e-1$, and its accuracy decreases with the decrease of regularization parameters. In terms of Purdue data, KCRC-bagging is greatly affected by parameters, and its accuracy is slightly higher than KCRC at 1e-3, 1e-4, 1e-8. In terms of Yellow River Delta, the OA of KCRC-bagging is higher than that of CRC and CRC-bagging as the first two groups, but the overall performance is slightly better than that of KCRC.

According to Fig. 10(a), the OA of KCRC-bagging first increases with the number of integrations in the Pavia University data set. When the number of integrations reaches 30, it changes and fluctuates between 30 and 50. When the number of integrations $k=40$, the accuracy is close to 88%. The second data set, Purdue Campus, with the increase of integration times, the OA of KCRC-bagging shows a trend of fluctuating growth, while the OA of CRC-bagging shows an obvious increase between 5 and 15, and then shows a trend of first decreasing, then increasing and then decreasing. It can be shown in Fig. 10(b).

The results of Yellow River Delta are shown in Fig. 10(c). The overall performance of KCRC-bagging is relatively stable. After the number of integrations reaches 30, KCRC-bagging almost shows a stable performance, remaining between 95.35% and 96%, while CRC-bagging shows a significant increase in the number of integrations ranging from 15 to 25. After the number of K reaches 25, it tends to be stable.

Fig. 10(b) and (c) shows that KCRC-bagging algorithm has more advantages in classification effect and stability. In addition, the three groups of experimental data also show that increasing the number of ensemble times does not necessarily improve the accuracy, and may even present a downward trend. After the number of ensembles reaches 15–25, it gradually tends to be stable.

B. Parameters Analysis for DIV-KCRC

For the DIV-KCRC algorithm, the performance is mainly affected by diversity values. From Fig. 11, there is no obvious linear relationship between the four diversity indices and the OA. From the distribution of the two, the evaluation results of the four indexes are consistent.

As can be seen from Fig. 11, the variation trend of Cor and Q is relatively consistent, and the distribution of DF and DIS is

