Random Subspace Ensembles for Hyperspectral Image Classification With Extended Morphological Attribute Profiles

Junshi Xia, *Student Member, IEEE*, Mauro Dalla Mura, *Member, IEEE*, Jocelyn Chanussot, *Fellow, IEEE*, Peijun Du, *Senior Member, IEEE*, and Xiyan He

***Abstract*—Classification is one of the most important techniques to the analysis of hyperspectral remote sensing images. Nonetheless, there are many challenging problems arising in this task. Two common issues are the curse of dimensionality and the spatial information modeling. In this paper, we present a new general framework to train series of effective classifiers with spatial information for classifying hyperspectral data. The proposed framework is based on the two key observations: 1) the curse of dimensionality and the high feature-to-instance ratio can be alleviated by using random subspace (RS) ensembles; and 2) the spatial–contextual information is modeled by the extended multiattribute profiles (EMAPs). Two fast learning algorithms, i.e., decision tree (DT) and extreme learning machine (ELM), are selected as the base classifiers. Six RS ensemble methods, namely, RS with DT, random forest (RF), rotation forest, rotation RF (RoRF), RS with ELM (RSELM), and rotation subspace with ELM (RoELM), are constructed by the multiple base learners. Experimental results on both simulated and real hyperspectral data verify the effectiveness of the RS ensemble methods for the classification of both spectral and spatial information (EMAPs). On the University of Pavia Reflective Optics Spectrographic Imaging System image, our proposed approaches, i.e., both RSELM and RoELM with EMAPs, achieve the state-of-the-art performances, which demonstrates the advantage of the proposed methods. The key parameters in RS ensembles and the computational complexity are also investigated in this paper.**

Manuscript received October 19, 2014; revised January 14, 2015; accepted February 16, 2015. This work was supported in part by the Natural Science Foundation of China under Grant 41471275, by the Priority Academic Program Development of Jiangsu Higher Education Institutions, by the Fundamental Research Funds for the Central Universities, and by the Project XIMRI ANR-BLAN-SIMI2-LS-101019-6-01. *(Corresponding author: Peijun Du.)*

J. Xia is with the Key Laboratory for Satellite Mapping Technology and Applications of National Administration of Surveying, Mapping and Geoinformation of China, Nanjing University, Nanjing 210023, China, and also with the Grenoble Image Speech Signals and Automatics Laboratory (GIPSA-lab), Grenoble Institute of Technology, 38400 Grenoble, France (e-mail: xiajunshi@gmail.com).

M. Dalla Mura and X. He are with the Grenoble Images Speech Signals and Automatics Laboratory (GIPSA-lab), Grenoble Institute of Technology, 38400 Grenoble, France (e-mail: mauro.dalla-mura@gipsa-lab.grenoble-inp.fr; greenhxy@gmail.com).

J. Chanussot is with the Grenoble Image Speech Signals and Automatics Laboratory (GIPSA-lab), Grenoble Institute of Technology, 38400 Grenoble, France, and also with the Faculty of Electrical and Computer Engineering, University of Iceland, 101 Reykjavik, Iceland (e-mail: jocelyn.chanussot@gipsa-lab.grenoble-inp.fr).

P. Du is with the Key Laboratory for Satellite Mapping Technology and Applications of National Administration of Surveying, Mapping and Geoinformation of China, Nanjing University, Nanjing 210023, China (e-mail: dupjrs@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2015.2409195

***Index Terms*—Classification, extended multiattribute profiles (EMAPs), hyperspectral data, random subspace (RS).**

I. Introduction

In THE context of hyperspectral image analysis, classification is an intense field of research and development [1]–[5]. The difficulties of the supervised classification of high spatial resolution hyperspectral data come from at least three sources.

- The ratio between features (spectral bands) and available training samples is large.
- The feature set might show some redundancy.
- Existence of many approaches to exploit the spatial information in classification, but unavailability of a reliable approach.

A considerable amount of literature has focused on hyperspectral image classification [6]–[11]. One of the most widely used approaches is based on kernel methods, such as the support vector machines (SVMs), due to their good generalization capability, ability in addressing small-size-sample problems, and the curse of dimensionality [9], [12], [13]. In addition, kernel methods can efficiently define nonlinear decision boundaries, dealing with cases in which the data are not linearly separable [12]. However, the selection of kernels and parameters is still an open question, which needs to be further investigated.

Another alternative strategy for providing enhanced classification performance is classifier ensembles, which are deemed to be better than individual classifiers [14]–[18]. A popular ensemble method is the random subspace (RS) ensembles [19]. The idea is intuitive and simple: subsets of feature set are used in the ensemble instead of using all features for all the individual classifiers, and ensemble classifier integrates the outputs of all individual classifiers using a majority voting rule to obtain the final result. Each classifier in the ensemble is constructed on a different feature subset by randomly sampling the original feature set. The rationale behind the RS ensembles is to break down a complex high dimensional problem into several lower dimensional subproblems. Thus, they can address such problems as the curse of dimensionality and high feature-to-instance ratio [20].

The most popular RS ensemble method for high-dimensional data (hyperspectral and multidate images) classification is random forest (RF) [21]–[24]. In addition, Waske *et al.* [25]

TABLE I
INDIVIDUAL AND ENSEMBLE CLASSIFICATION APPROACHES CONSIDERED FOR THE STUDY

Individual classifiers	(Notation)	Classifier ensembles	(Notation)
Decision tree	DT	Random subspace with DT	RSDT
		Random Forest	RF
		Rotation Forest	RoF
		Rotation Random Forest	RoRF
Extreme learning machine	ELM	Random subspace with ELM	RSELM
		Rotation subspace with ELM	RoELM

developed a random-selection-based SVM for the classification of hyperspectral images. Recently, Xia *et al.* [26], [27] have used rotation forest (RoF), which was originally developed by Rodriguez and Kuncheva [26], to classify hyperspectral remote sensing images. In comparison with RF, RoF [26]–[28] promotes both the interclassifier diversity and accuracy of individual classifiers by using a feature extraction approach. Therefore, it can produce more accurate result than RF in most cases [26], [27].

Recent studies have demonstrated that *spectral–spatial* approaches can provide more accurate classification results by integrating the spatial and spectral information together [29]. The motivation is due to the fact that spatial features are discriminant features that can well complement the spectral ones. Different approaches can be used to extract spatial features [30]–[34]. Some of them are based on mathematical morphology, which is a powerful tool for the analysis and processing of geometrical structures in the spatial domain [35]. Pesaresi and Benediktsson introduced the morphological profile (MP) for classifying very high spatial resolution images using a sequence of geodesic opening and closing operations [30]. The derivative of the MP was also defined in their study. Furthermore, Benediktsson *et al.* proposed the extended MP (EMP), in which an MP is computed on each component after reducing the dimensionality of the data [36]. The first few components and the EMP are stacked together and then classified by a neural network. The main drawback of the method in [36] is that it is constructed for classification of urban structures and it cannot fully use the spectral information of hyperspectral data [37]. Fauvel *et al.* developed a spectral and spatial fusion methods based on EMP and the original hyperspectral data to overcome this problem [37].

In the work of Dalla Mura *et al.* [38], [39], attribute profiles (APs) were proposed for extracting additional spatial features for the classification of remote sensing imagery, extending the MP and EMP concepts. APs have proven to extract more informative spatial features than MPs in the classification of high-resolution images. Since then, the AP and its extensions have been widely used for the classification and change detection of multispectral/hyperspectral and LiDAR data. Dalla Mura *et al.* presented a technique based on extended APs (EAPs) and independent component analysis for the classification of urban hyperspectral images [40]. Reddy Marpu *et al.* explored the use of APs based on three supervised and two unsupervised feature extraction techniques for the classification of hyperspectral data with SVM and RF classifiers [41]. Pedergnana *et al.* proposed a classification approach of features extracted with EAPs computed on both optical and LiDAR images, leading to the integration of spectral, spatial, and elevation data [42]. Pedergnana *et al.* proposed a novel iterative technique based on genetic algorithm to select the optimal features from the extended multiattribute profiles (EMAPs) [43]. Falco *et al.* investigated the performance of change detection in very high resolution image based on APs [44]. Li *et al.* presented a generalized composite kernel framework for hyperspectral image classification by combining spectral and spatial information (EMAPs) [45]. Bernabe *et al.* proposed a new strategy combining EMAPs and kernel principal component analysis (PCA) for the classification of multispectral/hyperspectral images [46]. Song *et al.* applied a sparse-representation-based learning approach to classify EMAPs extracted from hyperspectral data [47].

From the preceding literature review, it can be seen that, when the EMAPs are used for hyperspectral data classification, two strategies are often adopted, as follows:

- applying feature selection/extraction [43] or advanced classifiers to EMAPs [47];
- integrating EMAPs with spectral information to formulate the composite kernel for kernel-based methods [45].

In this paper, we propose an advanced classification scheme based on RS ensembles applied to EMAPs features. Decision tree (DT) and artificial neural network (ANN) are usually adopted as a base learner in RS ensemble because they are unstable as weak learners. Small changes in the training data can lead to potentially large variations in the results, making a high diversity within the ensemble. Considering the computational cost, we construct the RS ensembles with two fast learning algorithms: classification and regression tree (CART) and a recently proposed neural network classifier, i.e., extreme learning machine (ELM) [48], [49]. EMAPs are generated by the combination of APs and the first several components extracted by PCA. Six classifier ensembles, namely, RSDT, RF, RoF, RoRF, RSELM, and RoELM, are considered, as shown in Table I.

The novelty of this work lies in the following:

- proposing an ensemble classifier using ELM base learner and two possible strategies for building the ensembles (i.e., random and rotation subspace);
- introducing rotation RF (RoRF) [50] in the field of hyperspectral remote sensing;
- defining spectral–spatial classification techniques based on the proposed ensembles and on the spatial features computed by EMAPs.

In particular, the performances in a scenario with limited training samples and high input dimensions and the computational complexity are investigated in this paper. It is should be noted that the spectral information and EMAPs are directly applied to the RS ensemble methods without any preprocessing technique (e.g., feature extraction/selection or whitening).

The overall structure of this paper takes the form of seven sections, including this introductory section. Section II presents an introduction to DT and its ensembles. The proposed ELM ensemble methods are detailed in Section III. The main description of EMAPs is presented in Section IV. Section V reports classification results based on simulated hyperspectral data. We report the experimental results on two real hyperspectral data sets in Section VI. Section VII contains the conclusion of the presented work and its perspectives.

II. DT and Its Ensembles

Let $\{\mathbf{X}, \mathbf{Y}\} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$ be a set of labeled samples, where $\mathbf{x}_i \in \mathbb{R}^D$ is a pixel, and y_i contains the label information.[1] Let $\mathbb{F}$ be the set of D features. In order to construct an RS ensemble, we collect T classifiers based on the subsets of the original features. Each feature set in the ensemble defines a subspace of features of cardinality M, and a classifier is trained on this feature set using the whole training samples [19]. The final result is generated by a majority voting rule. Two parameters, namely, the ensemble size T and the cardinality of the feature set M, are required in the RS ensemble.

A. DT

DT is a nonparametric supervised learning algorithm used for classification and regression [51].

It is composed of a root node, a set of internal nodes (split), and a set of terminal nodes (leaves). In classification, a root node and each internal node have a splitting decision and splitting features associated with it. Class labels can be then assigned to the leaves. The creation of a DT from training samples involves two phases. At first, a splitting measure and a splitting attribute should be chosen. In the second phase, the records among the child nodes are split based on the decision made in the first phase. This process is applied recursively until a stopping criterion is met [52]. Then, the DT can be used to predict the class label of a new sample. The prediction process starts at the root, and a path to a leaf is traced by performing a splitting decision at each internal node. The class label attached to the leaf is then assigned to the new sample [52].

A critical component of the DT induction process is the selection of the split. Different algorithms use various metrics to split the nodes. The most widely used splitting criteria rely on the minimization of the Gini index of the splits [53].

B. DT Ensembles

1) RS With DT: The RS ensemble, which was introduced by Ho [19], was proposed for constructing multiple DTs. The objective of the RS ensemble is to sample a feature set into low-dimensionality subspaces from the whole original high-dimensional feature space, construct a classifier on each smaller subspace, and, finally, apply a majority voting rule for the final decision.

[1] y_i is different between DT and ELM. In DT, y_i is a scalar with classes of interest $\mathcal{Q} = \{1, \ldots, Q\}$, where Q is the total number of classes. In ELM, y_i is a vector of label in which the jth column is set to be 1 if the sample belongs to class j, whereas the other columns are set to be 0.

2) RF: RF, which was developed by Breiman [21], combines bagging [54] and random subspace [19] together to produce the DT ensemble. RF is a particular implementation of bagging in which each model is a random tree. A random tree is grown according to the CART algorithm with one exception: for each split, only a small subset of features of randomly selected splits is considered, and the best split is chosen from this subset. Since only a portion of the input features is used to split and no pruning on the tree is done, the computational complexity of RF is relatively light [21]. The computing time is approximately of $T\sqrt{M}n\log(n)$, where T, M, and n represent the number of classifiers, the number of features in a subset, and the number of training samples, respectively.

3) RoF: RoF has been a recently proposed ensemble method for building classifier ensembles using independent DTs built on a different set of extracted features [26]. The main heuristic of RoF is to apply feature extraction and to subsequently reconstruct a full different feature set for each classifier in the ensemble. To do this, the feature space is randomly split into K subsets, and each subset contains M features; then, PCA is applied to each K subset, and a new set of M linear extracted features in each subset is constructed by all principal components. Furthermore, new training data are formed by concatenating M linear extracted features in each subset.

An individual DT classifier is trained with this training data. A series of individual classifiers is generated by repeating the preceding steps several times. The final classification result is produced by integrating the results from individual classifiers using a majority voting rule. Different splits of the features will lead to different extracted features, thereby further increasing the diversity already introduced by the bootstrap sampling.

4) RoRF: RoRF is a variant of RoF, which uses RFs as the base classifiers instead of DTs [50]. This method was already evaluated on genomic and proteomic data sets [50], but it was not yet used for remote sensing image classification. The main training and prediction steps of RoRF are presented in Algorithm 1.

