

Random Forest Ensembles and Extended Multi-Extinction Profiles for Hyperspectral Image Classification

Junshi Xia, *Member, IEEE*, Pedram Ghamisi, *Member, IEEE*, Naoto Yokoya, *Member, IEEE*, Akira Iwasaki

Abstract—Classification techniques for hyperspectral images based on random forest (RF) ensembles and extended multi-extinction profiles (EMEPs) are proposed as a means of improving performance. To this end, five strategies—bagging, boosting, random subspace, rotation-based, and boosted rotation-based—are used to construct the RF ensembles. Extinction profiles (EPs), which are based on an extrema-oriented connected filtering technique, are applied to the images associated with the first informative components extracted by independent component analysis, leading to a set of EMEPs. The effectiveness of the proposed method is investigated on two benchmark hyperspectral images, University of Pavia and Indian Pines. Comparative experimental evaluations reveal the superior performance of the proposed methods, especially those employing rotation-based and boosted rotation-based approaches. An additional advantage is that the CPU processing time is acceptable.

Index Terms—Ensemble learning, Extended multi-extinction profiles (EMEPs), Hyperspectral Image Classification, Random Forest (RF).

I. INTRODUCTION

Hyperspectral imaging sensors generate tens or hundreds of narrow bands with very fine spectral resolution. Therefore, they can provide excellent discriminability of different materials [1]. With the development of hyperspectral technology, the sensors can also make further improvements on spatial resolution, which allows us to describe the spatial structures in the scene with clarity.

Supervised classification is one of the most relevant topics in the analysis of remote sensing images today, making it widely used in thematic applications such as environmental mapping [2], [3]. An increase in the number of spectral bands, when only a limited number of training samples are available, can potentially cause the “curse of dimensionality” (the so-called Hughes phenomenon) [4]. Therefore, there is an urgent need for advanced techniques, particularly kernel-based methods [5], [6] and feature extraction/selection [7], [8], to alleviate this problem.

Manuscript received; revised. This work was supported in part by the Grants-in-Aid for Scientific Research (KAKENHI) under Grant 15K20955, 24360347 and 16H04587, and the Japan Society for the Promotion of Science through KAKENHI under Grant 16F16053.

J. Xia, N. Yokoya and A. Iwasaki are with the Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan (e-mail: xiajunshi@gmail.com; yokoya@sal.rcast.u-tokyo.ac.jp; aiwasaki@sal.rcast.u-tokyo.ac.jp).

P. Ghamisi is with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), 82234 Wessling, Germany, and also with the Signal Processing in Earth Observation, Technische Universität München (TUM), 80333 Munich, Germany (e-mail: pedram.ghamisi@dlr.de)

Kernel-based methods (e.g., support vector machines, or SVMs) [5], [9] and deep learning (e.g., convolution neural networks, or CNNs) [10], [11] are powerful techniques for the classification of hyperspectral data. For SVMs, the kernels are empirically defined, and the parameters need to be optimized by cross-validation techniques. For CNNs, the number of convolutional and pooling layers should be carefully determined. The main disadvantage of SVMs and CNNs is high computational complexity.

Recently, researchers have shown an increased interest in the random forest (RF) classifiers [12] due to the following advantages:

- they are insensitive to high-dimensional features;
- they perform out-of-sample prediction rapidly;
- they require only slight parameter tuning; and
- they are capable of ranking the importance of features.

RF is constructed by several individual decision trees (DTs), in which each tree is trained on a bootstrapped sample, while randomly selected features are used to split a leaf on each tree [12]. Researchers have exploited RF on a variety of remote sensing resources in various applications, such as hyperspectral data [13], high spatial resolution data [14], LiDAR datasets [15], and time-series datasets [16].

