Evolutionary Multitask Ensemble Learning Model for Hyperspectral Image Classification

Jiao Shi, *Member, IEEE*, Tao Shao, Xiaodong Liu, Xi Zhang, Zeping Zhang, and Yu Lei, *Member, IEEE*

***Abstract*—Recently, ensemble learning paradigm has shown great potential to achieve better prediction performance in the hyperspectral image classification. However, in the traditional methods, each classifier independently searches for the optimal spectral feature subspace in series and some important features are searched repeatedly, which leads to high computing redundancy and low effective utilization of features. In this article, an evolutionary multitask ensemble learning model (EMT_EL) for hyperspectral image classification is designed. First, the model formulates the spectral feature subspaces generation into a multitask optimization problem to concurrently search for optimal feature subspaces for multiple classifiers, which would be capable to select more informative and representative feature subspaces effectively. Second, seeking the optimal feature subspace for one base classifier can assist in the optima-seeking process for some other base classifiers via sharing the useful features, which can accelerate converge toward the direction of the optimal feature subspace, avoid trapping in local optimal subspace and improve searching capability. Third, randomization-enhanced genetic operators are designed for effective and reasonable feature selection, which can facilitate the exchange of information and improve the joint searching efficiency of the feature subspace. Eventually, the quality of generated spectral feature subspaces for each base classifier is improved and the feature sharing can parse HSI data by knowing which spectral features are important. Experimental results demonstrate that the proposed method can generate the appropriate feature subspace for each base classifier, thus it has outstanding classification performance on the different hyperspectral datasets.**

***Index Terms*—Ensemble learning, evolutionary multitasking, feature subspace, hyperspectral images.**

I. Introduction

REMOTE sensing (RS) images contain a wealth of land-cover information [1], therefore, they occupy an important position in earth observation. Hyperspectral imagery (HSI) is a branch of remote sensing images. HSI obtains the spatial properties of the target concurrently with spectral representations, dozens or even hundreds of narrow bands are concluded, which is the most distinctive feature of HSI. And finally three-dimensional data cube based on the fusion of 2-D spatial information and spectral information is formed. In recent years, HSI has become an important tool for resource exploration, environmental monitoring, and target recognition, to name a few.

More recently, HSI plays a vital role in addressing the issue of land-cover classification due to their multiple spectral bands, abundant information, and high-resolution [2]–[5]. Although HSI contains a great deal of continuous spectral bands, the high redundancy between adjacent bands and high-dimensional features have posed great challenges to the classification of HSI [6], such as difficult storing data, numerous transferring data, and complex calculating. In the HSI classification, it will be prone to the curse of dimensionality if the number of features exceeds special value [7], and the classification accuracy reduces rather than has significant improvement. A great deal of previous researches into HSI classification demonstrated that traditional classification methods, for instance, support vector machines (SVM) and K-Nearest Neighbor (KNN) [8]–[10], are likely to involve small data scale, low processing speed, and take no full advantages of HSI information. In particular, these methods are difficult to meet the demand of current classification efficiency.

Ensemble learning has been well studied and widely applied to the classification of hyperspectral data, because ensemble methods have the potential to increase classification accuracy and decrease the generalization error [11]–[15]. In order to further improve the classification performance of ensemble learning-based HSI classification methods, many extensions have been researched. Random feature subspace-based studies [16]–[18] optimize parameters of multiple classifiers or integrate some ascendant classifiers, which have presented more effective and robust performances instead of training a single classifier to classify hyperspectral images. However, few of these methods comprehensively consider the useful and redundant band information, which may lead to low robustness and poor classification performance. The ensemble learning-based feature selection methods with specific criteria are proposed to resolve these issues [19]–[21], in which they train base classifiers on selected feature subspaces derived from the original spectral space. Most of these methods transform the spectral feature subspace selection into an evolutionary optimization. Although they devote to maintaining significant information, remove redundant information and reduce computational burden

Manuscript received August 7, 2020; revised October 17, 2020; accepted November 3, 2020. Date of publication November 16, 2020; date of current version January 6, 2021. This work was supported in part by National Natural Science Foundation of China under Grant 62076204 and Grant 61806166, in part by the National Natural Science Foundation of Shaanxi Province under Grant 2018JQ6003 and Grant 2018JQ6030), in part by the China Postdoctoral Science Foundation under Grant 2017M613204 and Grant 2017M623246, in part by the Fundamental Research Funds for the Central Universities, and in part by the Seed Foundation of Innovation and Creation for Graduate Students in Northwestern Polytechnical University. *(Corresponding author: Jiao Shi.)*

The authors are with the School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: jiaoshi@nwpu.edu.cn; taoshao@mail.nwpu.edu.cn; lxd@mail.nwpu.edu.cn; zhang_xi@mail.nwpu.edu.cn; zhangzeping@mail.nwpu.edu.cn; leiy@nwpu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2020.3037353

[22]–[25], the relatedness among multiple feature selection processes has received little notice. Thus, some important features are underutilized, low searching effectiveness and local optimal are inevitable. Therefore, searching optimal spectral feature subspaces play a key role in a successful ensemble learning based HSI classification method.

In ensemble learning-based HSI classification mechanism, majority voting is a simple and rapid voting strategy to integrated multiple prediction results [26]. Nevertheless, the relative majority voting mechanism will randomly select a pseudo label in a case that the predictions given by each classifier are different from each other. Then a few isolated labels may appear in the final HSI classification maps [27], [28]. Therefore, some works have explored a strategy that identifies informative and representative samples from the abundant unlabeled data, then inducted unlabeled data into a very limited set of training samples. These methods offer possibility to reduce the number of noisy points and grow the training dataset in a systematic way [14], [29]–[31]. However, many of these methods are likely to have tedious process, complicated calculation, and relatively low computational efficiency.

On the basis of the foregoing, it is necessary to design a method which can take into full consideration of spectral information, search for multiple optimal spectral feature subspaces interactively and minimize the possibility of noisy map. In this article, an evolutionary multitask ensemble learning (EMT_EL) model for HSI classification is proposed to address aforementioned limitations. The main contributions of this article are listed as follows.

1) The feature selection problem is modeled as a multitask optimization (MTO) problem. Multiple tasks that search for the optimal spectral feature subspaces are handled in parallel. It is capable to select more informative and representative feature subspaces effectively, the classification performance is improved accordingly.

2) The significant spectral features sharing happens between each classification method to assist each other, which will accelerate converge toward the direction of the optimal feature subspace, avoid trapping in local optimal subspace and improve searching capability.

3) The randomization-enhanced genetic operators are designed in the EMT_EL model, which can avoid the invalid and repetitive spectral feature subspaces, facilitate the exchange of information, and improve the joint searching efficiency of the feature subspace.

4) The dynamic updated training set strategy is adopted in which pseudo labels are assigned to raw data by semisupervised label generation strategy and update the training set in terms of the classification performances of base classifiers. Thereby, it avoids the inconsistency and uncertainty of the predictions among base classifiers in HSI classification.

The reminder of the article is organized as follows. Section II introduces the related background. Section III concentrates mainly on the proposed classification optimization model for HSI. The various experiments and discussions on three datasets: Indian Pines, Pavia University, and Salinas are given in Section IV. Section V concludes the article.

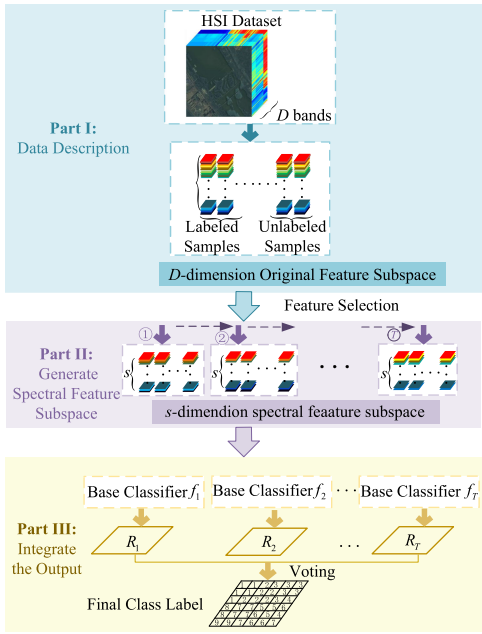


Fig. 1. Diagram of ensemble learning based HSI classification.

II. Background and Motivation

A. Problem Statement and Literature Review

One of traditional ensemble learning-based HSI classification frameworks is to manipulate different feature subspace for each classifier. In this type, spectral feature selection is an indispensable process of selecting feature subsets for diversifying the base classifiers [32]–[35]. The diagram of this framework is shown in Fig. 1, and the framework is described in detail as follows.

1) *Data Description:* As Fig. 1 Part I indicates, in the dataset $\boldsymbol{D}$, a small part of samples $\boldsymbol{D}_L$ with s dimendions ($s < D$) as training set $\boldsymbol{D}_{\text{train}}$ to build base classifiers, which is performed to label the abundant unlabeled samples $\boldsymbol{D}_U$.

2) *Generate Spectral Feature Subspace:* As shown in Fig. 1 Part II, multiple spectral feature subspaces in s-dimensions for classifiers are sequentially generated by feature selection [36]–[39]. Seeking the optimal spectral feature subspace for one base classifier may assist in the optima-seeking process for some other base classifiers, however, this relatedness is seldom noticed in these traditional methods. Thus, some important features cannot be shared among classifiers.

3) *Combine the Predictions:* As shown in Fig. 1 Part III, once the spectral feature subspaces are obtained, we train the classifiers on the subspaces and the predictions of base classifiers are combined by the relative-majority-voting scheme to generate the classification results $\boldsymbol{R}$ [40]. If more than one class label get the most number of votes, randomly select one. Thus, due to the uncertainty of class labels in multiclass HSI classification, there would appear a number of isolated class points whose class labels are different from those surrounding pixels, as shown in Fig. 2.

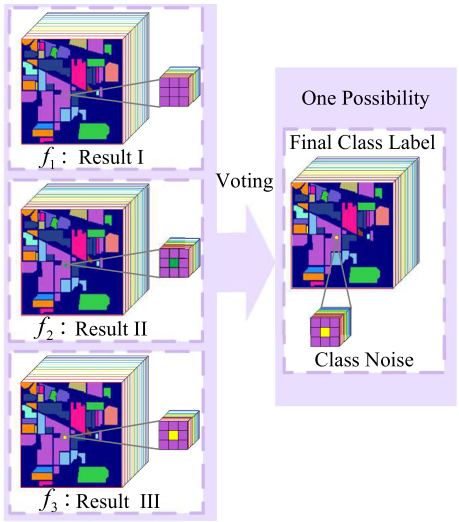


Fig. 2. Isolated class points in classification result.