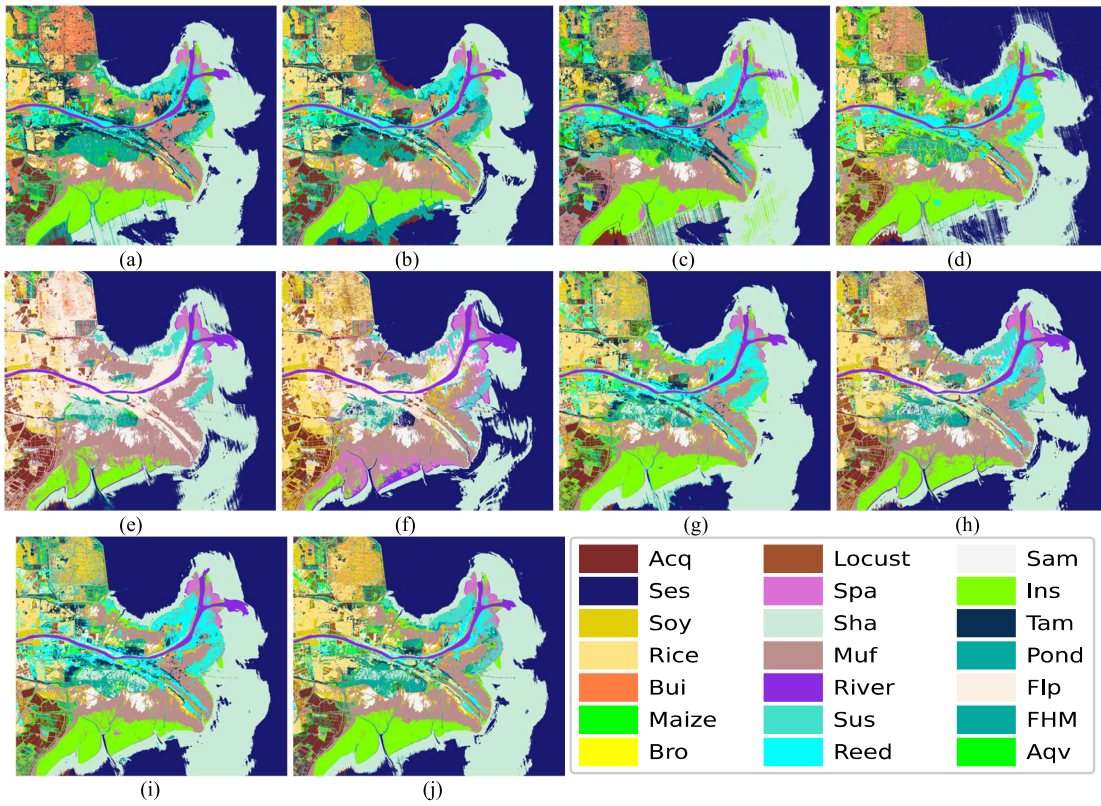


Fig. 8. Classification maps resulting from the classification for the Yellow River Delta using eight labeled samples per class. (a) RF (b) SVM. (c) XGBOOST (d) LightGBM. (e) CRC. (f) CRC-Bagging. (g) KCRC. (h) KCRC-ALL. (i) KCRC-Bagging. (j) DIV-KCRC.

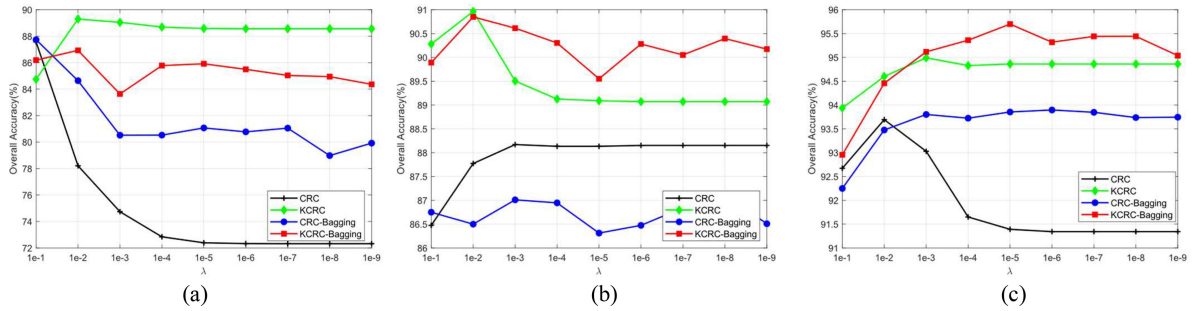


Fig. 9. OA of different algorithms for three data set with varying λ . (a) Pavia University. (b) Purdue campus. (c) Yellow River Delta.

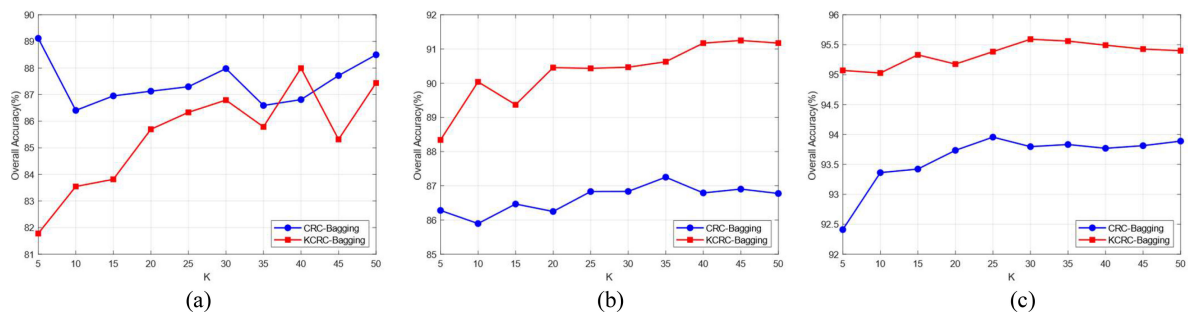


Fig. 10. OA of different algorithms for three data set with varying K . (a) Pavia University. (b) Purdue campus. (c) Yellow River Delta.