Algorithm 1 Rotation RF

Training phase

Input: $\{\mathbf{X}, \mathbf{Y}\} = \{\mathbf{x}_i, y_i\}_{i=1}^n$: training samples, T: number of classifier, K: number of subsets (M: number of features in each subset), L: base classifier. The ensemble $\mathcal{L} = \varnothing$. $\mathbb{F}$: Feature set

Output: The ensemble $\mathcal{L}$

```
for i = 1 : T do
    randomly split the features F into K subsets F_j^i
    for j = 1 : K do
        extract from X the new training set X_{i,j} with the corresponding features F_j^i
        generate a subset X̂_{i,j} by selecting with the bootstrap algorithm, the 75% of the initial training samples in X_{i,j}
        transform X̂_{i,j} to get the coefficients v_{i,j}^(1), ..., v_{i,j}^(M_k)
    end for
```

8: sparse matrix $\mathbf{R}_i$ is composed of the above coefficients

$$\mathbf{R}_i = \begin{bmatrix} v_{i,1}^{(1)}, \ldots, v_{i,1}^{(M_1)} & 0 & \cdots & 0 \\ 0 & v_{i,2}^{(1)}, \ldots, v_{i,2}^{(M_2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & v_{i,j}^{(1)}, \ldots, v_{i,j}^{(M_K)} \end{bmatrix}$$

9: with respect to the original feature set, rearrange $\mathbf{R}_i$ to $\mathbf{R}_i^a$

10: obtain the new training samples $\{\mathbf{X}\mathbf{R}_i^a, \mathbf{Y}\}$

11: build RF classifier L_i using $\{\mathbf{X}\mathbf{R}_i^a, \mathbf{Y}\}$

12: Add the classifier to the current ensemble, $\mathcal{L} = \mathcal{L} \cup L_i$.

13: **end for**

Prediction phase

Input: The ensemble $\mathcal{L} = \{L_i\}_i^T$. A new sample $\mathbf{x}^*$. Rotation matrix: $\mathbf{R}_i^a$.

Output: class label y^*

1: get the output ensemble with $\mathbf{x}^*\mathbf{R}_i^a$.

2: $y^* = \underset{q \in \{1,2,\ldots,Q\}}{\arg\max} \sum_{j:L_j(\mathbf{x}^*\mathbf{R}_i^a)=q}^{T} 1$

In the training phase, the feature space is first divided into K disjoint subspaces. PCA is performed on each subspace with the bootstrapped samples of 75% of the original training set. A transformed training set is generated by rotating with a sparse matrix $\mathbf{R}_i^a$ the original training set. An individual classifier is trained on this rotated training set. In the prediction phase, a new sample $\mathbf{x}^*$ is rotated by $\mathbf{R}_i^a$. Then, the transformed set, i.e., $\mathbf{x}^*\mathbf{R}_i^a$, is classified by the ensemble, and the class with the maximum number of votes is chosen as the final class. It is important to notice Step 5 in Algorithm 1, in which 75% of the original size of training samples are selected to avoid obtaining the same coefficients of the transformed components if the same features are selected, thus enhancing the diversity among the member classifiers.

RoRF can improve the performance of RF by introducing further diversity performing a feature extraction within the ensemble. The base classifiers in RoRF are more diverse and accurate with respect to RoF, and this could be beneficial for the ensemble.

III. ELM and Its Ensembles

ANN is another base learner used for the construction of classifier ensembles. However, the main drawback of conventional ANN is the high computation complexity and low efficiency. To address the shortcoming, ELM was proposed for the learning of generalized single hidden layer feedforward neural networks (SLFNs) without tuning the hidden layers [48], [49].

A. ELM

For generalized SLFNs, the output function of ELM is defined as

$$f(\mathbf{x}_i) = \sum_{j=1}^{\delta} \beta_j h_j(\mathbf{x}_i) = \mathbf{h}(\mathbf{x}_i)\boldsymbol{\beta} \tag{1}$$

where $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_\delta]^\top$ is the vector of weights between the hidden layer of δ nodes and the output node, and $\mathbf{h}(\mathbf{x}_i) = [h_1(\mathbf{x}_i), h_2(\mathbf{x}_i), \ldots, h_\delta(\mathbf{x}_i)]$ is the vector of hidden layer of $\mathbf{x}_i$. Specifically, $\mathbf{h}(\cdot)$ is the feature mapping from the D-dimensional input space to the δ-dimensional hidden layer feature space.

The standard SLFNs can approximate these n samples with zero error means that $\sum_i \|f(\mathbf{x}_i) - y_i\| = 0$. Thus, the n equations can be compactly written as

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{Y} \tag{2}$$

where $\mathbf{Y}$ is the target matrix, and $\mathbf{H}$ is the hidden layer output matrix, i.e.,

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}(\mathbf{x}_1) \\ \vdots \\ \mathbf{h}(\mathbf{x}_n) \end{bmatrix} = \begin{bmatrix} h_1(\mathbf{x}_1) & \cdots & h_\delta(\mathbf{x}_1) \\ \vdots & \vdots & \vdots \\ h_1(\mathbf{x}_n) & \cdots & h_\delta(\mathbf{x}_n) \end{bmatrix}. \tag{3}$$

The output weights in (2) are given by the following smallest norm least squares solution [49]:

$$\boldsymbol{\beta} = \mathbf{H}^+\mathbf{Y} \tag{4}$$

where $\mathbf{H}^+$ is the Moore–Penrose generalized inverse of the hidden layer output matrix $\mathbf{H}$.

In ELM, a feature mapping $\mathbf{H}$ from input space to a higher dimensional space is needed. The work of [55] and [56] demonstrated that almost all nonlinear piecewise continuous functions can be used as output functions of the hidden nodes. In this paper, the sigmoid function is adopted as the nonlinear piecewise continuous function, i.e.,

$$g(\omega, b, \mathbf{x}_i) = \frac{1}{1 + \exp\left(-(\omega \cdot \mathbf{x}_i + b)\right)} \tag{5}$$

where $\{\omega_j, b_j\}_{i=1}^{\delta}$ are randomly generated values that can define a continuous probability distribution (i.e., $\int g = 1$). Thus, $\mathbf{h}(\mathbf{x}_i)$ is defined based on the nonlinear piecewise continuous function $g(\omega_i, b_i)$, i.e.,

$$\mathbf{h}(\mathbf{x}_i) = [g(\omega_1, b_1, \mathbf{x}_i), \ldots, g(\omega_\delta, b_\delta, \mathbf{x}_i)]. \tag{6}$$

The training and prediction steps of ELM are listed in Algorithm 2.

Algorithm 2 Extreme Learning Machine

Training phase

Input: $\{\mathbf{X}, \mathbf{Y}\} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$: training samples, δ: number of nodes in a hidden layer. g: the sigmoid function.

Output: The output weight $\boldsymbol{\beta}$.

1: Randomly select the $\{\omega_1, \ldots, \omega_\delta\}$ and $\{b_1, \ldots, b_\delta\}$

2: For each training sample $\mathbf{x}_i$, calculate the output layer matrix: $\mathbf{h}(\mathbf{x}_i) = [g(\omega_1, b_1, \mathbf{x}_i), \ldots, g(\omega_\delta, b_\delta, \mathbf{x}_i)]$

3: Calculate the output weight: $\boldsymbol{\beta} = \mathbf{H}^+\mathbf{Y}$

Prediction phase

Input: A new sample $\mathbf{x}^*$. The output weight $\boldsymbol{\beta}$. The sigmoid function g. $\{\omega_1, \ldots, \omega_\delta\}$ and $\{b_1, \ldots, b_\delta\}$

Output: Class label of $\mathbf{x}^*$.

1: Calculate the output layer matrix: $\mathbf{h}(\mathbf{x}^*) = [g(\omega_1, b_1, \mathbf{x}^*), \ldots, g(\omega_\delta, b_\delta, \mathbf{x}^*)]$.
2: $y^* = \mathbf{h}(\mathbf{x}^*)\boldsymbol{\beta}$. Assign the number of column that gets the greatest value among the columns to the class label of $\mathbf{x}^*$.

Compared with conventional feedforward ANNs, ELM offers significant advantages such as the following: 1) fast learning speed; 2) no need to tune the parameters; 3) better generalization performance; and 4) ease of implementation [48], [49].

B. Proposed ELM Ensembles

ELM decreases the learning time dramatically with respect to a conventional ANN due to the random selection of weights and biases for hidden nodes [55], [56]. However, these parameters are not optimized. In this case, ELM will not be able to incorporate prior knowledge of the inputs; thus, the generalization error might increase. Consequently, we propose to construct an ensemble of several predictors on the training set using the RS method, in which the parameters in each predictor are randomly selected. In this paper, two implementations of the ELM ensembles, namely, RS based and rotation subspace based, are developed for hyperspectral image classification.

1) RS With ELM: Given a training set, the parameters of ELM (activation function and number of hidden nodes), the number of features in a subset M, and the number of classifiers T, the RS with ELM algorithm can be summarized by the following three steps (see Algorithm 3): 1) generate a subset of M features from the entire feature set for T times; 2) apply these features to ELM classifier and obtain T classification results; and 3) produce the final classification map by combining the T predictions using a majority voting rule.

Algorithm 3 Random Subspace With ELM

Training phase

Input: $\{\mathbf{X}, \mathbf{Y}\} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$: training samples. T: number of classifiers. L: Base classifier. $\mathcal{L} = \varnothing$: the ensemble. M: number of features in a subspace ($M < D$). $\mathbb{F}$: feature set.

Output: The ensemble $\mathcal{L}$.

1: **for** $i = 1$ to T **do**
2: Randomly selected from $\mathbb{F}$ without replacement to form a new training set composed of M features.
3: Train an ELM classifier L_i using a new training set.
4: Add the classifier to the current ensemble, $\mathcal{L} = \mathcal{L} \cup L_i$.
5: **end for**

Prediction phase

Input: The ensemble $\mathcal{L} = \{L_i\}_i^T$. A new sample $\mathbf{x}^*$.

Output: class label y^*

1: run each classifier in the ensemble using $\mathbf{x}^*$.
2: $y^* = \underset{q \in \{1,2,\ldots,Q\}}{\arg\max} \sum_{j: L_j(\mathbf{x}^*) = q}^{T} 1$

2) RoELM: The main steps of rotation subspace ELM (RoELM) can be summarized as follows.

- Divide the feature space into K disjoint subspaces.
- Perform PCA to each subspace with the bootstrapped samples of 75% of the original training set.
- The new training set, which is obtained by rotating the original training set, is treated as input to the individual classifier.
- The final result is generated by combining the individual classification results using a majority voting rule.

The main difference between RoELM and RoRF is that we use the ELM classifier instead of the RF classifier as base learner (see Step 8 in training phase of Algorithm 1). Diversity in RoELM is promoted in three aspects: 1) random selection of features; 2) feature extraction applied to the selected features using bootstrap sampling technique; and 3) random selection of parameters in each ELM classifier. When the number of training samples is less than the number of features, the covariance matrix will be singular and cannot be inverted. In order to avoid the problem of singularity of the covariance matrix, the value of $0.75 \times n$ should be larger than M in the ensembles of RoF, RoRF, and RoELM.

IV. EMAPs

Mathematical morphology is a powerful framework for the analysis of spatial information in remote sensing imagery [30], [35]. In particular, APs have been successfully applied to produce classification maps of remote sensing data [38], [39]. A sequence of attribute filters (AFs) are applied to a scalar image to obtain APs. AFs are connected operators, that is, they process a gray-level image by keeping or merging their connected components at different gray levels.

Respectively denoting with ϕ and γ an attribute thickening and thinning based on the arbitrary criterion P_λ, an AP of an image f is obtained by applying several attribute thickening and thinning operators, given a sequence of thresholds $\{\lambda_1, \lambda_2, \ldots, \lambda_\epsilon\}$ for the predicate P as follows [39]:

$$\mathrm{AP}(f) = \left\{\phi^{\lambda_\epsilon}(f), \phi^{\lambda_{\epsilon-1}}(f), \ldots, \phi^{\lambda_1}(f), f, \gamma^{\lambda_1}(f), \ldots, \gamma^{\lambda_{\epsilon-1}}(f), \gamma^{\lambda_\epsilon}(f)\right\}. \quad (7)$$

AP deals with only one spectral band. If we apply the full spectral bands of hyperspectral data to extract APs, the dimensionality of APs becomes extremely high. In order to address the problem, Dalla Mura *et al.* proposed to consider few of the first several principal components of the hyperspectral data [39]. However, any feature extraction and selection could be also used [42]. Thus, the expression of an EAP computed on the first C principal components from the original hyperspectral data [39] is given by

$$\mathrm{EAP} = \left\{\mathrm{AP}(\mathrm{PC}_1), \mathrm{AP}(\mathrm{PC}_2), \ldots, \mathrm{AP}(\mathrm{PC}_C)\right\}. \quad (8)$$

An EMAP is composed of m different EAPs based on different attributes $\{a_1, a_2, \ldots, a_m\}$, i.e.,

$$\mathrm{EMAP} = \left\{\mathrm{EAP}_{a_1}, \mathrm{EAP}'_{a_2}, \ldots, \mathrm{EAP}'_{a_m}\right\} \quad (9)$$

where $\mathrm{EAP}'_a = \mathrm{EAP}_a / \{\mathrm{PC}_1, \mathrm{PC}_2, \ldots, \mathrm{PC}_C\}$.