In order to further improve the performance of RF, many extensions have been suggested. One is to reduce the original feature space via feature selection or extraction prior to application of the RF classifier [17]. A second extension defines an efficient aggregation method in RF, such as alternating decision forests [18]. A third extension defines an ensemble of RF classifiers by manipulating the training set and training features. Representative methods for building the ensembles are bootstrap aggregating (bagging) [19], boosting [20], random subspace [21], and rotation-based [22], [23] approaches. In [24], the rotation-based scheme was proposed to construct the ensemble of RF and an extreme learning machine (ELM) with extended multi-attribute profiles (EMAPs), which achieved high classification performance. The main objective of a rotation-based scheme is related to data transformation (e.g., principal component analysis, PCA), applied to the node level in order to identify the best split, which leads to improvements in accuracy and diversity. Furthermore, in order to increase the diversity within the ensemble, we proposed a rotation-based random forest ensemble via kernel PCA [25]. In this work, we propose combining the boosting and rotation-based methods to improve the performance.

Due to the spatial variability observed in the image, traditional pixel-wise classifiers only consider the spectral measurements without spatial regularization, generating a low classification performance [1]. In order to reduce the labeling uncertainty and the salt-and-pepper noise appearing on the classification maps, a joint spectral and spatial classifier is needed in the analysis. Thus, a reliable approach to exploiting the spatial information is imperative.

In order to extract spatial information, fixed and adaptive neighborhood systems are always taken into account. In the fixed neighborhood system, the center pixel is classified by considering the information from its neighbors according to one of a number of approaches, such as Markov random fields (MRFs) modeling [26]–[28]. Tarabalka *et al.* [29], Li *et al.* [30], and Xia *et al.* [31] integrated MRFs with probability SVMs, sparse multinomial logistic regression (SMLR), and rotation forest (RoF) classifiers, respectively. However, the main disadvantage of the fixed neighborhood system is that it does not always accurately reflect information about spatial structures. To solve the problem delineated above, an adaptive neighborhood system, such as morphological processing, should be taken into account.

Pesaresi and Benediktsson [32] proposed using a sequence of geodesic opening and closing operations to build a morphological profile (MP), as well as the derivative of the MP. Furthermore, the extension of MPs [33] is proposed to handle hyperspectral datasets. The main limitation of MP lies in the partial analysis that is performed with the computation of the profile. In order to solve this limitation, Dalla Mura *et al.* proposed attribute profiles (APs) [34] and extended APs (EAPs) [35] by using the sequential application of morphological attribute filters (AFs), extending the concepts of MPs and EMPs. A comprehensive survey of MPs, APs, and their modifications can be found in [36], [37]. The main weakness of constructing APs is the difficulty of determining the optimal range of parameters in the filtering step, if no prior knowledge of the study area is available.

Recently, Ghamisi *et al.* proposed the concept of extinction profiles (EPs) to further improve the classification performance of the APs [38]. EPs are based on extrema-oriented connected extinction filters. These filtering operators are automatic by nature and, as a result, they solve the main shortcoming of the conventional APs, i.e., the manual setting of threshold values. Specifically, the thresholds used by APs vary greatly based on the attribute being analyzed, the dataset being used, and sensor-related information, including spatial resolution. Therefore, the thresholds for APs are more difficult to set [38]. In addition, EPs are able to preserve the height of the extrema kept, while attribute filters erode the height of the tree. In the experiments conducted on panchromatic remote sensing data using EPs [38], the capability of EPs has been evaluated through experiments where EPs show better performance than APs in terms of simplification for recognition and classification accuracy. In [39], the concept of EPs has successfully been generalized to handle hyperspectral data.

In this paper, we propose advanced classification schemes based on RF ensembles applied to extended multi-extinction profiles (EMEPs). EMEPs are produced by applying an

extrema-oriented connected filtering to the images associated with the first informative components extracted by the independent component analysis (ICA) [40]. Although the capability of the RF classifier is already proven for the classification of the features extracted by EAPs and EMEPs [36], [37], its performance should be further improved by introducing an ensemble strategy. In particular, five strategies—bagging, boosting, random subspace, rotation-based, and boosted rotation-based—are used to construct the RF ensembles.