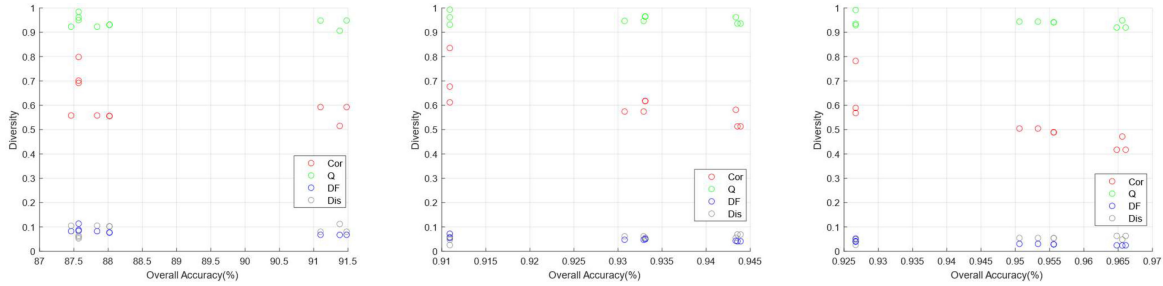


Fig. 11. OA of different algorithms for three data set with Diversity. (a) Pavia University. (b) Purdue campus. (c) Yellow River Delta.

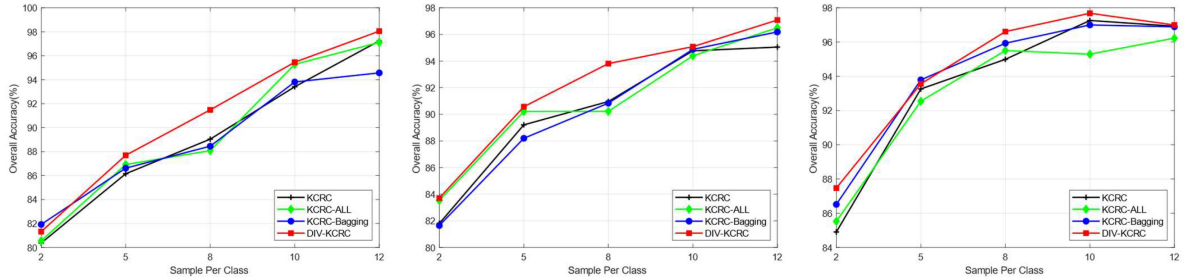


Fig. 12. OA of different algorithms for three data set with Different Training Sets. (a) Pavia University. (b) Purdue campus. (c) Yellow River Delta.

similar. Among them, when the classification accuracy is high, the values of Q , Cor , and DF are at a low level, and the value of Dis is at a high level, that is, the classifier combination is diverse large. In addition, the distribution of the two diversity measures Cor and Q is more dispersed, and the diversity of the ensemble system can be judged more intuitively. For example, in Fig. 11(c), it can be seen that the classifier combination with precision at (0.96, 0.97) is more diverse than the classifier combination with precision at (0.925, 0.95). In other words, diversity results in higher accuracy. However, when the number of classifier combination is too high, the OA decreases. This proves that the diversity between classifiers is limited in ensemble learning. In addition, Fig. 11(c) clearly shows that the more the overall distribution of DF is inclined to 1, the worse the performance of the classifier is, and the distribution of the OA is at a lower level.

By comparing the results of the three groups of experimental data, it is found that more complex data include more diversity. Specifically, under the condition of ensuring the accuracy of the base classifier, increasing the difference between individual classifiers can effectively improve the ensemble accuracy, while reducing the number of individual classifiers in ensemble learning. Analyzing the relationship between the difference and precision of different combinations, it can be found that the classifier combination selected by the four diversity indicators is consistent and has the highest accuracy among all combinations. The experimental results of the three sets of data all prove the effectiveness of the diversity measures in ensemble learning.

Moreover, the final trend of the four diversity indexes is consistent, which can ensure that the selected combination is optimal. Among them, the trend of Cor and DF is more obvious than the other two indicators, and the accuracy is higher.

C. Sizes of Training Set

Fig. 12(a)–(c) shows the effect of different sample sizes on classification accuracy in three sets of data. The Yellow River Delta achieved the best classification effect when the number of samples reached 12. With the continuous increase of the number of samples, the classification accuracy showed a downward trend. The data of Purdue and Pavia University are greatly affected by the number of samples, and with the increase of the number of samples, the accuracy is obviously improved. In the three sets of data, when the number of samples is 8, the proposed algorithm has the most obvious improvement over other algorithms.

