Dynamic Ensemble Learning With Multi-View Kernel Collaborative Subspace Clustering for Hyperspectral Image Classification

Hongliang Lu, Hongjun Su, Senior Member, IEEE, Jun Hu, Student Member, IEEE, and Qian Du, Fellow, IEEE

Abstract—Recently, a series of collaborative representation (CR) methods have attracted much attention for hyperspectral images classification. In this article, two CR-based dynamic ensemble selection (DES) methods using multiview kernel collaborative subspace clustering (MVKCSC) and random subspace MVKCSC (RSMVKCSC) are proposed. In order to combine spectral and spatial information to construct a region of competence (RoC), the multiview learning strategy is used in the general DES method. Compared with traditional clustering methods, the MVC can more effectively utilize multi-feature information. Moreover, a new method of constructing the Laplacian matrix using kernel CR coefficients is proposed for clustering based on subspace clustering and CR theory. This method is called MVKCSC, which can obtain the clustering results by using kernel CR self-representation coefficients. In addition, to increase the diversity of samples, the random subspace method (RSM) and MVKCSC are combined for RSMVKCSC. Moreover, the algorithm can obtain better clustering results by constraining samples and features simultaneously. The effectiveness of the proposed methods is validated using three hyperspectral data sets with few samples. The experimental results show that both DES-MVKCSC and DES-RSMVKCSC outperform their single classifier counterparts. In particular, the proposed methods provide superior performance compared with the state-of-the-art DES methods.

Index Terms—Collaborative representation (CR) classifiers, dynamic ensemble learning, hyperspectral imagery, multiview clustering (MVC), subspace clustering.

I. INTRODUCTION

Hyperspectral image provides abundant spectral information in hundreds of contiguous spectral bands [1]. This property has led to hyperspectral images widely applied in many applications [2], such as environmental monitoring [3], agriculture [4], mineral exploration [5], and military [6]. Hyperspectral image classification is required in most of these applications. However, several factors such as high dimensionality, spectral redundancy, noise bands, and limited labeled samples make hyperspectral image classification challenging [7]–[9]. Classification accuracy can be improved by using more advanced classifiers [10]–[14]. For example, random forests [15], multilayer perceptron [16], collaborative representation (CR) classifier [17], and the more popular deep learning models [18], [19] in recent years have more applications in the fields of hyperspectral remote sensing, energy, and natural disasters. However, because a single classifier often cannot obtain an optimal classification result, some researchers have proposed to improve the classification accuracy of hyperspectral images based on ensemble learning methods [20]–[22]. Therefore, using ensemble learning to take advantage of multiple classifiers is promising in addressing the challenges in hyperspectral image classification.

Ensemble learning methods can be divided into two categories: static methods and dynamic methods. The static ensemble methods usually assume that each base classifier is independent of each other and has higher accuracy than random guessing. Then a certain strategy is adopted to combine multiple classifiers for higher classification accuracy. The most common static ensemble framework is boosting, which can promote weak classifiers to be powerful. The main idea of boosting is to assign a larger weight to previously misclassified samples when training a new classifier [23]. Specifically, a base classifier is first trained from an initial training set, and the training sample distribution is adjusted according to the classification accuracy. That is, a larger weight is given to the misclassified samples. Then a new base classifier is trained based on the adjusted sample distribution, and this step is repeated until the model converges [24], [25]. The most famous algorithm in boosting methods is AdaBoost, which has many applications in hyperspectral image classification. For example, from the perspective of base classifier construction, Kawaguchi et al. [26] proposed to use AdaBoost with stumps functions as base classifiers for hyperspectral image classification. Ramzi et al. [27] combined the SVM algorithm with the AdaBoost ensemble framework and proposed an AdaBoostSVM algorithm for hyperspectral image classification, and the results demonstrate that the proposed algorithm has higher classification accuracy than other classification models. Both of the above two algorithms show the effectiveness of AdaBoost in the field of hyperspectral image classification. However, AdaBoost is susceptible to abnormal samples, which may cause outlier samples to obtain higher weights in iterations and affect the model's accuracy [28]. Another simple and effective static ensemble
algorithm is Bagging [29]. The basic idea of bootstrap is to train the base classifier by obtaining different training subsets through the bootstrap sampling method. Aggregation refers to the fusion of the results from the base classifier through a certain method, such as the use of majority voting in the classifier task [30], [31]. In addition, Su et al. [32] proposed a new ensemble learning framework combining Bagging and tangent CR models for hyperspectral image classification, where the representation learning classifier is applied to ensemble learning for the first time. The result shows that the Bagging ensemble strategy is also applicable to the representation learning model. However, there are duplicate data between the training subset and the original training data. If the base classifier is not sensitive to the distribution of the training data, the result of the base classifier will be very similar, which makes the improvement of classification accuracy limited. Stacking is another widely adopted static ensemble learning algorithm based on the ideas of meta-learning [33]. Its main idea is to use the result of the initial classifier as a new training set to learn a new classifier [34]–[36]. This algorithm can also be regarded as a representation learning method that can conduct effective feature extraction from the original data through multiple base classifiers. Chen et al. [37] proposed to use the SVM as a base classifier combined with a stacking model for hyperspectral image classification and demonstrate that this method can effectively improve the classification accuracy. However, the stacking algorithm requires the base classifier to offer accuracy and diversity simultaneously, and it is not suitable for the condition of a small training size. To summarize, the above static ensemble learning methods are effective in hyperspectral image classification. However, a specific weak classifier may produce better classification accuracy for a particular local area of unknown samples than a strong classifier and static ensemble methods often discard such classifiers directly. Therefore, hyperspectral image classification based on static ensemble learning is difficult to fully use the local classification advantages of each base classifier in unknown samples.