B. Brief Introduction to Evolutionary Multitask Optimization

The evolutionary MTO algorithm focuses on the transmission of complex developmental traits to offspring through the interactions of genetic and cultural factor τ (also called skill factor) [41]–[45]. The cultural factor of an individual indicates the one task, amongst all other tasks in multifactorial optimization, on which the individual is most effective. And τ is viewed as a computational representation of an individual's cultural bias.

The MTO does not impose any strict constraint on the intertask relationship. Its efficiency is derived by implicit parallelism offered by a population in a way that transfers the positive information among tasks [41]. The index, $t = 1, 2, \ldots, T$, indicates the task index and T is the number of tasks. A MTO can be stated as the following:

$$\begin{aligned} &\text{minimize } \{\mathrm{F}_1(\boldsymbol{x}_1), \ldots, \mathrm{F}_\mathrm{t}(\boldsymbol{x}_\mathrm{t}), \ldots, \mathrm{F}_\mathrm{T}(\boldsymbol{x}_\mathrm{T})\} \\ &\boldsymbol{x}_t \in \boldsymbol{\Omega}_t \end{aligned} \tag{1}$$

where $\boldsymbol{x}_t$ is a feasible solution, $F_t(\cdot)$ is the fitness of individuals on tth task T_t and $\boldsymbol{\Omega}_t$ is the search space of T_t. The search space of each task is the solution space which contains representations of all possible feature subsets. The solution space of each task is mapped to a unified space by a coding strategy, and corresponding decoding operation would be carried out to obtain the final spectral feature subspaces.

C. Motivation of the Proposed Method

As shown in Fig. 3, the motivation for designing the multitask ensemble learning model for HSI classification is as follows.

1) *Search for Spectral Feature Subsets in Parallel:* Searching for the corresponding optimal spectral feature subspaces of multiple base classifiers in parallel can hardly be achieved in the traditional HSI classification methods based on ensemble learning [46], [47]. Therefore, the spectral feature subspaces generation problem can be formulated as an MTO problem [41], which can select concurrently different feature subsets from the original spectral feature space of HSI data according to different base classifiers. An evolutionary multitask ensemble learning model would make individuals evolved, which provides an efficient way to find optimal spectral feature subspaces. It will optimize multiple classifiers concurrently and improve the classification accuracy finally.
2) *Share the Significant Spectral Features:* When single task optimization algorithms are used in HSI classification based on ensemble learning, they individually seek the optimal spectral feature subspace for each base classifier. Thus some spectral features are searched repeatedly and the classification result of the classifiers on the feature subspaces is less than satisfactory [48], [49]. Therefore, evolutionary multitask ensemble learning algorithm is designed to take advantage of the relatedness between two or more searching procedure in HSI classification [50], [51]. The proposed algorithm encourages the spectral features which are significant to the interclass separability of HSI data to be shared between each searching procedure of each classifier, as noted by the red line in Fig. 3. Therefore, to some extent, the design of this algorithm in HSI classification will accelerate convergence to the optimal feature subspaces, and the classification accuracy would be improved. It can help us understand the HSI data by knowing which features are important.

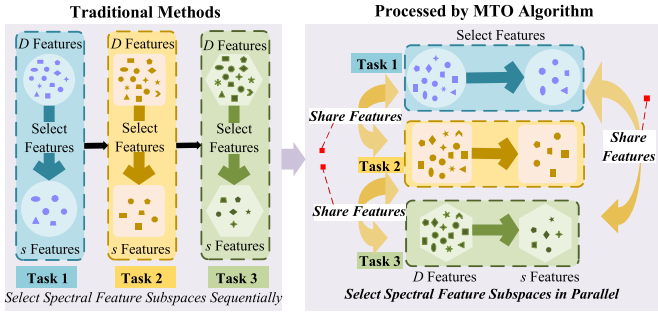


Fig. 3. Comparison of the proposed method and traditional approaches.

III. Proposed Approach

The proposed EMT_EL model for HSI classification, together with designed generators, population evaluation under the interaction of individuals and dataset, dynamic training set selection approach, is detailed in this section.

A. Proposed Overall Model

The proposed EMT_EL method can address T tasks that seek optimal spectral feature subspaces for T classifiers. The pseudocode of the proposed algorithm is shown in Algorithm 1, and the overall frame is illustrated in Fig. 4.

1) *Data Description:* First, the HSI dataset $\boldsymbol{D} \in \Re^{N*D}$ has N samples in D dimension, as presented in Part I of Fig. 4, The $\boldsymbol{D}$ can be represented as

$$\begin{aligned} &\boldsymbol{D} = (\boldsymbol{s}_i, y_i), i = 1, 2, \ldots, N \\ &\text{subject to } y_i \in 1, 2, \ldots, M \end{aligned} \tag{2}$$

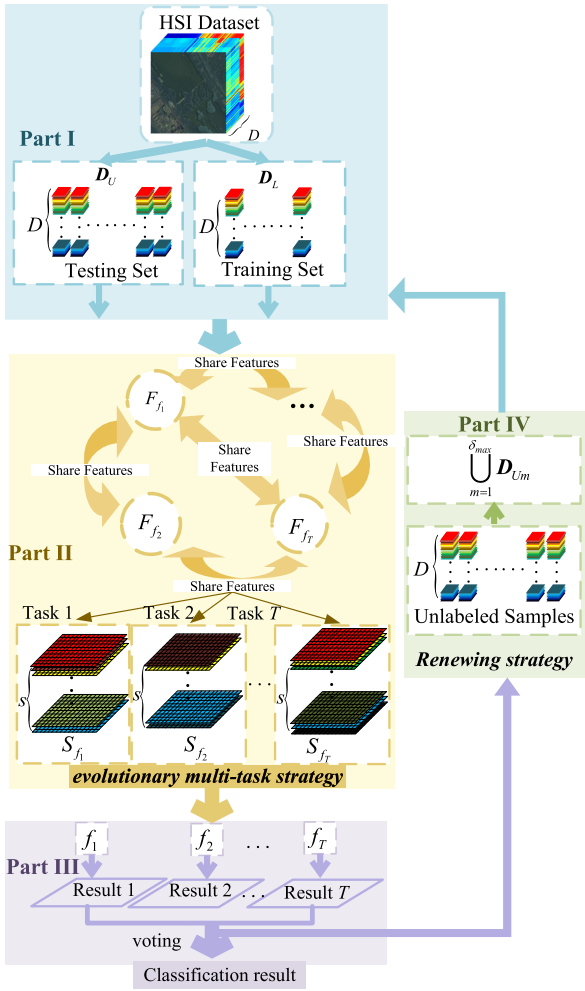


Fig. 4. Framework of the proposed EMT_EL method for HSI classification.

herein, M is the total number of land cover classes, $\boldsymbol{s}_i = [s_1^i, s_2^i, \ldots, s_D^i]$ represents a sample with D spectral features, the features are numbered as d, where $d = 1, 2, \ldots, D$. First, the hyperspectral image dataset are divided into two subsets: A small amount of samples are selected as training set $\boldsymbol{D}_{\text{train}}$, and the rest is used as testing set $\boldsymbol{D}_{\text{test}}$.

2) *Spectral Feature Subspaces Generation:* Second, as Fig. 4 Part II shows, the population of individuals is initialized and each individual represents the selected spectral feature subspace, then the feature subspaces are updated by nonrandom or assortative mating, mutation operation, and selective evaluation. In this process, some spectral features which are significant for the separability of HSI data are encouraged to be shared among T tasks, then T spectral feature subspaces are generated in parallel.

3) *Ensemble the Predictions:* Third, as shown in Fig. 4 Part III, T classifiers $\boldsymbol{f} = \{f_1, f_2, .., f_T\}$ are trained on these selected spectral feature subspaces $\boldsymbol{S}$ in s-dimension, then T predictions from T classifiers are obtained. Next, the classification results are obtained by relative-majority-voting scheme, and unlabeled samples also get pseudo labels. The final ensemble classifier EC for HSI classification are designed as follows:

$$EC(\boldsymbol{s}_{\text{train}}) = \arg\max_{y=\phi} \sum_{t=1}^{T} \boldsymbol{D}_{\text{train}}(f_t(\boldsymbol{s}_{\text{train}}) = y) \quad (3)$$

where $\phi = \{1, 2, \ldots, M\}$, $\boldsymbol{s}_{\text{train}}$ represents the samples in training set $\boldsymbol{D}_{\text{train}}$.

4) *Dynamic Selection of Training Set:* Fig. 4 Part IV illustrates the dynamic selection of training set strategy which is applied to construct the model. Unlabeled samples subset $\boldsymbol{D}_{U_m}$ with unanimous labeling results from T classifiers are selected and added to $\boldsymbol{D}_{\text{train}}$, then these samples are removed from $\boldsymbol{D}_{\text{test}}$.

Repeat step 2)–4) until the maximum number of iterations. Finally, The T classifiers will obtain better predictions and the classification results will generate by relative majority voting scheme.

Algorithm 1: Proposed EMT_EL method.

Input $\boldsymbol{D}$: HSI dataset, T: The number of classifiers, D: the number of spectral features, s: the number of selected features, N_p: Population size, rmp_1: Random mating probability.
Output Classification result.
Step 1) Set related parameters.
Step 2) Select part of samples as training set $\boldsymbol{D}_{\text{train}}$, the rest is used as testing set $\boldsymbol{D}_{\text{test}}$.
Step 3) Initialize the population $\boldsymbol{P}$.
Step 4) Cycling
  while (stopping conditions are not satisfied) **do**
    i. Train base classifiers on the feature subspaces.
    ii. Evaluate individuals with respect to classification accuracy.
    iii. Compute the skill factor τ of each individual.
    iv. //Evolution
      for $g = 1$ to $g_{\max}$
        Mutation operation, crossover operation based on τ and rmp_1, $\boldsymbol{P}_C$ is generated.
        Concatenate $\boldsymbol{P}$ and $\boldsymbol{P}_C$ to form $\boldsymbol{P}_I$.
        Evaluate individuals for selected tasks only.
        Update the fitness φ and τ of individuals in $\boldsymbol{P}_I$.
        Select the fittest individuals to form the next $\boldsymbol{P}$.
      end
    v. The pseudo labels are obtained by voting scheme.
    vi. Add some unlabeled samples $\boldsymbol{D}_{U_m}$ to $\boldsymbol{D}_L$.
  end while
Step 5) Achieve the classification result

B. Evolutionary Multitask Spectral Feature Subspaces Generation

As Fig. 4 Part II shows, the spectral feature subspace of each individual classifier can be optimally selected in designed evolutionary multitask method. The proposed evolutionary multitask ensemble learning model with designed generators: Population

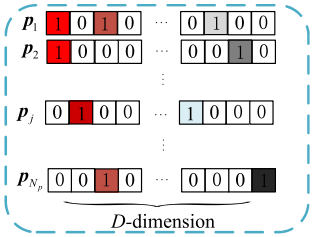


Fig. 5. Population initialization.

