Weighted Residual Dynamic Ensemble Learning for Hyperspectral Image Classification

Hongliang Lu, Hongjun Su, *Senior Member, IEEE*, Pan Zheng, Yihan Gao, and Qian Du, *Fellow, IEEE*

***Abstract*—Recently, collaborative representation classifiers have been extensively studied as an essential method for the hyperspectral image. However, how to comprehensively utilize the classification advantages of multiple collaborative classifiers has not been well investigated. In this article, two new dynamic ensemble learning methods using local weighted residual (LWR-DEL) and double-weighted residual (DWR-DEL) of multicollaborative representation classifiers are proposed. First, the dynamic ensemble learning method based on clustering is utilized to introduce prior knowledge for the collaborative representation classifier. Then, with prior knowledge, the local weights of each classifier for a different region of competence are obtained. To consider the global information of hyperspectral data, the K-nearest neighbor algorithm is adopted to achieve validation samples with global information. The global weights for each classifier can be obtained and then used to constrain the locally weighted residuals. Similar to LWR-DEL, the global information is also used to constrain residual, and then double-weighted constrained residual fusion obtains the final classifier result. The effectiveness of the proposed methods is validated using three hyperspectral data sets. The experimental results show that both LWR-DEL and DWR-DEL outperform their single-classifier counterparts. In particular, the proposed methods provide superior performance compared with the state-of-the-art methods.**

***Index Terms*—Dynamic ensemble learning, hyperspectral imagery, multicollaborative representation, weighted residual ensemble.**

I. Introduction

HYPERSPECTRAL images have abundant spectral information in hundreds of contiguous narrow spectral bands [1]. Based on these properties, hyperspectral images have many applications in many fields [2], [3], [4], [5], [6]. Among them, hyperspectral image classification is one of the most critical tasks for real applications. With high spectral resolution, fine classification can be achieved. However, the vast amount of hyperspectral data, redundant data, few labeled samples, and correlation between bands have become essential factors restricting the classification performance of hyperspectral images [7], [8], [9], [10], [11], [12], [13]. To solve the above problems, some advanced hypersectral image (HSI) classification algorithms have been proposed. Song et al. [14] proposed a new band selection method to deal with the redundant information problem in HSI classification. In this progressive band selection method, classification is performed incrementally in multiple stages. The experimental results show that this method performs better than other HSI classification methods that use full bands. Yu et al. [15] investigated the feedback attention modules in an HSI classification network and proposed a spatial-spectral dense convolutional neural network (CNN) framework with a feedback attention mechanism. Experimental results based on real HSIs demonstrate the superiority of the proposed methods over other state-of-the-art algorithms. Meanwhile, some advanced algorithms were developed to solve the limited samples problem in HSI classification [16], [17], [18], [19]. However, a single classifier often cannot solve the above problems comprehensively. Therefore, how to develop a classifier or classifier ensemble that can overcome the above limitations is a crucial problem in hyperspectral image classification.

Ensemble learning can be divided into two categories according to whether prior information is used to measure its competence. The first category is the static ensemble [20], [21], [22], [23], [24], [25], [26]. It assumes that each base classifier is independent and has higher accuracy than random guessing. Then, a specific strategy is adopted to combine multiple classifiers for higher classification accuracy. The most famous static ensemble algorithms include boosting [27], [28], bagging [29], and random subspace [30], which has many applications in hyperspectral image classification [31], [32], [33], [34]. Su et al. [35] proposed a new ensemble fusion strategy that first uses collaborative representation (CR) based classifiers as base classifiers for hyperspectral image classification. The result shows that the traditional ensemble strategies such as bagging and boosting are also suitable for the CR-based models. Bao et al. [36] first proposed ensemble learning from the perspective of the feature layer and applied it to hyperspectral image classification. The results show that ensemble learning from the feature perspective is effective for hyperspectral image classification. Pan et al. [37] first proposed using an ensemble strategy to combine hierarchical guidance filtering and matrix of spectral angle distance for hyperspectral image classification, which can effectively improve classification accuracy. However, there are many ground objects. Some classifiers have higher classification accuracy for a specific object but lower overall accuracy (OA). Moreover, all static ensemble learning requires the high accuracy of the base classifier. These restrictions make the static ensemble

Manuscript received 28 June 2022; revised 1 August 2022; accepted 9 August 2022. Date of publication 19 August 2022; date of current version 29 August 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 42122008 and Grant 41871220 and in part by the Jiangsu Province Graduate Research and Innovation Program (KYCX220662). *(Corresponding author: Hongjun Su.)*

Hongliang Lu, Hongjun Su, Pan Zheng, and Yihan Gao are with the School of Earth Sciences and Engineering, Hohai University, Nanjing 211100, China (e-mail: hllu_hhu@163.com; hjsu@hhu.edu.cn; zhengpan1013@163.com; yihangao_hhu@163.com).

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.2022.3200042

TABLE I
SIXTEEN CLASSES OF THE INDIAN PINES DATA SET

Class	Train	Validation	Test	Name
1	20	20	17	Alfalfa
2	20	20	1389	Corn-no till
3	20	20	790	Corn-min till
4	20	20	201	Corn
5	20	20	443	Grass/pasture
6	20	20	690	Grass/trees
7	20	20	9	Grass/pasture-mowed
8	20	20	441	Hay-windrowed
9	20	20	4	Oats
10	20	20	932	Soybean-no till
11	20	20	2415	Soybean-min till
12	20	20	556	Soybeans-clean
13	20	20	172	Wheat
14	20	20	1225	Woods
15	20	20	348	Buildings-grass
16	20	20	62	Stone-steel-towers

TABLE II
NINE CLASSES OF THE UNIVERSITY OF PAVIA DATA SET

Class	Train	Validation	Test	Name
1	20	20	6591	Asphalt
2	20	20	18609	Meadows
3	20	20	2059	Gravel
4	20	20	3024	Trees
5	20	20	1306	Painted metal sheets
6	20	20	4989	Bare Soil
7	20	20	1290	Bitumen
8	20	20	3642	Self-Blocking Bricks
9	20	20	908	Shadows

TABLE III
THIRTEEN CLASSES OF THE YELLOW RIVER DATA SET

Class	Train	Validation	Test	Name
1	20	20	355	Acquaculture (Acq)
2	20	20	756	Seep sea (Ses)
3	20	20	76	Soybean (Soy)
4	20	20	154	Rice
5	20	20	54	Building (Bui)
6	20	20	63	Maize
7	20	20	62	Broomcorn (Bro)
8	20	20	176	Locust
9	20	20	165	Spartina (Spa)
10	20	20	896	Shallow (Sha)
11	20	20	515	Mud flat (Muf)
12	20	20	430	River
13	20	20	322	Suaeda salsa (Sus)
14	20	20	202	Reed
15	20	20	556	Salt marsh (Sam)
16	20	20	418	Intertidal saltwater (Ins)
17	20	20	100	Tamarix(Tam)
18	20	20	338	Pond
19	20	20	37	Flood plain(Flp)
20	20	20	41	Freshwater herbaceous marsh (FHM)
21	20	20	15	Aquatic vegetation (Aqv)

unable to fully play the advantages of these classifiers with higher local accuracy.

The second category is the dynamic ensemble selection (DES) [38], which is to obtain the prior information of the classification by using a specific region division and classifier selection strategy [39], [40], [41]. The best-fit classifier is assigned to unknown samples in each region based on these priors. DES assumes that each classifier, including weak classifiers with extremely weak classification accuracy, is an expert for specific testing samples [42], [43], [44], [45]. Therefore, compared with the

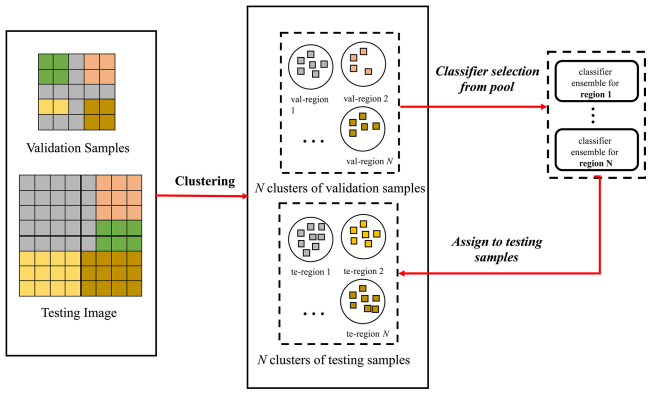

Fig. 1. Illustration of DES process based on clustering method.

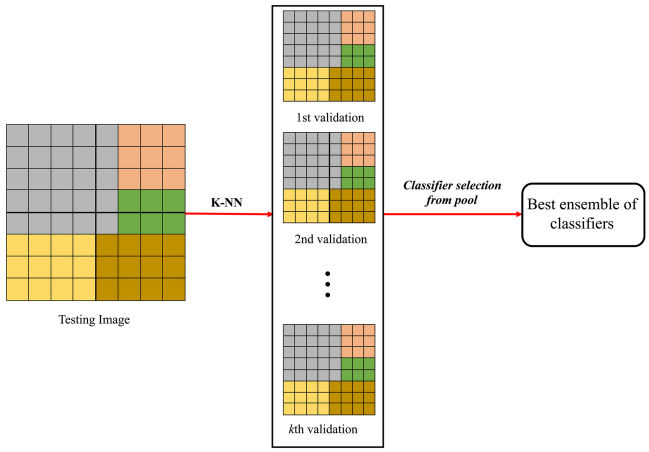

Fig. 2. Illustration of DES process based on K-NN method.

static ensemble, the DES method can better utilize the local classification advantages of weak classifiers. Recently, the DES has also been introduced into hyperspectral image classification. Damodaran et al. [46] first proposed using DES for hyperspectral image classification and further improved the method [47]. It first combines a dimensionality reduction process with the dynamic selection method to construct a DES framework. Then, the random subspace method, Markov random field, and extreme learning machine are introduced into DES. The experimental results show that the two proposed methods can obtain better classification performance compared with the traditional ensemble learning methods. However, the typical DES method is a classifier selection strategy, and some very weak classifiers with high classification accuracy for specific regions will still be eliminated. Therefore, the local advantage of the classifier is still not fully utilized. Meanwhile, DES models directly fuse the classification results without considering the residuals difference of these classifiers.

Based on the above analysis, the static ensemble has a specific improvement in classification accuracy compared with a single-classifier method. However, most traditional ensemble methods directly fuse the classification results without considering the