The main innovative features of this new work are as follows:

- defining spectral-spatial classification techniques based on the RF ensembles as well as the spatial features extracted by EMEPs;
- proposing a new strategy (i.e., the boosted rotation-based method) for building the ensemble of RF classifier; and
- performing a detailed comparison of different RF ensemble strategies.

The proposed classification method is validated on two well-known datasets, one acquired by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) sensor over the agricultural area of Indian Pines in northwestern Indiana in the United States, and the other from the reflective optics spectrographic imaging system (ROSIS) sensor over the city area of Pavia, Italy. Experimental results demonstrated that the proposed framework can classify hyperspectral images effectively and efficiently in terms of classification accuracies.

The remainder of this paper is structured as follows: Section II describes the proposed framework, including the specific components of EMEP construction and RF ensembles. The experimental results on the two real hyperspectral data sets are reported in Section III. Section IV contains conclusions about the presented work and the implications.

II. PROPOSED METHODOLOGY

In the proposed methodology, the hyperspectral image is transformed by ICA. The first three independent components (ICs) are used as the base image to construct the EPs. EPs with area, height, volume, diagonal of bounding box, and standard deviation attributes are concatenated into a stacked vector to construct EMEPs. Then, the EMEPs features are classified by the proposed RF ensembles, i.e., bagging RF (BagRF), boosting RF (BoostRF), random subspace RF (RSRF), rotation-based RF (RoRF), and boosted rotation-based RF (BRoRF), and a final classification map is obtained. In the following, the specific parts of the proposed framework will be detailed.

A. Extended Multi-Attribute Extinction Profiles (EMEPs)

1) *Tree representation*: This representation increases the efficiency of the subsequent filtering step by dividing the transformation process into three steps: (1) tree creation, (2) filtering, and (3) image restitution. It highlights the inclusion relation between the connected components of any input gray scale image (obtained from the upper or lower sets). In this manner, two components are either nested or disjointed. Therefore, a set of connected components can construct a tree whose nodes are the components; the links between nodes are

the inclusion relations between components. The tree derived by the components in the upper (resp. lower) level set is called *max-tree* (resp. *min-tree*) [41]. The leaves of such trees are considered to be regional extremum for any increasing attribute whose extinction values can be estimated.

2) *Extinction values*: Extinction values are able to measure the persistence of extrema (leaves of the trees for any increasing attribute), which was initially proposed by Vachier for increasing attributes [42]. Let M represent a regional maximum of a gray scale image f , and $\Psi = (\psi_\lambda)_\lambda$ is a family of decreasing connected anti-extensive transformations. The extinction value corresponding to M with respect to Ψ and denoted by $\varepsilon_\Psi(M)$ is the maximal λ value, such that M still is a regional maxima of $\psi_\lambda(f)$. This definition can be expressed using the following equation:

$$\varepsilon_\Psi(M) = \sup\{\lambda \geq 0 \mid \forall \mu \leq \lambda, M \subset \text{Max}(\psi_\mu(f))\}. \quad (1)$$

3) *Extinction filters (EFs) for increasing attributes*: This approach is a connected filter that is able to preserve the relevant extrema of the image. EFs can be estimated as follows: let $\text{Max}(f) = \{M_1, M_2, \dots, M_N\}$ be the set of regional maxima of the image f (the leaves of the trees for any increasing attribute). Each regional maxima M_i has its corresponding extinction value ε_i . The EF of f set to preserve the n maxima with the highest extinction values is estimated by $EF^n(f) = R_f^\delta(g)$ where $R_f^\delta(g)$ is the reconstruction by dilation [43] of the mask image f from marker image g . The marker image g is given by $g = \max_{i=1}^n \{M'_i\}$ where \max is the pixel-wise maximum operation. M'_1 is the maximum with the highest extinction value, M'_2 has the second highest extinction value, and so on.