Unlike the static ensemble learning algorithms, the dynamic ensemble selection (DES) methods divide unknown data into different regions and select optimal classifiers for each region [38]. Ho et al. [39] first proposed a dynamic classifier selection method, and the main idea is to divide the training set into some regions. The base classifier is evaluated on each region and the best-performing classifier is determined in each region. Then the testing set is classified into different regions through a specific method. Finally, the corresponding optimal classifier is used to classify the unknown samples. The dynamic selection framework has three main components: classifier pool generation, dynamic selection, and aggregation. When dynamic selection chooses multiple classifiers instead of a single classifier, it is called DES. The most significant difference between dynamic and static ensembles is that the dynamic ensemble has a dynamic selection step. Dynamic selection includes two steps: defining the region of competence (RoC) and determining the selection criteria [40]. Most of the existing studies are improving and innovating in these two parts. For example, some work proposed to use the K-NN technique [41], clustering methods [42], the proposed a dynamic linear classifier for hyperspectral image classification. Damodaran et al. [48] proposed a dynamic linear classifier for hyperspectral image classification. The main idea of this method is to combine a dimensionality reduction process with the dynamic selection method to construct a new dynamic classifier selection framework. Damodaran and Nidamanuri [49] also introduced random subspace method (RSM), Markov random field (MRF), and extreme learning machine (ELM) into dynamic integration options to construct a new hyperspectral image classification framework. The experimental results show that compared with the traditional multiclassifier selection system and SVM method, the two proposed methods can offer higher classification accuracy. However, the base classifiers used in the above methods are all traditional machine learning classifiers, which do not have advantages when using a small training set. Therefore, these methods cannot solve limited label samples in hyperspectral classification. Meanwhile, in the construction of RoC, the spatial information of hyperspectral data is not fully used.

It can be seen that although the classification method based on static ensemble makes full use of the diversity of classifiers, the fusion strategy does not take into account the local classification advantages of different classifiers for unknown samples. The method based on dynamic ensemble adopts the concept of the RoC. First, it divides the testing data into multiple regions and then finds a locally optimal classifier set for each region. This method can make better use of the advantages of multiple classifiers. However, the traditional method of defining RoC has some shortcomings. First, the construction of dynamic ensemble only uses spectral features without considering spatial information of the hyperspectral data. Meanwhile, traditional K-NN methods for dividing RoC are often not accurate enough, and clustering methods are rarely used for DES. At last, the redundancy and noise of hyperspectral bands also cause the formed RoC to be inaccurate. Multiview [50]–[53] and subspace clustering [54]–[57] can solve the above problems well. Multiview clustering (MVC) considers spectral information and other spatial information simultaneously to obtain complementary clustering results. Subspace clustering can effectively solve the problem of data redundancy. Hyperspectral image classification based on a representation model [58]–[61] is simple, computationally efficient, and has good classification results under a few sample conditions, but it is rarely used for ensemble learning. How to generate a representation model classifier pool suitable for a few samples and to define a more reliable RoC combining with spatial information should be addressed based on the dynamic ensemble of hyperspectral image classification.

Therefore, this article first proposes a dynamic ensemble method based on MVC [62]. This method uses MVC instead of K-means clustering in the traditional DES. Then a new clustering method called multiview kernel collaborative subspace clustering (MVKSC) is proposed to define RoC for DES. MVKSC first proposed to use the kernel collaborative method to construct a similarity matrix from the representation coefficients, and then
use it for spectral clustering. It is a new MVC subspace clustering method. Finally, random subspace [63]–[66] is added to the proposed algorithm to construct a new ensemble framework combining static ensemble and dynamic ensemble. Because CR-based classifier has higher classification accuracy in the case of small samples, the paper considers using limited samples for experiments. The proposed algorithms utilize spatial and spectral information of hyperspectral data to divide the RoC required for DES, assign different sets of optimal classifiers to each RoC. The characteristic of this method is that it utilizes both spatial and spectral information to construct an RoC of dynamic ensemble combined with spatial-spectral information, which is more reliable than the traditional RoC construction method. The major contributions are summarized as follows.

1) This is the first time that MVC is introduced into DES. Unlike the traditional DES method of constructing RoC based on a single feature, the MVC method can divide the image into homogeneous regions by using the consistency between different views (features).

2) Considering the collaborative relationship among the classes of hyperspectral images, the self-representation of the samples is used to learn the collaborative embedding representation of each data point through the kernel collaborative constraint. Then the similarity is constructed with the self-cooperative representation coefficients, and spectral clustering is adopted for clustering. This method double-constrained the data at the feature and sample level simultaneously, which can provide better clustering results.

3) Bootstrap sampling is proposed in the feature dimension because of the high dimensionality and band correlation of hyperspectral data. The resampled local features are regarded as different views and subspaces to improve the clustering accuracy of the MVKCSC algorithm.

The rest of this article is organized as follows. Section II introduces related work. Section III proposes MVKCSC and RSMVKCSC based dynamic CR ensemble classification algorithms. In Section IV, experiments and analysis with three real hyperspectral data are presented. Finally, Section V concludes this article.

II. RELATED WORK

A. CR-Based Classifiers Using in DES

Given the training set $X$ and the testing set $Y$, the linear combination of training samples $X$ can represent the approximation $\tilde{y}$ of a testing sample $y$ in CRC as [67]

$$y = X\alpha + e$$

where $\alpha$ stands for representation coefficients.