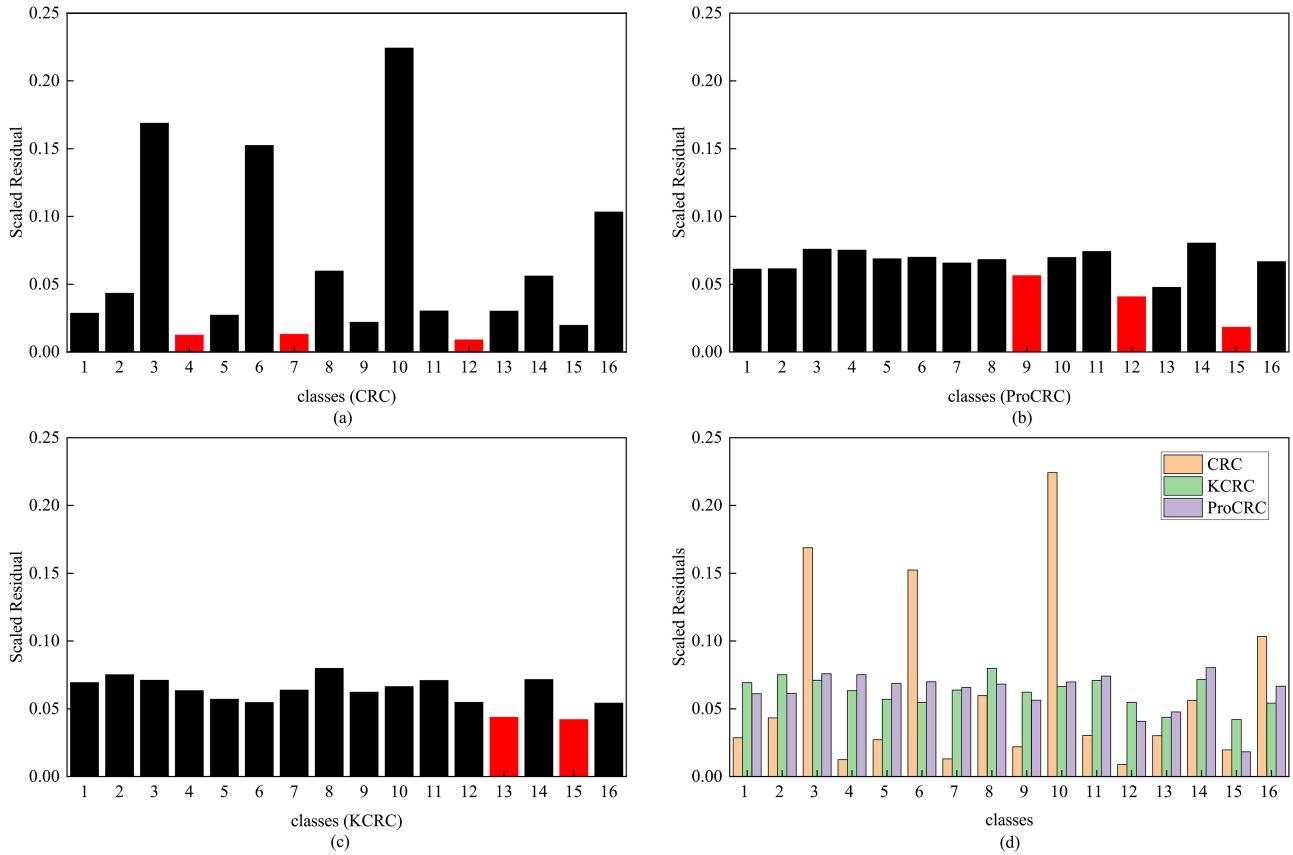

Fig. 3. Scaled residuals of the Indian Pines dataset by different CR-based classifiers. (a) CRC ($\lambda = 1e{-}1$), (b) ProCRC ($\lambda = 1e{-}1$, $\gamma = 1e{-}1$), and (c) KCRC ($\lambda = 1e{-}1$). (d) Residuals between the three classifiers.

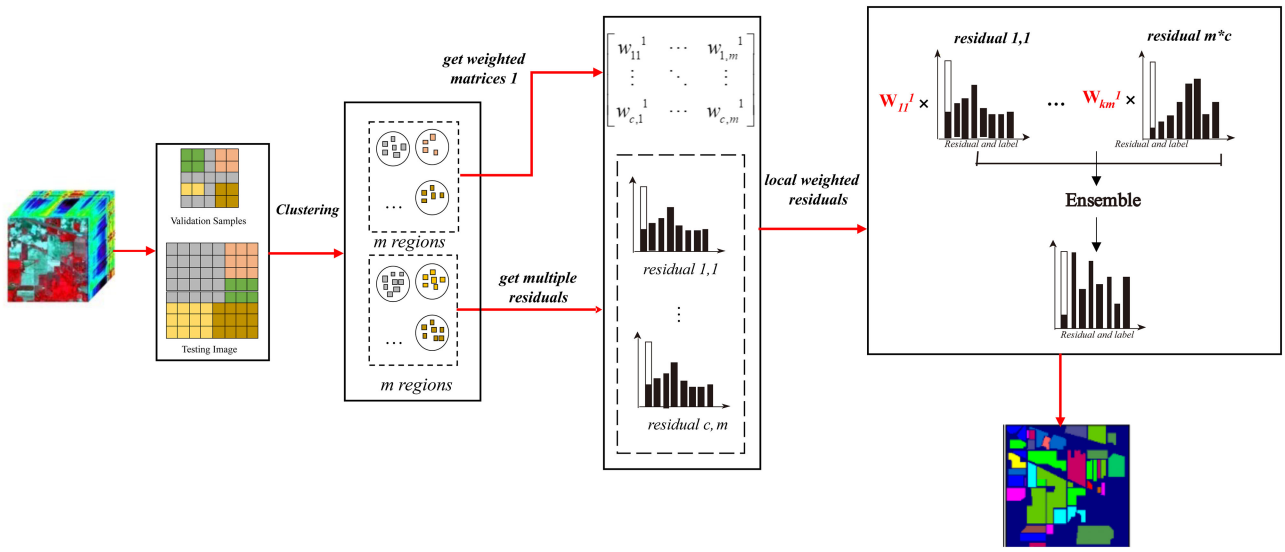

Fig. 4. Flowchart of LWR-DEL.

local performance. The typical DES methods often use clustering or K-nearest neighbor (K-NN) methods to divide the classification target into different regions. However, local and global information are not both considered in DES methods. Ensemble learning based on a representation learning model is concise and computationally efficient and has strong generalization ability. However, the existing representation learning-based ensemble methods are still based on traditional ensemble strategies, which do not fully apply the intrinsic principles of models. Therefore, how to fully consider the possible local classification accuracy of extremely weak classifiers and the diversity of CR-based classifiers for ensemble learning is still an open problem.

Therefore, the goal of the article is to make full use of the ability of DES to obtain *a priori* information. Then, the CR classifier (CRC) is directly improved by weighting residuals. This method considers the inherent discrepancies of different representation learning through residual analysis. The idea of the DES is introduced to use the prior information of classifier behavior. Meanwhile, the unique advantages of each base classifier are fully utilized compared with traditional ensemble learning, double-weighting the residuals directly, and taking into account both local and global information of the classifier. The methods proposed in this article have the following advantages. First, then the two methods make full use of the representation learning model for the ensemble. Meanwhile, the local and prior global information is used for weighting. At last, the proposed methods also consider the inherent differences of the classifiers and use the specific advantages of each base classifier, which can yield better classification results. Notably, different from the traditional DES method, the region of competence (RoC) used in the article is only to obtain the prior information of the classifier without the selection process. Since the residuals of different CR-based classifiers are different, this prior information is used to construct a weight matrix to constrain the final result of the ensemble. That is, a new classifier-weighted learning strategy is proposed based on the DES strategy. The two methods proposed in the article do not perform classifier selection but weigh each classifier according to prior knowledge. Two classifier ensemble strategies called local weighted residuals (LWR) and double weighted residuals (DWR) dynamic ensemble learning are proposed. The major contributions are summarized as follows:

1) The multiple CR-based classifiers are combined through the use of residuals. The prior behavior information of the classifier is obtained by constructing validation samples to constrain the residuals of CR. Ensemble classifiers from the perspective of residuals can make better use of the differences between individual classifiers. Meanwhile, the misclassification problem due to insignificant residuals can be avoided by using the prior information to weigh the residuals.
2) The article proposes to obtain the prior information of the classifiers and use it directly for a residual ensemble to better exploit the unique advantage of each classifier. The prior information on the classification behavior in multiple target regions is obtained by K-NN and clustering. The residuals for each classification are then constrained with local priors' weights. Then, the behavior of each classifier is also constrained by weighting the residuals. This method is different from the traditional ensemble method that fuses classification results but directly weights the residuals from individual classifiers. This makes better use of the local classification advantages of each classifier.
3) The article proposes to use both local and global information to double constraints on the classifier to obtain ensemble results. Local weights are used to constrain the behavior of the classifier while also considering global information. Using the clustering method to obtain the local information of the classifier, the global behavior information of the classifier is also obtained by K-NN. Finally, a more reliable dual-weight constrained classifier fusion result is obtained. This method simultaneously considers each classifier's regional and global behavior on the unknown testing samples while ensuring that weak classifiers can also be fully utilized.

The remainder of this article is organized as follows. Section II introduces related work. Section III proposes the LWR-DEL and DWR-DEL algorithms. In Section IV, experiments and analyses with three real hyperspectral data are presented. Finally, Section V concludes this article.

II. Related Work

A. Residuals of CR-Based Classifiers

The basic idea of sparse representation comes from compressed sensing. It is assumed that the fewest samples can represent the testing data. When the samples are highly correlated, the projection of the sample $\mathbf{y}$ in each class may be roughly the same, and the result of the sparse representation classifier is unstable. Therefore, CRC is proposed, which is described in detail as follows. Given a matrix of training samples $\boldsymbol{X} = [x_1, x_2, \ldots, x_k] \in \mathbf{R}^{m\times n}$ for k classes and a testing sample $\boldsymbol{y} \in \mathbf{R}^m$, the objective is to solve the ℓ^2-minimization problem

$$\hat{\boldsymbol{x}}_1 = \arg\min_x \|\boldsymbol{x}\|_2 \text{ subject to } \|\boldsymbol{\alpha X} - \boldsymbol{y}\|_2 < \varepsilon. \quad (1)$$

The residual is computed as

$$r_i(\boldsymbol{y}) = \|\boldsymbol{y} - \alpha \boldsymbol{x}_i\|_2 \text{ for } i = 1, \ldots, k. \quad (2)$$

Finally, y can be classified as

$$\text{class}(\boldsymbol{y}) = \arg\min_i r_i(\boldsymbol{y}). \quad (3)$$

Based on CRC, improved variants are developed from kernel tricks (KCRC) [48], probabilistic interpretation (ProCRC) [49], and other perspectives. Fundamentally, the basic principle of all CR-based classifiers is still based on the ℓ^2- minimization problem.

However, the significant difference between the improved methods of CR-based classifiers makes their residuals dissimilar in value. Therefore, a suitable residual fusion method for optimal CR-based classifiers ensemble results is to be found in the article. The normalized residuals were used in the article because large difference among the residuals obtained by multiple CR-based classifiers.

B. Dynamic Selection

DES is an ensemble strategy that can choose an expert classifier for each testing region in the feature space. Unlike the traditional ensemble methods, the DES method assumes that each classifier has its own advantage. Even a weak classifier may have a better classification performance for some classes and some testing samples. Therefore, the most competent classifiers can be chosen for specific instances. For DES, the most essential concept is competence. It should be noted that the most competent here refers to the classification ability of each base classifier for a specific area.

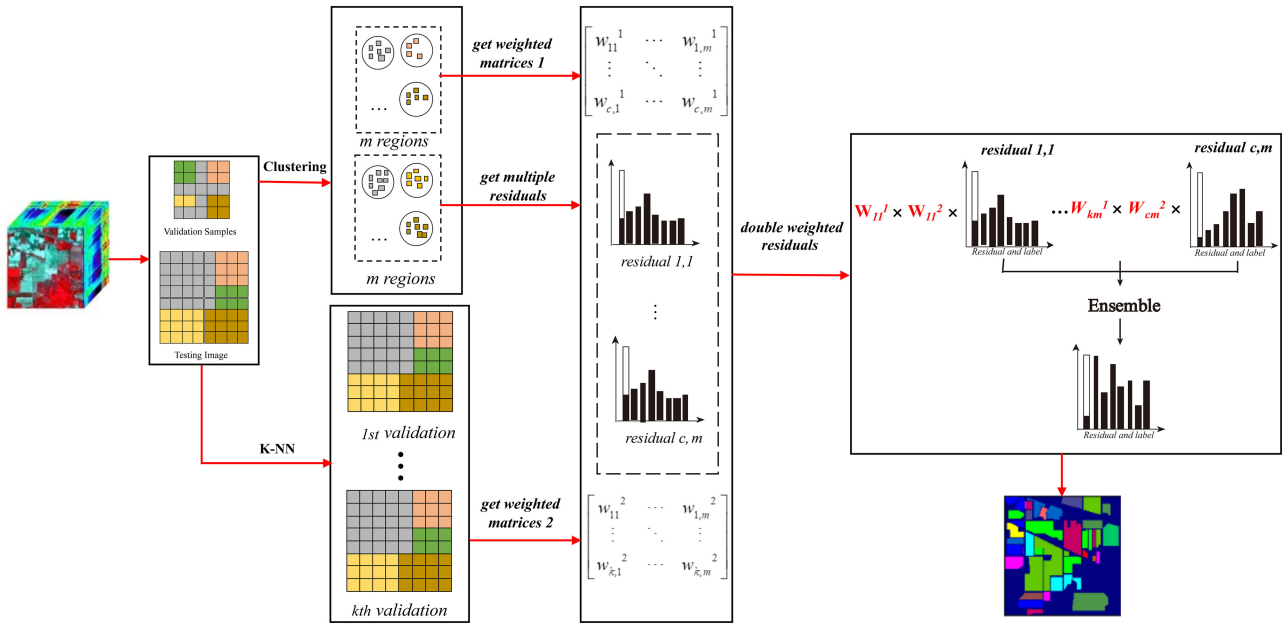

Fig. 5. Flowchart of DWR-DEL.