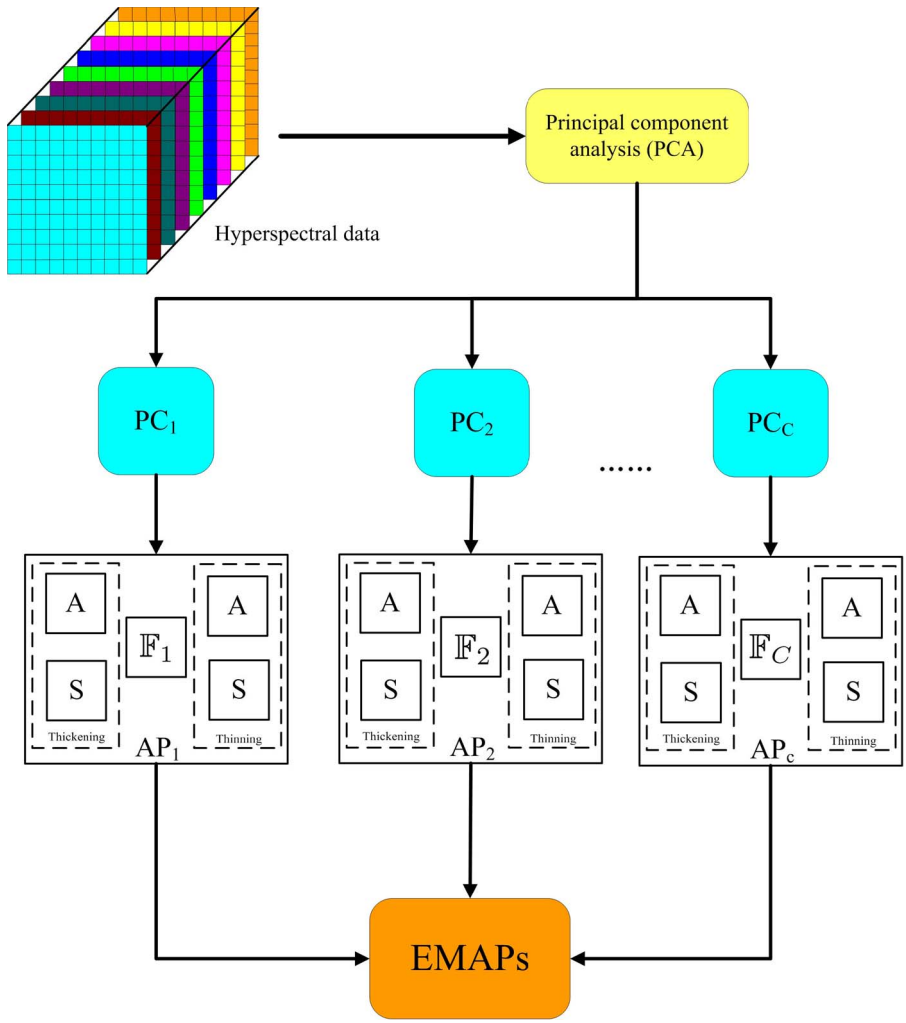


Fig. 1. Construction of EMAPs using the area (A) and standard (S) deviation attributes. First, PCA is performed on the original hyperspectral image, and the first features with cumulative eigenvalues over 99% are kept. Then, APs with attribute and standard deviation attributes are computed on the first features, and the output features are concatenated into a stacked vector to construct EMAPs.

Although a wide variety of attributes can be used to construct APs, only the area and standard deviation attributes are considered in this study. Fig. 1 presents the general steps of the construction of EMAPs using the area and standard deviation attributes. First, PCA is performed on the original hyperspectral image, and the first components with cumulative eigenvalues over 99% are retained. Then, APs with attribute and standard deviation attributes are computed on the first retained features, and the output features are concatenated into a stacked vector to construct an EMAP.

According to [43], λ_s is initialized to cover a reasonable amount of deviation in the individual feature, which is mathematically given by

$$\lambda_s(\mathbb{F}_i) = \frac{\mu_i}{100}\{\tau_{\min}, \tau_{\min} + \epsilon_s, \tau_{\min} + 2\epsilon_s, \ldots, \tau_{\max}\} \quad (10)$$

where $\mathbb{F}_i$ is the ith feature of the image, and μ_i is the mean value of the ith feature. The values of $\tau_{\min}$, $\tau_{\max}$, and ϵ_s are 2.5%, 27.5%, and 2.5%, respectively, which leads to 11 thinning and 11 thickening operations.

The construction of the attribute area is established in the following:

$$\lambda_a(\mathbb{F}_i) = \frac{100}{\nu}\{\alpha_{\min}, \alpha_{\min} + \epsilon_a, \alpha_{\min} + 2\epsilon_a, \ldots, \alpha_{\max}\} \quad (11)$$

where ν is the spatial resolution of the remote sensing image. The values of $\alpha_{\min}$, $\alpha_{\max}$, and ϵ_a are 1, 14, and 1. The EAP for the area attribute contains 14 thinning and 14 thickening operations for each feature.

V. Experiments With Simulated Data

Here, simple simulated hyperspectral data are used to evaluate the classification performance of the proposed methods, namely, RoRF ensemble, ELM ensembles, and spectral–spatial strategy.

A synthetic image is generated by a linear mixture model with $Q = 4$ spectra, i.e.,

$$\mathbf{x}_i = \sum_{q=1}^{Q} \mathbf{m}^q \mathbf{s}_i^q + \mathbf{n}_i \quad (12)$$

where $\mathbf{x}_i$ is the simulated mixed pixel. $\{\mathbf{m}^q\}$ are the spectral signatures obtained from the U.S. Geological Survey digital library.[2] The spatial information is generated by using a multilevel logistic (MLL) distribution with a value of smoothness parameter equal to 2. The simulated image is composed of 128×128 pixels with 224 spectral bands. Assume that $\mathbf{x}_i$ has class label $y_i = q_q$; then, we define $s_i^{q_q}$ as the abundance of the objective class and s_i^q as the abundance of the remaining features, which contribute to the mixed pixel, where s_i^q is constructed by the uniform distribution over the simplex. Here, we take $s_i^{q_q} = s$, $\sum_{q \in \mathcal{Q}, q \neq q_i} s_i^q = 1 - s$, and we use the same value of $s = 0.7$ for all pixels.

Furthermore, zero-mean Gaussian noise with variance $\sigma^2\mathbf{I}$, i.e., $\mathbf{n}_i \sim \mathcal{N}(0, \sigma^2\mathbf{I})$, is added to the simulated image. In particular, σ^2 is set to be 0.8 in order to make a very challenging classification problem. More details about how this data set is generated can be found in [57].

We conducted four different experiments with the simulated hyperspectral image in order to investigate several relevant aspects of our proposed framework. In all experiments, 40 samples per class (a total number of 160 samples) are selected as training set. In order to increase the statistical significance of the results, the reported means and standard deviations are obtained from ten Monte Carlo runs. Experiments for the simulated data consist of the following.

- In the first experiment, we evaluate the classification performance of RoRF with respect to other DT ensembles, including RSDT, RF, and RoF. The impact of the parameters, such as the number of trees in RF and RoRF and the number of features in a subset M, is analyzed.
- In the second experiment, we compare the ELM ensembles with a standard ELM classifier. The key parameters are also analyzed.
- In the third experiment, we compare the ELM ensembles with the DT ensembles.
- In the fourth experiment, we give the classification performance of the proposed spectral–spatial classification strategy.

In order to analyze the ensembles clearly, we present the following measures to evaluate the performance.

- Overall accuracy (OA) of the ensemble.
- Average of OAs (AOA) of the individual classifiers.
- Diversity among the individual classifiers within the ensemble. In this paper, we select *coincident failure diversity (CFD)* as a diversity measure [58].[3] Please refer to the details of CFD in [58]. Higher values of CFD means the stronger diversity within the ensemble.

A. Experiment 1

In this experiment, the number of base classifiers T and the number of features in a subset M are set to be 20 and 10, respectively. A nonparametric DT learning technique, i.e., CART, is used to construct the DT ensembles [59]. The impurity measure used in selecting the variables in CART is the Gini index. Table II shows OAs, AOAs of individual classifiers, and diversities obtained for different DT ensemble classifiers. Greater values of AOA and diversity usually lead to better performances of the ensemble. From Table II, RoRF gets the highest values of AOA and diversity, resulting in the best classification result. Then, we analyze the performance of the DT ensembles for different values of M with $T = 20$. Fig. 2 shows the OAs, AOAs, and diversities obtained by RSDT, RF, RoF, and RoRF as a function of M. The best performance was achieved by the proposed RoRF ensemble approach, which yields the best OA, AOA, and diversity with all different values of M. From the figure, we can observe that, when M becomes larger, all the DT ensembles tend to have better performances. Furthermore, we also studied the classification performances of RoRF with different numbers of trees in RF. The statistics are reported in Table III. OA, AOA, and diversity of RoRF increase as the number of trees increases. Notice that a good performance was achieved by RoRF with only ten DTs in RF, which gets an OA of 82.84%, even higher than the one of RoF with 20 trees. Therefore, in order to save time and memory, we can build the RoRF ensemble with a small number of trees in RF.

TABLE II
OAs (IN PERCENT), AOAs (IN PERCENT), AND DIVERSITIES OBTAINED FOR DIFFERENT DT ENSEMBLES WHEN APPLIED TO THE SIMULATED HYPERSPECTRAL DATA ($T = 20$ AND $M = 10$)

Classifiers	RSDT	RF	RoF	RoRF
OA	53.52±1.15	57.81±0.80	79.50±1.09	85.15±0.73
AOA	35.83±0.42	37.12±0.51	57.07±1.24	67.75±1.06
Diversity	0.38±0.0044	0.39±0.0053	0.61±0.0120	0.70±0.0089

B. Experiment 2

In this experiment, T, M, and δ are set to be 20, 10, and 256, respectively. The sigmoid function is selected as the functions of hidden nodes. OAs, AOAs, and diversities obtained by ELM and the proposed ELM ensemble classifiers are shown in Table IV. Note that the proposed RoELM ensemble classifier produced the best result. This is reasonable since RoELM introduces rotation strategy into the ensemble, which can both increase the accuracies of the member classifiers and the diversity within the ensemble. Fig. 3 plots the OAs, AOAs, and diversities of ELM ensembles with respect to the number of features in a subset (M). For these experiments, T and δ are fixed to be 20 and 256, respectively. Fig. 4 reports the OAs, AOAs, and diversities with different parameter δ for ELM ensembles. T and M are fixed to be 20 and 10, respectively. From the two figures, we have the following observations: 1) RoELM achieves the best results in all the cases, which demonstrates that RoELM is an effective ensemble method; 2) the effects of M are consistent with the preliminary test in Fig. 2. The higher the OAs, AOAs, and diversities of ELM ensembles, the higher the number of features in a subset (M); and 3) for these simulated data, ELM and its ensemble produce the best classification performances when $\delta = 64$. In this case, larger δ results in lower AOAs and diversities of RSELM and diversities

[2] https://engineering.purdue.edu/biehl/MultiSpec/.

[3] http://pages.bangor.ac.uk/mas00a/book_wiley/matlab_code/diversity/demo_diversity.html.

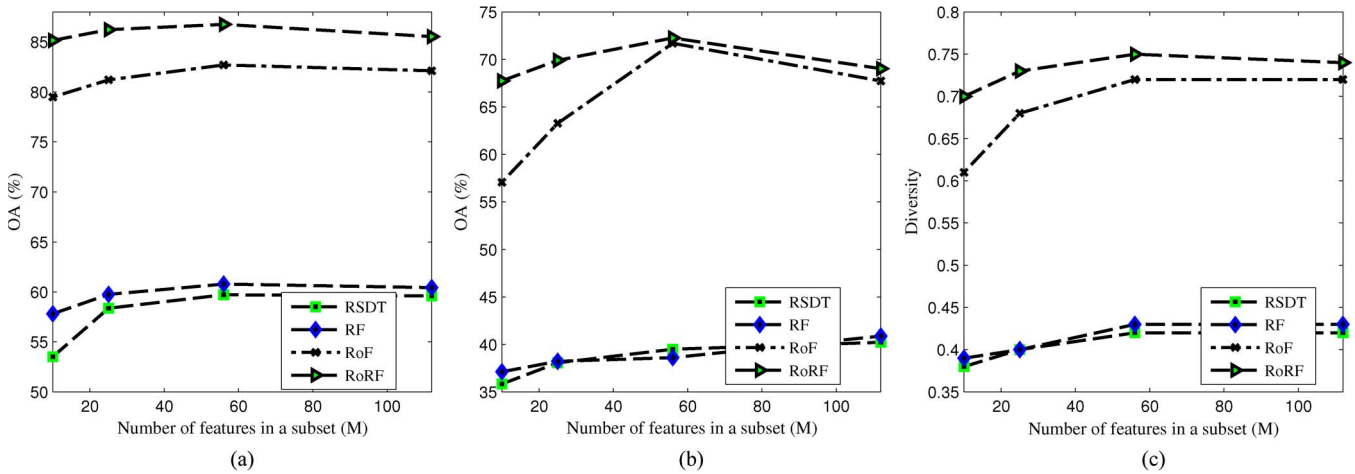


Fig. 2. (a) OAs obtained by RSDT, RF, RoF, and RoRF with different values of M. (b) AOAs obtained by RSDT, RF, RoF, and RoRF with different values of M. (c) Diversities obtained by RSDT, RF, RoF, and RoRF with different values of M ($T = 20$).

TABLE III
OAs (IN PERCENT), AOAs (IN PERCENT), AND DIVERSITIES OBTAINED FOR RoRF WITH DIFFERENT NUMBERS OF TREES IN RF ($T = 20$ AND $M = 10$)

Classifiers	Number of trees in RF					
	1		5		10	
	RF	RoRF	RF	RoRF	RF	RoRF
OA	39.31±1.64	72.80±1.59	44.36±1.58	79.85±0.95	52.25±0.91	82.84±0.87
AOA	N/A	46.54±0.80	36.97±1.24	57.25±0.82	37.64±0.96	60.60±0.89
Diversity	N/A	0.49±0.0084	0.46±0.0106	0.62±0.0087	0.42±0.0040	0.68±0.0088

TABLE IV
OAs (IN PERCENT), AOAs OF INDIVIDUAL CLASSIFIERS (IN PERCENT), AND DIVERSITIES OBTAINED FOR DIFFERENT ELM ENSEMBLES WHEN APPLIED TO THE SIMULATED HYPERSPECTRAL DATA ($M = 10$)

Classifiers	ELM	RSELM	RoELM
OA	53.52±1.15	67.75±1.00	79.73±1.67
AOA	N/A	43.80±0.54	58.98±1.19
Diversity	N/A	0.46±0.0057	0.62±0.0115

of RoELM. In contrast, AOAs of RoELM increase as δ increases. The reason is that a complex network (high value of δ) may overfit the training data of the member classifier in RSELM.

C. Experiment 3

Table V gives the OAs, AOAs, and diversities obtained for RSDT, RSELM, RoF, and RoELM ensembles. In order to make fair comparisons of DT ensembles and ELM ensembles, RSDT shares the same subspace with RSELM, and RoF shares the same rotations with RoELM. In these experiments, T, M, and δ are set to be 20, 10, and 256, respectively. Compared with DT ensembles (RSDT and RoF), the proposed ELM ensembles (RSELM and RoELM) adapt two strategies to improve the classification performance: one is to use ELM as the base learner to improve the individual classification accuracies, and the other is to utilize random selection of parameters in each ELM classifier to promote diversity within the ensemble. In addition, RoELM and RoF are superior to RSELM and RSDT. Sensitivity of the parameters to the classification performance is shown in Section V-B and C.