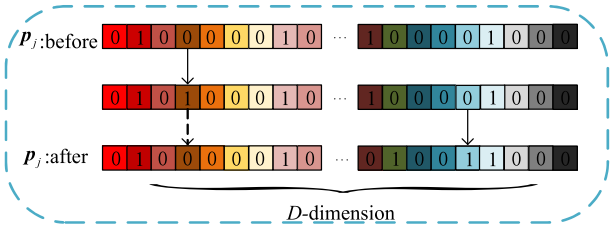


Fig. 6. Mutation operation.

initialization, crossover, mutation operator as well as the evaluation criteria of population individuals are detailed.

1) Population Initialization: First, an initial population of individuals is generated, and every individual represents a selected feature subspace. The initialization of population is shown as below.

Each sample in HSI has D_t spectral features, thus the dimensionality of the tth task is D_t. Accordingly, we define a unified search space with dimensionality D equal to $\max_t\{D_t\}$. Binary representation is commonly used to describe solutions in feature selection of HSI dataset. Therefore, in the population initialization, every individual is endowed with a vector of D binary variables. This vector constitutes the chromosome (the complete genetic material) of that individual. The initialization of N_p individuals is shown in Fig. 5. For jth individual, a randomly generated solution is allocated, which can be represented as a vector

$$\boldsymbol{p}_j = \left\{p_j^1, p_j^2, \ldots, p_j^d, \ldots, p_j^D\right\}, j = 1, 2, \ldots, N_p \tag{4}$$

where the variable p_j^d equals to 1 if the correspondent feature is selected; otherwise, it equals to 0. And particular attention is paid that the number of values 1 equals to the established number of features s. Then we compute the skill factor $\tau = \text{argmin}_t\{r_t\}$ of each individual on T tasks, where r_t is simply the index of the individual in the list of population members sorted in ascending order with respect to fitness value.

2) Crossover Operator: Second, the designed crossover operator is analogous to their biological namesakes, which as with the difference set based crossover operator proposed in [25]. However, a key feature of evolutionary multitask spectral feature subspaces generation is that two randomly selected parent candidates must satisfy certain conditions to undergo crossover. The principle is individuals prefer to mate with those belonging to the same cultural background. Two randomly selected parent candidates can freely undergo crossover if they possess the same skill factor. Conversely, if their skill factors differ, crossover only occurs as per a prescribed random mating probability, or else mutation kicks in [41].

3) Mutation Operator: Third, mutation occurs when conditions of crossover operation are not satisfied. In order to maintain the number of selected spectral features number, the mutation operator based on single-point mutation is designed. Taking individual $\boldsymbol{p}_j$ for example, Fig. 6 depicts the mechanism of the redesigned mutation operator. When there is a mutation in one or more of this chromosome bits, single-point mutation mechanism cannot guarantee the number of selected spectrums to be s. To tackle this problem, the single-point mutation operator with a restriction is designed. The chromosomes are first manipulated by single-point mutation operator, assume N_m bits have mutated. Second, N_m chromosome bits randomly selected from remaining genes which without mutation are forced to mutate in the opposite direction. The specialty of the redesigned mutation operator is individual's dimensions with value 1 and 0 remain unchanged, i.e., the number of value 1 is s and the total dimensions of individuals is D.

4) Selective Evaluation Based on HSI Dataset: T component classifiers take the tasks that maximize their classification accuracy on the selected spectral feature subspaces $\boldsymbol{S} = \{\boldsymbol{S}_{f_1}, \ldots, \boldsymbol{S}_{f_T}\}$ as the objective of the population evolution. The dth dimension binary value of each population individual chromosome determines whether the dth dimension feature of all samples in HSI dataset $\boldsymbol{D}$ is selected.

Only labeled samples are available to evaluate the classification performance of each base classifier before unlabeled samples obtain the pseudo labels. Unlabeled samples can obtain pseudo labels from predictions of T classifiers by relative majority voting scheme. Next, the labeled samples and unlabeled samples with highest confidence are added to $\boldsymbol{D}_{\text{train}}$, and then the classification accuracy of classifiers could be calculated. Details are described as follows. The labeled samples and unlabeled samples with pseudo labels in $\boldsymbol{D}_{\text{train}}$ are divided into two subsets $\boldsymbol{D}_{L_1}$ and $\boldsymbol{D}_{L_2}$ in each task: $\boldsymbol{D}_{L_1}$ is used to carry on the training of the classifiers and $\boldsymbol{D}_{L_2}$ is used to calculate the fitness of each individual according to the selected feature subset. Assume $\boldsymbol{F}_{f_t}$ is the fitness of individuals. The fitness of $\boldsymbol{p}_j$ on task $\boldsymbol{T}_t$ is calculated as follows:

$$\boldsymbol{F}_{f_t}^j = 1 - G_{L_2}/N_{L_2} \tag{5}$$

where G_{L_2} and N_{L_2} are the number of correctly classified samples and the total number of samples in $\boldsymbol{D}_{L_2}^{f_t}$ which meets the restriction with an original spectral value feature subspace $\boldsymbol{S}_{f_t}$. T classifiers search for optimal feature subspaces at the same time, thus the ultimate aim is to find a solution which describes the selected spectral feature subspace $\boldsymbol{S}_{f_t}^j$ for classifier f_t

$$\begin{aligned}\boldsymbol{S}_{f_t}^j = \text{argmin}&\left\{\boldsymbol{F}_{f_t}(\boldsymbol{S}_{f_t}^1), \ldots, \boldsymbol{F}_{f_t}(\boldsymbol{S}_{f_t}^{N_p})\right\}\\ &\text{subject to } \boldsymbol{S}_{f_t}^j \in \left\{\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_{N_p}\right\}.\end{aligned} \tag{6}$$

As a result, T different optimal spectral feature subspaces $\boldsymbol{S}$ for T classifiers can be generated.

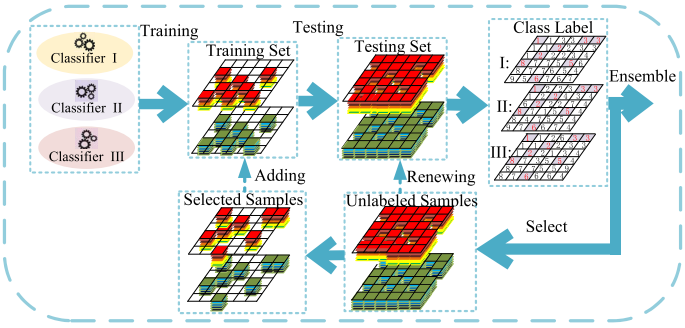


Fig. 7. Diagram of renewing the training set.

Fig. 8. Indian Pines image. (a) False-color image. (b) Reference image.

C. Dynamic Selection of Training Set for Reducing Isolated Points

As shown in Fig. 4 Part IV, the dynamic training set selection strategy is adopted in ensemble model for HSI classification to provide the predictions, and the diagram is shown in Fig. 7. First, depending on the classification results, the mth subset $\boldsymbol{D}_{U_m}$ of sample set with pseudo labels is selected, which is subject to the condition that the unlabeled samples get unanimous labeling results from T base classifiers are selected and added to the $\boldsymbol{D}_{\text{train}}$ and $\boldsymbol{D}_{\text{test}}$ is renewed simultaneously [52]. Repeat the above steps until the maximum number of iterations M or all samples with pseudo labels are added to $\boldsymbol{D}_{\text{train}}$. Specifically, in the mth iteration, $m = 1, 2, ..., M$, the labeled sample set $\boldsymbol{D}_{\text{train}}$ and sample set with pseudo labels $\boldsymbol{D}_{\text{test}}$ as follows:

$$\begin{aligned} \boldsymbol{D}_{\text{train}} &= \boldsymbol{D}_{\text{train}} \cup \boldsymbol{D}_{U_m} \\ \boldsymbol{D}_{\text{test}} &= \boldsymbol{D}_{\text{test}} - \boldsymbol{D}_{U_m}. \end{aligned} \tag{7}$$

The great advantage of this strategy is obvious for HSI classification to reduce the possibility of the isolated noises appearing in final classification maps.

D. Computational Complexity of EMT_EL

An analysis of computational complexity of EMT_EL is calculated. All the methods are implemented by MATLAB R2017b and executed on the same computer which has I7-4790 CPU, 3.60Ghz, and 8 GB RAM. Much more computational load is inevitably required for combining multiple optimal spectral features and selecting unlabeled samples with the highest confidence [53]. Thus, for each meta heuristic, the computational complexity is related to the number of variables and number of iterations. The computational complexity of EMT is $O(l(D \times T + Cof \times T))$, where l is the number of generations, D is number of variables, T is the number of solutions, and Cof is the cost of objective function. For ensemble learning with T classifiers is $O(T)$. When the number of iterations is M, the computational complexity of dynamic selection of training set strategy is $O(M)$. Therefore, the computational complexity of EMT_EL is $O(M(l(D \times T^2 + Cof \times T^2)))$.

IV. Experiments

In this section, three HSI datasets are used in experiments to test the performance of the proposed method. The original spectrum value feature is used to describe the features of samples. The overall accuracy (OA), average accuracy (AA), and Kappa coefficient measure (Kappa) are adopted to evaluate the classification performance of the proposed method. In order to obtain the statistical characteristics of the algorithm, each experiment repeated 30 times with different initial random subspaces, and simultaneously calculated the mean of OA, AA, and Kappa.

TABLE I
Sixteen Reference Classes in the Indian Pines Image

Class		Samples	
No	Name	#Train	#Test
1	Alfalfa	5	41
2	Corn-notill	137	1291
3	Corn-min	80	750
4	Corn	21	216
5	Grass-mowed	46	437
6	Grass-trees	69	661
7	Grass-pasture	3	25
8	Hay-windrowed	45	433
9	Oats	2	18
10	Soybeans-notill	93	879
11	Soybeans-min	237	2218
12	Soybeans-clean	56	537
13	Wheat	20	185
14	Woods	123	1142
15	Buildings-grass	36	350
16	Stones-towers	8	85
Total		981	9268

To verify the effectiveness of the proposed EMT_EL method, experiments are conducted on three hyperspectral images: The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Indian Pines image, the Reflective Optics System Imaging Spectrometer (ROSIS) University of Pavia image, and Salinas image.