The general process for DES can be divided into three main steps: RoC definition, competence estimation, and selection strategy. RoC definition is the critical step, and it is also the main idea introduced by the algorithm proposed in the article. The summary is detailed as follows:

1) RoC Definition Based on Clustering: The first common RoC definition method is mainly based on clustering. Given validation and testing samples, they are assumed to have the same classes and feature space. The number of verification samples and test samples and the corresponding class are described in detail in Tables I–III, respectively. Then, according to the concept of multiview clustering, the validation and the testing set can be divided into the same homogeneity regions according to a specific pattern similarity measure. As shown in Fig. 1, given the validation sample set $\boldsymbol{X}_v = [x_v^1, \ldots, x_v^n]$ and testing samples set $\boldsymbol{X}_t = [x_t^1, \ldots, x_t^m]$, they are divided into the same homogeneous testing set. It can be considered that each base classifier has the same classification ability for the corresponding roc_i. Therefore, the prior competence information of each base classifier can be obtained through the validation set. Finally, assign the most suitable base classifier for each RoC of the testing set. In fact, by using the clustering method to divide the RoC, the local prior information of the samples in each cluster can be obtained. The method is shown to be effective in selection classifiers with local classification ability.

2) RoC Definition Based on K-NN: The K-NN method is another standard partitioning method for RoC in DES. This method uses the K-NN to construct the k nearest samples for each instance in the testing set for validation. Unlike the clustering method, it defines $\boldsymbol{N}$ regions $\text{RoC} = [\text{roc}_1, \ldots, \text{roc}_N]$ with homogeneity. The divided region of validation t and the RoC through K-NN can be regarded as a region partitioning method with global information. Because this method constructs k validation samples for each instance in the testing set. As shown in Fig. 2, given a testing sample $\boldsymbol{X}_t = [x_t^1, \ldots, x_t^m]$, k global validation samples can be obtained through the K-NN method. Similarly, it can be considered that the k validation sample sets are homogeneous with the testing set. As with the clustering-based method, the classification accuracy of each base classifier is obtained the validation set. The difference is that the K-NN method can get the global prior classification information of each base classifier in the pool.

Algorithm 1: LWR-DEL.

Input: the pool of CR-based classifiers ***Clf***; the spectral feature dataset ***Va*** and ***Te*****;** the cluster size ***n_r***;
for each validation sample ***v*** in ***Va*****, do**
Find $\boldsymbol{RoC}_{val}$ as the *K-RoC* via clustering
for each testing sample ***t*** in ***Te*** **do**
Find $\boldsymbol{RoC}_{te}$ *as the K-RoC* via clustering based on $\boldsymbol{RoC}_{val}$
for each classifier $\boldsymbol{clf}_i$ in ***Clf*** **do**
get the local weight matrix ***LWM*** from $\boldsymbol{RoC}_{val}$
get the residual of each class from $\boldsymbol{RoC}_{te}$
end for
Get local weighted residuals ***WR*** according to formula (9)
end for
Obtain the final results ***C*** via minimize the ***WR***
end for
Output: Classification results ***C*** for each unknown sample *t* in *Te*.

III. Proposed Methods

The details of the proposed LWR-DEL and DWR-DEL algorithms are mainly described from three aspects: classifier pool construction, weight matrices calculation, residual weighting

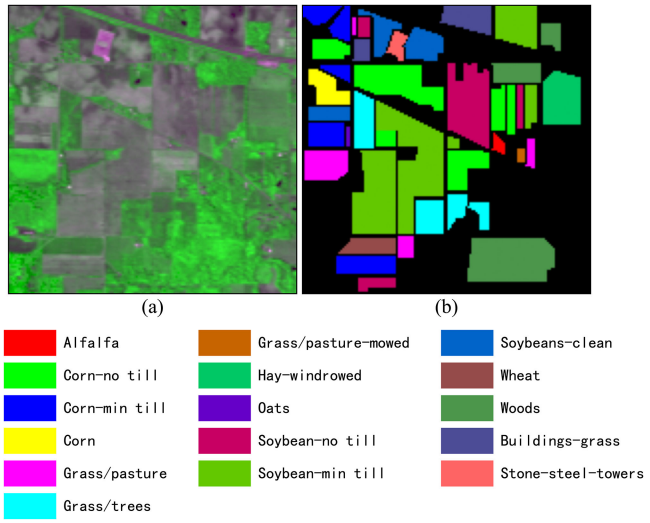

Fig. 6. (a) False-color image and (b) ground truth of the Indian Pines data set.

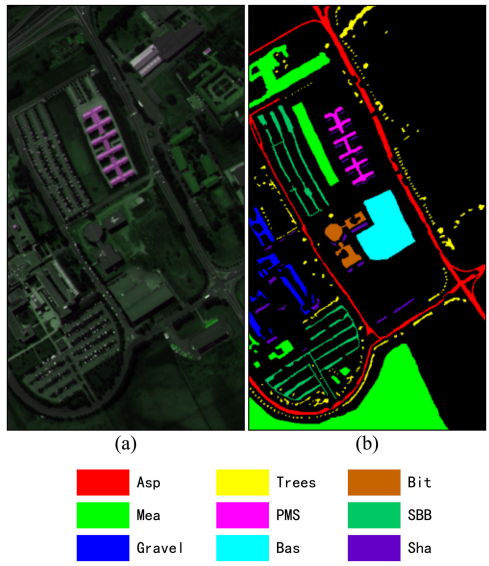

Fig. 7. (a) False-color image and (b) ground truth of the University of Pavia data set.

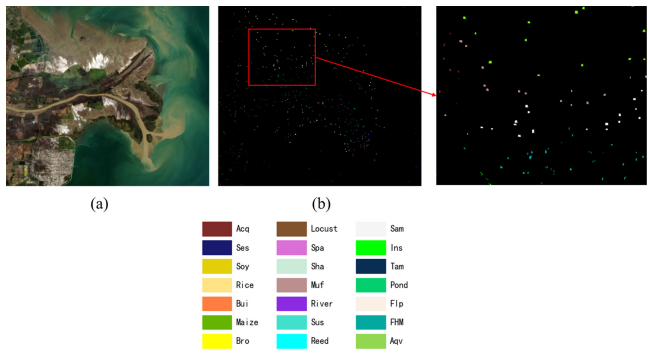

Fig. 8. (a) False-color image and (b) ground truth of the Yellow River data set.

and fusion. The structure of LWR-DEL and DWR-DEL is shown in Figs. 4 and 5, respectively

Algorithm 2: DWR-DEL

Input: the pool of CR-based classifiers ***Clf***; the spectral feature dataset ***Va*** and ***Te***;the cluster size ***n_r***; the number neighborhood ***k***
for each validation sample ***v*** in ***Va***, **do**
Find $\boldsymbol{RoC}_{val}$ as the *K-RoC* via clustering
for each testing sample ***t*** in ***Te*** **do**
Find $\boldsymbol{RoC}_{te}$ *as the K-RoC* via clustering based on $\boldsymbol{RoC}_{val}$
for each classifier $\boldsymbol{clf}_i$ in ***Clf*** **do**
get the local weight matrix $\boldsymbol{W}^1$ from $\boldsymbol{RoC}_{val}$according to formula (4)
get the residual of each class from $\boldsymbol{RoC}_{te}$
end for
for each testing sample ***t*** in ***Te*** **do**
Find *K-validations* via K-NN based on ***Va***
for each classifier $\boldsymbol{clf}_i$ in ***Clf*** **do**
get the global weight matrix $\boldsymbol{W}^2$ from ***Te***according to formula (11)
end for
Obtain double-weighted residuals ***DWR*** according to formula (12)
Obtain the final results ***C*** via minimize the ***DWR***
end for
Output: Classification results ***C*** for each unknown sample *t* in *Te*;

A. Necessity of Weighted Residual Fusion

However, as can be seen from Fig. 3, the normalized residual results of CRC ($\lambda = 1e{-}1$), ProCRC ($\lambda = 1e{-}1$, $\gamma = 1e{-}1$), and KCRC ($\lambda = 1e{-}1$) for the same pixel are quite different. For a single classifier, as shown in Fig. 3(a), the CRC classifier has little difference for the residuals of classes 4, 7, and 12. As can be seen from Fig. 3(b) and (c), the overall residual discrimination of ProCRC and KCRC is very low. If converted into probability output, the probabilities of classification results for the pixel belonging to these 16 classes are similar. Meanwhile, it can be seen from Fig. 3(d) that the residuals between the three classifiers are also different. Therefore, the results of residual fusion are also unreliable if they are simply added. In summary, different CR-based classifiers are not discriminative for the single classification results of certain pixels, and cannot use a simple addition method for residual fusion. Therefore, we introduce the local and global prior information obtained by DS and propose the following two weighted residual fusion algorithms.

B. Local Weighted Residuals Dynamic Ensemble

1) Classifier Pool Generation: Three different CR classifiers, i.e., CRC, KCRC, and ProCRC, are selected to meet the differences of the models, and various parameters are set to meet

TABLE IV
CLASSIFICATION ACCURACY (%) FOR THE INDIAN PINES DATA SET

Class	SVM	RF	CRC	ProCRC	GBDT	CatBoost	LightGBM	XGboost	DES-MI	DES-Cluster	Meta-DES	LWR-DEL	DWR-DEL
1	100.00	95.45	95.45	100.00	95.45	95.45	95.45	95.45	95.45	95.45	95.45	95.45	95.45
2	51.33	64.55	61.96	65.92	63.83	71.18	80.04	70.68	75.22	74.86	71.83	71.47	72.19
3	74.56	86.73	83.82	81.67	83.57	88.24	87.86	81.04	90.64	91.40	88.24	84.58	84.45
4	82.09	83.92	77.39	86.93	82.41	94.97	93.97	95.98	85.93	86.93	88.94	80.40	81.41
5	89.84	87.61	87.16	86.71	77.70	87.61	86.94	84.68	89.86	89.41	90.54	88.29	88.29
6	94.20	88.89	93.51	98.56	64.50	94.52	91.77	79.37	92.93	92.21	96.97	99.42	99.42
7	100.00	100.00	100.00	100.00	66.67	100.00	100.00	100.00	83.33	100.00	83.33	100.00	100.00
8	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	99.77	100.00	100.00	100.00	100.00
9	100.00	100.00	100.00	0.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
10	77.47	82.62	84.44	87.77	79.94	79.40	79.72	81.65	83.69	83.26	85.19	86.16	86.80
11	72.38	69.19	90.72	93.13	69.77	76.85	77.47	79.79	74.33	67.37	79.21	92.17	92.46
12	58.81	81.08	71.71	82.16	73.87	92.43	94.23	92.79	75.50	81.26	79.46	79.28	79.46
13	94.19	99.40	99.40	100.00	92.81	99.40	99.40	94.01	97.60	96.41	97.60	100.00	100.00
14	78.78	78.12	99.51	99.92	91.02	92.65	92.33	89.14	89.55	80.65	88.98	99.43	99.43
15	92.82	76.66	89.05	75.50	93.08	96.83	98.85	93.37	70.03	74.35	89.91	94.81	94.81
16	93.55	98.31	93.22	100.00	94.92	91.53	89.83	100.00	100.00	100.00	98.31	93.22	93.22
OA (%)	75.21	78.07	85.76	87.83	77.02	84.55	85.85	83.05	82.58	80.13	84.52	88.89	89.13
AA (%)	85.00	87.03	89.21	84.89	83.10	91.32	91.74	89.87	87.74	88.35	89.62	91.54	91.70
Kappa	0.52	0.75	0.84	0.86	0.74	0.82	0.84	0.81	0.80	0.78	0.82	0.79	0.88
F1-score	0.73	0.72	0.77	0.75	0.74	0.79	0.79	0.77	0.75	0.73	0.78	0.87	0.79
Times (s)	0.25	0.46	0.87	0.91	18.29	218.58	0.74	0.69	3.43	3.52	4.95	0.96	6.74

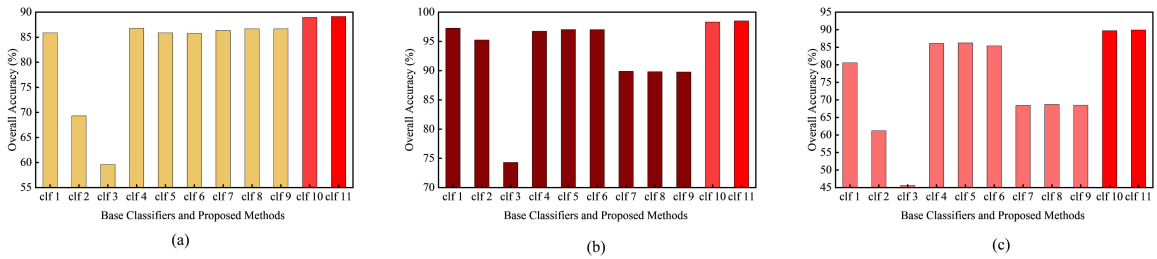

Fig. 9. Classification accuracy comparison of base classifiers versus proposed methods. Clf 1 to clf 9 are base classifiers used in the ensemble. Clf 10 and clf 11 are proposed algorithms in the article. (a) Indian Pines data set. (b) University of Pavia data set. (c) Yellow River data set.