The coefficient vector $\alpha$ is the solution to the optimization problem with an $l_2$-norm regularization

$$\alpha = \arg \min_{\|\alpha\|_2^2 + \lambda } \left\{ \|y - X\alpha\|_2 \right\}$$

where $\lambda$ is the regularization parameter used to weight the residual term and the regularization term. According to formula (2), a sample $y$ can be classified to the best approximation in the subspace as

$$\text{class}(y) = \arg \min_{C} \left\{ \|y - X\alpha_C\|_2 \right\}$$

where $\alpha_C$ is $n \times 1$ vector represented as the weighted coefficient, $X$ is $n_c \times 1$ vector, and $\alpha_C$ is the corresponding coefficient.

The analytical solution to (2) is

$$\alpha = \left( X^T X + \lambda I \right)^{-1} X^T y.$$  

where $\Phi(\bullet)$ projects $X$ and $y$ to a high-dimensional space, respectively.

Similar to (4), the coefficient $\alpha$ can be calculated as

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

and, the label of sample $y$ is determined according to (2).

Apart from the kernel method, Cai et al. [70] proposed a CR classifier with probabilistic interpretation and called it a probabilistic CR classifier (ProCRC).

B. Dynamic Ensemble Selection

DES is an ensemble method that selects multiple base classifiers from a pool and integrates them for classification. This method assumes that each base classifier in the classifier pool is an expert in a different local area in the feature space. The process of dynamic integration selection is shown in Fig. 1. It contains the following three components.

1) Classifier Pool Construction: The objective of classifier pool generation is to obtain a set of classifiers, $C = \{c_1, c_2, ..., c_l\}$ with diversity and accuracy. The base classifiers in the classifier pool need to have sufficient differences. In this way, the prediction results of each classifier can be as different as possible.

2) Dynamic Ensemble Selection: The DES is mainly divided into the two following.

a) RoC division: The testing data can be divided into different regions, and each region can measure the local classification accuracy of the base classifier. An RoC construction method mainly includes the K-NN technique, clustering method, the decisions spaces, or the use of a potential function.

b) Classifier selection index determination: It is to select an index to measure the classification ability of each classifier in a local region.

3) Aggregation: The aggregation step mainly uses some fusion rules to combine the results from the selected classifiers to obtain the final result. This step includes three strategies, namely, nontrainable, trainable, and dynamic weighting.
C. Multiview Clustering

MVC is a clustering method that searches for the same and different topics in multiple views, grouping them into a group, and searching for consistency among different views. Because of the low computational cost of the k-means method and considering the large volume of hyperspectral data, we use the k-means-based method for MVC.

The nonnegative matrix factorization-based formulation for the k-means clustering is

$$
\min_{U, V} \|X^T - UV^T\|_F^2
\text{s.t. } U_{i,k} \in \{0, 1\}, \sum_{k=1}^{K} U_{i,k} = 1, \forall i = 1, 2, \ldots, N
$$

where the columns of $V$ give the cluster centroids. On this basis, for a large multiview data, a new MVC clustering method based on k-means was proposed [50]. In the proposed method, the $\ell_{21}$ norm is used to replace the Frobenius norm. The new formulation obtained from (1) is

$$
\min_{V^{(v)}, \alpha^{(v)}, U} \sum_{v=1}^{m} \alpha^{(v)} \gamma \|X^{(v)T} - U V^{(v)T}\|_{2,1}
\text{s.t. } U_{i,k} \in \{0, 1\}, \sum_{k=1}^{K} U_{i,k} = 1, \sum_{v=1}^{m} \alpha^{(v)} = 1, \forall i = 1, 2, \ldots, N
$$

where $\alpha^{(v)}$ is the weight of the $v$th view, and the parameter $\gamma$ controls the weight distribution. Furthermore, for high-dimensional problems, Xu et al. [50] proposed a new k-means-based MVC method. The proposed method introduces a projection matrix for each view and then performs MVC by forcing the common index matrix. It can be formulated as

$$
\min_{V^{(v)}, \alpha^{(v)}, U} \sum_{v=1}^{m} \alpha^{(v)} \gamma \|X^{(v)T} - U V^{(v)T}\|_F
\text{s.t. } W^{(v)T} W^{(v)} = I, U_{i,k} \in \{0, 1\}, \sum_{k=1}^{K} U_{i,k} = 1
$$

where $W^{(v)}$ is the projection matrix.

D. Subspace Clustering Based on Spectral Clustering

The subspace clustering algorithm based on spectral clustering is one of the subspace clustering algorithms. It intends to use the dictionary data in the space of each type of data as much as possible to represent that type of data. The general form of the algorithm is

$$
\min \lambda f(e) + \|\alpha\|_p
\text{s.t. } X = X\alpha + e
$$

where $\lambda$ is the regularization parameter, $e$ is the noise, and $\alpha$ is the self-representation coefficient. The core of the algorithm is to obtain the self-representation coefficient through a certain method, such as using sparse representation, low-rank representation, and so on. According to $\alpha$, the similarity matrix can be constructed as

$$
S = |\tilde{\alpha}| + |\tilde{\alpha}|^T/2
$$

Finally, the Laplacian matrix is calculated, as

$$
\min_F Tr \left( F^T (D - S) F \right) \text{ s.t. } F^T F = I
$$

where $F$ is the cluster indicator matrix.

III. PROPOSED METHODS

The details of the proposed DES-MVKCSC and DES-RSMVKCSC algorithms are mainly described from four aspects: classifier pool construction, RoC division, classifier selection, and classifier fusion. The structure of DES-MVKCSC and DES-RSMVKCSC is shown in Figs. 1 and 2, respectively.