D. Experiment 4

Although RS ensemble can provide better classification results than single classifiers, the map [see Fig. 5(b)] looks still noisy due to the use only of spectral information. In order to further improve the results, the spatial information should be considered. In this paper, we propose to use EMAPs, which can offer the potential to model structural information in great details according to the use of different types of attributes. In this experiment, the proposed RoELM for spectral and spatial classification is considered. The first four components resulting from PCA (which comprise more than 99% of the data variance) were used, and the EMAPs consisted of 204 features. Fig. 5 shows the ground truth of the simulated image and classification maps of RoELM with spectral and spatial information, respectively. The OAs, AOAs, and diversities are also listed in Table VI. The classification accuracy of RoELM with EMAPs is significantly higher than the one of RoELM with spectral information, and the map produced by RoELM with EMAPs is smoother than the one produced by RoELM with spectral information when compared with the ground truth. In addition, we also studied the effect of M and δ on the classification performances of RoELM with EMAPs. For this data set, RoELM with EMAPs can obtain high OA with $M = 10$ (OA $= 99.30\%$). Larger values of M cannot improve the accuracies significantly. Similar to the results of RoELM with spectral information, classification accuracy of RoELM with EMAPs increases gradually in the beginning as δ increases, but decreases dramatically when δ reaches 64.

Summarizing, the experiments conducted with simulated data sets indicated that RS ensembles with or without EMAPs

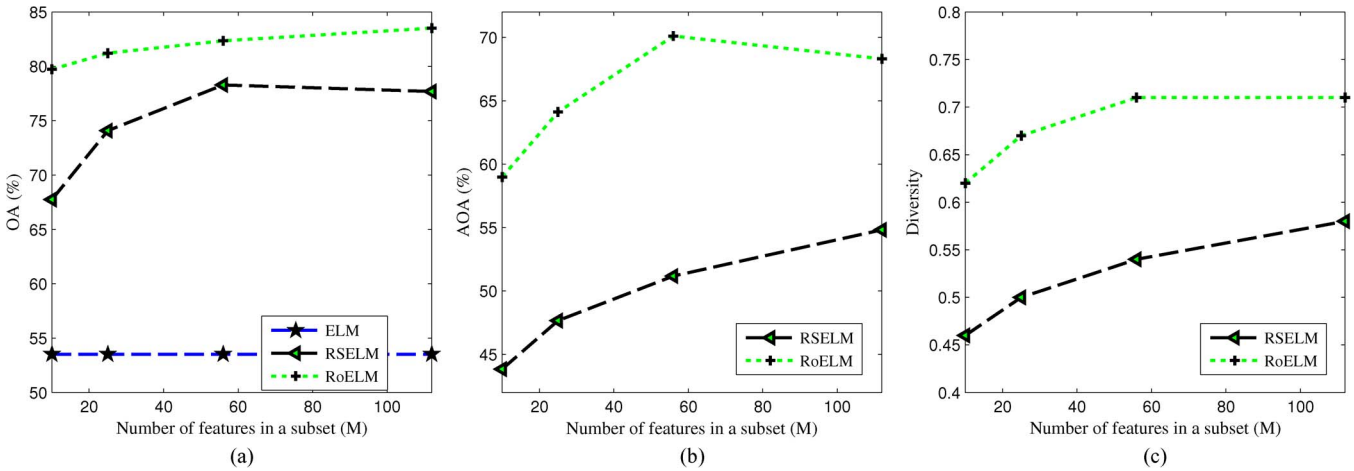


Fig. 3. (a) OAs obtained by ELM, RSELM, and RoELM with different values of M. (b) AOAs obtained by RSELM and RoELM with different values of M. (c) Diversities obtained by RSELM and RoELM with different values of M ($T = 20$ and $\delta = 256$).

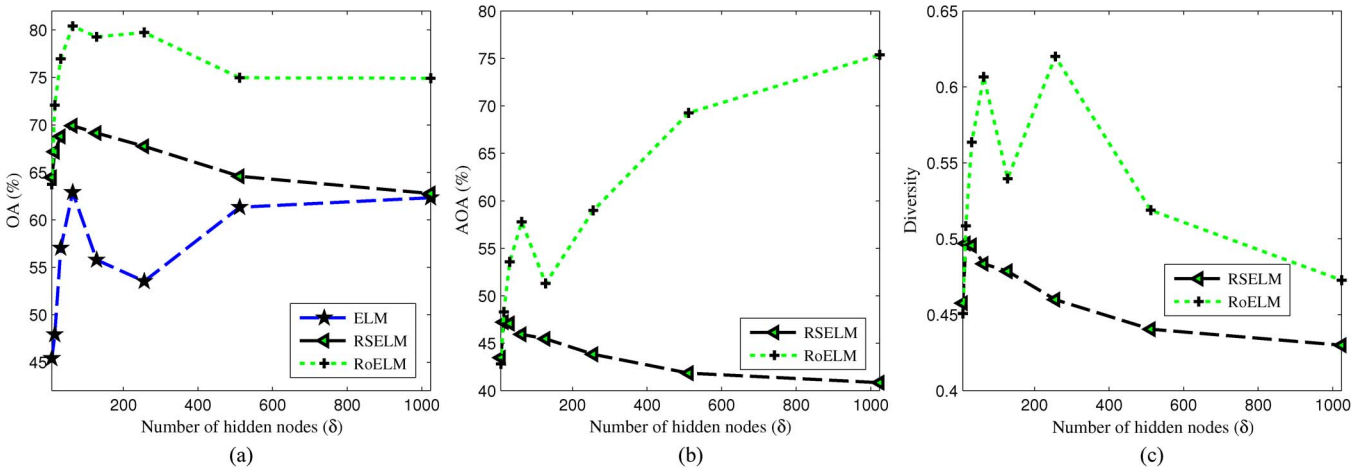


Fig. 4. (a) OAs obtained by ELM, RSELM, and RoELM with different values of δ. (b) AOAs obtained by RSELM and RoELM with different values of δ. (c) Diversities obtained by RSELM and RoELM with different values of δ ($T = 20$ and $M = 10$).

TABLE V
OAs (IN PERCENT), AOAs OF MEMBER CLASSIFIERS (IN PERCENT), AND DIVERSITIES OBTAINED FOR RSDT, RSELM, RoF, AND RoELM ENSEMBLES WHEN APPLIED TO THE SIMULATED HYPERSPECTRAL DATA ($T=20$, $M=10$, AND $\delta=256$)

Classifiers	RSDT	RSELM	RoF	RoELM
OA	53.52±1.15	67.75±1.00	79.50±1.09	79.73±1.67
AOA	35.83±0.42	43.80±0.54	57.07±1.24	58.98±1.19
Diversity	0.38±0.0044	0.46±0.0057	0.61±0.0120	0.62±0.0115

achieve better performance than single classifiers in highly mixed and noisy environments and with a limited number of training samples. In particular, the proposed RoRF and ELM ensembles show good performance in the classification of hyperspectral data.

However, the performances of the RS ensembles have been shown to be dependent on the setting of parameters M and δ. In order to get good characterization results, these parameters should be optimized. In particular, fine-tuning of these parameters can be done by manual manipulation.

Although it is encouraging to observe positive classification results using the proposed methods on simulated hyperspectral data, we have performed further analysis with real hyperspectral scenes and comparisons with other state-of-the-art methods in the next section in order to fully substantiate the proposed methods.

VI. EXPERIMENT WITH REAL HYPERSPECTRAL DATA SETS

Here, the proposed approaches are evaluated using two real hyperspectral data sets. Two individual classifiers, i.e., DT and ELM, and six ensemble methods, namely, RSDT, RF, RoF, RoRF, RSELM, and RoELM, are applied to classify the spectral information and EMAPs of hyperspectral data. CART is used to build DT ensembles, and the Gini index is adopted as the impurity measure used in selecting the variable [59]. The sigmoid function is selected as the functions of hidden nodes in ELM and its ensembles.

In this paper, only the first four components resulting from PCA (which comprise more than 99% of the data variance) were used, and the EMAPs consisted of 204 features. The reported results in this work are achieved by the mean of ten Monte Carlo runs.

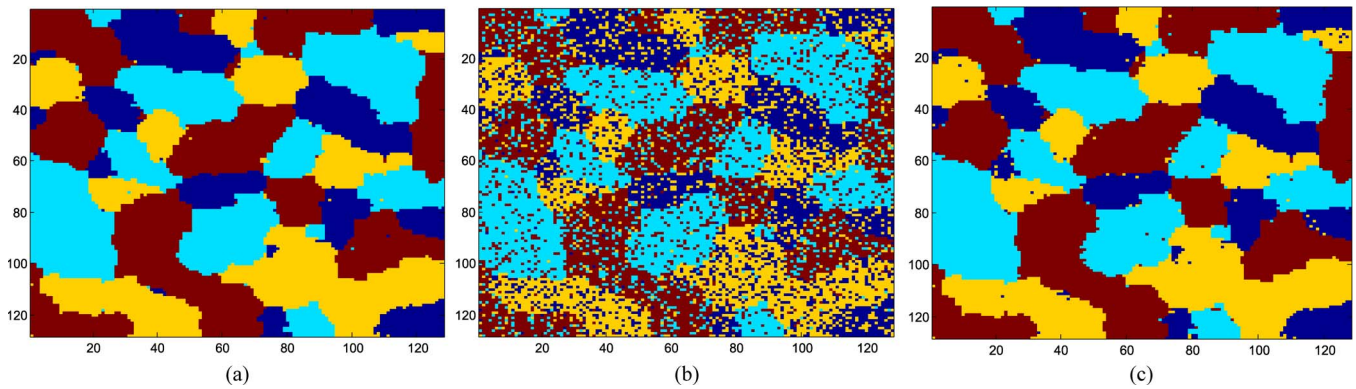


Fig. 5. (a) Image of class labels for a simulated image. (b) Classification map of RoELM with spectral information (OA = 79.67%). (c) Classification map of RoELM with EMAPs (OA = 99.42%).

TABLE VI
OAs (IN PERCENT), AOAs OF MEMBER CLASSIFIERS (IN PERCENT), AND DIVERSITIES OBTAINED FOR ROELM ENSEMBLES WHEN APPLIED TO SPECTRAL AND SPATIAL INFORMATION OF THE SIMULATED HYPERSPECTRAL DATA

Classifiers	RoELM with spectral information	RoELM with EMAPs
OA	79.73±1.67	99.30±0.08
AOA	58.98±1.19	97.12±0.13
Diversity	0.62±0.0115	0.74±0.0221

We used the following measures to evaluate the performances of different classification methods.

- OA: The percentage of correctly classified samples.
- Average accuracy (AA): Average percentage of correctly classified samples for individual class.
- Kappa coefficient (κ): The percentage agreement corrected by the level of agreement that could be expected to chance alone.
- Computation time: All methods were implemented in MATLAB on a computer having Intel Xeon 2 CPU, 2.8 GHz, and 12 GB of memory. RF implementation was downloaded from the website http://code.google.com/p/randomforest-matlab/. The source code of ELM can be assessed from the website http://www.ntu.edu.sg/home/egbhuang/elm_codes.html.

A. Hyperspectral Data Sets

Two hyperspectral remote sensing images are used to assess the performance of the proposed methods.

1) Indian Pines AVIRIS Image: The first hyperspectral image is recorded by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over the Indian Pines in Northwestern Indiana, USA. This scene, which consists of 220 spectral bands in the wavelength range from 0.4 to 2.5 μm with a spectral resolution of 10 nm, is composed of 145 × 145 pixels, and the spatial resolution is 20 m/pixel. Fig. 6 shows the three-band color composite image and the ground truth of AVIRIS hyperspectral data. Table VII gives the class name and the number of ground truth of the AVIRIS hyperspectral data.

2) University of Pavia ROSIS Image: The second experiment was carried out on the University of Pavia image of an urban area, which was acquired by the Reflective Optics Spectrographic Imaging System (ROSIS)-03 optical airborne sensor. Nine land cover classes were considered for classification. The original image is composed of 610 × 340 pixels, with a spatial resolution of 1.3 m/pixel and 115 spectral bands. In this paper, 12 noisy channels were removed, and the remaining 103 spectral bands are used for the investigation. Fig. 7 shows the three-band color composite image and the reference map of the University of Pavia data. Class name and the number of training and test samples of the ROSIS image are presented in Table VIII.

B. Results of the Indian Pines AVIRIS Image

Tables IX and X present the classification results obtained for the individual classifiers and RS ensemble methods using different numbers of training samples when the spectral and EMAPs are used as input, respectively. AAs for each classifier are also given in parentheses. According to the studies in [27] and [60], the parameters used for each ensemble classifier are shown in Table XI.[4] As shown in Tables IX and X, the RS ensemble methods exhibit the potential to improve the classification performance by using both spectral and spatial information. The proposed RoELM outperforms ELM, RSELM, and the other DT ensemble in terms of achieving higher classification accuracies in all cases. With the help of promoting diversity using feature extraction approaches, rotation subspace classifiers, namely, RoF, RoRF, and RoELM, are superior to the ensemble classifiers RF, RSDT, and RSELM.

In order to show the performance of the RS ensemble methods under different training conditions and scenarios, in the second experiment, we evaluated the classification accuracies of the RS ensemble approaches using a fixed number of training samples in which 10% of the labeled samples per class have been used for training (a total number of 1036 samples) and the remaining labeled samples are used for testing. Tables XII and XIII provide the OAs, AAs, κ, and class-specific accuracies obtained from the individual and ensemble classifiers using

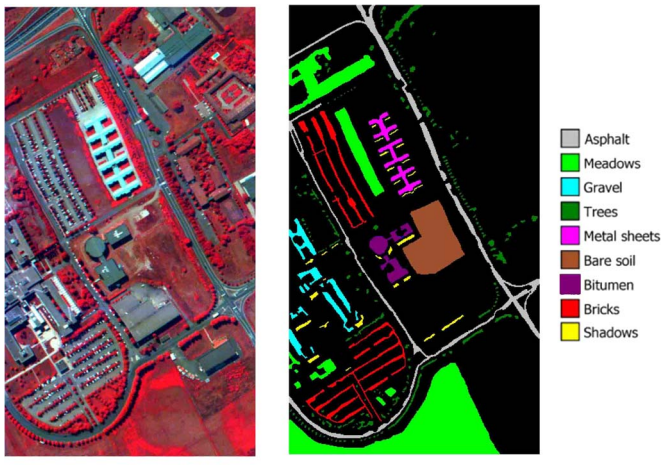
[4]For five samples per class, M is set to be 55 in RoF, RoRF, and RoELM ensembles with spectral information.