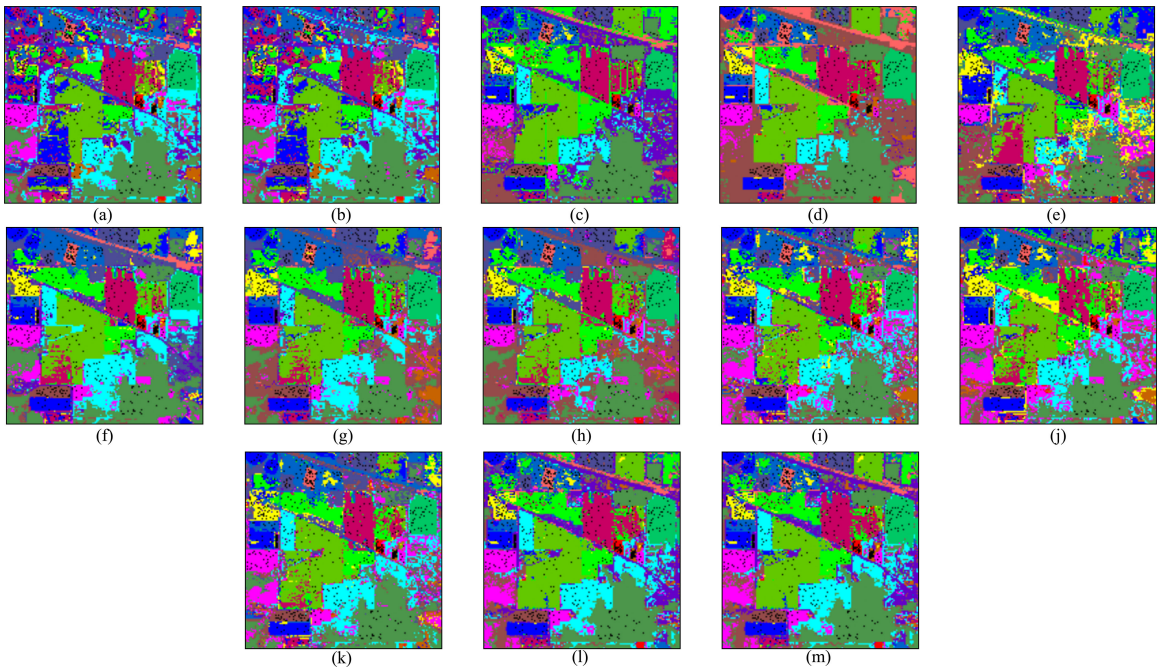

Fig. 10. Classification maps resulting from the classification for the Indian Pines data set using 20 labeled samples per class. (a) SVM. (b) RF. (c) CRC. (d) ProCRC. (e) GBDT. (f) CatBoost. (g) LightGbM. (h) XGboost. (i) DES-MI. (j) DES-Cluster. (k) Meta-DES. (l) LWR-DEL. (m) DWR-DEL.

the differences of the parameters. Let the resulting CR classifier pool be denoted as $P = \{\text{clf}_1, \text{clf}_2, \ldots, \text{clf}_c\}$.

2) Local Weight Matrices Calculating: Fig. 4shows the validation sample set $\boldsymbol{X}_v = [x_v^1, \ldots, x_v^n]$ and testing samples set $\boldsymbol{X}_t = [x_t^1, \ldots, x_t^m]$. The validation set $\boldsymbol{X}_v$is divided into m regions by multiview clustering. Then, the distance from the testing sample x_t^i to the m cluster centers is computed, and the testing samples into m regions are partitioned according to the

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

Class	SVM	RF	CRC	ProCRC	GBDT	CatBoost	LightGBM	XGboost	DES-MI	DES-Cluster	Meta-DES	LWR-DEL	DWR-DEL
1	90.97	94.72	95.99	96.31	86.18	96.43	96.65	93.92	96.09	97.54	95.43	97.65	97.69
2	73.53	78.45	98.14	99.19	73.97	73.06	84.39	69.55	98.90	98.96	98.86	97.72	98.01
3	54.59	92.42	92.52	99.61	84.26	93.88	96.26	87.47	97.18	95.82	96.55	98.54	98.54
4	98.25	98.08	98.88	99.57	86.90	92.06	93.72	85.65	91.83	90.21	95.44	98.97	99.07
5	99.85	99.69	99.92	100.00	100.00	100.00	99.16	95.33	99.77	99.92	99.77	100.00	100.00
6	79.78	99.70	93.25	96.15	88.27	99.76	99.86	97.90	95.93	84.39	97.47	99.56	99.52
7	99.84	100.00	99.92	100.00	99.38	99.92	99.53	99.53	99.84	99.22	100.00	100.00	100.00
8	98.79	95.36	96.40	86.19	94.23	96.40	98.43	96.35	90.97	93.82	94.62	98.02	98.05
9	99.89	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
OA (%)	82.16	88.78	97.01	97.39	82.84	86.43	91.81	83.34	96.93	95.87	97.53	98.27	**98.41**
AA (%)	88.39	95.38	97.23	97.45	90.36	94.61	96.45	91.74	96.72	95.54	97.57	98.94	**98.99**
Kappa	0.77	0.86	0.96	0.97	0.78	0.83	0.89	0.79	0.96	0.94	0.97	0.98	**0.99**
F1-score	0.87	0.92	0.97	0.97	0.87	0.92	0.94	0.87	0.97	0.96	0.97	0.98	0.98
Times (s)	3.60	2.14	4.36	6.86	15.66	63.46	1.07	0.54	29.65	36.96	27.21	10.24	12.24

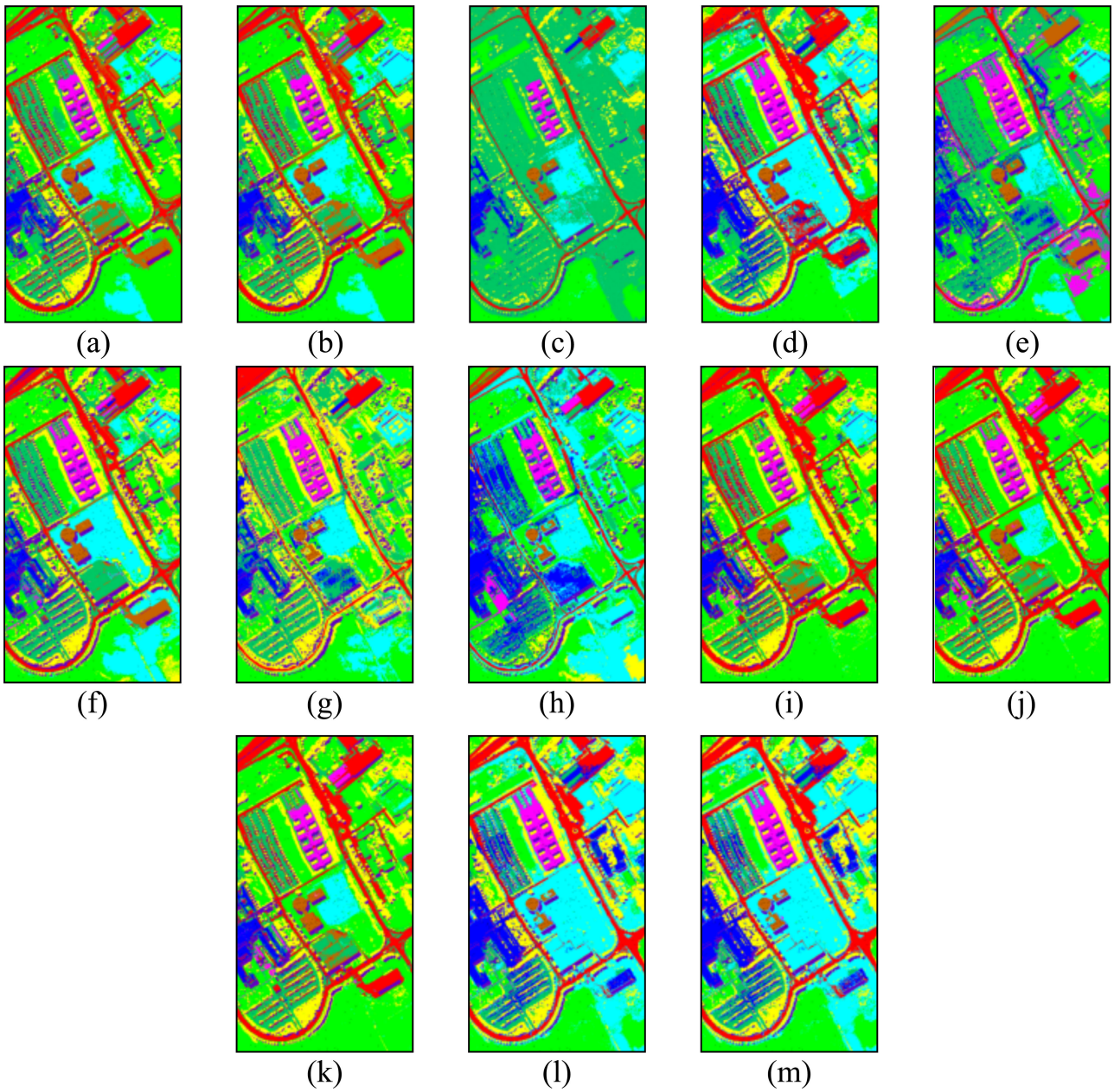

Fig. 11. Classification maps resulting from the classification for the University of Pavia data set using ten labeled samples per class. (a) SVM. (b) RF. (c) CRC. (d) ProCRC. (e) GBDT. (f) CatBoost. (g) LightGbM. (h) XGboost. (i) DES-MI. (j) DES-Cluster. (k) Meta-DES. (l) LWR-DEL. (m) DWR-DEL.