A. Classifier Pool Generation Based on CR

In this section, a dynamic CR classifiers ensemble learning framework based on the MVC method is proposed. The key of the proposed algorithm is mainly to use spectral features and EMP features as different views through MVC to construct the RoC. The RoC first learns from the validation set, that is, uses the clustering method to divide the validation set into $K$ regions...
and then assigns the unknown samples in the testing set to these $K$ regions.

First, construct a CR-based classifier pool, and the classifiers in the classifier pool need to be diverse. In this article, the classifier pool is generated by different classifiers and parameters. 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 the differences of the parameters. Let the resulting CR classifier pool be denoted as $C = \{c_1, c_2, \ldots, c_i\}$.

\section*{B. RoC Based on MVKCSC (DES-MVKCSC)}

Single-view subspace clustering uses the self-representation of data. Each point can be represented by a linear combination of other points in the data set, that is, $\tilde{X} = X\alpha + e$. The proposed method is to consider the self-representation of different views, which can be expressed as

$$\tilde{X}^{(v)} = X^{(v)}\alpha^{(v)} + e^{(v)}$$

(13)

where $\alpha^{(v)}$ is the self-representation coefficient of each view. Unlike the traditional method, this article innovatively proposed using kernel CR to obtain self-representation coefficients.

The coefficient vector $\alpha^{(v)}$ is the solution to the optimization problem with an $l2$-norm regularization

$$\alpha^{(v)} = \operatorname{arg\,min}_{\alpha^{(v)}} \left\{ \| X^{(v)} - X^{(v)}\alpha^{(v)} \|^2_2 + \lambda \|\alpha^{(v)}\|^2_2 \right\}.$$ 

(14)

Use the average value of the self-representation coefficient to bring the self-representation coefficient representing the multi-view, which can be expressed as

$$\bar{\alpha}^{(1)} = \frac{1}{v} \sum_{v=1}^{v} \alpha^{(v)}.$$ 

(15)

Then we calculate the mean of the representation coefficients and the similarity matrix based on formulas (11), (12).

$$S^{(1)} = \frac{\bar{\alpha}^{(1)} + \bar{\alpha}^{(1)\top}}{2}$$ 

(16)

Algorithm 1: DES-MVKCSC.

\begin{algorithm}
\begin{footnotesize}
Input: The pool of CR-based classifiers $C$; the spectral feature dataset $Va$ and $Te$; the EMP feature dataset $emp-Va$ and $emp-Te$; the cluster size $K$; the empty ensemble of classifiers $EoC\_t$.

\For{each validation sample $v$ in $Va$ and $emp-Va$}{
Calculating the self-representation coefficients of the spectral feature $\alpha^s$ and the emp feature $\alpha^e$ separately according to (14)
}

\For{each testing sample $t$ in $Te$ and $emp-Te$}{
Similar to $Va$, and use MVKCSC to divide the testing data into $K$ capability regions $RoC_{te}$

\For{each classifier $c_i$ in $C$}{
if ($c_i$ correctly recognizes all samples in $RoC_{val}$) then $EoC_{\_t} = EoC_{\_t} \cup c_i$
end if
}

end for

end for

Calculating the Laplacian matrix and obtain indicator vector $F$ according to (18)

end for

Taking the $F$ as input, and use k-means to divide the validation data into $K$ capability regions $RoC_{val}$

\For{each classifier $c_i$ in $C$}{
if ($c_i$ correctly recognizes all samples in $RoC_{val}$) then $EoC_{\_t} = EoC_{\_t} \cup c_i$
end if
}

end for

end for

Output: $EoC\_t$, an ensemble of classifiers for each unknown sample $t$ in $Te$;
\end{footnotesize}
\end{algorithm}
Algorithm 2: DES-RSMVKCSC.

Input: The pool of CR-based classifiers $C$; the spectral feature dataset $Va$ and $Te$; the EMP feature dataset $emp-Va$ and $emp-Te$; the spectral view size $s_n$; the EMP view size $e_n$; the cluster size $K$; the empty ensemble of classifiers $EoC^*_t$.

for each validation sample $v$ in $Va$ and $emp-Va$, do

Get the set of sub-validation sets $Sub-Va$ via RSM according to (21)

for each testing sample $t$ in $Te$, do

Get the set of sub-testing sets $Sub-Te$ via RSM according to (21)

for each sub-validation sample $sub-v$ in $Sub-Va$, do

Find $RoC_{val}$ as the $K$-RoC via MVKCSC according to (22)–(25)

for each sub-testing sample $sub-t$ in $Sub-Te$, do

Find $RoC_{te}$ as the $K$-RoC via MVKCSC according to (22)–(25)

for each classifier $c_i$ in $C$, do

if ($c_i$ correctly recognizes all samples in $RoC_{val}$) then

$EoC^*_t = EoC^*_t \cup c_i$

end if

if ($c_i$ cannot correctly recognizes all samples in $RoC_{val}$) then

$EoC^*_t = EoC^*_t \cup$ the classifier with the highest prediction accuracy

end if

Use the ensemble $EoC^*_t$ to classify $RoC_{te}$

end for

Output: $EoC^*_t$, an ensemble of classifiers for each unknown sample $t$ in $Te$;

\[
\min_F \text{Tr} \left( F^T \left( D - S^{(1)} \right) F \right) \quad \text{s.t.} \quad F^T F = I. \quad (17)
\]

Taking the feature vector $F$ as input, and using k-means to divide the original data into $K$ capability regions, result in:

$RoC_i = \text{kmeans}(F, k), \quad i = 1, \ldots, K. \quad (18)$

Finally, sample splitting is performed in each cluster to obtain training set, validation set, and testing set in different regions.