(a) (b)

Fig. 6. (a) Three-band color composite of AVIRIS image. (b) Ground truth.

TABLE VII
INDIAN PINES AVIRIS IMAGE: CLASS NAME AND NUMBER OF SAMPLES IN GROUND TRUTH

Class		Number
Number	Name	Ground Truth
1	Alfalfa	54
2	Corn-no till	1434
3	Corn-min till	834
4	Bldg-Grass-Tree-Drives	234
5	Grass/pasture	497
6	Grass/trees	747
7	Grass/pasture-mowed	26
8	Corn	489
9	Oats	20
10	Soybeans-no till	968
11	Soybeans-min till	2468
12	Soybeans-clean till	614
13	Wheat	212
14	Woods	1294
15	Hay-windrowed	380
16	Stone-steel towers	95



(a) (b)

Fig. 7. (a) Three-band color composite image of AVIRIS data. (b) Reference map.

spectral information and EMAPs, respectively. The processing times in seconds are also included for reference.

It can be seen from the results in Tables XII and XIII that the performance of ELM is superior to CART in both testing accuracy and learning time. When the spectral information is treated as input, RoELM and RoF share the top position. The OAs (AAs) of the two methods are 84.70% (75.33%) and 81.31% (73.62%), higher than those of the other methods. Class 9 produces bad results in all classifiers; the reason may be that there is insufficient information provided by class 9 using only two samples in the training. Compared with the results reported in Table XII, the classification accuracies in Table XIII involving the spatial information are much better than those obtained only with the spectral information, demonstrating that EMAP can accurately model spatial–contextual information in all cases. For the EMAPs as the input for this scene, RF and RSELM are slightly better than RoF and RoELM. Among them, RoRF yields the highest OA, AA, and κ. Feature extraction techniques in the processing of RoF, RoRF, and RoELM classifiers will lead the longer computation time than those of RSDT, RF, and RSELM. The computational complexity of the ELM ensemble is lower than that of the DT ensemble. The computation time of the RF ensemble is extremely low (less than 1 s).

TABLE VIII
UNIVERSITY OF PAVIA ROSIS IMAGE: CLASS NAME AND NUMBER OF TRAINING AND TEST SAMPLES

Class		Number of samples	
Number	Name	Train	Test
1	Bricks	524	3682
2	Shadows	514	947
3	Metal Sheets	375	1345
4	Bare Soil	540	5029
5	Trees	231	3064
6	Meadows	532	18649
7	Gravel	265	2099
8	Asphalt	548	6631
9	Bitumen	392	1330

Fig. 8 presents the classification maps (one of the ten Monte Carlo runs) obtained for the individual and ensemble learning methods with 10% labeled sample as the training samples in Tables XII and XIII. As shown in the two figures, RS ensemble can improve the classification performance and reduce the classification noise. The classification methods based on EMAPs spatial features result in classification maps with more homogeneous regions when compared with the classification result using spectral information.

TABLE IX
OAs AND AAs (IN PARENTHESES) OBTAINED FOR DIFFERENT CLASSIFICATION ALGORITHMS USING DIFFERENT SIZES OF TRAINING SET WHEN APPLIED TO THE SPECTRAL INFORMATION OF THE INDIAN PINES AVIRIS HYPERSPECTRAL DATA

Samples per class	DT	RSDT	RF	RoF	RoRF	ELM	RSELM	RoELM
5	29.64±3.61(39.62)	36.41±3.49(47.17)	42.87±3.65(53.79)	47.79±3.23(61.05)	51.14±3.22(62.66)	51.15±2.58(65.8)	55.39±3.08(69.69)	**58.72±2.06(71.87)**
10	38.85±4.03(49.87)	46.57±2.48(58.33)	49.89±3.16(60.86)	57.33±2.26(69.95)	58.39±1.78(69.58)	57.17±1.92(71.06)	65.11±1.83(76.35)	**69.84±1.27(79.67)**
15	40.43±2.42(51.79)	49.62±1.18(61.91)	51.13±1.52(63.73)	63.03±1.65(73.79)	61.49±1.71(73.01)	58.51±1.80(71.99)	68.69±1.59(79.7)	**72.93±1.07(83.43)**
20	43.13±2.35(55.15)	53.82±2.07(65.78)	55.52±2.50(66.92)	68.98±2.71(79.89)	67.56±2.55(78.54)	57.68±2.02(70.75)	70.73±1.18(80.36)	**75.95±0.82(85.22)**
25	46.59±1.32(57.54)	55.71±1.18(66.78)	57.23±1.56(68.52)	71.81±1.80(81.19)	70.17±1.46(79.43)	62.03±1.58(76.34)	72.67±0.97(83.71)	**77.13±0.94(86.51)**
30	47.49±1.35(58.24)	58.23±2.36(67.98)	60.07±2.10(70.47)	72.65±2.27(81.36)	70.79±1.77(80.03)	61.88±1.10(73.96)	75.61±1.11(85.49)	**78.24±0.67(86.87)**
35	47.90±2.56(57.75)	59.82±1.31(69.12)	61.48±1.59(71.10)	74.36±0.58(82.54)	72.78±1.17(81.23)	66.73±1.62(76.98)	74.86±1.53(84.31)	**78.89±1.07(87.17)**
40	49.05±2.06(58.33)	60.85±1.27(70.10)	62.66±1.52(71.95)	74.46±1.19(82.97)	73.45±1.53(81.64)	66.44±1.13(77.79)	75.46±0.80(85.05)	**80.08±0.5(88.24)**
45	50.42±2.14(60.79)	62.22±1.24(72.08)	63.68±0.93(73.25)	76.58±1.26(84.90)	74.81±1.37(83.37)	67.46±1.05(77.31)	77.43±0.31(86.64)	**80.34±0.25(88.08)**
50	50.51±2.63(59.48)	62.63±0.99(71.60)	64.20±0.58(77.51)	75.96±1.06(84.39)	74.56±1.26(82.21)	67.65±1.69(77.11)	77.85±1.09(87.57)	**81.19±0.7(89.00)**

TABLE X
OAs AND AAs (IN PARENTHESES) OBTAINED FOR DIFFERENT CLASSIFICATION ALGORITHMS USING DIFFERENT SIZES OF TRAINING SET WHEN APPLIED TO THE EMAPs OF THE INDIAN PINES AVIRIS HYPERSPECTRAL DATA

Samples per class	DT	RSDT	RF	RoF	RoRF	ELM	RSELM	RoELM
5	55.07±6.68(65.95)	57.48±5.51(70.24)	65.69±5.58(77.51)	66.22±4.41(76.91)	70.81±4.59(81.64)	73.24±5.28(81.44)	74.24±5.38(81.55)	**75.97±4.01(83.59)**
10	70.22±5.01(80.19)	73.84±4.59(82.36)	77.21±3.43(85.77)	78.29±2.94(76.91)	80.98±2.45(88.31)	82.47±2.71(87.17)	83.36±2.33(87.69)	**85.19±2.32(89.43)**
15	76.04±1.12(83.34)	80.34±2.28(85.96)	83.18±1.78(89.26)	83.26±1.9(88.25)	86.01±1.65(90.78)	83.92±2.94(88.02)	85.76±2.67(89.25)	**87.45±2.05(91.4)**
20	80.82±2.32(85.98)	82.69±1.75(87.84)	84.46±1.80(89.44)	85.36±1.53(89.29)	87.41±1.45(91.55)	83.98±2.41(87.31)	87.67±1.49(90.15)	**89.35±1.09(92.19)**
25	81.62±2.24(87.64)	84.37±2.14(89.95)	87.54±1.34(92.1)	87.92±1.31(92.10)	89.30±1.83(93.48)	87.02±2.60(90.7)	88.37±1.69(91.61)	**90.70±1.33(94.02)**
30	84.17±2.23(88.38)	87.07±2.82(90.27)	88.66±1.06(92.94)	89.32±1.67(92.60)	91.05±1.60(94.46)	89.65±1.78(92.28)	90.73±1.60(92.82)	**92.14±1.45(94.71)**
35	85.13±3.11(89.15)	87.51±2.25(91.72)	88.93±1.33(92.66)	89.84±1.06(93.07)	91.29±0.74(94.00)	89.48±1.70(91.73)	90.76±1.20(92.45)	**92.77±0.98(94.99)**
40	85.14±1.71(88.52)	87.72±1.82(90.91)	89.85±1.23(93.29)	90.44±0.98(92.98)	91.99±1.20(94.45)	88.76±1.17(91.03)	91.01±1.25(92.54)	**92.99±0.64(94.8)**
45	85.68±1.48(89.31)	89.12±1.91(92.42)	90.81±0.97(93.69)	91.34±0.96(93.87)	92.71±0.93(95.16)	91.4±1.24(93.73)	93.71±0.65(95.41)	**93.97±0.60(95.61)**
50	87.08±1.53(90.43)	90.27±1.21(92.82)	91.61±0.79(94.01)	92.20±0.85(94.59)	93.31±0.42(95.33)	91.32±0.86(93.35)	94.29±0.43(95.31)	**94.53±0.41(95.25)**

TABLE XI
PARAMETERS USED FOR ELM AND RS ENSEMBLE CLASSIFIERS (INDIAN PINES AVIRIS IMAGE)

Features	Methods	T	M	δ	Features	Methods	T	M	δ
	RSDT	20	110	–		RSDT	20	102	–
	RF	20	15	–		RF	20	15	–
	RoF	20	110	–		RoF	20	3	–
Spectral	RoRF	20	110	–	EMAPs	RoRF	20	3	–
	ELM	–	–	256		ELM	–	–	256
	RSELM	20	110	256		RSELM	20	102	256
	RoELM	20	110	256		RoELM	20	3	256

TABLE XII
CLASSIFICATION ACCURACIES OBTAINED FOR DIFFERENT CLASSIFICATION ALGORITHMS USING 10% OF THE NUMBER OF SAMPLES IN GROUND TRUTH AS TRAINING SAMPLES WHEN APPLIED TO THE SPECTRAL INFORMATION OF THE INDIAN PINES AVIRIS HYPERSPECTRAL DATA

Class	DT	RSDT	RF	RoF	RoRF	ELM	RSELM	RoELM
1	38.16	28.77	15.71	53.06	29.80	15.11	25.71	**49.39**
2	50.76	65.79	61.82	79.86	74.38	73.01	79.95	**83.13**
3	45.09	53.86	50.55	69.31	62.88	59.01	61.89	**71.85**
4	26.87	34.69	30.76	**63.32**	50.76	36.59	54.17	62.09
5	67.63	79.04	79.98	88.37	83.06	90.56	**93.17**	91.48
6	78.23	92.14	92.37	94.72	93.99	95.55	**97.49**	94.64
7	20.43	16.09	14.78	**59.13**	50.87	3.04	8.02	46.52
8	85.93	92.77	96.16	97.73	98.20	99.43	**99.57**	98.32
9	1.67	0.56	6.11	0	2.78	4.44	12.22	**37.77**
10	47.01	63.85	61.76	**77.19**	76.87	64.02	68.56	75.67
11	62.02	80.56	83.88	87.35	**90.53**	80.77	87.72	87.87
12	29.73	42.28	47.25	69.58	68.72	67.09	75.51	**82.64**
13	80.94	90.94	92.93	97.91	97.07	**99.63**	99.58	98.53
14	89.57	93.67	94.13	96.22	96.36	95.55	96.94	**97.36**
15	35.96	39.77	39.33	55.85	45.09	59.97	**60.26**	54.77
16	54.59	71.88	81.06	88.35	**89.53**	46.12	70.59	73.29
OA	59.77	72.53	72.84	83.14	81.31	77.46	82.38	**84.70**
AA	50.79	57.98	59.29	73.62	69.12	61.89	68.23	**75.33**
κ	54.13	68.44	68.70	80.70	78.49	74.11	79.74	**82.44**
Time(s)	1.49	9.25	0.85	26.95	18.11	0.22	6.18	14.63

More classification results for the Indian Pines AVIRIS image based on EMAPs and other spatial–contextual information can be found in [34] and [45]–[47]. The accuracies in the previous studies are not directly compared with those given in this paper because different experimental settings (number of features, training and testing samples) are used in these studies. However, it can be concluded that RS ensembles with EMAPs perform well compared with other previously proposed classification approaches for hyperspectral data.