Traditional HSI classification methods like SVM method [8], random forest (RoF) method [54], random subspace ensemble classifiers (RSE) [31], and simultaneous orthogonal matching pursuit (SOMP) [55] method are compared as baselines. STO feature selection-based SVM (PSO_SVM) [38] and ensemble classifiers (STO_EL) [39], the bilateral filter with support vector machine (BF_SVM) [56], the intrinsic image decomposition, and support vector machine (SVM IID)-based method [57] are used to approximately classify the HSI. In HSI classification, the multiclass version of SVM with Gaussian kernel is used, KNN where $k = 3$ and discriminator as three base classifiers for the ensemble classifier, and the ensemble results are obtained by the relative voting scheme of them. For the sparsity-based algorithms, the sparsity level is chosen as 30, since the HSI data consist of homogenous regions, a window of size 9×9 is used in SOMP.

TABLE II
CLASSIFICATION ACCURACY (%) FOR THE INDIAN PINES IMAGE. BEST RESULTS ARE REPORTED IN BOLD

Class	SVM	RoF	RSE	SOMP	PSO_SVM	STO_EL	BF_SVM	SVM_IID	EMT_EL
1	45.65	76.96	58.70	47.83	97.83	**100.00**	97.39	**100.00**	95.65
2	60.08	47.81	88.73	62.32	86.90	90.41	84.45	96.54	**97.62**
3	58.55	52.92	63.98	36.27	91.08	87.35	94.20	97.14	**98.31**
4	40.93	67.72	71.31	69.20	77.22	78.06	**98.80**	90.95	95.36
5	90.89	86.38	89.23	83.64	91.93	91.30	96.67	98.60	**100.00**
6	94.93	93.22	97.67	96.44	97.67	98.49	99.21	98.85	**99.32**
7	10.71	77.69	96.43	57.14	85.71	92.86	94.62	92.28	**100.00**
8	99.58	93.77	**100.00**	**100.00**	99.79	98.95	99.48	**100.00**	**100.00**
9	10.00	24.00	45.00	10.00	80.00	80.00	**100.00**	85.61	**100.00**
10	63.89	62.18	80.14	44.75	80.04	88.79	91.43	94.51	**95.68**
11	88.35	55.72	91.57	91.32	95.85	95.40	88.54	98.16	**99.59**
12	71.33	49.61	66.27	47.22	91.57	89.88	95.06	**96.93**	96.29
13	96.10	97.39	98.05	92.20	97.56	98.05	99.37	99.44	**100.00**
14	98.34	89.36	98.50	98.34	99.84	98.81	96.17	99.67	**99.92**
15	76.68	55.94	86.27	70.98	89.38	88.60	**98.26**	97.91	92.23
16	97.85	98.26	94.62	**100.00**	94.62	95.70	99.78	98.08	98.92
OA	79.19	65.72	87.23	75.50	92.39	93.10	92.77	97.56	**98.32**
AA	68.99	70.50	82.90	69.23	91.06	92.04	95.84	96.54	**98.06**
Kappa	79.07	61.09	87.20	75.36	92.38	92.79	91.3	97.19	**98.32**

In all evolutionary experiments that follow, a population of 100 individuals is evolved over 100 generations. In order to allow sufficient intertask communication, both the random mating probability and random mutation probability is 0.1. The s in each subspace are initialized using 10–50% of features. Suppose a small part of pixels are labeled, which can be used to construct the dataset $\boldsymbol{D}_L$, and the rest for dataset $\boldsymbol{D}_U$. The training samples are randomly divided into two sets in the process of individual evaluation. The experiments are conducted on a personal computer with 2-GHz Core2 Duo CPU and 4 GB-RAM using MATLAB.

A. Experimental Results on Three Datasets

1) AVIRIS Indian Pines Dataset: First, the Indian Pines image is used to evaluate the proposed method. The size of the Indian Pines image is 145 × 145, and spatial resolution is 20 × 20 m. The image has 220 bands across the spectral range from 0.2 to 2.4 μm, after removing 20 atmospheric, water absorption, and noisy bands, 200 bands are used in our experiments. Fig. 8(a) and (b) show the false-color image, and the image covers 16 ground-truth classes, respectively. As mentioned above, it has 200 spectral features per sample. In this experiment, around 10% of the samples for each class are chosen randomly for training (total 981 samples) and the remaining samples (9268 samples) are used to test the method (see Table I). The quantitative classification accuracy results are listed in Table II and the classification maps of compared methods on the Indian Pines image are visually shown in Fig. 9.

As Table II shows, in each class, the traditional classification methods like SVM, RoF, RSE, and SOMP perform worse than other classification methods, and it is obvious that these methods are challenging to identify samples belonging to the ninth class consisting of oats. This is partly due to quite similar spectral characteristics, and the lack of training samples (20 samples in total, and only two are used for training) is another possible cause. The OA of SVM method is 79.19%, yet the OA of PSO_SVM, BF_SVM SVM_IID method have been greatly improved, so it is proved that the performance of a single classifier can be optimized under a certain method. Moreover, compared with SVM classifier, the ensemble classifier based on RSE method has higher classification accuracy. From the comparison between RSE and STO_EL, we can clearly observe that the OA is increased to 93.10% from 87.23%. Above all, the optimization algorithm and ensemble learning show the advantages in improving the classification performance of classifiers. By using the proposed method, the OA, AA, and Kappa coefficient increased greatly. As shown in Table II, the proposed method obtains an OA of 98.32%, an AA of 98.06%, and a Kappa of 98.32%. Our method can achieve satisfying classification performance on "Oats" and "Grass-P", which are challenging tasks for the traditional methods. As can be observed from Fig. 9, the classification map of the proposed methods can achieve a better classification result than some of the classification methods which have a very noisy appearance. This experiment reveals that the proposed method successfully complete the implicit information transfer across multitasks and the information of samples without class label is used to full advantage.

2) ROSIS University of Pavia Dataset: The second HSI is the University of Pavia images whose spatial resolution is 1.3-m per pixel. It contains 610 × 340 pixels. After removing 12 noisy bands, there are 103 bands range from 0.43 to 0.86 μm for classification. The false-color image and 9 ground-truth classes are included in the reference image as shown in Fig. 10(a) and (b), respectively. We choose around 160 samples from each class randomly to construct the dataset $\boldsymbol{D}_L$ (total 1440 samples, which represents approximately 3.5% of the samples) and use the rest (total 41336 samples) for validation (see Table III). Table IV reports the OA, AA, Kappa coefficient, and the class-specific accuracy of different methods. And corresponding classification maps are presented in Fig. 11.

In this experiment, the BF_SVM method improves 13.58% than SVM in term of OA and the PSO_SVM has a gain of 1.5% over SVM. It shows that optimizing a single classifier can improve the classification accuracy to a certain extent. The SOMP

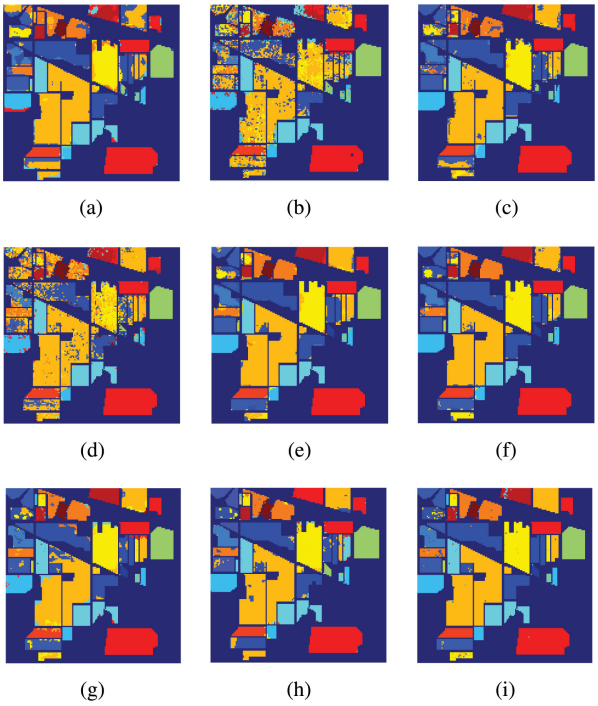


Fig. 9. Indian Pines image: The classification maps obtained by (a) SVM; (b) RoF; (c) RSE; (d) SOMP; (e) PSO_SVM; (f) STO_EL; (g) BF_SVM; (h) SVM_IID; (i) EMT_EL. (a) OA = 79.19. (b) OA = 65.72. (c) OA = 87.23. (d) OA = 75.50. (e) OA = 92.39. (f) OA = 93.10. (g) OA = 92.77. (h) OA = 97.56. (i) OA = 98.32.

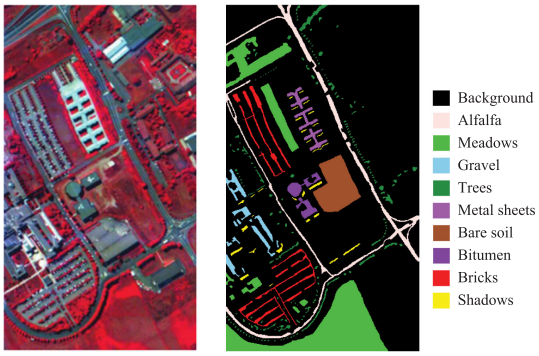


Fig. 10. Pavia-U image. (a) False-color image. (b) Reference image.