C. RoC Based on Random Subspace MVKCSC

(DES-RSMVKCSC)

The RSM is proposed to be added to the MVKCSC algorithm, thereby reducing the dimensionality of features and increasing the diversity of classification results. The method is used in feature dimensions and generates feature subspaces

\[
X_s^{(v)} = \text{rsm}(X^{(v)}). \quad (19)
\]

With the sample subspace as input, other steps are the same as MVKCSC.

According to formula (16) and (20), the kernel CR coefficients of each view are calculated separately, as

\[
\alpha_s = \arg \min_{\alpha^{(v)}} \left\{ \left\| \Phi(X_s^{(v)}) - \Phi(X_{s_n^{(v)}}) \alpha_s^{(v)} \right\|^2_2 + \lambda \left\| \alpha_s^{(v)} \right\|^2_2 \right\}. \quad (20)
\]

According to formula (15), the average value of the self-representation coefficient is calculated as

\[
\bar{\alpha}^{(2)} = \frac{1}{\nu} \sum_{i=1}^{\nu} \alpha_s. \quad (21)
\]

Then we calculate the mean of the representation coefficients of different views for the similarity matrix

\[
S_{(2)} = \frac{1}{\nu} |\bar{\alpha}^{(2)}| + |\bar{\alpha}^{(2)}|^{T} \quad (22)
\]

\[
\min_F \text{Tr} \left( F^T \left( D - S^{(2)} \right) F \right) \quad \text{s.t.} \quad F^T F = I. \quad (23)
\]

Taking the feature vector $F$ as input, and using k-means to divide the original data into $K$ capability regions, result in

$RoC_i = \text{kmeans}(F, k), \quad i = 1, \ldots, K. \quad (24)$

Finally, sample splitting is performed in each cluster to obtain training set, validation set, and testing set in different regions.

D. Dynamic Selection and Aggregation

For the $K$ regions of the validation set, each base classifier is used to classify, and the classification accuracy of each region is obtained. If all the base classifiers in the pool with the highest classification accuracy are selected as the final classifier set.

Finally, different base classifier combinations are dynamically assigned to the $K$ regions of each validation set. Meanwhile, these different classifier combinations are correspondingly used to classify unknown pixels in the $K$ area of the testing set.

In the final aggregation part, for each $K$ testing set region, use the classifier set selected in (3) to make predictions. The final result is obtained via majority voting.

IV. EXPERIMENTS

A. Experiment Setup

All experiments are implemented on the platform of Python 3.6.13. In order to ensure fairness, all classifiers use multifeatures (spectral feature and EMP feature) as training sets.

1) Pool of Classifiers: To evaluate the proposed dynamic heterogeneous representation ensemble models, three CR classifiers from different types in classifier pool are applied, i.e., CRC, KCRC, and ProCRC. The range of regularization parameters in three classifiers are all set as $\{1e-1, 1e-2, 1e-3\}$. The regularization parameters $\gamma$ in ProCRC is set as $\{1e-1, 1e-2, 1e-3\}$. The classifier pool finally contains 15 base classifiers.

2) Parameter Settings: The range of clusters numbers $c_n$ for DES-MVKCSC and DES-RSMVKCSC are set $\{2, 3, 4, 5\}$. The Spectral View Numbers $s_n$ and EMP View Numbers $e_n$ of DES-RSMVKCSC in both sets are $\{5, 10, 15, 20, 25\}$. In order...
to verify that the proposed algorithms have better classification accuracy in the case of small samples, this article compares each group classification accuracy of experiments with small samples. The detailed parameters information is shown in Table V.

3) Comparison Algorithms: To evaluate the performance of the proposed DES models, a variety of classification algorithms are used for comparison. For example, support vector machine (SVM) [71] and extremely randomized trees [72] are the baseline. In order to fairly measure the classification performance of multiple algorithms, all algorithms classify three hyperspectral images with multifeatures. Moreover, the newest DES algorithm named DES-MI [40] is added for comparison. It is the state-of-the-art DES algorithm proposed in recent years. At last, some state-of-the-art ensemble learning models are also conducted for comparison, including XgBoost [73] and CatBoost [74].

B. Hyperspectral Data Sets

The Indian Pines data set contains 224 spectral bands and is collected by the AVIRIS sensor located in northwest Indiana. The wavelengths of spectral bands range from 0.4 to 2.5 um. After removing water absorption bands, there are 200 effective bands remaining in the data. The image includes 16 classes, the pixels are $145 \times 145$, and the spatial resolution is 20 m. The detailed information of each class is described in Table I, and images of this data set are shown in Fig. 3.

The second data set is the University of Pavia data set, collected by the ROSIS sensor. This data set contains 103 spectral bands with wavelengths ranging from 0.43 to 0.86 um. The spatial size of the University of Pavia is 610 $\times$ 340. The details of nine classes in this HSI are described in Table II, and the images are shown in Fig. 4.