TABLE XIII
CLASSIFICATION ACCURACIES OBTAINED FOR DIFFERENT CLASSIFICATION ALGORITHMS USING 10% OF THE NUMBER OF SAMPLES IN GROUND TRUTH AS TRAINING SAMPLES WHEN APPLIED TO THE EMAPs OF THE INDIAN PINES AVIRIS HYPERSPECTRAL DATA

Class	DT	RSDT	RF	RoF	RoRF	ELM	RSELM	RoELM
1	82.65	83.67	87.14	87.35	87.14	74.08	86.94	**87.76**
2	86.14	91.01	91.00	91.11	**91.55**	90.21	90.12	90.33
3	92.65	95.5	95.31	95.63	96.51	97.48	98.75	**98.95**
4	74.55	87.11	89.15	87.57	91.66	88.34	94.27	**95.02**
5	89.53	92.37	92.51	93.31	93.20	91.28	**94.63**	94.36
6	94.15	95.33	97.22	96.95	98.07	97.75	99.12	**99.32**
7	23.48	23.04	73.91	40.43	84.78	88.28	**96.09**	**96.09**
8	**100**	99.77	99.77	99.75	99.80	96.23	99.39	99.55
9	69.44	61.11	92.77	62.78	**98.33**	80.56	89.44	95.56
10	84.43	86.89	88.43	87.50	89.06	92.61	90.55	91.56
11	94.67	96.18	97.94	96.74	98.55	96.53	98.49	**98.66**
12	85.14	90.29	92.28	90.22	**93.06**	86.20	89.19	89.17
13	99.16	98.84	99.11	99.42	**99.53**	98.95	99.48	99.48
14	98.57	99.22	99.25	99.23	99.24	96.29	99.16	**99.42**
15	93.27	96.40	97.63	96.35	**98.63**	75.38	92.40	94.36
16	96.12	97.41	98.00	97.65	**98.12**	0.71	50.51	32.71
OA	91.57	94.05	95.17	94.56	**95.83**	92.59	95.46	95.40
AA	85.23	87.13	93.21	90.00	**94.83**	84.43	92.32	91.39
κ	90.41	93.22	94.36	93.80	**95.24**	91.64	94.82	94.73
Time(s)	0.77	4.55	0.63	13.99	14.07	0.21	4.08	13.59

C. Results of the University of Pavia ROSIS Image

RS ensembles both with spectral and spatial information are performed on the University of Pavia ROSIS image. For all the ensemble classifiers, the number of classifiers (iterations) is fixed to be 20. Following the studies in [27] and [60], the number of features in a subset (M) in each ensemble classifier is set as follows. For the RF algorithm, the number of features in a

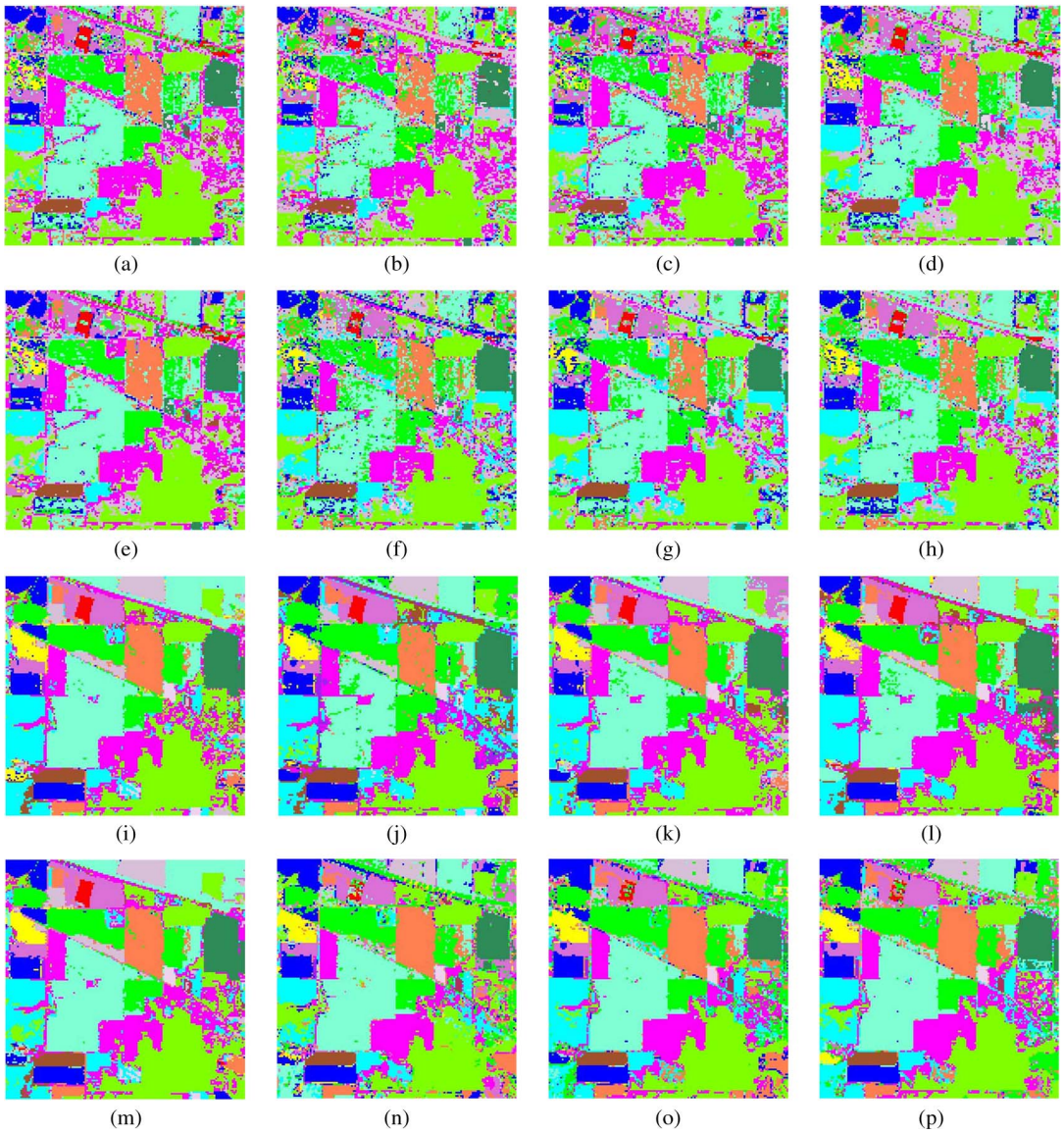


Fig. 8. Classification results of the Indian Pines AVIRIS image (only one Monte Carlo run). OAs of the classifiers are also given. (a) DT (59.69%). (b) RSDT (72.1%). (c) RF (73.23%). (d) RoF (83.40%). (e) RoRF (81.19%). (f) ELM (77.15%). (g) RSELM (82.23%). (h) RoELM (84.13%). (i) DT_EMAP (91.46%). (j) RSDT_EMAP (94.06%). (k) RF_EMAP (95.01%). (l) RoF_EMAP (94.55%). (m) RoRF_EMAP (96.06%). (n) ELM_EMAP (93.37%). (o) RSELM_EMAP (95.59%). (p) RoELM_EMAP (95.62%).

TABLE XIV
CLASSIFICATION ACCURACIES OBTAINED FOR DIFFERENT CLASSIFICATION ALGORITHMS USING THE ENTIRE TRAINING SET WHEN APPLIED TO THE SPECTRAL INFORMATION OF THE UNIVERSITY OF PAVIA ROSIS IMAGE

Class	DT	RSDT	RF	RoF	RoRF	ELM	RSELM	RoELM
1	83.46	91.68	90.17	92.55	**93.29**	90.7	95.05	91.76
2	92.93	97.42	97.44	98.30	99.60	99.65	99.69	**99.89**
3	96.95	98.99	98.82	99.58	**99.55**	85.64	98.95	99.85
4	76.29	81.56	77.80	95.60	95.55	94.92	96.14	**97.62**
5	97.75	98.67	98.58	95.62	**98.79**	96.68	97.11	95.33
6	52.35	53.12	56.10	**74.61**	65.38	58.76	64.32	69.19
7	54.79	51.32	53.79	58.49	57.54	**70.18**	68.06	63.43
8	71.93	79.60	80.07	84.55	**85.34**	77.21	80.50	76.02
9	76.62	83.68	84.63	89.93	90.39	88.01	**91.08**	90.90
OA	67.30	70.44	71.37	**82.66**	79.04	74.56	78.45	79.44
AA	78.11	81.78	81.93	87.69	87.27	84.64	**87.88**	87.11
κ	60.15	63.90	64.79	**78.09**	73.98	68.75	72.84	74.25
Time(s)	1.98	20.50	2.33	44.74	53.41	1.56	34.65	51.45

TABLE XV
CLASSIFICATION ACCURACIES OBTAINED FOR DIFFERENT CLASSIFICATION ALGORITHMS USING THE ENTIRE TRAINING SET WHEN APPLIED TO THE EMAPS OF THE UNIVERSITY OF PAVIA ROSIS IMAGE

Class	DT	RSDT	RF	RoF	RoRF	ELM	RSELM	RoELM
1	98.02	98.95	98.94	98.61	99.16	98.96	99.51	**99.58**
2	85.96	92.47	97.33	97.00	**99.32**	98.37	99.31	98.38
3	99.55	99.58	99.62	99.62	99.62	96.51	**99.67**	99.56
4	98.95	96.55	96.34	99.39	97.35	97.61	**99.96**	99.90
5	89.69	97.10	99.12	94.55	**99.23**	94.54	98.52	97.45
6	90.85	91.66	97.28	93.65	97.42	96.42	98.35	**98.55**
7	67.13	80.63	73.05	85.45	75.15	87.88	98.08	**99.38**
8	91.34	94.26	95.14	93.52	95.32	97.16	**97.69**	97.54
9	99.32	100.00	100.00	100.00	100.00	99.92	**99.93**	99.92
OA	91.67	93.61	96.08	94.85	96.47	96.49	98.67	**98.69**
AA	91.20	94.58	95.20	95.75	95.84	96.37	**99.00**	98.92
κ	89.14	91.71	94.83	93.26	95.34	95.37	98.21	**98.25**
Time(s)	1.38	29.72	2.82	37.93	70.24	1.83	37.66	71.59

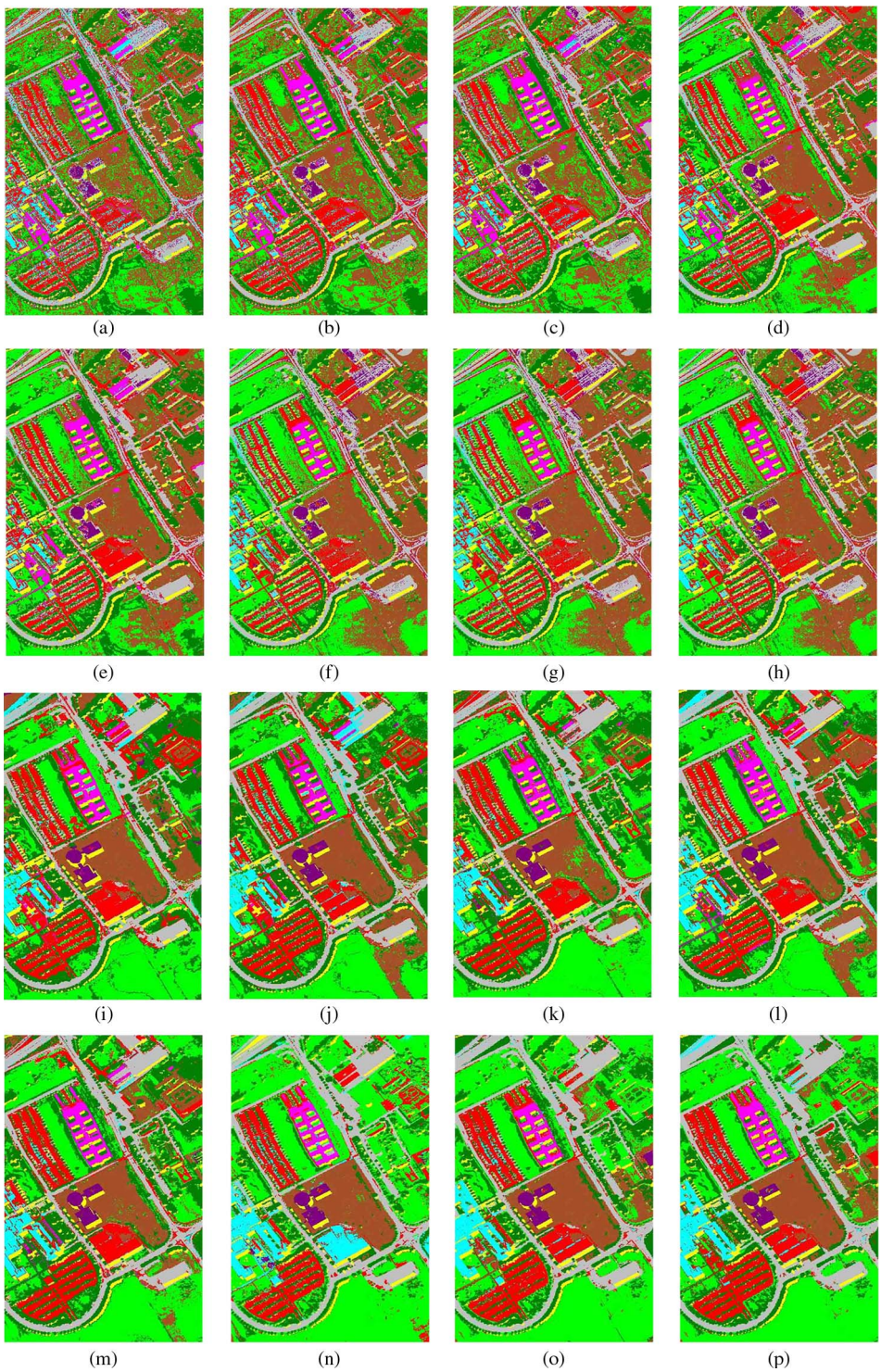


Fig. 9. Classification results of the University of Pavia ROSIS image (only one Monte Carlo run). OAs for each classifier are given. (a) DT (67.30%). (b) RSDT (70.58%). (c) RF (71.12%). (d) RoF (83.61%). (e) RoRF (79.65%). (f) ELM (74.43%). (g) RSELM (78.15%). (h) RoELM (79.95%). (i) DT_EMAP (91.67%). (j) RSDT_EMAP (93.79%). (k) RF_EMAP (96.23%). (l) RoF_EMAP (95.74%). (m) RoRF_EMAP (96.73%). (n) ELM_EMAP (96.52%). (o) RSELM_EMAP (98.74%). (p) RoELM_EMAP (99.11%).

subset is set to be the default value $\sqrt{N}$ of the software package (10 for this scene). For the RSDT and RSELM approaches (spectral and spatial information), M is set to 52 and 102, respectively. The numbers of features in a subset for RoF, RoRF, and RoELM used for spectral and spatial information are set to 10 and 3, respectively. The number of hidden nodes in ELM and its ensemble is fixed to be 128.

Table XIV gives the OA, the AA, and the class-specific accuracy obtained for different classification algorithms using the entire training set when applied to the spectral information of the University of Pavia ROSIS image. The computational times are also given in this table. From this table, it is clear that RoF provided the best results in terms of global and individual class accuracies, followed by RoRF and RoELM. In order to enhance

TABLE XVI
CLASSIFICATION ACCURACIES OBTAINED FROM THE PROPOSED METHODS (ROELM_EMAPS AND RSELM_EMAPS) IN COMPARISON WITH OTHER SPATIAL–SPECTRAL CLASSIFIERS FOR THE UNIVERSITY OF PAVIA ROSIS IMAGE

Classifier	SVM+Clustering [4]	MLRsubMLL [57]	GCK [45]	Mixed_lasso_3D-DWT [34]	RSELM_EMAPs	RoELM_EMAPs
OA	94.68	94.10	98.09	98.15	98.67	**98.69**
AA	95.21	93.45	97.76	97.56	**99.00**	98.92
κ	92.92	92.24	97.46	97.48	98.21	**98.25**

the classification results, the RS ensemble with EMAPs is further applied to classify hyperspectral data, and the global and class-specific accuracies are reported in Table XV.