4) *Extinction filters (EFs) for non-increasing attributes*: Xu et al. [44] proposed a methodology to construct a max-tree of a max-tree or a max-tree of a tree of shapes [45]. The second tree representation takes the image to the space of shapes which makes it possible to construct a novel class of connected operators from the leveling family. After producing the second max-tree using a non-increasing attribute in the space of shapes, the height of the attribute used to compute the second max-tree becomes increasing. As a result, one can then compute extinction values and EFs for non-increasing attributes. It should be noted that EFs for non-increasing attributes do not follow the same extrema preservation property of the EFs for increasing attributes. They can also be considered as second max-tree increasing attribute EFs, which belong to a class of filters known as shape-based filters [44].

5) *Extinction profiles (EPs)*: In order to extract informative features from any gray scale image, instead of one filtering step, a sequence of filtering steps with progressively higher threshold values can be taken into account. In this way, the EPs can be produced composed of a sequence of thinning and thickening transformations defined with a sequence of progressively stricter criteria. An EP for the input gray scale

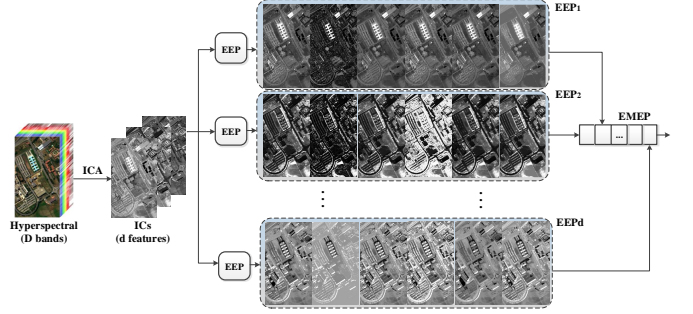


Fig. 1: Construction of EMEPs. First, the number of dimensions is decreased from D to d using ICA. Then, each IC is considered as one base image to construct its corresponding EP including five extinction attributes (i.e., area, volume, standard deviation, diagonal of the bounding box, and height). Finally, all EEPs extracted from three different ICs are concatenated to construct the EMEP.

image, f , can be defined as follows:

$$EP(f) = \underbrace{\{\phi^{P_{\lambda_s}}(f), \phi^{P_{\lambda_{s-1}}}(f), \dots, \phi^{P_{\lambda_1}}(f)\}}_{\text{thickening profile}}, \quad (2)$$

$$\underbrace{f, \gamma^{P_{\lambda_1}}(f), \dots, \gamma^{P_{\lambda_{s-1}}}(f), \gamma^{P_{\lambda_s}}(f)}_{\text{thinning profile}},$$

where $P_\lambda : \{P_{\lambda_i}\}$ ($i = 1, \dots, s$) a set of s ordered predicates. The terms γ and ϕ are thinning and thickening operators, respectively. For EPs, the number of extrema can be considered as the predicates.

6) *Extended EPs (EEP)*: The concept of EPs was originally proposed for gray scale images; in [39], it has been generalized and adapted for spatial information extraction from hyperspectral data, a method known as *extended extinction profiles* (EEP). To produce the EEP, a feature reduction approach such as PCA or ICA can be performed on the input data. Then, the most informative features [37] can be preserved as base images to produce the profiles (in this work, EPs have been applied to the first three independent components extracted by ICA). More precisely, the dimensionality of the image from $E \subseteq \mathbf{Z}^D$ to $E' \subseteq \mathbf{Z}^d$ ($d \leq D$) is reduced with a generic transformation $\Psi : E \rightarrow E'$ applied to an input image $\mathbf{f} = \{f_i\}_{i=1}^D$ (i.e., $\mathbf{q} = \Psi(\mathbf{f})$, where $\mathbf{q} = \{q_i\}_{i=1}^d$). The EP can then be performed on the most informative features \mathbf{q}_i ($i = 1, \dots, d$) of the transformed image, which can be mathematically described as follows:

$$EEP(\mathbf{q}) = \{EP(q_1), EP(q_2), \dots, EP(q_d)\}. \quad (3)$$

7) *Extended multi-EPs (EMEPs)*: Unlike MPs, which are only able to extract information about the size and structure of different objects with respect to the size and shape of structuring elements, EPs have a more flexible concept and can model different types of information, such as area, height, volume, diagonal of bounding box, and standard deviation. By concatenating all that information into a single stacked vector, the concept of *extended multi-EPs* (EMEPs) can be derived, as shown in Fig. 1, and mathematically defined as follows:

$$EMEP(\mathbf{q}) = \{EEP_{\mathcal{A}_1}(\mathbf{q}), EEP'_{\mathcal{A}_2}(\mathbf{q}), \dots, EEP'_{\mathcal{A}_k}(\mathbf{q})\}, \quad (4)$$

where $EEP_{\mathcal{A}_i}$ is an EEP built with a set of predicates evaluating the attribute \mathcal{A}_i . To avoid including the input image $\{q_i\}$ in the EMEPs that are present in each EEP, $EEP' = EEP \setminus \{q_i\}_{i=1, \dots, d}$ is considered. To generate the EMEP, five different extinction profiles ($k = 5$) have been taken into account [38], including area (a), diagonal of the bounding box (bb), volume (v), height (h), and standard deviation (std).

The original concept of the EP was applicable to computing increasing attributes (e.g., a , bb , v , and h). In order to generalize the concept of the EP to deal with other types of attributes (such as std , which is not an increasing attribute), a further modification is needed. In 2012, Xu *et al.* [44] proposed an approach, called space of shapes, to construct max-trees of tree-based image representations (i.e., to build a max-tree of a max-tree or a max-tree of a tree of shapes [45]). The second tree construction takes the image to the space of shapes, enabling the creation of a novel class of connected operators from the leveling family and more complex morphological analysis, such as the computation of extinction values for non-increasing attributes. More specifically, the attribute used to build the second tree becomes an increasing attribute in this space, which enables the computation of extinction values and extinction filters for non-increasing attributes such as std [38]. It is important to note that the creation of the second max-tree is much faster than for the first, since the complexity of the second max-tree construction depends on the number of nodes in the initial max-tree, while the first max-tree construction depends on the number of image pixels, which in general is much higher than the number of max-tree nodes [46].

The extrema values used to generate the profile for different attributes are automatically set using $\lambda = \alpha^j$ $j = 0, 1, \dots, s - 1$, where s is the number of predicates (thresholds) and α is a constant value. The total EP size is $2s + 1$, in which s is the number of threshold values, since the original image is also included in the profile. As mentioned in [39], the larger the α , the larger the differences between consecutive images. As the α becomes smaller, the profile will concentrate on keeping a few extrema, where most of the image information is usually present [47]. In this context, if an image is filtered out using an EF set to preserve 1 extrema and another filter set to preserve 2 extrema, the difference between these two images will be higher than applying an EF set to preserve 1000 extrema and the other set to preserve 1001 extrema. That occurs because the extrema with the highest extinction values are the ones that contain most of the image structural information [47]. In summary, our reason for defining the equation for selecting the values of n was to generate a profile with more images with few extrema, where most of the changes occur and which have more information for the classifier to learn, but also keep some images with a higher number of extrema, where less changes occur, but nonetheless they contain important information for the classifier to learn. We suggest using an α between 2 and 5. The size of each EP is $2s + 1$, which includes the input data and s filtering steps on the max-tree and s filtering steps on

the min-tree. Since the EMEP used in this paper is composed of five different profiles, the number of features are $10s + 1$. In this paper, we retained only three independent components, as they include almost all variations of the input data. Therefore, the total number of features extracted by EMEP is $3(10s + 1)$. In this paper, we used $\alpha = 3$, and set $s = 7$. The profiles have been computed considering the 4-connected connectivity rule.