TABLE III
NINE REFERENCE CLASSES IN THE UNIVERSITY OF PAVIA IMAGE

Class		Samples	
No	Name	#Train	#Test
1	Asphalt	160	6002
2	Meadows	160	16871
3	Gravel	160	1899
4	Trees	160	2772
5	Metal sheet	160	1220
6	Bare soil	160	4550
7	Bitumen	160	1204
8	Brick	160	3337
9	Shadows	160	857
Total		1440	41336

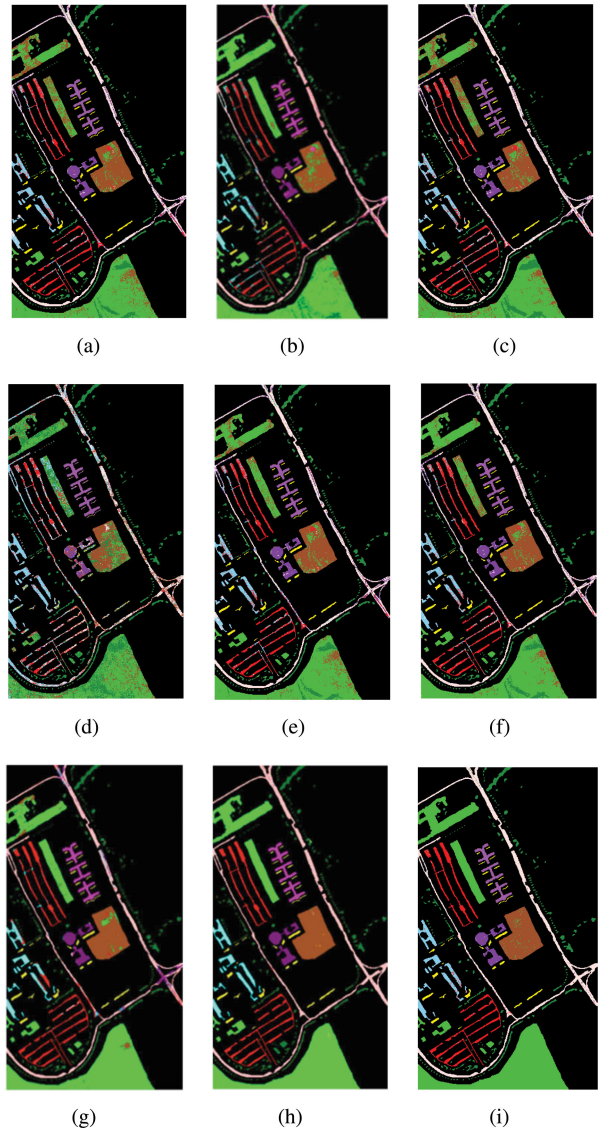


Fig. 11. Pavia-U image: The classification maps obtained by (a) SVM; (b) RoF; (c) RSE; (d) SOMP; (e) PSO_SVM; (f) STO_EL; (g) BF_SVM; (h) SVM_IID; (i) EMT_EL. (a) OA = 82.96. (b) OA = 85.98. (c) 81.33. (d) OA = 59.50. (e) OA = 84.43. (f) OA = 85.86. (g) OA = 91.59. (h) OA = 98.98. (i) OA = 98.87.

method incorporates the spatial information to original spectral information, but the overall accuracy is as low as 59.50%. The RoF and RSE methods also have high classification accuracy. However, comparing the results between the proposed method and other classification methods, we can clearly observe that the local continuity of pixels in local patches is greatly improved from Fig. 11. Although the SVM_IID method outperforms our proposed method with about 0.11% gain in terms of OA, the accuracy of AA and Kappa values are lower than our proposed method. And the classification results in Table IV presents that the proposed method obtains the best result with an average accuracy of 98.58%, a Kappa coefficient of 98.83%. The overall accuracy of the proposed method is 98.87%, which is 13.01% and 14.44% higher than STO_EL and PSO_SVM, respectively. The proposed method offers competitive superiority, especially

TABLE IV
CLASSIFICATION ACCURACY(%) FOR UNIVERSITY OF PAVIA IMAGE. BEST RESULTS ARE REPORTED IN BOLD

Class	SVM	RoF	RSE	SOMP	PSO_SVM	STO_EL	BF_SVM	SVM_IID	EMT_EL
1	75.49	81.33	81.84	40.66	76.72	79.40	81.04	**99.71**	97.33
2	80.89	85.47	75.42	59.09	84.62	85.38	92.66	99.75	**100.00**
3	81.61	78.00	84.23	90.38	81.13	86.90	81.61	**99.89**	95.57
4	91.61	95.29	91.97	95.86	90.63	93.18	98.85	93.19	**99.18**
5	99.63	99.35	99.63	**100.00**	99.63	99.63	99.97	**100.00**	**100.00**
6	86.22	89.56	83.79	61.36	84.41	86.42	94.89	**99.71**	97.18
7	93.31	90.54	93.68	62.86	93.83	94.59	96.84	**100.00**	98.87
8	81.83	79.84	80.53	39.95	81.12	80.64	93.90	97.21	**99.13**
9	100.00	99.97	99.89	17.42	100.00	99.89	99.98	96.21	**100.00**
OA	82.96	85.98	81.33	59.50	84.43	85.86	91.59	**98.98**	98.87
AA	87.84	88.81	87.89	63.06	88.01	89.56	93.30	98.41	**98.58**
Kappa	82.90	81.70	81.25	58.96	84.38	85.82	88.91	98.64	**98.83**

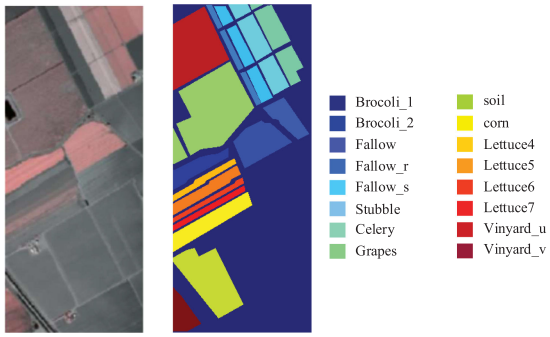


Fig. 12. Salinas image. (a) False-color image. (b) Reference image.

in identify samples belonging to the second class consisting of Meadows, the fifth class consisting of Metal sheets, and the ninth class consisting of Shadows. Information shared between different classifiers avoids the spectral feature sets falling into local optimizations and enlarges the training set, which can improve the performance of classification.

3) AVIRIS Salinas Dataset: The third HSI in our experiments is the Salinas image which was collected by the 224-band AVIRIS sensor over Salinas Valley, California. The Salinas scene data contains 512 × 217 pixels and it is characterized by high spatial resolution (3.7 m). As with Indian Pines scene, 204 bands are used after the 20 water absorption bands are discarded. It includes vegetables, bare soils, and vineyard fields. The false color image is included in Fig. 12(a) and ground truth map is included in Fig. 12(b), respectively. Around 1% samples from each class are selected randomly for training, where the remaining 99% for evaluate the method. The number of labeled and unlabeled samples for each class is shown in Table V. The classification accuracies of different methods evaluated by OA, AA, and Kappa coefficient are listed in Table VI. In order to visually display the results, the pixel-wise classification maps are generated for the Salinas dataset shown in Fig. 13.

As shown in Table VI, similar to the cases of Indian Pines and University of Pavia, SVM with an optimization-based techniques perform better than traditional SVM. In this comparison, the improvement of PSO_SVM, BF_SVM, and SVM_IID toward SVM demonstrates the advantage of optimization algorithms when the size of labeled training set is very small. The RSE and STO_EL methods outperform SVM with about

TABLE V
SIXTEEN REFERENCE CLASSES IN THE SALINAS IMAGE

Class		Samples	
No	Name	#Train	#Test
1	Brocoli1	20	1989
2	Brocoli2	37	3689
3	Fallow	20	1956
4	Fallow_r	14	1380
5	Fallow_s	27	2651
6	Stubble	39	3920
7	Celery	35	3544
8	Grapes	113	11158
9	Soil	62	6141
10	Corn-s	33	3245
11	Lettuce-romaine-4wk	11	1057
12	Lettuce-romaine-5wk	19	1908
13	Lettuce-romaine-6wk	9	907
14	Lettuce-romaine-7wk	11	1059
15	Vinyard-u	73	7195
16	Vinyard-v	18	1789
Total		541	53588

6.1% gain and 7.83% gain in terms of OA, AA, respectively. Especially, the STO_EL method provides higher classification performance than that of RSE and SVM, this result suggests the superiority of evolutionary feature subspaces generation and ensemble learning. By analyzing the experimental quantitative results and classification maps, concludes that the proposed method outperforms most of the other methods in terms of OA, AA, and Kappa. The proposed method obtains an OA of 99.99%, an AA of 99.99%, and a Kappa of 99.99%. The proposed method can achieve the best classification performance in each class of the Salinas image. The classification results in Table VI and Fig. 13 indicate that the proposed method can select better spectral features to improve the final classification result. Every part of the proposed algorithm is helpful to the classification of HSI dataset.

B. Analysis of the Parameters

The parameters (the number of training samples δ_s, the number of iteration of renewing the training set δ_i, and the number of the selected number of features δ_f) on HSI classification performance are worthy of research. This article examines the effect of the δ_s, δ_i, and δ_f in the proposed algorithms on Indian Pines.

TABLE VI
CLASSIFICATION ACCURACY (%) FOR SALINAS IMAGE. BEST RESULTS ARE REPORTED IN BOLD

Class	SVM	RoF	RSE	SOMP	PSO_SVM	STO_EL	BF_SVM	SVM_IID	EMT_EL
1	96.81	97.43	96.67	**100.00**	96.91	97.31	98.09	**100.00**	**100.00**
2	96.05	99.29	97.53	97.64	95.52	98.74	99.76	**100.00**	**100.00**
3	79.05	61.78	92.97	74.90	98.38	96.61	99.80	99.80	**100.00**
4	99.00	97.39	99.07	75.54	99.57	99.64	99.13	**100.00**	**100.00**
5	94.14	94.65	96.71	83.68	96.30	94.47	95.55	98.91	**100.00**
6	98.69	99.62	99.24	**100.00**	98.71	99.44	99.95	99.92	**100.00**
7	98.60	99.24	99.22	99.69	99.16	99.11	99.44	99.89	**100.00**
8	85.78	86.06	92.64	87.31	87.27	89.09	90.81	99.89	**100.00**
9	95.53	99.15	98.13	99.98	97.07	96.39	98.86	99.71	**100.00**
10	68.88	79.88	83.77	90.24	92.53	89.96	90.85	99.72	**99.97**
11	66.95	80.81	80.15	97.85	88.76	86.33	91.68	**100.00**	**100.00**
12	90.19	96.49	98.18	53.19	99.69	99.64	99.32	**100.00**	**100.00**
13	68.45	98.90	98.03	13.97	97.82	98.69	98.24	98.68	**100.00**
14	86.73	84.72	86.73	97.94	91.21	89.91	93.58	89.91	**100.00**
15	30.60	36.49	41.06	52.53	52.85	49.6	55.88	99.24	**99.94**
16	68.29	82.17	74.43	99.11	97.29	96.85	94.52	**100.00**	**100.00**
OA	80.81	83.89	86.91	84.62	88.87	88.64	94.09	99.59	**99.99**
AA	82.37	87.13	89.66	82.72	93.06	92.61	90.57	99.11	**99.99**
Kappa	80.68	81.96	85.97	83.57	88.02	88.6	89.46	99.53	**99.99**