TABLE VI
CLASSIFICATION ACCURACY (%) FOR YELLOW RIVER DATA SET

Class	SVM	RF	CRC	ProCRC	GBDT	CatBoost	LightGBM	XGboost	DES-MI	DES-Cluster	Meta-DES	LWR-DEL	DWR-DEL
1	96.90	91.55	78.03	83.38	90.14	96.34	85.92	88.73	94.08	94.37	95.49	84.23	84.23
2	100.00	97.62	100.00	100.00	97.49	98.68	89.55	89.55	98.02	99.74	98.41	100.00	100.00
3	100.00	90.79	97.37	97.37	80.26	93.42	89.47	78.95	92.11	92.11	97.37	98.68	100.00
4	85.71	86.36	78.57	87.66	74.68	88.96	72.73	80.52	85.71	90.91	87.66	87.01	87.66
5	62.96	75.93	90.74	98.15	57.41	81.48	72.22	61.11	75.93	72.22	77.78	94.44	96.30
6	100.00	90.48	100.00	100.00	88.89	85.71	84.13	80.95	92.06	95.24	92.06	100.00	100.00
7	100.00	88.71	88.71	98.39	87.10	100	95.16	100.00	100.00	100.00	96.77	100.00	100.00
8	99.43	98.30	100.00	100.00	93.75	100	97.16	92.61	96.59	96.59	98.30	100.00	100.00
9	98.79	97.58	100.00	100.00	100.00	99.39	100	99.39	98.79	98.79	98.79	100.00	100.00
10	50.45	88.17	83.59	87.95	66.07	88.06	80.92	75.67	84.26	89.06	95.54	88.28	88.39
11	96.12	85.83	99.61	100.00	78.64	86.6	82.72	80.39	85.63	85.63	86.02	100.00	100.00
12	100.00	100.00	100.00	100.00	98.37	100	96.51	96.28	100.00	100.00	100.00	100.00	100.00
13	100.00	93.79	100.00	100.00	72.98	100	81.99	84.16	91.30	94.41	94.10	100.00	100.00
14	85.64	82.67	53.96	37.62	85.64	82.67	80.69	79.70	79.70	84.16	78.22	69.80	77.72
15	100.00	100.00	97.48	99.28	100.00	100	97.84	96.58	100.00	100.00	100.00	100.00	100.00
16	93.30	90.91	89.71	100.00	84.69	87.56	89.23	87.32	90.19	90.19	91.15	95.93	94.02
17	95.00	91.00	70.00	96.00	74.00	87	89	85.00	91.00	95.00	86.00	88.00	93.00
18	95.86	89.05	82.25	74.26	80.47	91.72	84.62	79.29	80.47	82.25	78.70	84.91	87.57
19	100.00	97.30	100.00	100.00	97.30	97.3	97.3	94.59	97.30	94.59	94.59	97.30	100.00
20	87.80	92.68	58.54	78.05	92.68	82.93	90.24	85.37	75.61	90.24	92.68	90.24	92.68
21	100.00	100.00	100.00	100.00	100.00	100	100	100.00	100.00	100.00	100.00	100.00	100.00
OA (%)	89.51	92.48	90.73	92.67	85.10	93.42	87.63	85.97	91.26	92.99	93.44	94.22	**94.71**
AA (%)	92.76	91.84	88.98	92.29	85.74	92.75	88.45	86.48	90.89	92.64	92.36	94.23	**95.31**
Kappa	0.89	0.92	0.90	0.92	0.84	0.93	0.87	0.85	0.90	0.92	0.93	0.94	0.94
F1-score	0.90	0.89	0.86	0.89	0.80	0.89	0.83	0.81	0.88	0.90	0.89	0.92	0.93
Times (s)	30.43	27.14	152.20	174.29	150.93	650.95	142.95	3.40	295.91	274.06	289.46	8.70	7.62

distance. For each classifier in the classifier pool P, their classification accuracy in the m regions RoC $= [\text{roc}_1, \text{roc}_2, \ldots, \text{roc}_m]$ is computed. The weight matrices are obtained based classification accuracy and denoted as

$$\boldsymbol{W}^1 = \begin{bmatrix} w_{11}^1 & \cdots & w_{1m}^1 \\ \vdots & \ddots & \vdots \\ w_{c1}^1 & \cdots & w_{cm}^1 \end{bmatrix} \tag{4}$$

where w_{ij} represents the weight of the classification accuracy of the ith classifier in the pool for the jth region.

3) Residuals Weighting and Fusing: According to (1), the representation coefficient of RoC can be obtained by using the classifier clf_i.

$$\hat{\boldsymbol{x}}_{\text{roc}_i} = \arg\min_x \|\boldsymbol{x}_{\text{roc}_i}\|_2 \text{ subject to } \left\|\boldsymbol{\alpha}_{\text{roc}_i}\boldsymbol{X}_{\text{roc}_i} - \boldsymbol{y}_{\text{roc}_i}\right\|_2 < \varepsilon \quad i = 1, 2, .., c, \quad j = 1, 2, \ldots, m \tag{5}$$

which is solved as

$$\alpha_{\text{roc}_{ij}} = \left(X_{\text{roc}_{ij}}^T X_{\text{roc}_{ij}} + \lambda I\right)^{-1} X_{\text{roc}_{ij}}^T y_{\text{roc}_{ij}} \quad i = 1, 2, \ldots, c, \quad j = 1, 2, \ldots, m. \tag{6}$$

Then, the coefficient matrix is

$$\text{RoC}(\alpha) = \begin{bmatrix} \alpha_{11}^1 & \cdots & \alpha_{1c}^1 \\ \vdots & \ddots & \vdots \\ \alpha_{m1}^1 & \cdots & \alpha_{mc}^1 \end{bmatrix}. \tag{7}$$

The residual matrix is

$$r_{\text{RoC}_i}(\boldsymbol{y}) = \|\boldsymbol{y} - \alpha \boldsymbol{x}_{\text{RoC}_i}\|_2 \text{ for } i = 1, \ldots, k \quad \boldsymbol{R} = \begin{bmatrix} r_{11}^1 & \cdots & r_{1c}^1 \\ \vdots & \ddots & \vdots \\ r_{m1}^1 & \cdots & r_{mc}^1 \end{bmatrix}. \tag{8}$$

Finally, using the weight matrix in (4), the LWR is obtained

$$WR = \sum_i^c \left(\begin{bmatrix} w_{11}^1 & \cdots & w_{1m}^1 \\ \vdots & \ddots & \vdots \\ w_{c1}^1 & \cdots & w_{cm}^1 \end{bmatrix} * \begin{bmatrix} r_{11} & \cdots & r_{1m} \\ \vdots & \ddots & \vdots \\ r_{1m}^1 & \cdots & r_{cm}^1 \end{bmatrix} \right) \tag{9}$$

where w_{ij} represents the weight of the classification accuracy of the ith classifier in the pool for the jth region and r_{ij} is the corresponding residual. Then, the final classification result is obtained according to (3)

$$\text{class}(\boldsymbol{y}) = \arg\min WR(y). \tag{10}$$

C. Double-Weighted Residuals Dynamic Ensemble

1) Classifier Pool Generation: The process of classifier pool construction is the same as the LWR-DEL algorithm.

2) Double-Weight Matrices Calculating: According to the basic idea of RoC definition based on K-NN, construct k global validation samples from the testing samples $V = [v_1, v_2, \ldots, v_k]$. For each classifier in the classifier pool, compute their classification accuracy on k validation samples. Similar to LWR-DEL, a global weight matrix is obtained according to

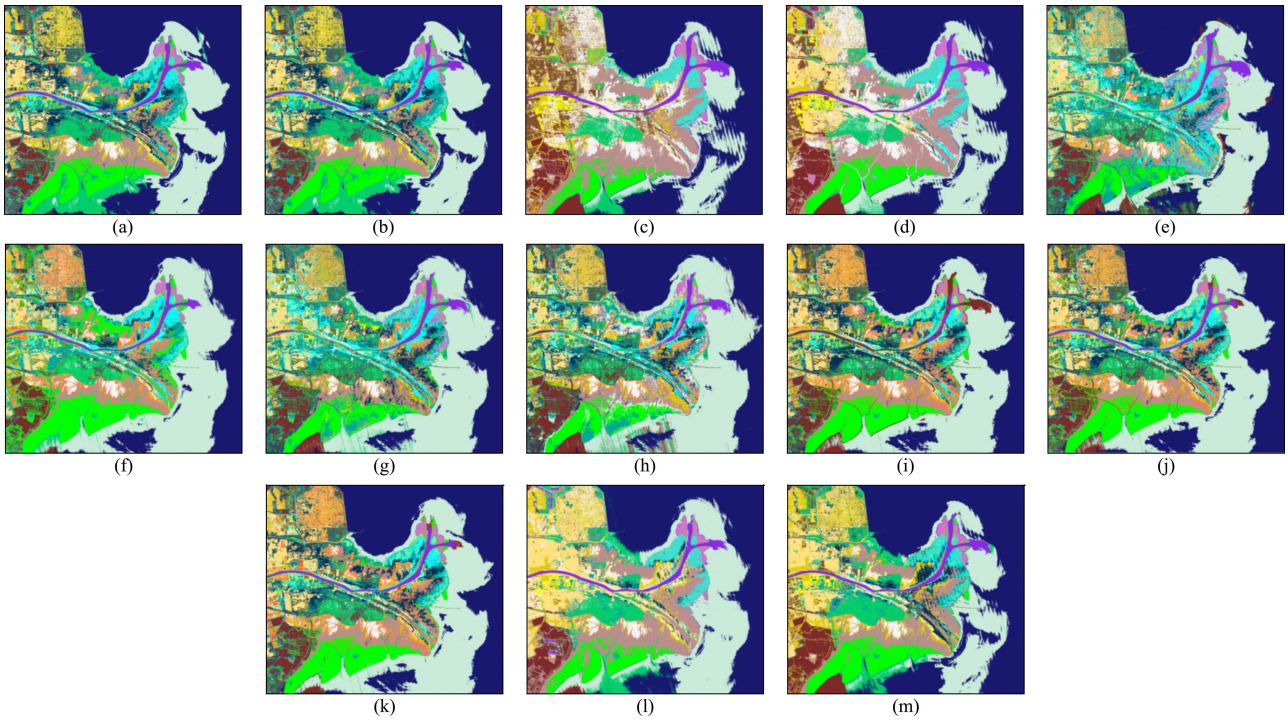

Fig. 12. Classification maps resulting from the classification for the Yellow River data set using fifteen labeled samples per class. (a) SVM. (b) RF. (c) CRC. (d) ProCRC. (e) GBDT. (f) CatBoost. (g) LightGbM. (h) XGboost. (i) DES-MI. (j) DES-Cluster. (k) Meta-DES. (l) LWR-DEL. (m) DWR-DEL.

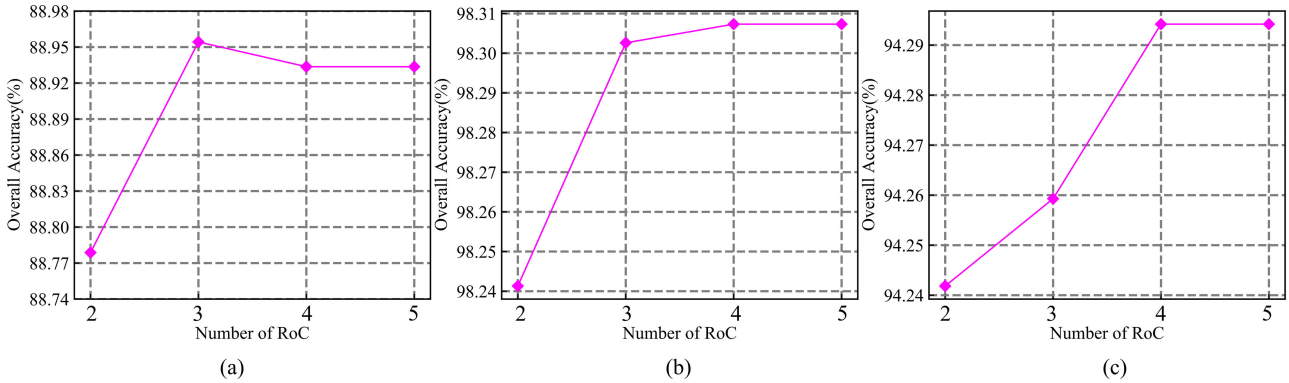

Fig. 13. Classification accuracy of three data set versus varying *n_r* of LWR-DEL. (a) Indian Pines data set. (b) University of Pavia data set. (c) Yellow River data set.

the classifier accuracy

$$\boldsymbol{W}^2 = \begin{bmatrix} w_{11}^2 & \cdots & w_{1k}^2 \\ \vdots & \ddots & \vdots \\ w_{c1}^2 & \cdots & w_{ck}^2 \end{bmatrix}. \tag{11}$$