The third image used in this article is the KSC data set, acquired by the AVIRIS sensor. This image contains 224 spectral bands with wavelengths ranging from 0.4 to 2.5 um. After removing the absorbent band, the scene consists of 13 classes and
TABLE IV
POOL OF CLASSIFIERS AND PARAMETERS FOR SUBCLASSIFIER

| Sub-classifier | $\gamma$ | $\lambda$ |
|---------------|----------|----------|
| CRC           | (1e-1, 1e-2, 1e-3) | - |
| KCRC          | (1e-1, 1e-2, 1e-3) | - |
| ProCRC        | (1e-1, 1e-2, 1e-3) | (1e-1, 1e-2, 1e-3) |

TABLE V
PARAMETERS SETTING

| Algorithms     | Parameters          |
|----------------|---------------------|
| DES-MVKCSC     | $c_n$ (2, 3, 4, 5)  |
| DES-RSMVKCSC   | $c_n$ (2, 3, 4, 5)  |
| $s_n$          | (5, 10, 15, 20, 25) |
| $e_n$          | (5, 10, 15, 20, 25) |

TABLE VI
OPTIMAL PARAMETERS

| Algorithms     | Data set          | Parameters |
|----------------|-------------------|
| DES-MVKCSC     | Indian Pines      | $c_n = 5$  |
| University of Pavia | $c_n = 2$  |
| Kennedy Space Center | $c_n = 5$  |
| DES-RSMVKCSC   | Indian Pines      | $c_n = 2$, $s_n = 5$, $e_n = 25$ |
| University of Pavia | $c_n = 2$, $s_n = 25$, $e_n = 15$ |
| Kennedy Space Center | $c_n = 2$, $s_n = 5$, $e_n = 5$ |

C. Classification Performance

In the first experiment, the overall accuracy (OA), average accuracy (AA), per-class accuracy, kappa statistic, F1-score, and running time (s) calculated from experiments are shown in Table VI. The classification maps obtained by various models are shown in Fig. 7(a)–(j). Similar to the AVIRIS dataset, the classification accuracy of DES-MVKCSC and DES-RSMVKCSC methods are higher than other classifiers used for comparison. The accuracy of the proposed DES-RSMVKCSC algorithm can reach 96.83% in the case of small samples, which is the highest accuracy among all classifiers. Compared with the optimal CR-based methods, DES-RSMVKCSC methods yield 2% and 3% improvements. Notably, DES-RSMVKCSC produces the best AA for the ROSIS data set. The trend of f1-score is the same as OA for all the classifiers. The runtime of the two proposed algorithms is also higher than these of other comparison models, but not more than 1 min.

In the second experiment, the University of Pavia data set was used to verify the classification performance of DES-MVKCSC and DES-RSMVKCSC. The best parameters are shown in Table VI. The classification maps obtained by various models are shown in Fig. 7(a)–(j). Similar to the AVIRIS dataset, the classification accuracy of DES-MVKCSC and DES-RSMVKCSC methods are higher than other classifiers used for comparison. The accuracy of the proposed DES-RSMVKCSC algorithm can reach 96.83% in the case of small samples, which is the highest accuracy among all classifiers. Compared with the optimal CR-based methods, DES-RSMVKCSC methods yield 2% and 3% improvements. Notably, DES-RSMVKCSC produces the best AA for the ROSIS data set. The trend of f1-score is the same as OA for all the classifiers. The runtime of the two proposed algorithms is also higher than these of other comparison models, but not more than 1 min.

In the third experiment, the detailed results of the proposed algorithms and other methods for the KSC data set are listed in Table IX. The best parameters are listed in Table VI. The classification maps obtained are shown in Fig. 8(a)–(j). Similar to AVIRIS and ROSIS images, the accuracy of our methods can also better reach about 91.38% and 92.07%. Compared with other dynamic ensemble classifiers such as DES-MI, DES-Cluster, and the state-of-the-art Meta-DES, our methods yield nearly 13% and 13% improvements. For f1-score, the DES-RSMVKCSC has the highest value. Similar to the first dataset, MVC-RSMVKCSC takes the longest time, which costs 42.67 s.

D. Parameter Analysis

1) Parameters Analysis for DES-MVC and DES-MVKCSC:

For the DES-MVC and DES-MVKCSC algorithm, the parameter $c_n$ significantly impact the algorithm performance. The experiment set 2, 3, 4, and 5 clusters to study their influence on classification accuracy. In this experiment, when $c_n$ change in range, the accuracy of DES-MVC for three datasets are shown in Fig. 9(a)–(c).

For the Indian Pines dataset, the classification results of DES-MVC and DES-MVKCSC algorithm, the parameter $c_n$ significantly impact the algorithm performance. The experiment set 2, 3, 4, and 5 clusters to study their influence on classification accuracy. In this experiment, when $c_n$ change in range, the accuracy of DES-MVC for three datasets are shown in Fig. 9(a)–(c).

For the Indian Pines dataset, the classification results of DES-MVC and DES-MVKCSC algorithm, the parameter $c_n$ significantly impact the algorithm performance. The experiment set 2, 3, 4, and 5 clusters to study their influence on classification accuracy. In this experiment, when $c_n$ change in range, the accuracy of DES-MVC for three datasets are shown in Fig. 9(a)–(c).

For the Indian Pines dataset, the classification results of DES-MVC and DES-MVKCSC algorithm, the parameter $c_n$ significantly impact the algorithm performance. The experiment set 2, 3, 4, and 5 clusters to study their influence on classification accuracy. In this experiment, when $c_n$ change in range, the accuracy of DES-MVC for three datasets are shown in Fig. 9(a)–(c).
the rise of $c_n$. Meanwhile, when $c_n$ is 2, the DES-MVKCSC algorithm can obtain the highest accuracy, which is 96.60%.

Similar to the Indian Pines dataset, the parameter $c_n$ has a more significant impact on the classification accuracy for DES-MVKCSC.

Overall, the experimental results show that the parameter $c_n$ has no obvious effect on the classification performance of the DES-MVKCSC algorithm, which indicates that 1) the model is not sensitive to the parameters; 2) the proposed methods have better classifier results without adjusting the parameters.