D. Time Complexity Analysis

According to Tables VIII, X, and XII for the three experimental data, DIV-KCRC significantly reduces the operation time compared with the traditional bagging algorithm and improves the computational efficiency. In other words, DIV-KCRC with diversity evaluation can effectively reduce the scale of ensemble and achieve better classification results with as few ensemble times as possible.

VI. CONCLUSION

In this article, ensemble learning method with kernel CR and multiple kernel CR classifier based on diversity measures are proposed for hyperspectral image classification, i.e., KCRC-bagging and DIV-KCRC. For the proposed method, the diversity of the ensemble system is increased with different training sets. In addition, the similarity evaluation index is used to screen classifiers for ensemble and reduce the scale. Experimental results show that KCRC-bagging can obtain better performance

than a single classifier in most cases, and the multiclassifier ensemble based on diversity measures (DIV-KCRC) can not only improve the performance but also greatly improve the computing efficiency, which demonstrates that the generating diversity and evaluation diversity are equally important in ensemble learning.

ACKNOWLEDGMENT

The authors would like to thank Prof. P. Gamba from the University of Pavia, Pavia, Italy, for providing the ROSIS data and Prof. D. Landgrebe for providing the AVIRIS data.

REFERENCES

- [1] D. Landgrebe, "Hyperspectral_image_data_analysis," *IEEE Signal Process. Mag.*, vol. 10, no. 1, pp. 17–28, Jan. 2002.
- [2] H. Su, Z. Wu, H. Zhang, and Q. Du, "Hyperspectral anomaly detection: A survey," *IEEE Geosci. Remote Sens. Mag.*, to be published, doi: [10.1109/MGRS.2021.3105440](https://doi.org/10.1109/MGRS.2021.3105440).
- [3] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. M. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.
- [4] J. G. Carbonell, "Machine learning research," *ACM SIGART Bull.*, vol. 18, no. 77, p. 29, 1981.
- [5] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, Aug. 2016, vol. 42, no. 8, pp. 785–794, doi: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785).
- [6] G. Ke *et al.*, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 3147–3155.
- [7] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing," *J. Sensors*, vol. 42, pp. 1778–1790, 2004.
- [8] P. Du, J. Xia, Z. Xue, K. Tan, H. Su, and R. Bao, "Review of hyperspectral remote sensing image classification," *Yaogan Xuebao/J. Remote Sens.*, vol. 20, no. 2, pp. 236–256, 2016, doi: [10.11834/jrs.20165022](https://doi.org/10.11834/jrs.20165022).
- [9] A. Samat, P. Du, S. Liu, J. Li, and L. Cheng, "E2LMs: Ensemble extreme learning machines for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1060–1069, Apr. 2014, doi: [10.1109/JSTARS.2014.2301775](https://doi.org/10.1109/JSTARS.2014.2301775).
- [10] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, 2019.
- [11] Q. S. Ul Haq, L. Tao, and S. Yang, "Neural network based adaboosting approach for hyperspectral data classification," in *Proc. Int. Conf. Comput. Sci. Netw. Technol.*, 2011, vol. 1, pp. 241–245, doi: [10.1109/ICC-SNT.2011.6181949](https://doi.org/10.1109/ICC-SNT.2011.6181949).
- [12] M. D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 341–354, Jan. 2014.
- [13] W. Li, Q. Du, F. Zhang, and W. Hu, "Hyperspectral image classification by fusing collaborative and sparse representations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4178–4187, Sep. 2016.
- [14] P. Zheng, H. Su, and Q. Du, "Sparse and low-rank constrained tensor factorization for hyperspectral image unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1754–1767, 2021.
- [15] W. Li, Q. Du, F. Zhang, and W. Hu, "Collaborative-representation-based nearest neighbor classifier for hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 2, pp. 389–393, Feb. 2015.
- [16] H. Su, B. Zhao, Q. Du, and Y. Sheng, "Tangent distance-based collaborative representation for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 9, pp. 1236–1240, Sep. 2016.
- [17] W. Li, Q. Du, and M. Xiong, "Kernel collaborative representation with tikhonov regularization for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 48–52, Jan. 2015.
- [18] W. Li and Q. Du, "Joint within-class collaborative representation for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2200–2208, Jun. 2014.
- [19] J. Peng, L. Li, and Y. Y. Tang, "Maximum likelihood estimation-based joint sparse representation for the classification of hyperspectral remote sensing images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 6, pp. 1790–1802, Jun. 2019.
- [20] P. Du, L. Gan, J. Xia, and D. Wang, "Multikernel adaptive collaborative representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4664–4677, Aug. 2018.