3) Residuals Weighting and Fusing: Similar to LWR-DEL, using the weight matrix in (4) and (11), the double-weighted residual is obtained

$$\boldsymbol{DWR} = \sum_{i}^{c} \left\{ \sum_{j}^{k} \begin{bmatrix} w_{11}^2 & \cdots & w_{1k}^2 \\ \vdots & \ddots & \vdots \\ w_{c1}^2 & \cdots & w_{ck}^2 \end{bmatrix} * \left(\begin{bmatrix} w_{11}^1 & \cdots & w_{1m}^1 \\ \vdots & \ddots & \vdots \\ w_{c1}^1 & \cdots & w_{cm}^1 \end{bmatrix} * \begin{bmatrix} r_{11} & \cdots & r_{c1} \\ \vdots & \ddots & \vdots \\ r_{1m}^1 & \cdots & r_{cm}^1 \end{bmatrix} \right) \right\} \tag{12}$$

where w_{ij} represents the weight of the classification accuracy of the *i*th classifier in the pool for the *j*th region and r_{ij} is the corresponding residual. Finally, the final classification result is obtained as

$$\text{class}(\boldsymbol{y}) = \arg\min DWR(y). \tag{13}$$

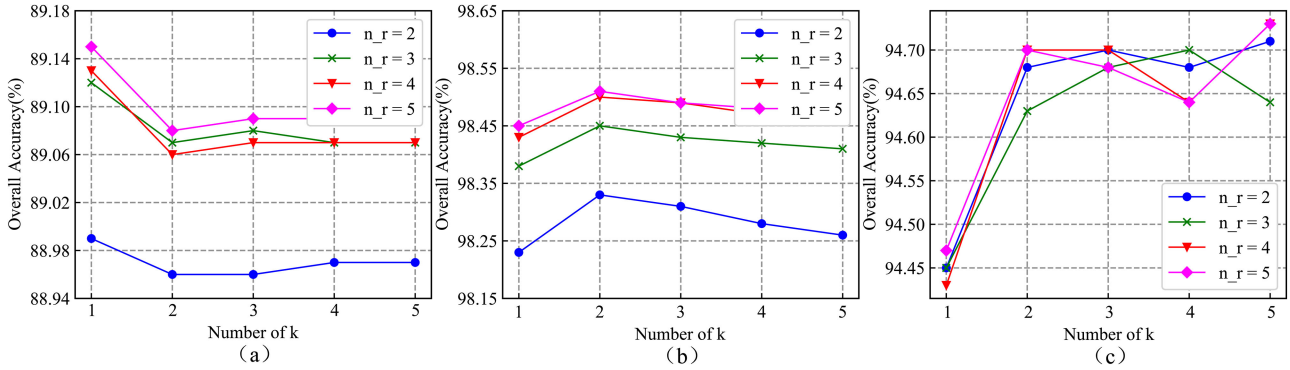

Fig. 14. Classification accuracy of Indian Pines Data set versus varying *n_k* of DWR-DEL. (a) Indian Pines data set. (b) University of Pavia data set. (c) Yellow River data set.

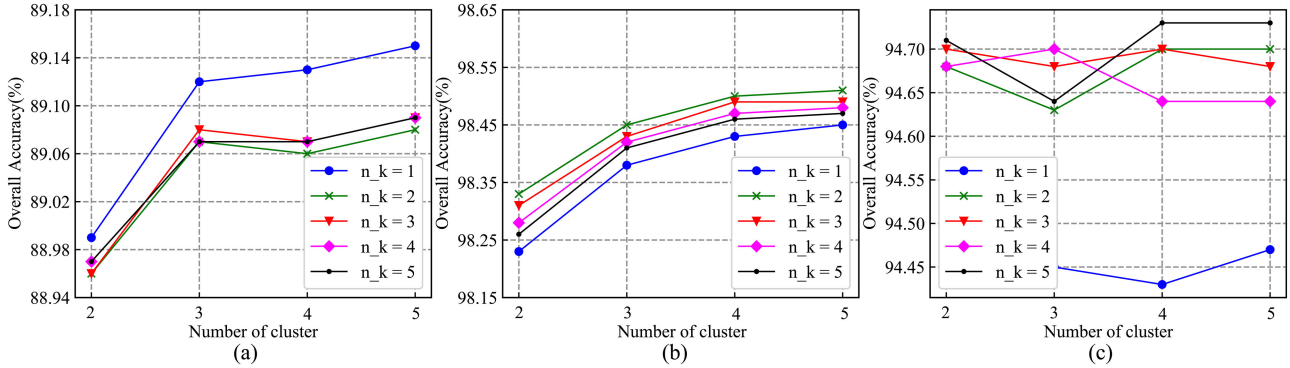

Fig. 15. Classification accuracy of three data sets versus varying *n_r* of DWR-DEL. (a) Indian Pines data set. (b) University of Pavia data set. (c) Yellow River data set.

IV. Experiments

A. Experiment Setup

All experiments are implemented using the platform of Python 3.6.13. To ensure comparability and fairness, the compared and proposed algorithms use the same training, validation, and testing samples.

1) Pool of Classifiers: To evaluate the proposed LWR-DEL and DWR-DEL, three CR classifiers from different types in the classifier pool are applied, i.e., CRC, KCRC, and ProCRC. The range of regularization parameters λ in the three classifiers are all set as {1e-2, 1e-4, 1e-6}. The regularization parameters γ in PoCRC are {1e-1}. The classifier pool finally contains 15 base classifiers.

2) Parameter Settings: The range of cluster numbers *n_r* in LWR-DEL and DWR-DEL is set {2, 3, 4, 5}. The range of *k* numbers *n_k* in DWR-DEL is set {1, 2, 3, 4, 5}. It is worth noting that the method of repeated replacement sampling for the selection of training and testing samples is used in the article. This method can increase the randomness of samples and avoid the impact of sample importance on classification accuracy.

3) Comparison Algorithms: To evaluate the performance of the proposed algorithm, multiple classification algorithms were used for comparison. For example, the classic machine learning algorithms support vector machine (SVM) and random forest (RF) are the baselines. Moreover, the advanced ensemble algorithm GBDT, CatBoost [28], LightGBM [50], and XGboost [51] are also used as comparison algorithms. In addition, the two state-of-the-art DES algorithms, namely DES-MI and META-DES algorithms, are used as comparative algorithms in the article.

B. Hyperspectral Data Sets

The performance of the LWR-DEL and DWR-DEL is evaluated by three real HSI data sets.

The first data set is the Indian Pines data set, collected by the AVIRIS senor. This data set contains 224 spectral bands with wavelengths ranging from 0.4 to 2.5 μm. There are 200 effective bands remaining in the data after removing water absorption bands. The spatial size is 145 $\times$ 145. The details of 16 classes in this HSI are described in Table I, and the images are shown in Fig. 6.

The second image used in this article is the University of Pavia data set, acquired by the ROSIS sensor. This image contains 103 spectral bands with wavelengths ranging from 0.43 to 0.86 μm. The scene consists of nine classes, containing 512 $\times$ 614 pixels, and the spatial resolution is 20 m. The descriptions of classes in this data are listed in Table II, and images are shown in Fig. 7.

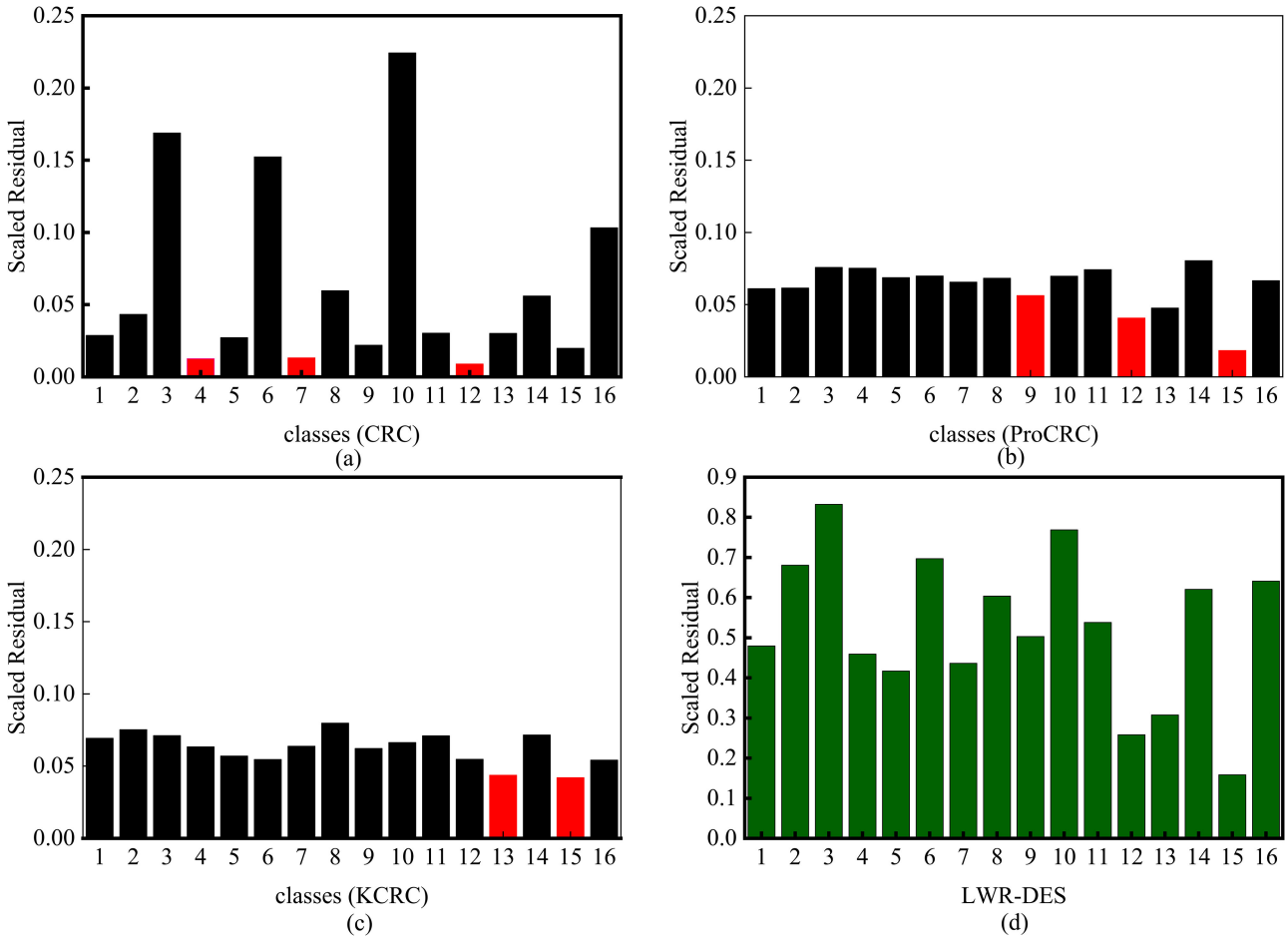

Fig. 16. Scaled Residuals of LWR-DEL versus scaled residuals of multiple CR.

The third is the real HSI Yellow River, which is collected on January 7, 2019, by Gaofen-5 senor [52], [53]. This image contains 285 spectral bands. The spatial size of the Yellow River is 1185 × 1324. There are 21 classes in this data, as shown in Table III, and the false-color image and ground truth image are shown in Fig. 8.

C. Ensemble Classification Performance Analysis

For the ensemble learning method, the key is that the final classification result is better than all base classifiers in the pool. Traditional ensemble learning methods often have higher requirements for the base classifier used. For example, the classification accuracy should be higher than 50%. To verify whether the ensemble performance of the proposed method is effective, we compare the accuracy of the base classifier and the classification accuracy of the proposed methods. As shown in Fig. 9(a), for the first data set, the accuracy of the two proposed algorithms is higher than that of all base classifiers. For the second data set, the same conclusion can be drawn, but the final classification performance is better than the former. For the Yellow River data set, LWR-DEL and DWR-DEL classification accuracy are much higher than that of most base classifiers. Overall, the ensemble results of the two proposed models are higher than those of the base classifier for the Indian Pines and HSI the University of Pavia data. Meanwhile, for the last data set, the proposed methods also can obtain great performance. According to Fig. 9, the proposed models can still achieve high accuracy even if the base classifier accuracy is low. Even when the base classifier accuracy is lower than 85%, the final ensemble accuracy is still higher than 95% [see Fig. 9(c)]. The results show that two ensemble models can fully utilize each base classifier's advantages in different regions. That is, the prior information effectively constrains the behavior of each base classifier. For the Indian Pines data set, the OA, average accuracy (AA), per-class accuracy, kappa statistic, F1-score, and running time (s) of different models are shown in Table IV. The classification maps are shown in Fig. 10(a)–(j).

For the first experiments, the OA of LWR-DEL and DWR-DEL reached 88.89% and 89.13%. Two proposed algorithms have better classification performance than other comparative models. Moreover, the proposed methods are superior to the state-of-the-art DES methods DES-MI and Meta-DES, whereas the LWR-DEL and DWR-DEL do not take too much running time. Compared with CatBoost and LightGBM, the OA of the proposed algorithm is higher, but the AA is not significantly improved. The time complexity of the proposed algorithm is much lower than that of LightGBM. The classification accuracy of LWR-DEL and DWR-DEL is also much higher than that of the

XGboost algorithm. From the F1-score index, the performance of several ensemble algorithms is relatively similar. Therefore, the proposed two dynamic ensemble algorithms outperform the existing baselines and state-of-the-art ensemble learning methods.