2) Parameters Analysis for DES-RSMVKCSC: Since the proposed DES-RSMVKCSC algorithm is susceptible to different views and clusters, the impact of spectral view numbers ($s_n$), EMP view numbers ($e_n$), and cluster numbers on the classification performance are investigated. The experiment studies the influence of the third parameter on the OA of classification by fixing two of the three parameters for all three datasets. Finally, control the parameter $c_n$, analyze the combined effect of the changes of the parameters $s_n$ and $e_n$ on the classification accuracy.
As shown in Fig. 10, the OA of the classifier gradually decreases as the parameter $s_n$ increases. It can be seen from Fig. 11 that the accuracy increases as the parameter $e_n$ increases. The parameter $c_n$ impacts the accuracy of classification, and the overall trend of classification accuracy is sensitive. Fig. 12(a)–(d) describes the common effect of parameters $s_n$ and $e_n$ on classification accuracy. It can be seen from Fig. 12 that no matter how the parameter $c_n$ changes, when the parameter $e_n$ is greater than $s_n$, the classification accuracy will gradually increase. This means the more extensive the EMP feature weight, the higher the classification accuracy.

Figs. 13(a)–(e) and 14(a)–(e) show the overall classification accuracy tendency with different $s_n$ and different $e_n$ in the ROSIS University of Pavia data set, respectively. Similar to AVIRIS Indian Pines data set, it can be seen from Figs. 13 and 14 that the parameter $s_n$ and $c_n$ have great effect on the overall trend of classification accuracy. Fig. 15(a)–(d) describes the combined effect of the parameters $s_n$ and $e_n$ on the University
Fig. 8. Classification maps resulting from the classification for the KSC data set using 15 labeled samples per class. (a) SVM. (b) ET. (c) CRC. (d) XgBoost. (e) CatBoost. (f) DES-MI. (g) DES-Cluster. (h) Meta-DES. (i) DES-MVKCSC. (j) DES-RSMVKCSC.

Fig. 9. Classification accuracy of three data sets versus different parameters of DES-MVKCSC. (a) Indian Pines data set. (b) University of Pavia data set. (c) KSC data set.

Fig. 10. Classification accuracy of Indian Pines data set versus varying $s_n$.

Fig. 11. Classification accuracy of Indian Pines data set versus varying $e_n$.

of Pavia data set. When $c_n$ is 2 and 3, there is no apparent trend of the accuracy change trend. But when $c_n$ is 4 and 5, the classification accuracy of the algorithm changes similarly. Figs. 16(a)–(e) and 17(a)–(e) show the overall classification accuracy tendency with different $s_n$ and different $e_n$ in the KSC data set, respectively. Finally, it can be seen from Figs. 16 and 17 that the parameters $s_n$ and $e_n$ have a negligible impact on the classification accuracy. Similar to the AVIRIS Indian Pines data set, the parameter $c_n$ has a more significant impact on the tendency of classification accuracy. Fig. 18(a)–(d) shows the impact of the three parameters on the classification accuracy of the KSC data set. It can be seen that when the values of the
Fig. 12. Classification accuracy of Indian Pines data set versus multiparameters of DES-RSMVKCSC. (a) Number of cluster center is 2. (b) Number of cluster center is 3. (c) Number of cluster center is 4. (d) Number of cluster center is 5.

Fig. 13. Classification accuracy of the University of Pavia Data set versus varying $s_n$.

Fig. 14. Classification accuracy of the University of Pavia Data set versus varying $e_n$.

Fig. 15. Classification accuracy of the University of Pavia data set versus multiparameters of DES-RSMVKCSC. (a) Number of cluster center is 2. (b) Number of cluster center is 3. (c) Number of cluster center is 4. (d) Number of cluster center is 5.

Fig. 16. Classification accuracy of KSC data set versus varying $s_n$. 
parameters $s_n$ and $e_n$ are the same, the OA changes less. The sensitivity to the two parameters is minor.

In summary, the third algorithm proposed in this article has two more parameters than the first two methods. However, the parameters $s_n$ and $e_n$ have certain regularity for the classification accuracy of different data. Therefore, in practical application, the parameters can be adjusted according to the characteristics of different data.

V. Conclusion

In this article, a new dynamic ensemble framework based on CRC using MVC and subspace clustering to construct an ROC is proposed. The MVC method combines spectral features and EMP features to define the ROC for DES, dividing regions more reliably for improved classification accuracy. Moreover, DES-MVKCSC uses kernel CR coefficients to build a Laplacian matrix, performs subspace clustering to construct ROC, and proposes a more reliable new clustering algorithm. DES-RSMVKCSC model introduces an RSM based on the proposed MVKCSC algorithm, which improves the diversity of classifiers and the accuracy of the ROC. The experiments show that the proposed algorithms provide competitive classification performance compared to the classic classifiers and the state-of-the-art DES classifiers. Moreover, compared with others, the two proposed methods offer higher accuracy with limited training samples. In future research, we will design a dynamic ensemble method specifically for CR in hyperspectral image classification.