- [21] B. Tu, C. Zhou, D. He, S. Huang, and A. Plaza, "Hyperspectral classification with noisy label detection via superpixel-to-pixel weighting distance," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 4116–4131, Jun. 2020.
- [22] J. Liu, Z. Wu, J. Li, A. Plaza, and Y. Yuan, "Probabilistic-Kernel collaborative representation for spatial-spectral hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 2371–2384, Apr. 2016.
- [23] A. C. Karaca, "Spatial aware probabilistic multi-kernel collaborative representation for hyperspectral image classification using few labelled samples," *Int. J. Remote Sens.*, vol. 42, no. 3, pp. 839–864, 2021, doi: [10.1080/01431161.2020.1823516](https://doi.org/10.1080/01431161.2020.1823516).
- [24] Y. Ma, C. Li, H. Li, X. Mei, and J. Ma, "Hyperspectral image classification with discriminative kernel collaborative representation and Tikhonov regularization," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 4, pp. 587–591, Apr. 2018.
- [25] H. Su, B. Zhao, Q. Du, and P. Du, "Kernel collaborative representation with local correlation features for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1230–1241, Feb. 2019.
- [26] B. Tu, C. Zhou, X. Liao, G. Zhang, and Y. Peng, "Spectral-Spatial hyperspectral classification via structural-kernel collaborative representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 5, pp. 861–865, May 2021.
- [27] L. Gupta, S. Kota, D. L. Molfese, and R. Vaidyanathan, "Diversity-based selection of components for fusion classifiers," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2010, pp. 6304–6307, doi: [10.1109/IEMBS.2010.5628090](https://doi.org/10.1109/IEMBS.2010.5628090).
- [28] N. Li, Y. Yu, and Z. H. Zhou, "Diversity regularized ensemble pruning," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discov. Databases*, 2012, pp. 330–345, doi: [10.1007/978-3-642-33460-3_27](https://doi.org/10.1007/978-3-642-33460-3_27).
- [29] J. Xia, J. Chanussot, P. Du, and X. He, "Rotation-based support vector machine ensemble in classification of hyperspectral data with limited training samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1519–1531, Mar. 2016.
- [30] B. Tu, X. Yang, X. Ou, G. Zhang, J. Li, and A. Plaza, "Ensemble entropy metric for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5513617.
- [31] B. Tu, X. Zhang, X. Kang, G. Zhang, and S. Li, "Density peak-based noisy label detection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1573–1584, Mar. 2019.
- [32] G. Tuysuzoglu, N. Moarref, and Y. Yaslan, "Ensemble based classifiers using dictionary learning," in *Proc. Int. Conf. Syst., Signals Image Process.*, May 2016, pp. 1–4, doi: [10.1109/IWSSIP.2016.7801373](https://doi.org/10.1109/IWSSIP.2016.7801373).
- [33] B. Tu, X. Yang, C. Zhou, D. He, and A. Plaza, "Hyperspectral anomaly detection using dual window density," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8503–8517, Dec. 2020.
- [34] Y. Yu and H. Su, "Collaborative representation ensemble using bagging for hyperspectral image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2019, pp. 2738–2741, doi: [10.1109/IGARSS.2019.8898684](https://doi.org/10.1109/IGARSS.2019.8898684).
- [35] H. Su, Y. Yu, Q. Du, and P. Du, "Ensemble learning for hyperspectral image classification using tangent collaborative representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 3778–3790, Jun. 2020.
- [36] Y. Chen, X. Zhao, and Z. Lin, "Joint adaboost and multifeature based ensemble for hyperspectral image classification," in *Proc. Int. Geosci. Remote Sens. Symp.*, 2014, pp. 2874–2877, doi: [10.1109/IGARSS.2014.6947076](https://doi.org/10.1109/IGARSS.2014.6947076).
- [37] B. Tu, C. Zhou, X. Liao, Q. Li, and Y. Peng, "Feature extraction via 3-D block characteristics sharing for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10503–10518, Dec. 2021.
- [38] Y. Chen, X. Zhao, and Z. Lin, "Optimizing subspace SVM ensemble for hyperspectral imagery classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1295–1305, Apr. 2014.
- [39] H. Su, Y. Yu, Z. Wu, and Q. Du, "Random subspace-based k -nearest class collaborative representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6840–6853, Aug. 2021.
- [40] P. Cunningham and J. Carney, "Diversity versus quality in classification ensembles based on feature selection," in *Proc. Mach. Learn.: ECML 2000*. Berlin, Heidelberg, Germany: Springer, 2000, pp. 109–116.
- [41] K. Cheng, Z. Song, Y. Yue, F. Shan, and X. Guo, "Classifier selection method based on multiple diversity measures," in *Proc. IEEE Int. Conf. Softw. Eng. Serv. Sci.*, 2019, pp. 47–53, doi: [10.1109/ICSESS47205.2019.9040785](https://doi.org/10.1109/ICSESS47205.2019.9040785).