To evaluate the performance of the two new DEL models, the University of Pavia data set was used in the second experiment. The best parameters are described in Table V, and the thematic maps of various models are displayed in Fig. 11(a)–(j). Similar to the Indian Pines data set, LWR-DEL and DWR-DEL obtain the best classification performance compared with other methods. The best OA and AA for the ROSIS data set are obtained by the DWR-DEL algorithm, which can reach 98.41%. Compared with the classic ensemble learning method GBDT, LWR-DEL and DWR-DEL yield 15.43% and 15.57% improvements. Moreover, compared with the three DES methods, the two proposed algorithms also have great performance.

For the Yellow River data set, the classification performance of all classifiers is the list in Table VI. The thematic maps are shown in Fig. 12(a)–(j). Compared with other classic machine learning classifiers such as SVM and RF, our methods yield nearly 5% and 3% improvements. Compared with the CatBoost and LightGBM algorithms, the classification accuracy of LWR-DEL and DWR-DEL did not significantly improve but the required running time is greatly reduced. Meanwhile, the classification accuracy of LWR-DEL and DWR-DEL is much better than the XGboost algorithm. Experiments on the Yellow River data set demonstrate that the time cost and classification accuracy of LWR-DEL and DWR-DEL are superior to that of other comparison algorithms.

To verify that weak classifiers contribute to the proposed method, a set of comparative experiments is set up in the article. As shown in Fig. 18(a)–(c), the classification accuracy with no weak classifier is better than all classifiers' ensemble results for the three real HSI data sets. It can be seen from the experimental results that the accuracy of the classifier is lower than using all the classifiers when some weak classifiers are removed. Therefore, this result indicates that the proposed method does take advantage of the classification advantages of the weak classifiers to a certain extent.

D. Discussion

1) Sensitivity in Relation to Region Size for LWR-DEL: For the LWR-DEL, the number of ROC (cluster) n_r significantly impacts the classification performance. To evaluate the influence of various n_r. Fig. 13(a)–(c) shows the accuracy of LWR-DEL when n_r changes in range.

For the first dataset, the classification accuracy increases first and then decreases as n_r increases. The best classification accuracy can be obtained when $n\_r = 3$. For the University of Pavia data set, when n_r is gradually increased, the classification performance of LWR improved, and the optimal classification accuracy is reached when $n\_r = 4$. Unlike the other two datasets, for the Yellow River dataset, the effect of n_r on the classification accuracy has no apparent regularity. The best results are obtained when $n\_r = 5$.

In summary, the experimental results show that when the parameter n_r is set to different values, the accuracy of other datasets is affected differently. However, with the change of n_r, the accuracy of the three data sets does not change much. This result shows that this algorithm is sensitive to the parameter n_r in a limited range but has little effect on classification performance. So, the proposed method has strong robustness as the parameter does not need to be tuned carefully in practical applications.

2) Sensitivity in Relation to Region and k *Size for DWR-DEL:* Since the proposed DWR-DEL adds a global weight constraint based on K-NN. The combined effect of the parameters n_r and n_k on the classification accuracy is investigated in the article. First, it can be seen from Figs. 14(a)–(c) and 15 that after adding the prior global information, the variation trend of the classifier accuracy with the parameters change has changed significantly, which is different from that of LWR-DEL (see Fig. 13). Simultaneously, the classification accuracy of three datasets have been significantly improved. It shows that the double-weight constrained method proposed in DWR-DEL effectively changes the behavior of the classifier in different regions.

For the Indian Pines dataset, when the parameter $n\_r = 2$, the classification accuracy is generally lower, and when $n\_r = 5$, the classification accuracy is overall higher. Notably, this trend should be a combined effect of n_k and n_r. When $n\_k = 1$ and $n\_r = 5$, Algorithm 2 can get the highest accuracy. The results fully illustrate the impact of DWR-DEL on the behavior of the classifier.

Similar to the first dataset, when the parameter $n\_r = 2$, the classification accuracy of the University of Pavia data set is overall lower. The difference is that when the parameter $n\_k = 2$, DWR performs better for classifying the University of Pavia data set. When $n\_k = 2$ and $n\_r = 5$, Algorithm 2 can get the highest accuracy.

As with the first two datasets, the parameters n_r and n_k have a minor impact on the classification accuracy of the Yellow River dataset. The classifier accuracy has a slight trend of change. When $n\_k = 5$ and $n\_r = 4$, Algorithm 2 can get the highest accuracy.

To sum up, compared with LWR-DEL, the new method improves the classification accuracy of images, but the influence of parameter changes on the accuracy is not apparent. Overall, when the same parameters are selected for the three data sets, the variation in accuracy is not very large. Experimental results on this data set present the high robustness of the DWR-DEL algorithm.

3) Comparison of Original and Weighted Residuals: Since the two algorithms proposed in this article are based on the conclusion that the residual discrimination of multiple CR-based classifiers is not obvious enough. Therefore, it is very important to show whether the proposed residual fusion algorithm can effectively increase the final residual discrimination. To validate the effectiveness of the proposed weighted residual ensemble methods, the article compares the difference between the residual distribution of the base classifiers and the final ensemble result. As shown in Fig. 16, it is obvious that the residual distribution of LWR-DEL is more discriminative than the base

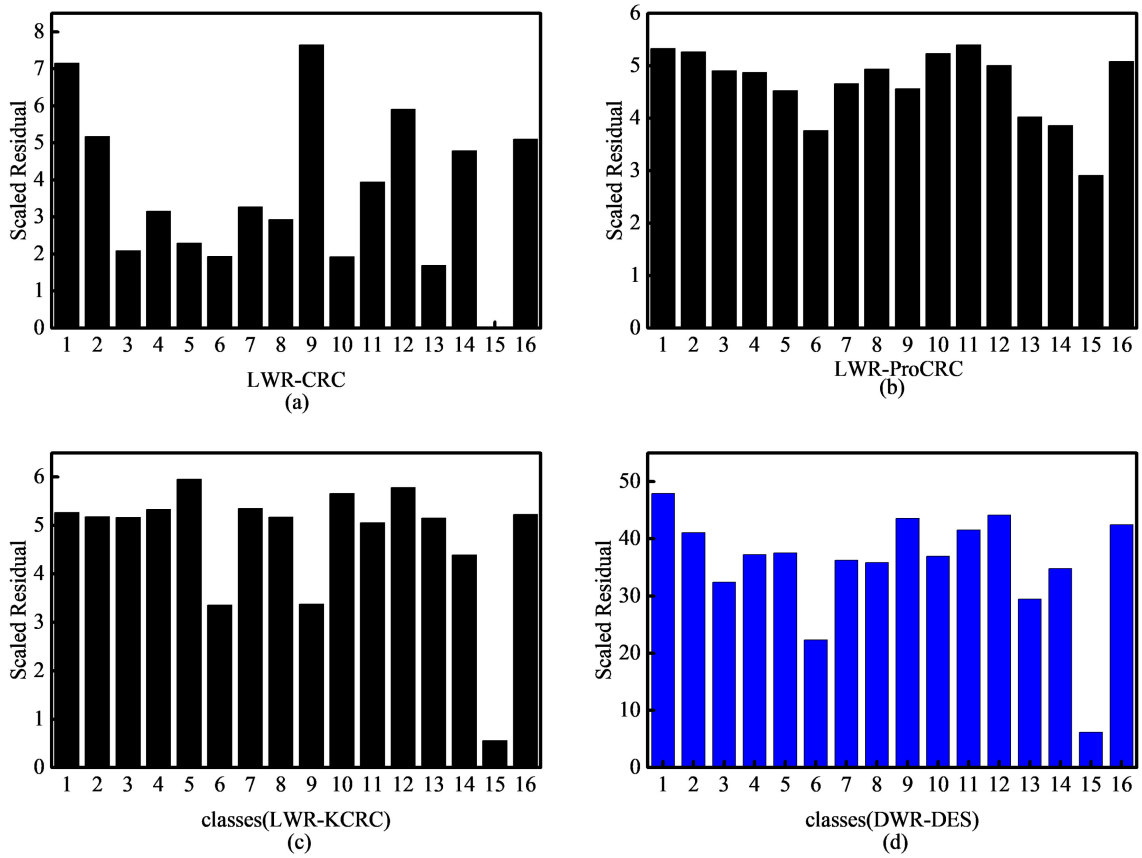

Fig. 17. Scaled residuals of DWR-DES versus scaled residuals of multiple CR.

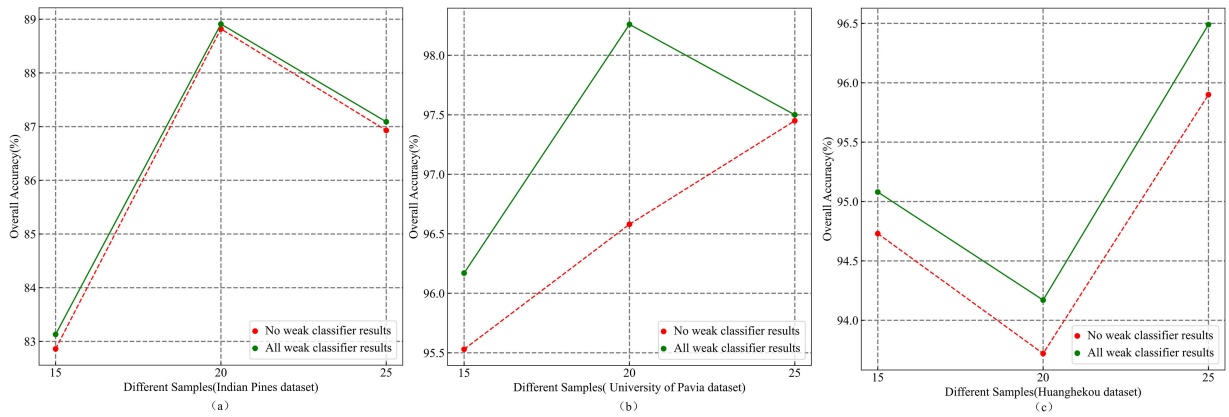

Fig. 18. Classification accuracy of no weak classifier ensemble results versus all classifiers ensemble results.

classifier. Meanwhile, it should be noted that the range of residuals after local weighting is also wider than of all base classifiers. Similar conclusions can be drawn from Fig. 17 , the DWR-DEL method can also increase the discrimination of residuals. The experimental results fully demonstrate the effectiveness of the proposed LWR-DEL and DWR-DEL algorithms.

V. Conclusion

In this article, two new weighted residuals ensemble learning strategies with CR are proposed, which introduces the idea of DES. First, a locally weighted dynamic ensemble algorithm is proposed, which uses the prior accuracy information of the classifier for different clusters as constraints. Residual weighting is then performed on different CR-based classifiers. Furthermore, the article uses the nearest neighbors of the testing samples to obtain the prior global information. Then, a dynamic ensemble method with local and global double-weighting constraints is proposed. Unlike traditional static and dynamic ensemble methods, the two algorithms weigh the residuals to obtain more distinguishable classification results. The experiments show that both the LWR-DEL and DWR-DEL algorithms provide better

classification performance compared to the state-of-the-art classifiers. Compared with the base classifiers in the classifier pool, the ensemble results of the two proposed models also have better accuracy. The experimental results fully demonstrate the feasibility and effectiveness of the proposed algorithms. However, only the Euclidean distance between samples is used in the LWR-DEL and DWR-DEL algorithms. The relationship between the spectral features is not fully considered in the article. Future research will focus on designing more suitable distance metrics that consider both spatial and spectral features.

ACKNOWLEDGMENT

The authors would like to thank Prof. P. Gamba from the University of Pavia, Pavia, Italy, for providing the ROSIS data, Prof. D. Landgrebe for providing the AVIRIS data, and Dr. W. Sun at Ningbo University, China for providing the GF-5 datasets.

REFERENCES

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

[2] M. B. Stuart, A. J. S. McGonigle, and J. R. Willmott, "Hyperspectral imaging in environmental monitoring: A review of recent developments and technological advances in compact field deployable systems," *Sensors*, vol. 19, no. 14. Jul. 2019, Art. no. 3071.