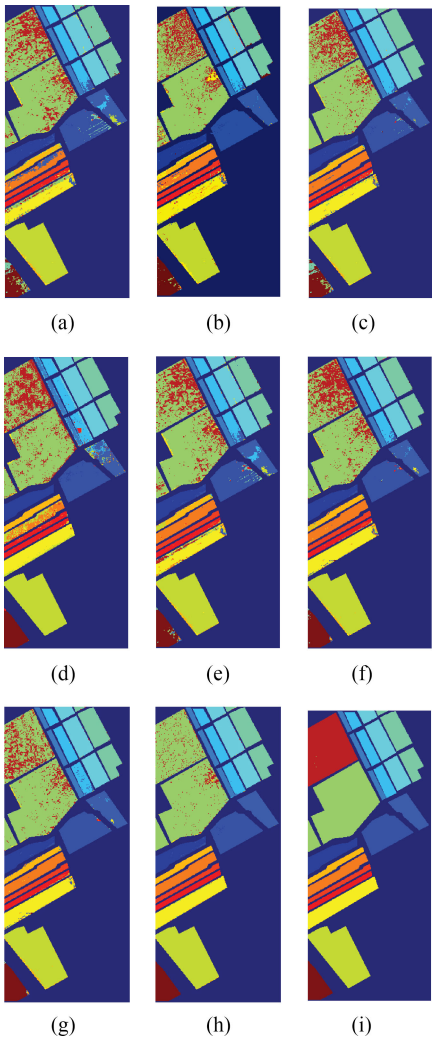


Fig. 13. Salinas image: The classification maps obtained by (a) SVM; (b) RoF; (c) RSE; (d) SOMP; (e) PSO_SVM; (f) STO_EL; (g) BF_SVM; (h) SVM_IID; (i) EMT_EL. (a) OA = 80.81. (b) OA = 83.89. (c) OA = 86.91. (d) OA = 84.62. (e) OA = 88.87. (f) OA = 88.64. (g) OA = 94.09. (h) OA = 99.62. (i) OA = 99.99.

When this article uses the same training set and testing set, that is about 10% samples from each class are chosen. The classification result of OA, AA, and Kappa value is analyzed in Fig. 14. One can observe from the experimental results in Fig. 14(a) that for the proposed method, $\delta_i = 5$, leads to the highest classification performance at all selected δ_i values. And the performance of the proposed method improves as δ_i increases, the number of iteration δ_i with the rise of $\delta_i = 1$ to $\delta_i = 5$, the OA value increased about 5%. This means that large number of iteration can yield good classification results. The appearance of significant increase partly because the information in unlabeled data is used for classification. When the parameter δ_i is more than 5, the classification accuracy of the proposed may be better, but this process would take a long time to finish. As shown in Fig. 14, the classification performance is good enough when the parameter δ_i is 5. Therefore, more iterative processes would rarely need.

Next, we demonstrate the effect of δ_s on the performance of the proposed algorithm. In the experiment, for a fixed $\delta_i = 1$, the number of training samples δ_s (in percent) ranges from 1% to 10% per class, the change of three indices (OA, AA, and Kappa) plotted on the entire test set for Indian Pines are shown in Fig. 14(b). The horizontal axis indicates the number of training samples and the vertical axis is the overall accuracy (%). For few training samples δ_s, the classification map cannot be faithfully approximated the ground-truth image. On the other hand, as δ_s decreases toward the small size, a small amount of information contained in the training set cannot obtain optimal parameters setup, leading to a classification performance degradation. One can also see that for sufficiently large number of training samples δ_s, with the increment of the number of training samples, OA, AA, and Kappa increased accordingly as a whole. There would certainly be a fall in classification accuracy, for instance, when the number of training samples increased from 8% to 9%, the Kappa value decreases.

According to the analysis of reality, in most cases, a large number of labeled samples for training is difficult to

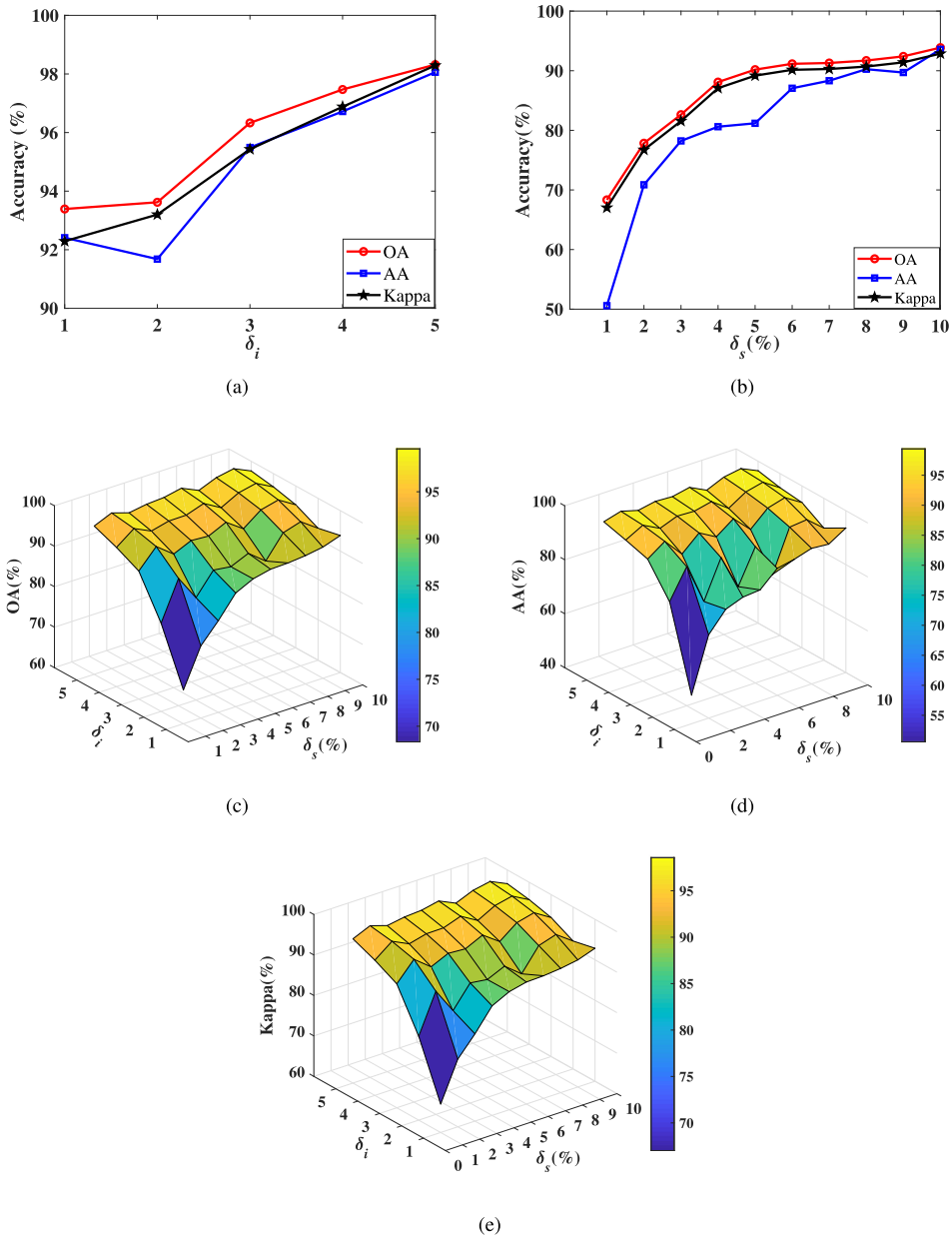


Fig. 14. Analysis of classification performance of parameters in 2-D and 3-D format on Indian Pines (the parameters include δ_i and δ_s). (a) Accuracy with the change of δ_i. (b) Accuracy with the change of δ_s. (c) Overall accuracy with the change of δ_i and δ_s. (d) Average accuracy with the change of δ_i and δ_s. (e) Kappa coefficient with the change of δ_i and δ_s.

obtain. When the size of training set is not large enough, the classification performance is increasing with δ_i growing, as show in Fig. 14(c)–(e). In a high number of iteration, the proposed method has the advantage of high classification accuracy with the limitation in the number of training samples, but the classification accuracy does not increase monotonously with the increase of δ_s. That would lead to a decrease in the classification accuracy once more samples with pseudo labels with incorrect classes information in the last classification result are selected.

Finally, the correlation between the number of the selected features δ_f and the classification accuracy is tested. It is apparent from Fig.15 that as the number of the selected features increases,

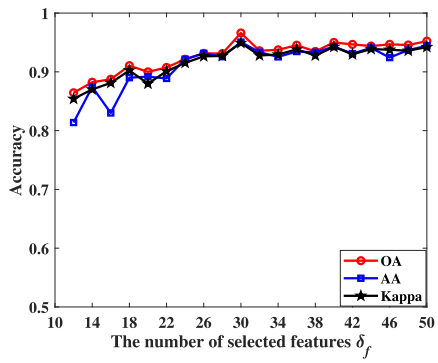


Fig. 15. Analysis of the δ_f on Indian Pines.