- [42] Y. Bi, "The impact of diversity on the accuracy of evidential classifier ensembles," *Int. J. Approx. Reasoning*, vol. 53, no. 4, pp. 584–607, 2012, doi: [10.1016/j.ijar.2011.12.011](https://doi.org/10.1016/j.ijar.2011.12.011).
- [43] L. I. Kuncheva, M. Skurichina, and R. P. W. Duin, "An experimental study on diversity for bagging and boosting with linear classifiers," *Inf. Fusion*, vol. 3, no. 4, pp. 245–258, 2002, doi: [10.1016/S1566-2535\(02\)00093-3](https://doi.org/10.1016/S1566-2535(02)00093-3).
- [44] N. Li and Z. H. Zhou, "Selective ensemble under regularization framework," in *Proc. Int. Workshop Mult. Classifier Syst.*, 2009, pp. 293–303, doi: [10.1007/978-3-642-02326-2_30](https://doi.org/10.1007/978-3-642-02326-2_30).
- [45] J. Zhao, "A classifier based on complementary information entropy is used to measure the ensemble difference," *J. XI'AN JIAOTONG Univ.*, vol. 50, no. 2, 2016.
- [46] X. C. Yin, K. Huang, H. W. Hao, K. Iqbal, and Z. Bin Wang, "A novel classifier ensemble method with sparsity and diversity," *Neurocomputing*, vol. 134, pp. 214–221, 2014, doi: [10.1016/j.neucom.2013.07.054](https://doi.org/10.1016/j.neucom.2013.07.054).
- [47] K. Tan, J. Zhu, Q. Du, L. Wu, and P. Du, "A novel tri-training technique for semi-supervised classification of hyperspectral images based on diversity measurement," *Remote Sens.*, vol. 8, no. 9, 2016, doi: [10.3390/rs8090749](https://doi.org/10.3390/rs8090749).
- [48] L. I. Kuncheva and C. J. Whitaker, "Ten measures of diversity in classifier ensembles: Limits for two classifiers," *IEE Colloq.*, no. 50, pp. 73–82, 2001, doi: [10.1049/ic:20010105](https://doi.org/10.1049/ic:20010105).
- [49] Z. H. Zhou, "Ensemble methods," *J. Phys. A, Math. Theor.*, vol. 44, no. 8, Feb. 2012, Art. no. 085201, doi: [10.1088/1751-8113/44/8/085201](https://doi.org/10.1088/1751-8113/44/8/085201).
- [50] K. Ren, W. Sun, X. Meng, G. Yang, and Q. Du, "Fusing China GF-5 hyperspectral data with GF-1, GF-2 and Sentinel-2A multispectral data: Which methods should be used?," *Remote Sens.*, vol. 12, no. 5, Mar. 2020, Art. no. 882, doi: [10.3390/rs12050882](https://doi.org/10.3390/rs12050882).
- [51] L. Jiao, W. Sun, G. Yang, G. Ren, and Y. Liu, "A hierarchical classification framework of satellite multispectral/hyperspectral images for mapping coastal wetlands," *Remote Sens.*, vol. 11, no. 19, 2019, Art. no. 2238, doi: [10.3390/rs11192238](https://doi.org/10.3390/rs11192238).
- [52] J. T. Peng, W. W. Sun, T. H. Wei, and W. Q. Fan, "A modified correlation alignment algorithm for the domain adaptation of GF-5 hyperspectral image, in Chinese," *J. Remote Sens.*, vol. 24, no. 4, pp. 417–426, 2020, doi: [10.11834/jrs.20209212](https://doi.org/10.11834/jrs.20209212).
- [53] H. Su, W. Yao, Z. Wu, P. Zheng, and Q. Du, "Kernel low-rank representation with elastic net for China coastal wetland land cover classification using GF-5 hyperspectral imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 171, pp. 238–252, 2021, doi: [10.1016/j.isprsjprs.2020.11.018](https://doi.org/10.1016/j.isprsjprs.2020.11.018).