It is shown in Table XV that the classification results with EMAPs significantly outperformed those only considering spectral information. All the RS ensembles yield the highest precision results. The proposed RSELM and RoELM outperform ELM, RoRF, and the other ensemble methods in terms of achieving higher global and class-specific accuracies. Rotation-subspace-based classifiers (RoRF, RoF, and RoELM) generate more accurate results than RF, RSDT, and RSELM because they introduced more diversity within the ensemble. Concerning the computational load, different observations can be made as in the former experiments. The computational cost of ELM and its ensembles is higher than those of DT and its ensembles, because of the large size of the data set. The spectral–spatial methods are less computationally efficient than the spectral-based methods due to the higher dimensionality of input features, but provide, in turn, higher accuracies. For illustrative purpose, Fig. 9 provides the classification maps of the individual and ensemble classifiers (one of ten Monte Carlo runs). Compared with the results using only spectral information presented in Fig. 9(a)–(f), the maps involving spatial information [see Fig. 9(g)–(p)] generate more homogeneous areas (particularly for the Class *Meadows* located at the lower left area) and reduce the classification noise.

In addition, Table XVI presents the comparisons of RSELM_EMAPs and RoELM_EMAPS against other state-of-the-art spectral–spatial classification methods, such as SVM+Clustering [4], MLRsubMLL [57], generalized composite kernels (GCK) [45], and Mixed_lasso with 3-D discrete wavelet transform (3D-DWT) features [34]. The SVM+Clustering approach combines the results of a pixelwise SVM classification and the segmentation map obtained by partitional clustering using majority voting [4]. MLRsubMLL is a Bayesian approach, which contains two main steps: 1) the posterior probability distributions are constructed by an a subspace MLR classifier, and 2) segmentation, which refers to an image of class labels from a posterior distribution built on the aforementioned classifier and on an MLL prior [57]. GCK combines the different kernels built on the spectral and spatial information of the hyperspectral data without any weight parameters [45]. The classifier in this work is the multinomial logistic regression, and the spatial information is modeled from EMAPs. Mixed_lasso with 3D-DWT features is to use structured sparse logistic regression (solved by Mixed_lasso) to classify 3D-DWT [34]. The results presented in Table XVI are obtained using the same training and testing sets. From Table XVI, we can conclude that both RSELM_EMAPs and RoELM_EMAPS outperform other spatial–spectral classifiers in terms of OA, AA, and κ. In particular, RoELM_EMAPs gains the highest OA and κ, and RSELM_EMAPs achieves the highest AA.

TABLE XVII
CLASSIFICATION ACCURACIES OBTAINED FOR DIFFERENT CLASSIFICATION ALGORITHMS USING TEN SAMPLES PER CLASS WHEN APPLIED TO THE SPECTRAL INFORMATION OF THE UNIVERSITY OF PAVIA ROSIS IMAGE

Class	DT	RSDT	RF	RoF	RoRF	ELM	RSELM	RoELM
1	64.06	66.11	69.78	73.67	**80.33**	52.83	61.17	54.08
2	87.71	91.22	96.58	95.21	**98.36**	92.83	93.36	97.56
3	92.92	96.39	95.73	98.61	**99.20**	40.37	13.37	94.59
4	44.66	51.76	51.06	82.51	77.54	52.54	55.26	**84.55**
5	75.78	84.35	88.54	89.24	**96.81**	55.41	41.34	89.13
6	36.08	39.71	46.02	47.65	48.54	50.08	**59.19**	48.29
7	37.66	40.96	43.83	48.64	46.34	50.27	56.91	**61.98**
8	58.84	60.97	65.06	63.85	**66.71**	47.99	54.97	49.75
9	64.08	68.71	78.63	84.31	**87.94**	61.36	67.21	73.96
OA	49.79	53.79	58.24	63.32	**64.77**	51.67	56.41	60.22
AA	62.45	66.69	70.58	75.97	**77.98**	55.97	55.86	72.63
κ	40.02	44.36	49.24	55.81	**57.51**	40.96	45.74	52.12
Time	0.29	2.51	0.63	10.13	19.74	2.6	41.86	80.67

TABLE XVIII
CLASSIFICATION ACCURACIES OBTAINED FOR DIFFERENT CLASSIFICATION ALGORITHMS USING TEN SAMPLES PER CLASS WHEN APPLIED TO THE EMAPS OF THE UNIVERSITY OF PAVIA ROSIS IMAGE

Class	DT	RSDT	RF	RoF	RoRF	ELM	RSELM	RoELM
1	84.63	87.66	89.04	89.14	90.85	56.65	71.11	**94.81**
2	82.48	89.59	98.26	90.61	98.59	95.77	**99.88**	97.42
3	85.92	91.23	94.71	90.21	97.70	88.91	**99.43**	91.44
4	84.85	85.74	82.26	87.52	87.11	73.24	**95.85**	95.27
5	84.35	88.55	92.84	92.25	94.62	87.00	**94.80**	93.18
6	79.48	82.18	87.37	86.81	87.72	81.80	**90.79**	88.38
7	63.08	67.38	76.10	77.73	76.21	66.61	85.83	**89.39**
8	80.62	85.15	90.71	87.01	91.18	88.61	95.26	**95.75**
9	98.66	99.44	99.10	99.21	99.45	99.62	**99.95**	99.91
OA	81.14	84.24	88.11	88.40	89.30	80.43	91.19	**91.93**
AA	82.67	86.32	90.04	89.20	91.49	82.02	92.55	**93.95**
κ	75.93	79.77	84.54	84.95	86.10	74.75	88.54	**89.55**
Time	0.32	2.47	0.79	15.02	27.56	2.87	58.35	114.35

In order to assess the effectiveness of the RS ensemble for a limited training set, we have randomly extracted a few training samples from the training set. Only ten samples for each class are used for this experiment. We have repeated the training sample selection and the classification process ten times, and the mean classification results are reported in this paper. Tables XVII and XVIII show the OA, the AA, and the class-specific accuracy obtained for individual and ensemble classifiers using ten samples per class when the spectral and spatial information of the University of Pavia ROSIS image was used as the input, respectively. The classification results in Tables XVII and XVIII are lower than those in Tables XIV and XV due to the limited training set. For instance, the OA and the AA of RoRF_EMAPs are 96.47% and 95.87% for the original

TABLE XIX
CLASSIFICATION ACCURACIES OBTAINED FROM THE PROPOSED METHODS (ROELM_EMAPS AND RSELM_EMAPS) IN COMPARISON WITH OTHER SPATIAL–SPECTRAL CLASSIFIERS FOR THE UNIVERSITY OF PAVIA ROSIS IMAGE (TEN SAMPLES PER CLASS)

Classifier	SVM+Clustering [4]	MLRsubMLL [57]	GCK [45]	Mixed_lasso_3D-DWT [34]	RSELM_EMAPs	RoELM_EMAPs
OA	61.83	73.68	89.38	87.73	91.19	**91.93**
AA	73.85	77.18	92.22	91.14	92.55	**93.95**
κ	57.14	66.41	86.53	84.36	88.54	**89.55**

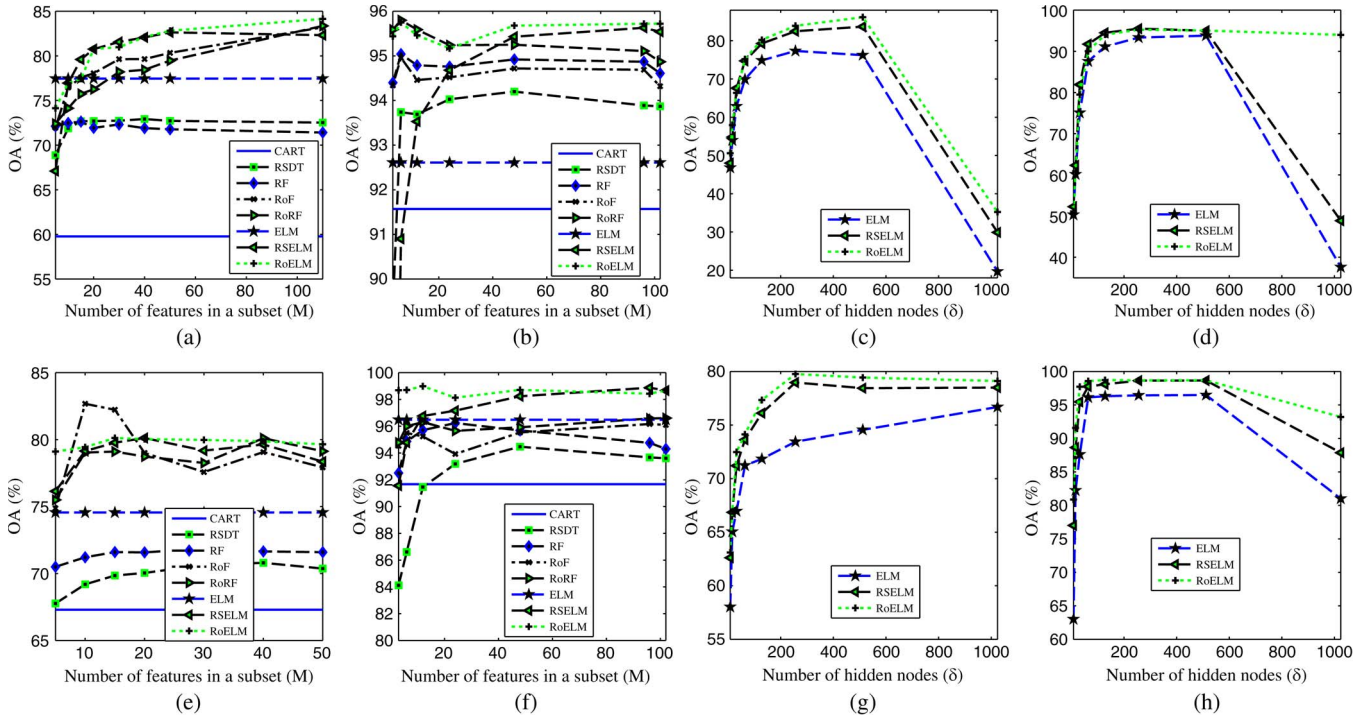


Fig. 10. Indian Pines AVIRIS image (10% of the labeled samples as training samples). (a) and (b) Sensitivity to the change of M with spectral information and M with EMAPs, respectively. (c) δ of ELM and its ensembles with spectral information. (d) δ of ELM and its ensembles with EMAPs. University of Pavia ROSIS image (entire training set). (e) and (f) Sensitivity to the change of M with spectral information and M with EMAPs. (g) δ of ELM and its ensembles with spectral information. (h) δ of ELM and its ensembles with EMAPs.

training set, respectively; whereas using limited training samples, the OA and the AA of RoRF_EMAPs are 89.30% and 91.49%, respectively. Nevertheless, with a very small training set, the results using the combination of RS ensembles and EMAPs are still very good. Furthermore, Table XIX gives the comparisons of RSELM_EMAPs and RoELM_EMAPS against the state-of-the-art spectral–spatial classification methods for a limited training set (ten samples per class). From this table, it can be found that our proposed method RoELM_EMAPs gains the best classification result. Considering the processing time, with the limited training set, the processing time of DT and its ensembles is significantly reduced. The computational cost of ELM and its ensembles with limited training samples is higher than that of ELM and its ensembles with the entire training set, because we used more hidden nodes ($\delta = 512$) to generate better performances.

The number of features in a subset (M) is the key parameter of the RS ensemble. In ELM and its ensembles, the number of hidden nodes (δ) plays an important role. The effects of parameters in RS ensemble are depicted in Fig. 10. It is observed from Fig. 10(a) and (b) and (e) and (f) that there is no pattern of dependence between M and the ensemble accuracy. Different RS ensemble classifiers gain the highest OA on different values of M. For the University of Pavia ROSIS image, RoF with spectral information gains the highest OAs when $M = 10$, and RoELM with EMAPs achieves the best classification result when $M = 6$. Fig. 10(c) and (d) and (g) and (h) depicts that a large number of hidden nodes may give higher accuracies in testing, but a complex network could also overfit the training data. For instance, the generalization performance decreases when the number of hidden nodes is larger than 512. In general, these parameters should be empirically selected in particular applications.

VII. CONCLUSION AND PERSPECTIVE

In this paper, we have developed a novel framework that combines RS ensembles and EMAPs for the spatial–spectral classification of remotely sensed hyperspectral data. Considering the computational cost, we selected two fast learning algorithms, i.e., DT and ELM, to build the RS ensembles. Several conclusions can be summarized based on our experimental results.

- Although RS ensembles require more training time than individual classifiers, their performance is superior to the individual classifiers using both spectral and spatial information as input. The computational load for RF algorithm

is very low (less than 1 s for the AVIRIS data set). In addition, the computation time for RS ensemble can be further reduced by decreasing the ensemble size.

- In most cases, rotation subspace classifiers, such as RoF, RoRF, and RoELM, outperform RSDT, RF, and RSELM. That is because we introduce more diversity in rotation subspace ensemble classifiers by using feature extraction and random selection strategies. However, it leads to increased computational complexity for rotation subspace approaches.
- In general, ELM and its ensembles can achieve higher accuracies than DT and its ensembles. The computation time of ELM and its ensembles depends on the hidden nodes when the ensemble size and training samples are fixed. Nevertheless, the efficiency of the ELM ensemble could be further improved by choosing smaller size of ensemble or using less hidden nodes.
- Spectral–spatial classification approaches computed by EMAPs based on the proposed ensembles achieve the state-of-the-art performances for two hyperspectral data sets.

However, RS ensembles are likely to have two limitations: 1) the number of features in a subset is required to be defined in advance (the optimal value for this parameter depends on the data set), and 2) the high computation time due to the high dimensionality of the input features. Therefore, our future work is to develop an effective scheme for automatically estimating the number of features in a subset for RS ensemble and to apply a dimensionality reduction step for both spectral and spatial information of hyperspectral data.

ACKNOWLEDGMENT

The authors would like to thank Prof D. Landgrebe from Purdue University, West Lafayette, IN, USA, and Prof P. Gamba for providing the hyperspectral remote sensing images.

REFERENCES

[1] D. A. Landgrebe, "Hyperspectral image data analysis as a high dimensional signal processing problem," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.

[2] C. I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. New York, NY, USA: Plenum, 2003.

[3] C. I. Chang, *Hyperspectral Data Exploitation: Theory and Applications*. Hoboken, NJ, USA: Wiley-Interscience, 2007.

[4] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral–spatial classification of hyperspectral imagery based on partitional clustering techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2973–2987, Aug. 2009.

[5] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using watershed transformation," *Pattern Recognit.*, vol. 43, no. 7, pp. 2367–2379, Jul. 2010.