Acknowledgment

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

References

[1] D. Landgrebe, “Hyperspectral image data analysis,” IEEE Signal Process. Mag., vol. 19, 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., vol. 99, no. 7, pp. 2–28, Aug. 2021.
[3] 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,” ISPRS J. Photogram. Remote Sens., vol. 169, pp. 195–211, Dec. 2020.
[4] H. Su, Z. Wu, A. Zhu, and Q. Du, “Low rank and collaborative representation for hyperspectral anomaly detection via robust dictionary construction,” ISPRS J. Photogram. Remote Sens., vol. 169, pp. 195–211, Dec. 2020.
[5] 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 twoband vegetation indices (HVI) for biophysical characterization and discrimination of crop types using field reflectance and hypersio/EO-1 data,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 6, no. 2, pp. 427–439, Apr. 2013.
[6] E. Saraloglou, E. T. Gormus, and O. Gungor, “Mineral exploration with hyperspectral image fusion,” in Proc. 24th Signal Process. Commun. Appl. Conf., May 2016, pp. 1281–1284.
[7] D. Lorts, “Real-time embedded hyperspectral image compression for tactical military platforms,” in Proc. Appl. Imagery Pattern Recognit. Workshop, Oct. 2002, p. 140.
[8] L. O. Jimenez and D. A. Landgrebe, “Supervised classification in high-dimensional space: Geometrical, statistical, and asymptotical properties of multivariate data,” IEEE Trans. Syst. Man Cybern. Part C Appl. Rev., vol. 28, no. 4, pp. 39–54, Feb. 1998.
[9] Z. Wu, W. Zhu, J. Chanussot, Y. Xu, and S. Osher, “Hyperspectral anomaly detection via global and local joint modeling of background,” IEEE Trans. Signal Process., vol. 67, no. 14, pp. 3858–3869, Jul. 2019, doi: 10.1109/TSP.2019.2922157.
[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, Sep. 2019.
[11] 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, doi: 10.1109/TGRS.2019.2961141.

[12] Y. Chen, X. Zhang, and M. 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, Mar. 2014, doi: 10.1109/JSTARS.2014.2307356.

[13] B. Tu, X. Zhang, X. Kang, J. Wang, and J. A. Benediktsson, “Spatial density peak clustering for hyperspectral image classification with noisy labels,” IEEE Trans. Geosci. Remote Sens., vol. 57, no. 7, pp. 5085–5097, Jul. 2019.

[14] Z. Wu et al., “Scheduling-guided automatic processing of massive hyperspectral image classification on cloud computing architectures,” IEEE Trans. Cybern., vol. 51, no. 7, pp. 3588–3601, Jul. 2021.

[15] 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.

[16] H. Wu, A. Lin, X. Xing, D. Song, and Y. Li, “Identifying core driving factors of urban land use change from global land cover products and POI data using the random forest method,” Int. J. Appl. Earth Observ. Geoinf., vol. 103, Aug. 2021, Art. no. 102475.

[17] Z. Wu, J. Sun, Y. Zhang, Z. Wei, and J. Chanussot, “Recent developments in parallel and distributed computing for remotely sensed big data processing,” Proc. IEEE, vol. 109, no. 8, pp. 1282–1305, Aug. 2021, doi: 10.1109/JPROC.2021.3087029.

[18] 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 GP-5 hyperspectral imagery,” ISPRS J. Photogramm. Remote Sens., vol. 171, pp. 235–252, Dec. 2020.

[19] T. Zhong et al., “A city-scale estimation of rooftop solar photovoltaic potential based on deep learning,” Appl. Energy, vol. 298, Jun. 2021, Art. no. 117132.

[20] T. Zhong et al., “Assessment of solar photovoltaic potentials on urban noise barriers using street-view imagery,” Renewable Energy, vol. 168, pp. 181–194, Dec. 2020.

[21] Y. Zhang, H. L. Yang, S. Prasad, E. Pasolli, J. Jung, and M. Crawford, “Ensemble multiple kernel active learning for classification of multisource remote sensing data,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 8, no. 2, pp. 845–858, Feb. 2015.

[22] A. Merentitsis, C. Debes, and R. Heremans, “Ensemble learning in hyperspectral image classification: Toward selecting a favorable bias-variance tradeoff,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 7, no. 4, pp. 1089–1102, Apr. 2014.

[23] R. Bao, J. Xia, M. D. Mura, P. Du, J. Chuanusot, and J. Ren, “Combining morphological attribute profiles via an ensemble method for hyperspectral image classification,” IEEE Geosci. Remote Sens. Lett., vol. 13, no. 3, pp. 359–363, May 2016.

[24] 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, Mar. 2021.

[25] S. Cao, H. Shen, S. Chen, and C. Li, “Boosting structure consistency for multispectral and multimodal image registration,” IEEE Trans. Image Process., vol. 29, pp. 5147–5162, Mar. 2020.

[26] D. Jin, M. Ji, L. Xu, G. Wu, L. Wang, and L. Fang, “Boosting single image super-resolution learnt from implicit multi-image prior,” IEEE Trans. Image Process., vol. 30, pp. 3240–3251, Feb. 2021.

[27] S. Kagawuchi and R. Nishii, “Hyperspectral image classification by bootstrap AdaBoost with random decision stumps,” IEEE Trans. Geosci. Remote Sens., vol. 45, no. 11, pp. 3845–3851, Nov. 2007.

[28] P. Ramzi, F. Samadzadegan, and P. Reinartz, “Classification of hyperspectral data using an AdaBoostSVM technique applied on band clusters,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 7, no. 6, pp. 2066–2079, Jun. 2014.

[29] Y. Freund and R. E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” J. Comput. Syst. Sci., vol. 55, no. 1, pp. 119–139,1, Aug. 1997.

[30] L. Breiman, “Bagging predictors,” Mach. Learn., vol. 24, no. 2, pp. 123–140, Jan. 1996.

[31] B. Fischer and J. M. Buhmann, “Bagging for path-based clustering,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 11, pp. 1411–1415, Nov. 2003.
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, respectively. 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, China. His main research interests include hyperspectral remote sensing dimensionality reduction, classification, and spectral unmixing.

Jun Hu (Student Member, IEEE) received the B.E. degree in geographic information science from China University of Mining and Technology, Xuzhou, 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 from 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. 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 and Advisory Committee.