Hongjun Su (Senior Member, IEEE) received the Ph.D. degree in cartography and geography information system from the Key Laboratory of Virtual Geographic Environment (Ministry of Education), Nanjing Normal University, Nanjing, China, in 2011.

He is currently a Full Professor with the School of Earth Sciences and Engineering, Hohai University, Nanjing, China. His main research interests include hyperspectral remote sensing dimensionality reduction, classification, and spectral unmixing.

Dr. Su is an Associate Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. He was the recipient of the 2016 Best Reviewer Award from the IEEE GEOSCIENCE AND REMOTE SENSING SOCIETY.



Yuzheng Hu received the B.E. degree from North China University of Water Resources and Electric Power, Zhengzhou, China, in 2019. She is currently working toward the M.E. degree in photogrammetry and remote sensing with the School of Earth Sciences and Engineering, Hohai University, Nanjing, China.

Her research interests include hyperspectral remote sensing images, collaborative representation, and ensemble learning in hyperspectral remote sensing image classification.



Hongliang Lu received the B.E. degree from Chuzhou University, Chuzhou, China, in 2015, and the M.E. degree from the School of Geomatics, Anhui University of Science and Technology, Huainan, China, in 2020. He is currently working toward the Ph.D. degree in in photogrammetry and remote sensing with the School of Earth Sciences and Engineering, Hohai University, Nanjing, China.

His research interests include ensemble learning, few shot learning, and collaborative representation in hyperspectral remote sensing imagery classification.



Weiwei Sun (Senior Member, IEEE) received the B.S. degree in surveying and mapping and the Ph.D. degree in cartography and geographic information engineering from Tongji University, Shanghai, China, in 2007 and 2013, respectively.

From 2011 to 2012, he studied with the Department of Applied Mathematics, University of Maryland, College Park, MD, USA, where he was a Visiting Scholar with the famous Professor John Benedetto to study on the dimensionality reduction of hyperspectral images. From 2014 to 2016, he studied with the State Key Laboratory for Information Engineering in Surveying, Mapping, and Remote Sensing (LIESMARS), Wuhan University, Wuhan, China, where he held a postdoctoral training to study intelligent processing in hyperspectral imagery. From 2017 to 2018, he worked with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA, where he worked as a Visiting Scholar in hyperspectral image processing. He is currently a Full Professor with the Department of Geography and Spatial Information Techniques, Ningbo University, Ningbo, Zhejiang, China. He has authored or coauthored more than 80 journal articles and his research interest includes hyperspectral image processing with machine learning.



Qian Du (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Maryland, Baltimore, MD, USA, in 2000.

She is currently the Bobby Shackouls Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA. Her research interests include hyperspectral image analysis and applications, pattern classification, data compression, and neural networks.

Dr. Du is a Fellow of the IEEE-Institute of Electrical and Electronics Engineers, and a Fellow of the SPIE-International Society for Optics and Photonics. She was a Co-chair of the Data Fusion Technical Committee of the IEEE Geoscience and Remote Sensing Society from 2009 to 2013, and the Chair of the Remote Sensing and Mapping Technical Committee of the International Association for Pattern Recognition from 2010 to 2014. She has served as an Associate Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, *Journal of Applied Remote Sensing*, and IEEE SIGNAL PROCESSING LETTERS. In 2016–2020, she was the Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. She currently is a member of IEEE Periodicals Review.