[6] J. C. Harsanyi and C. I. Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 4, pp. 779–785, Jul. 1994.

[7] L. O. Jimenez, A. Morales-Morell, and A. Creus, "Classification of hyperdimensional data based on feature and decision fusion approaches using projection pursuit, majority voting, and neural networks" *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1360–1366, May 1999.

[8] C. I. Chang, "An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis," *IEEE Trans. Inf. Theory.*, vol. 46, no. 5, pp. 1927–1932, Aug. 2000.

[9] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification" *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.

[10] T. V. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of hyperspectral images with regularized linear discriminant analysis" *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 862–873, Mar. 2009.

[11] A. Villa, J. A. Benediktsson, J. Chanussot, and C. Jutten, "Hyperspectral image classification with independent component discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 4865–4876, Dec. 2011.

[12] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag, 1995.

[13] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[14] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. New York, NY, USA: Wiley-Interscience, 2004.

[15] J. A. Benediktsson, J. Chanussot, and M. Fauvel, "Multiple classifier systems in remote sensing: From basics to recent developments," in *Proc. Int. Workshop Multiple Classifier Syst.*, Prague, Czech Republic, May 23–25, 2007, pp. 501–512.

[16] L. Rokach, *Pattern Classification Using Ensemble Methods*. Singapore: World Scientific, 2010.

[17] P. Du *et al.*, "Multiple classifier system for remote sensing image classification: A review," *Sensors*, vol. 12, no. 4, pp. 4764–4792, 2012.

[18] M. Wozniak, M. Grana, and E. Corchado, "A survey of multiple classifier systems as hybrid systems," *Inf. Fusion*, vol. 16, no. 1, pp. 3–17, Mar. 2014.

[19] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832–844, Aug. 1998.

[20] L. I. Kuncheva, J. J. Rodríguez, C. O. Plumpton, D. E. Linden, and S. J. Johnston, "Random subspace ensembles for fMRI classification," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 531–542, Feb. 2010.

[21] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[22] P. Gislason, J. A. Benediktsson, and J. Sveinsson, "Random forests for land cover classification," *Pattern Recognit. Lett.*, vol. 27, no. 4, pp. 294–300, Mar. 2006.

[23] J. C. Chan and D. Paelinckx, "Evaluation of random forest and AdBboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery," *Remote Sens. Environ.*, vol. 112, no. 6, pp. 2999–3011, 2008.

[24] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS J. Photogramm.*, vol. 67, no. 1, pp. 93–104, Jan. 2012.

[25] 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.

[26] J. J. Rodriguez and L. I. Kuncheva, "Rotation forest: A new classifier ensemble method" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1619–1630, Oct. 2009.

[27] J. Xia, P. Du, X. He, and J. Chanussot, "Hyperspectral remote sensing image classification based on rotation forest," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 239–243, Jan. 2014.

[28] J. Xia, J. Chanussot, P. Du, and X. He, "Spectral–spatial classification for hyperspectral data using rotation forests with local feature extraction and Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2532–2546, May 2015.

[29] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral–spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, Mar. 2013.

[30] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 309–320, Feb. 2001.

[31] J. A. Benediktsson, M. Pesaresi, and K. Amason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1940–1949, Sep. 2003.

[32] T. C. Bau, S. Sarkar, and G. Healey, "Hyperspectral region classification using a three-dimensional Gabor filterbank," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 9, pp. 3457–3464, Sep. 2010.

[33] F. Tsai and J. Lai, "Feature extraction of hyperspectral image cubes using three-dimensional gray-level cooccurrence" *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 6, pp. 3504–3513, Jun. 2013.
[34] Y. Qian, M. Ye, and J. Zhou, "Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features" *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2276–2291, Apr. 2013.
[35] J. Serra, *Image Analysis and Mathematical Morphology*. New York, NY, USA: Academic, 1982.
[36] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.
[37] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, Nov. 2008.
[38] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, Oct. 2010.
[39] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Extended profiles with morphological attribute filters for the analysis of hyperspectral data," *Int. J. Remote Sens.*, vol. 31, no. 22, pp. 5975–5991, Jul. 2010.
[40] M. Dalla Mura, A. Villa, J. Benediktsson, J. Chanussot, and L. Bruzzone, "Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 542–546, May 2011.
[41] P. Reddy Marpu *et al.*, "Classification of hyperspectral data using extended attribute profiles based on supervised and unsupervised feature extraction techniques," *Int. J. Image Data Fusion*, vol. 3, no. 3, pp. 269–298, 2012.
[42] M. Pedergnana, P. Reddy Marpu, M. Dalla Mura, J. A. Benediktsson, and L. Bruzzone, "Classification of remote sensing optical and lidar data using extended attribute profiles," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 7, pp. 856–865, Nov. 2012.
[43] M. Pedergnana, P. Reddy Marpu, M. Dalla Mura, J. A. Benediktsson, and L. Bruzzone, "A novel technique for optimal feature selection in attribute profiles based on genetic algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 6, pp. 3514–3528, Jun. 2013.
[44] N. Falco, M. Dalla Mura, F. Bovolo, J. A. Benediktsson, and L. Bruzzone, "Change detection in VHR images based on morphological attribute profiles," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 3, pp. 636–640, May 2013.
[45] J. Li, P. Reddy Marpu, A. Plaza, J. M. Bioucas-Dias, and J. A. Benediktsson, "Generalized composite kernel framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4816–4829, Sep. 2013.
[46] S. Bernabe, P. Reddy Marpu, A. Plaza, M. Dalla Mura, and J. A. Benediktsson, "Spectral–spatial classification of multispectral images using kernel feature space representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 288–292, Jan. 2014.
[47] B. Song *et al.*, "Remotely sensed image classification using sparse representations of morphological attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 5122–5136, Aug. 2014.
[48] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: A new learning scheme of feedforward neural networks," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Budapest, Hungary, 2004, vol. 2, pp. 985–990.
[49] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, Dec. 2006.
[50] G. Stiglic, J. J. Rodriguez, and P. Kokol, "Rotation of random forests for genomic and proteomic classification problems," *Softw. Tools Algorithms Biol. Syst.*, vol. 696, pp. 211–221, 2011.
[51] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, Mar. 1986.
[52] R. Narayanan, D. Honbo, G. Memik, A. Choudhary, and J. Zambreno, "Interactive presentation: An FPGA implementation of decision tree classification," in *Proc. Conf. Des. , Autom. Test Eur.*, San Jose, CA, USA, 2007, pp. 189–194.
[53] L. Rokach and O. Maimon, *Data Mining with Decision Trees: Theory and Applications*. Singapore: World Scientific, 2008.
[54] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, Aug. 1996.
[55] G. B. Huang and L. Chen, "Convex incremental extreme learning machine" *Neurocomputing*, vol. 70, no. 16–18, pp. 3056–3062, Oct. 2007.
[56] G. B. Huang and L. Chen, "Enhanced random search based incremental extreme learning machine," *Neurocomputing*, vol. 71, no. 16–18, pp. 3460–3468, Oct. 2008.
[57] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral–spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 809–823, Mar. 2012.
[58] L. I. Kuncheva and C. J. Whitaker, "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy," *Mach. Learn.*, vol. 51, no. 2, pp. 181–207, May 2003.
[59] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*. Boca Raton, FL, USA: CRC, 1984.
[60] J. Xia, J. Chanussot, P. Du, and X. He, "Rotation-Based Ensemble Classifiers for High-Dimensional Data," in *Fusion in Computer Vision*, B. Ionescu, J. Benois-Pineau, T. Piatrik, and G. Quénot, Eds. New York, NY, USA: Springer-Verlag, 2014, pp. 135–160.

Junshi Xia (S'11) received the B.S. degree in geographic information systems and the Ph.D. degree in photogrammetry and remote sensing from the China University of Mining and Technology, Xuzhou, China, in 2008 and 2013, respectively, and the Ph.D. degree in image processing from the Grenoble Institute of Technology, Grenoble, France, in 2014.

He is currently a Research Fellow with the Department of Geographic Information Sciences, Nanjing University, Nanjing, China. He is also currently with the Grenoble Images Speech Signals and Automatics Laboratory, Grenoble Institute of Technology. His research interests include multiple classifier system in remote sensing, hyperspectral remote sensing image processing, and urban remote sensing.

Mauro Dalla Mura (S'08–M'11) received the B.E. and M.E. degrees in telecommunication engineering from the University of Trento, Trento, Italy, in 2005 and 2007, respectively, and the joint Ph.D. degree in information and communication technologies (telecommunications area) from the University of Trento and in electrical and computer engineering from the University of Iceland, Reykjavik, Iceland, in 2011.

In 2011, he was a Research Fellow with Fondazione Bruno Kessler, Trento, conducting research on computer vision. He is currently an Assistant Professor with the Grenoble Institute of Technology (Grenoble INP), Grenoble, France, where he is conducting his research at the Grenoble Images Speech Signals and Automatics Laboratory (GIPSA-lab). His main research activities are in the fields of remote sensing, image processing, and pattern recognition. In particular, his interests include mathematical morphology, classification, and multivariate data analysis.

Dr. Dalla Mura is a member of the Geoscience and Remote Sensing Society (GRSS) and the IEEE GRSS Data Fusion Technical Committee and the Secretary of the IEEE GRSS French Chapter (2013–2016). He was a Lecturer at the Remote Sensing Summer School 2012 (organized by the IEEE GRSS), Munich, Germany. He is a Reviewer of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, the IEEE JOURNAL OF SELECTED TOPICS IN EARTH OBSERVATIONS AND REMOTE SENSING, the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, *Pattern Recognition Letters*, the *ISPRS Journal of Photogrammetry and Remote Sensing*, and *Photogrammetric Engineering and Remote Sensing*. He was a recipient of the IEEE GRSS Second Prize in the Student Paper Competition of the 2011 IEEE International Geoscience and Remote Sensing Symposium 2011 (Vancouver, CA, USA, July 2011).

Jocelyn Chanussot (M'04–SM'04–F'12) received the M.Sc. degree in electrical engineering from the Grenoble Institute of Technology (Grenoble INP), Grenoble, France, in 1995 and the Ph.D. degree from Savoie University, Annecy, France, in 1998.

In 1999, he was with the Geography Imagery Perception Laboratory for the Delegation Generale de l'Armement (French National Defense Department). Since 1999, he has been with Grenoble INP, where he was an Assistant Professor from 1999 to 2005, an Associate Professor from 2005 to 2007, and is currently a Professor of signal and image processing. He is conducting his research at the Grenoble Images Speech Signals and Automatics Laboratory (GIPSA-lab). Since 2013, he has been also an Adjunct Professor with the University of Iceland. His research interests include image analysis, multicomponent image processing, nonlinear filtering, and data fusion in remote sensing.

Dr. Chanussot is the Founding President of IEEE Geoscience and Remote Sensing French Chapter (2007–2010), which received the 2010 IEEE GRS-S Chapter Excellence Award, and a member of the Institut Universitaire de France (2012–2017). He was a member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society (2006–2008) and the Program Chair of the IEEE International Workshop on Machine Learning for Signal Processing (2009). He was the Cochair (2005–2008) and the Chair (2009–2011) of the GRS Data Fusion Technical Committee. He was a member of the IEEE Geoscience and Remote Sensing Society Administrative Committee (2009–2010), in charge of membership development. He was the General Chair of the first IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing. He was an Associate Editor of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS (2005–2007) and *Pattern Recognition* (2006–2008). Since 2007, he has been an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. Since 2011, he has been the Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. In 2013, he was a Guest Editor of the PROCEEDINGS OF THE IEEE, and in 2014, he was a Guest Editor of the IEEE SIGNAL PROCESSING MAGAZINE. He was a corecipient of the NORSIG 2006 Best Student Paper Award, the IEEE GRSS 2011 Symposium Best Paper Award, the IEEE GRSS 2012 Transactions Prize Paper Award, and the IEEE GRSS 2013 Highest Impact Paper Award.

Peijun Du (M'07–SM'12) received the Ph.D. degree from the China University of Mining and Technology, Xuzhou, China, in 2001.

After receiving the Ph.D. degree, he was employed by the China University of Mining and Technology. In 2011, he joined Nanjing University, Nanjing, China, where he is currently a Professor of remote sensing in the Department of Geographic Information Sciences and the Deputy Director of the Key Laboratory for Satellite Mapping Technology and Applications of National Administration of Surveying, Mapping and Geoinformation of China. From February 2002 to March 2004, he was a Postdoctoral Fellow with Shanghai Jiao Tong University, Shanghai, China. He has been a Senior Visiting Scholar with The University of Nottingham, Nottingham, U.K., and the Grenoble Images Speech Signals and Automatics Laboratory (GIPSA-lab), Grenoble Institute of Technology, Grenoble, France. He has authored or coauthored over 40 articles in international peer-reviewed journals and over 100 papers in international conferences and Chinese journals. His research interests focus on remote sensing image processing and pattern recognition, hyperspectral remote sensing, and applications of geospatial information technologies.

Dr. Du served as the Cochair for the Technical Committee of URBAN 2009, EORSA 2014, and IAPR-PRRS 2012; the Cochair of the Local Organizing Committee of JURSE 2009, WHISPERS 2012, and EORSA 2012; and the member of the scientific committee or the technical committee of other international conferences, for example, Spatial Accuracy 2008, ACRS 2009, WHISPERS (2010–2014), URBAN (2011, 2013, and 2015), MultiTemp (2011, 2013 and 2015), ISDIF 2011, and SPIE European Conference on Image and Signal Processing for Remote Sensing (2012–2014). He has been an Associate Editor of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS since 2009. He has been a Guest Editor of three special issues of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATION AND REMOTE SENSING.

Xiyan He received the Generalist Engineer degree from Ecole Centrale Paris, Paris, France, in 2006, the M.E. degree in pattern recognition and intelligent system from Xi'an Jiaotong University, Xi'an, China, and the Ph.D. degree in computer science from the University of Technology of Troyes, Troyes, France, in 2009.

He was a Teaching Assistant with the University of Technology of Troyes in 2009; a Postdoctoral Research Fellow with the Research Centre for Automatic Control of Nancy in 2010; and a Teaching Assistant with the University of Pierre Mendes France, Grenoble, France, in 2011. Since 2012, she has been a Postdoctoral Research Fellow with the Grenoble Images Speech Signal and Automatics Laboratory, Grenoble Institute of Technology, Grenoble. Her main research interests include machine learning, pattern recognition, and data fusion, with special focus on applications to remote sensed images.