[3] H. Su, Z. Wu, A. Zhu, and Q. Du, "Low rank and collaborative representation for hyperspectral anomaly detection via robust dictionary construction," *ISPRS J. Photogramm. Remote Sens.*, vol. 169, pp. 195–211, Dec. 2020.

[4] P. S. Thenkabail, I. Mariotto, M. K. Gumma, E. M. Middleton, D. R. Landis, and K. F. Huemmrich, "Selection of hyperspectral narrowbands (HNBs) and composition of hyperspectral two band vegetation indices (HVIs) for biophysical characterization and discrimination of crop types using field reflectance and Hyperion/EO-1 Data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 427–439, Apr. 2013.

[5] E. Saralıoğlu, E. T. Görmüş, and O. Güngör, "Mineral exploration with hyperspectral image fusion," in *Proc. 24th Signal Process. Commun. Appl. Conf.*, 2016, pp. 1281–1284.

[6] H. Su, Z. Wu, H. Zhang, and Q. Du, "Hyperspectral anomaly detection: A survey," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 1, pp. 64–90, Mar. 2022, doi: 10.1109/MGRS.2021.3105440.

[7] 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, doi: 10.1109/TGRS.2020.3042274.

[8] Z. Wu et al., "Scheduling-guided automatic processing of massive hyperspectral image classification on cloud computing architectures," *IEEE Tans. Cybern.*, vol. 51, no. 7, pp. 3588–3601, Jul. 2021, doi: 10.1109/TCYB.2020.3026673.

[9] Z. Gao, L. Tong, J. Zhou, B. Qian, J. Yu, and C. Xiao, "Stochastic depth residual network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5506113, doi: 10.1109/TGRS.2021.3090429.

[10] C. Yu et al., "Hyperspectral image classification method based on CNN architecture embedding with hashing semantic feature," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 6, pp. 1866–1881, Jun. 2019, doi: 10.1109/JSTARS.2019.2911987.

[11] S. Zhong et al., "Class feature weighted hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4728–4745, Dec. 2019.

[12] H. Su, B. Zhao, Q. Du, P. Du, and Z. Xue, "Multifeature dictionary learning for collaborative representation classification of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2467–2484, 2018, doi: 10.1109/TGRS.2017.2781805.

[13] 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.

[14] M. Song, C. Yu, H. Xie, and C.-I. Chang, "Progressive band selection processing of hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 10, pp. 1762–1766, Oct. 2020, doi: 10.1109/LGRS.2019.2953525.

[15] C. Yu, R. Han, M. Song, C. Liu, and C.-I. Chang, "Feedback attention-based dense CNN for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5501916, doi: 10.1109/TGRS.2021.3058549.

[16] L. Huang and Y. Chen, "Dual-path siamese CNN for hyperspectral image classification with limited training samples," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 3, pp. 518–522, Mar. 2021, doi: 10.1109/LGRS.2020.2979604.

[17] L. Zhao, W. Luo, Q. Liao, S. Chen, and J. Wu, "Hyperspectral image classification with contrastive self-supervised learning under limited labeled samples," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6008205, doi: 10.1109/LGRS.2022.3159549.

[18] J. Yue, D. Zhu, L. Fang, P. Ghamisi, and Y. Wang, "Adaptive spatial pyramid constraint for hyperspectral image classification with limited training samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5512914, doi: 10.1109/TGRS.2021.3095056.

[19] X. He, Y. Chen, and P. Ghamisi, "Dual graph convolutional network for hyperspectral image classification with limited training samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5502418, doi: 10.1109/TGRS.2021.3061088.

[20] Z. Zhu, Z. Wang, D. Li, Y. Zhu, and W. Du, "Geometric structural ensemble learning for imbalanced problems," *IEEE Trans. Cybern.*, vol. 50, no. 4, pp. 1617–1629, Apr. 2020, doi: 10.1109/TCYB.2018.2877663.

[21] Y. Miao, M. Chen, Y. Yuan, J. Chanussot, and Q. Wang, "Hyperspectral imagery classification via random multigraphs ensemble learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 641–653, 2022, doi: 10.1109/JSTARS.2021.3132993.

[22] J. Wang, J. Cao, S. Yuan, X. Zhou, and P. Zhou, "Spatiotemporal synergistic ensemble deep learning method and its application to S-Wave velocity prediction," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 8024705, doi: 10.1109/LGRS.2021.3123812.

[23] C. León et al., "Evaluation of maturation in preterm infants through an ensemble machine learning algorithm using physiological signals," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 1, pp. 400–410, Jan. 2022, doi: 10.1109/JBHI.2021.3093096.

[24] G. I. Webb and Z. Zheng, "Multistrategy ensemble learning: Reducing error by combining ensemble learning techniques," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 8, pp. 980–991, Aug. 2004, doi: 10.1109/TKDE.2004.29.

[25] T. Hussain, S. M. Siniscalchi, H.-L. S. Wang, Y. Tsao, V. M. Salerno, and W.-H. Liao, "Ensemble hierarchical extreme learning machine for speech dereverberation," *IEEE Trans. Cogn. Develop. Syst.*, vol. 12, no. 4, pp. 744–758, Dec. 2020, doi: 10.1109/TCDS.2019.2953620.

[26] Y. Chen, Y. Wang, Y. Gu, X. He, P. Ghamisi, and X. Jia, "Deep learning ensemble for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 6, pp. 1882–1897, Jun. 2019, doi: 10.1109/JSTARS.2019.2915259.

[27] J. Ding, Y. Wang, H. Si, S. Gao, and J. Xing, "Multimodal fusion-AdaBoost based activity recognition for smart home on WiFi platform," *IEEE Sens. J.*, vol. 22, no. 5, pp. 4661–4674, Mar. 2022, doi: 10.1109/JSEN.2022.3146137.

[28] A. Samat, E. Li, P. Du, S. Liu, and J. Xia, "GPU-Accelerated CatBoost-Forest for hyperspectral image classification via parallelized mRMR ensemble subspace feature selection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3200–3214, 2021, doi: 10.1109/JSTARS.2021.3063507.

[29] G. Fumera, F. Roli, and A. Serrau, "A theoretical analysis of bagging as a linear combination of classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 7, pp. 1293–1299, Jul. 2008, doi: 10.1109/TPAMI.2008.30.

[30] Q. Lv, W. Feng, Y. Quan, G. Dauphin, L. Gao, and M. Xing, "Enhanced-random-feature-subspace-based ensemble CNN for the imbalanced hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3988–3999, 2021, doi: 10.1109/JSTARS.2021.3069013.

[31] A. Samat, P. Du, S. Liu, J. Li, and L. Cheng, "E^2LMs: 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.

[32] R. Fan, R. Feng, L. Wang, J. Yan, and X. Zhang, "Semi-MCNN: A semisupervised multi-CNN ensemble learning method for urban land cover classification using submeter HRRS images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4973–4987, 2020, doi: 10.1109/JSTARS.2020.3019410.

[33] R. Bao, J. Xia, M. Dalla Mura, P. Du, J. Chanussot, and J. Ren, "Combining morphological attribute profiles via an ensemble method for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 13, no. 3, pp. 359–363, Mar. 2016, doi: 10.1109/LGRS.2015.2513002.
[34] J. Li, S. Wu, C. Liu, Z. Yu, and H.-S. Wong, "Semi-supervised deep coupled ensemble learning with classification landmark exploration," *IEEE Trans. Image Process.*, vol. 29, pp. 538–550, 2020, doi: 10.1109/TIP.2019.2933724.
[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] D. Cao, M. Zhang, W. Li, and Q. Ran, "Hyperspectral and infrared image collaborative classification based on morphology feature extraction," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4405–4416, 2021, doi: 10.1109/JSTARS.2021.3072843.
[37] B. Pan, Z. Shi, and X. Xu, "Hierarchical guidance filtering-based ensemble classification for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 4177–4189, Jul. 2017, doi: 10.1109/TGRS.2017.2689805.
[38] R. M. O. Cruz, R. Sabourin, and G. D. C. Cavalcanti, "Dynamic classifier selection: Recent advances and perspectives," *Inf. Fusion*, vol. 41, pp. 195–216, May 2018.
[39] K. Woods, W. P. Kegelmeyer, and K. Bowyer, "Combination of multiple classifiers using local accuracy estimates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 4, pp. 405–410, Apr. 1997.
[40] R. G. F. Soares, A. Santana, A. M. P. Canuto, and M. C. P. de Souto, "Using accuracy and diversity to select classifiers to build ensembles," in *Proc. IEEE Int. Joint Conf. Neural Netw. Proc.*, 2006, pp. 1310–1316.
[41] G. Giacinto and F. Roli, "Dynamic classifier selection based on multiple classifier behaviour," *Pattern Recognit.*, vol. 34, no. 9, pp. 879–1881.1, Jan. 2001.
[42] T. Woloszynski and M. Kurzynski, "A probabilistic model of classifier competence for dynamic ensemble selection," *Pattern Recognit.*, vol. 44, no. 10/11, pp. 2656–2668.1, Oct./Nov. 2011.
[43] R. M. O. Cruz, G. D. C. Cavalcanti, and T. Ing Ren, "A method for dynamic ensemble selection based on a filter and an adaptive distance to improve the quality of the regions of competence," in *Proc. Int. Joint Conf. Neural Netw.*, 2011, pp. 1126–1133.
[44] A. H. R. Ko, R. Sabourin, and A. S. Britto, "From dynamic classifier selection to dynamic ensemble selection," *Pattern Recognit.*, vol. 41, no. 5, pp. 1718–1731, May 2008.
[45] T. Kam Ho, J. J. Hull, and S. N. Srihari, "Decision combination in multiple classifier systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 1, pp. 66–75, Jan. 1994.
[46] B. B. Damodaran, R. R. Nidamanuri, and Y. Tarabalka, "Dynamic ensemble selection approach for hyperspectral image classification with joint spectral and spatial information," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens .*, vol. 8, no. 6, pp. 2405–2417, Jun. 2015.
[47] B. B. Damodaran and R. R. Nidamanuri, "Dynamic linear classifier system for hyperspectral image classification for land cover mapping," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2080–2093, Jun. 2014.
[48] 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, Dec. 2020.
[49] S. Cai, L. Zhang, W. Zuo, and X. Feng, "A probabilistic collaborative representation based approach for pattern classification," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2950–2959.
[50] X. Sun, L. Guo, W. Zhang, Z. Wang, and Q. Yu, "Small aerial target detection for airborne infrared detection systems using LightGBM and trajectory constraints," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 9959–9973, 2021, doi: 10.1109/JSTARS.2021.3115637.
[51] P. Liu, B. Fu, S. X. Yang, L. Deng, X. Zhong, and H. Zheng, "Optimizing survival analysis of XGBoost for ties to predict disease progression of breast cancer," *IEEE. Trans. Biomed. Eng.*, vol. 68, no. 1, pp. 148–160, Jan. 2021, doi: 10.1109/TBME.2020.2993278.
[52] 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.
[53] 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.

Hongliang Lu received the B.E. degree in surveying and mapping engineering from Chuzhou University, Chuzhou, China, in 2015, and the M.E. degree in geography and geographic information engineering 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 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.

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. His main research interests include hyperspectral remote sensing dimensionality reduction, classification, and spectral unmixing.

Dr. Su is an Associate Editor for 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.

Pan Zheng received the B.E. degree in geodesy and geomatics engineering and the M.E. degree in photogrammetry and remote sensing in 2018 and 2021, respectively, from the School of Earth Sciences and Engineering, Hohai University, Nanjing, China, where she is currently working toward the Ph.D. degree in surveying and mapping.

Her research interests include hyperspectral remote sensing imagery, collaborative representation, and machine learning in hyperspectral unmixing.

Yihan Gao received the B.E. degree in surveying and mapping engineering from Hohai University, Nanjing, China, in 2020. She is currently working toward the M.E. degree in surveying and mapping engineering from the School of Earth Sciences and Engineering, Hohai University, Nanjing, China.

Her research interests include hyperspectral remote sensing and soil moisture retrievals form remote sensing.

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 SPIE—International Society for Optics and Photonics and a member of IEEE Periodicals Review and Advisory Committee. She was the 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 was an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, *Journal of Applied Remote Sensing*, and IEEE SIGNAL PROCESSING LETTERS. During 2016–2020, she was the Editor-in-Chief for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.