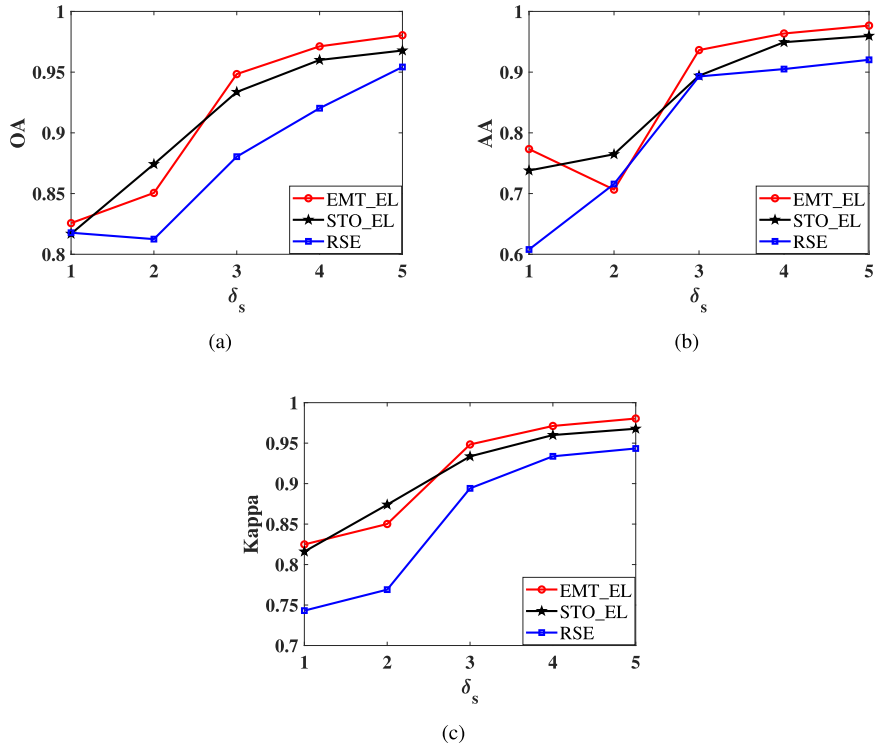


Fig. 16. Comparison of different model on Indian Pines dataset. (a) OA. (b) AA. (C) Kappa.

there will be an improvement of the classification accuracy in general when the number of the selected features δ_f is less than 30. However, an increase in δ_f of limited training samples will show no significant improvement and even a decrease in classification accuracy, this phenomenon is especially obvious when δ_f exceeds 30. Taken together, these results suggest that there is an appropriate δ_f which contains exactly all the information and would not cause redundancy, thus the classification accuracy can be guaranteed to some degree.

C. Analysis of the Quality of Selected Feature Subspaces in EL_EMT

In order to highlight the advantage of evolutionary multitask feature selection algorithm, Indian Pines dataset is chosen for comparison. Around 3% of the samples for each class are randomly selected for training and testing. The classification accuracy of evolutionary multitask ensemble learning model and other methods for hyperspectral image classification is shown in Fig. 16. For Indian Pine dataset, EL_EMT generally attains better classification performance than the RSE and STO_EL which is without evolutionary multitasking method. Compared with STO_EL, the OA of EL_EMT improved from 81.77% to 82.56% approximately when the δ_i is 1. As can be seen in Fig. 16, no matter how many iterations δ_i is selected, EL_EMT always achieves higher or similar classification accuracy to the RSE method. This suggests the feature subspaces generation benefits from multitask optimization, the cooperation among tasks has a positive influence on the classification performance.

TABLE VII
COMPUTATIONAL TIME OF DIFFERENT METHODS (SECONDS)

Model	PSO_SVM	STO_EL	EMT_EL
Computational Time	25.127624	94.133584	350.334532

D. Analysis of Computational Efficiency

In order to assess the computational efficiency, the computational time is listed in the Table VII and the convergence trends of computational objective functions are displayed in the Table VII and Fig. 17. From the data in Table VII, it is found that the proposed method takes longer computational time to find optimal spectral feature subspaces than the traditional HSI classification methods with evolutionary algorithm. From the comparison among PSO_SVM, STO_EL, and EMT_EL, it can see that STO_EL and EMT_EL cost more computational time than PSO_SVM. This result can be explained, in part, by the fact that training a base classifier spends less time than an ensemble classifier does. The difference of computational time between STO_EL and EMT_EL is likely to be related to the evaluation of initial population on the multiple tasks and feature sharing in the searching procedure.

However, although the computational time of the proposed method is long, an ensemble classifier with high robustness can be obtained in the proposed ensemble learning method and each objective function converges to the optimal solution more easily than other methods. As shown in Fig. 17, the convergence trends of the multiple classifiers' objective functions illustrate that the superiority on classification performance of the proposed method is at the cost of the increasing computational

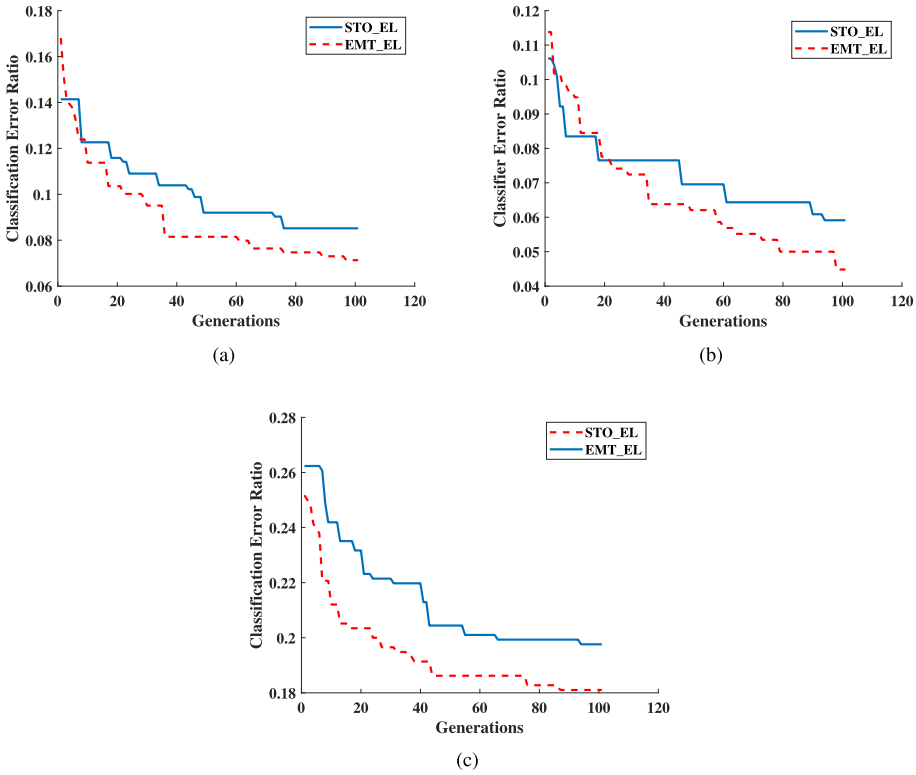


Fig. 17. Convergence trends of classification error ration in EMT_EL and STO_EL for Indian Pines dataset. (a) The first classifier: SVM. (b) The second classifier: Discriminator. (c) The third classifier: KNN.

time. Therefore, the computational efficiency of the proposed method is relatively higher than some other methods in terms of classification performance.

V. Conclusion

Traditional hyperspectral image classification methods based on ensemble learning optimize multiple base classifiers separately. However, little models pay attention to the correlation among searching processes of optimal spectral feature subspaces, which may cause some features to be searched repeatedly. Thus the effectiveness and quality of selected spectral feature subspace are not high. In summary, a method which can reduce many unnecessary search of features, avoid trapping in local optimal subspace, and achieve good quality in the classification performance is required. Therefore, in this article, an EMT_EL model for HSI classification was presented. The proposed model generated the optimal spectral feature subspace for each base classifier in parallel and feature sharing happens between each classifier to assist each other in searching processes. The advantages of the proposed method are as follows.

1) The proposed model searches for an optimal spectral feature subspace for each base classifier in parallel. In the unified search space, searching for feature subspaces for different base classifiers not only improves the utilization efficiency of spectral features but also selects more informative and representative feature subspaces.

2) The significant spectral features are encouraged to be shared among classification methods, which will accelerate convergence toward the direction of the optimal feature subspace, avoid trapping in local optimal subspace and improve searching capability.

3) The randomization-enhanced genetic operators are designed in evolutionary multitask ensemble learning model, which can avoid the invalid and repetitive spectral feature subspaces, facilitate the exchange of information, and improve the joint searching efficiency of the feature subspace.

4) The dynamic updated training set strategy is adopted in which pseudo labels are assigned to raw data by semisupervised label generation strategy and update the training set in terms of the classification performances of base classifiers. Thereby, the inconsistency and uncertainty of the predictions among base classifiers can be avoided, and a less noisy map is achieved finally.

On real three hyperspectral images, the experimental results reveal outstanding transferability among classification methods, which can avoid some features are searched repeatedly, converge to a better spectral feature subspace faster, and achieve a higher classification accuracy. Moreover, the experiment exhibits the noise reduction capacity of the proposed method.

However, the classification superiority is presented at the expense of computational burden, which leads to a long processing time. It is mainly caused by the complexity of multitask optimization mechanism, in which the feature sharing happens in the searching, and initial population is evaluated on all tasks. Thus, a proper balance between effectiveness and efficiency would form the central focus in our future work.

REFERENCES

[1] M. Borengasser, W. S. Hungate, and R. Watkins, *Hyperspectral Remote Sensing: Principles and Applications*, New York, NY, USA: Taylor & Francis, 2007.

[2] H. Su, S. Tian, Y. Cai, Y. Sheng, and M. Najafian, "Optimized extreme learning machine for urban land cover classification using hyperspectral imagery," *Frontiers Earth Sci.*, vol. 11, no. 4, pp. 765–773, 2017.

[3] S. U. Hongjun and L. Hao, "A novel dynamic classifier selection algorithm using spatial-spectral information for hyperspectral classification," *Remote Sens. Land Res.*, vol. 29, no. 2, pp. 15–21, 2017.

[4] J. Wang, J. Zhou, and W. Huang, "Attend in bands: Hyperspectral band weighting and selection for image classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4712–4727, Dec. 2019.

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

[6] E. T. Gormus, N. Canagarajah, and A. Achim, "Dimensionality reduction of hyperspectral images with wavelet based empirical mode decomposition," in *Proc. 18th IEEE Int. Conf. Image Process.*, 2011, pp. 1709–1712.

[7] M. Ghamary Asl, M. R. Mobasheri, and B. Mojaradi, "Unsupervised feature selection using geometrical measures in prototype space for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 3774–3787, Jul. 2014.

[8] Y. Bazi and F. Melgani, "Toward an optimal SVM classification system for hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3374–3385, Nov. 2006.

[9] B. Tu, J. Wang, X. Kang, G. Zhang, X. Ou, and L. Guo, "KNN-based representation of superpixels for hyperspectral image classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 4032–4047, Nov. 2018.

[10] B. Waske, S. Van Der Linden, J. A. Benediktsson, A. Rabe, and P. Hostert, "Sensitivity of support vector machines to random feature selection in classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 7, pp. 2880–2889, Jul. 2010.

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

[12] J. Xia, P. Ghamisi, N. Yokoya, and A. Iwasaki, "Random forest ensembles and extended multiextinction profiles for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 202–216, Jan. 2018.

[13] X. He and Y. Chen, "Transferring CNN ensemble for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, to be published.

[14] 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. Top. Appl. Earth Observ. Remote Sens.*, vol. 8, no. 2, pp. 845–858, Feb. 2015.

[15] Y. Gu, J. Chanussot, X. Jia, and J. A. Benediktsson, "Multiple kernel learning for hyperspectral image classification: A review," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6547–6565, Nov. 2017.

[16] A. Samat, P. Du, S. Liu, J. Li, and L. Cheng, "$\mathrm{E}^2\mathrm{lms}$: Ensemble extreme learning machines for hyperspectral image classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1060–1069, Apr. 2014.

[17] 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. Top. Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2405–2417, Jun. 2015.

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

[19] B. Zhang, A. K. Qin, and T. Sellis, "Evolutionary feature subspaces generation for ensemble classification," in *Proc. Genet. Evol. Comput. Conf.*, 2018, pp. 577–584.

[20] P. Gurram, H. Kwon, and C. Davidson, "Coalition game theory-based feature subspace selection for hyperspectral classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 2354–2364, Jun. 2016.

[21] J. Xia, J. Chanussot, P. Du, and X. He, "Rotation-based support vector machine ensemble in classification of hyperspectral data with limited training samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1519–1531, Mar. 2016.

[22] S. Patra, K. Bhardwaj, and L. Bruzzone, "A spectral-spatial multicriteria active learning technique for hyperspectral image classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 10, no. 12, pp. 5213–5227, Dec. 2017.

[23] J. Xia, L. Bombrun, Y. Berthoumieu, C. Germain, and P. Du, "Spectral-spatial rotation forest for hyperspectral image classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 10, no. 10, pp. 4605–4613, Oct. 2017.

[24] A. AlSuwaidi, B. Grieve, and H. Yin, "Feature-ensemble-based novelty detection for analyzing plant hyperspectral datasets," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1041–1055, Apr. 2018.

[25] M. Gong, M. Zhang, and Y. Yuan, "Unsupervised band selection based on evolutionary multiobjective optimization for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 544–557, Jan. 2016.

[26] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits Syst. Mag.*, vol. 6, no. 3, pp. 21–45, Jul./Sep. 2006.

[27] 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.

[28] J. Jiang, J. Ma, Z. Wang, C. Chen, and X. Liu, "Hyperspectral image classification in the presence of noisy labels," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 851–865, Feb. 2019.

[29] S. Jia, Z. X. Zhu, L. L. Shen, and Q. Q. Li, "A two-stage feature selection framework for hyperspectral image classification using few labeled samples," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1023–1035, Apr. 2014.

[30] Z. Wang, B. Du, L. Zhang, L. Zhang, and X. Jia, "A novel semisupervised active-learning algorithm for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3071–3083, Jun. 2017.

[31] J. Xia, M. Dalla Mura, J. Chanussot, P. Du, and X. He, "Random subspace ensembles for hyperspectral image classification with extended morphological attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 4768–4786, Sep. 2015.

[32] B. Kuo, C.-H. Chuang, C.-H. Li, and C.-T. Lin, "Subspace selection based multiple classifier systems for hyperspectral image classification," in *Proc. 1st Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens.*, 2009.

[33] Z. Yu *et al.*, "Multiobjective semisupervised classifier ensemble," *IEEE Trans. Cybern.*, vol. 49, no. 6, pp. 2280–2293, Jun. 2019.

[34] M. Ahmet, K. Niyazi, and B. Erdem, "Random subspace method with class separability weighting," *Expert Syst. Int. J. Knowl.*, vol. 33, no. 3, pp. 275–285, 2016.

[35] L. Hou, B. Zhang, X. Erlei, and J. Licheng, "Spectral-spatial hyperspectral image ensemble classification via joint sparse representation," *Pattern Recognit.*, vol. 59, pp. 42–54, 2016.

[36] D. W. Opitz, "Feature selection for ensembles," in *Proc. 16th Nat. Conf. Artif. Intell.*, 1999, pp. 379–384.

[37] C. Girish and S. Ferat, "A survey on feature selection methods," *Comput. Elect. Eng.*, vol. 40, no. 1, pp. 16–28, 2014.

[38] P. Ghamisi and J. A. Benediktsson, "Feature selection based on hybridization of genetic algorithm and particle swarm optimization," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 2, pp. 309–313, Feb. 2015.

[39] Y. Chen, X. Zhao, and Z. Lin, "Optimizing subspace SVM ensemble for hyperspectral imagery classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1295–1305, Apr. 2014.

[40] P. Du, Y. Chen, J. Xia, and K. Tan, "A novel remote sensing image classification scheme based on data fusion, multiple features and ensemble learning," *J. Indian Soc. Remote Sens.*, vol. 41, no. 2, pp. 213–222, 2016.

[41] A. Gupta, Y. Ong, and L. Feng, "Multifactorial evolution: Toward evolutionary multitasking," *IEEE Trans. Evol. Comput.*, vol. 20, no. 3, pp. 343–357, Jun. 2016.

[42] J. Tang, Y. Chen, Z. Deng, Y. Xiang, and C. P. Joy, "A group-based approach to improve multifactorial evolutionary algorithm," in *Proc. 27th Int. Joint Conf. Artifi. Intell.*, 2018, pp. 3870–3876.

[43] A. Gupta, Y. Ong, L. Feng, and K. C. Tan, "Multiobjective multifactorial optimization in evolutionary multitasking," *IEEE Trans. Cybern.*, vol. 47, no. 7, pp. 1652–1665, Jul. 2017.

[44] Lei Zhou, L. Feng, Jinghui Zhong, Y. Ong, Z. Zhu, and E. Sha, "Evolutionary multitasking in combinatorial search spaces: A case study in capacitated vehicle routing problem," in *Proc. IEEE Symp. Ser. Comput. Intell.*, 2016, pp. 1–8.

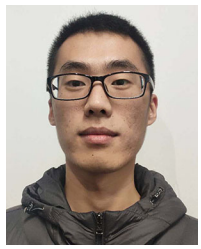
[45] H. T. Thanh Binh, P. Dinh Thanh, T. Ba Trung, and L. Phuong Thao, "Effective multifactorial evolutionary algorithm for solving the cluster shortest path tree problem," in *Proc. IEEE Cong. Evol. Comput.*, 2018, pp. 1–8.
[46] X. Liu, Q. Hu, Y. Cai, and Z. Cai, "Extreme learning machine-based ensemble transfer learning for hyperspectral image classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3892–3902, 2020.
[47] Y. Zhang, G. Cao, A. Shafique, and P. Fu, "Label propagation ensemble for hyperspectral image classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 12, no. 9, pp. 3623–3636, Sep. 2019.
[48] C. Qi, Z. Zhou, Y. Sun, H. Song, L. Hu, and Q. Wang, "Feature selection and multiple kernel boosting framework based on PSO with mutation mechanism for hyperspectral classification," *Neurocomput.*, vol. 220, no. 12, pp. 181–190, 2017.
[49] L. S. Oliveira, R. Sabourin, F. Bortolozzi, and C. Y. Suen, "Feature selection for ensembles: A hierarchical multi-objective genetic algorithm approach," in *Proc. 7th Int. Conf. Document Anal. Recognit.*, 2003, pp. 676–680.
[50] X. Zheng, A. K. Qin, M. Gong, and D. Zhou, "Self-regulated evolutionary multi-task optimization," *IEEE Trans. Evol. Comput.*, vol. 24, no. 1, pp. 16–28, Feb. 2020.
[51] Y. S. Ong, P. B. Nair, and A. J. Keane, "Evolutionary optimization of computationally expensive problems via surrogate modeling," *AIAA J.*, vol. 41, no. 4, pp. 687–696, 2003.
[52] Z. Zhang and M. M. Crawford, "An ensemble active learning approach for spectral-spatial classification of hyperspectral images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2015, pp. 4963–4966.
[53] S. A. Medjahed and M. Ouali, "A new hybrid SSA-TA: Salp swarm algorithm with threshold accepting for band selection in hyperspectral images," *Appl. Soft Comput.*, vol. 95, 2020, Art. no. 106534.
[54] J. Ham, Y. Chen, M. M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 492–501, Mar. 2005.
[55] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.
[56] A. S. Sahadevan, A. Routray, B. S. Das, and S. Ahmad, "Hyperspectral image preprocessing with bilateral filter for improving the classification accuracy of support vector machines," *J. Appl. Remote Sens.*, vol. 10, no. 2, 2016, Art. no. 025 004.
[57] X. Kang, S. Li, L. Fang, and J. A. Benediktsson, "Intrinsic image decomposition for feature extraction of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2241–2253, Apr. 2015.

Jiao Shi (Member, IEEE) received the B.Eng. degree in circuits and system and the Ph.D. degree in electrical science & technology from the School of Electronic Engineering, Xidian University, Xi'an, China, in 2009 and 2015, respectively.

From 2013 to 2014, she was a Visiting Scholar with the Leiden Institute of Advanced Computer Science, Leiden University, Leiden, The Netherlands. Since 2015, she has been a Teacher with the School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China, where she was promoted to Associate Professor in 2018. In 2016 and 2019, she was a Visiting Scholar with the College of Information Technology, Incheon National University, Incheon, South Korea.

Tao Shao received the B.S. degree in telecommunication engineering from ShanDong University, WeiHai, China, in 2019. She is currently working toward the master's degree in circuits and systems with Northwestern Polytechnical University.

Her current research interests include computational intelligence and image understanding.



Xiaodong Liu received the B.S. degree in electrical engineering and automation from Xidian University, Xi'an, China, in 2018. He is currently working toward the master's degree in circuits and systems with Northwestern Polytechnical University.

His current research interests include computational intelligence and image understanding.

Xi Zhang received the B.S. degree in integrated circuit design and integrated system from Xi'an University of Posts and Telecommunications, Xi'an, China, in 2019. She is currently working toward the master's degree in electronic and communication engineering with Northwestern Polytechnical University.

Her current research interests include computational intelligence and image understanding.

Zeping Zhang received the B.S. degree in electronic information engineering from South China Agriculture University, GuangZhou, China, in 2019. He is currently working toward the master's degree in electronic and communication engineering with Northwestern Polytechnical University.

His current research interests include computational intelligence and image understanding.

Yu Lei (Member, IEEE) received the B.Eng. degree in circuits and systems and the Ph.D. degree in electrical science & technology from the School of Electronic Engineering, Xidian University, Xi'an, China, in 2009 and 2015, respectively.

Since 2015, he has been a Teacher with the School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China, where he was promoted to Associate Professor, in 2019. In 2016 and 2019, he was a Visiting Scholar with the College of Information Technology, Incheon National University, Incheon, South Korea.