B. Random Forest Ensembles

Let $\mathcal{S} \equiv \{1, \dots, N\}$ denote a set of integers indexing the N pixels of a hyperspectral image; let $\mathcal{C} \equiv \{1, \dots, C\}$ be a set of C labels; let $\mathbf{x} \equiv \{\mathbf{x}_1; \dots; \mathbf{x}_N\} \in \mathbb{R}^{N \times D}$ denote an image of D -bands; let $\mathbf{y} \equiv \{y_1, \dots, y_N\} \in \mathcal{C}$ be a set of labels for the N pixels; and let $\{\mathbf{X}, \mathbf{Y}\} \equiv \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$ be the training set, where n is the number of training samples. The objective of classification is to assign a label $y_i \in \mathcal{C}$ to each pixel $i \in \mathcal{S}$, based on the vector \mathbf{x}_i , resulting in an image of class label y_i .

RF is an ensemble of \hat{T} decision trees developed by Breiman [12]. During the training stage, decision trees are independently constructed on a bootstrap training set with randomly chosen features using bagging and random subspace selection [12]. Each decision tree is constructed by the following steps:

- Selection of subset training samples with replacement from the original training set \mathcal{S} . Out-of-bag (OOB) samples, which are not included in the bootstrap sample, are used to estimate the misclassification error and variable importance;
- Random selection of $M \leq D$ features and consequent identification of the best split using the *Gini* index;
- Growing of the tree to the maximum depth, without pruning.

During the classification stage, a given sample \mathbf{x}^* is classified by going through each decision tree until a leaf node is reached. A classification result (the decision function h) is assigned to each leaf node. The final class label y^* is determined by choosing the class with the most votes:

$$y^* = \underset{c \in \{1, 2, \dots, C\}}{\operatorname{argmax}} \sum_{t: h_t(\mathbf{x}^*)=c} \hat{T} \quad (5)$$

In order to further improve the performance of RF, we propose to construct RF ensembles. Since bagging, boosting, random subspace, and rotation-based are the typical ensemble strategies, we create the RF ensemble using the four existing frameworks. Moreover, we propose combining two different ensemble methods, boosting and rotation-based, to further improve the capacity of RF ensembles. In the following, these five frameworks are discussed in detail (seen in Fig 2).

1) *Bagging random forest (BagRF)*: In the bagging method, the input training samples are selected at random from an original training set by replacement, instructive iteration is exerted to create some different training sets, and each training set is classified by vote to predict its class [19]. The main steps of bagging random forest (BagRF) are shown in Algorithm 1. For the classification phase, a new sample \mathbf{x}^* is run on the

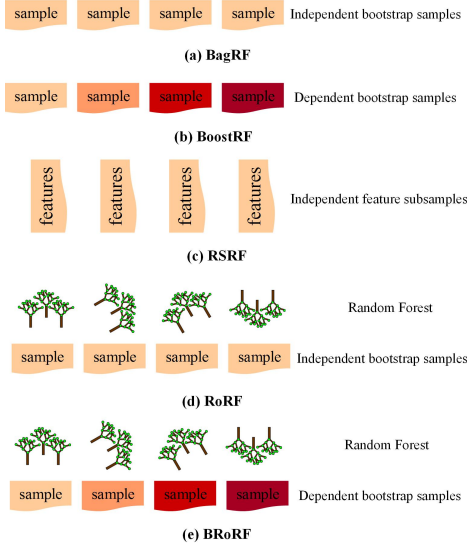


Fig. 2: Framework of random forest ensembles. Same color represents independent (different) samples and features. Different colors mean dependent samples with different weights in the classification process.

Algorithm 1 BagRF

Training phase

Input: $\{\mathbf{X}, \mathbf{Y}\} = \{\mathbf{x}_i, y_i\}_{i=1}^n$: training samples, T : number of RF classifiers, L : RF classifier.

Output: The ensemble \mathcal{L} .

- 1: **for** $t = 1 : T$ **do**
- 2: bootstrap sample from the original training set to form a new training set with size of n' , $n' = n$.
- 3: train L_t using a new training set.
- 4: add the classifier to the current ensemble, $\mathcal{L} = \mathcal{L} \cup L_t$.
- 5: **end for**

Prediction phase

Input: The ensemble $\mathcal{L} = \{L_t\}_{t=1}^T$. A new sample \mathbf{x}^* .

Output: class label y^*

- 1: $y^* = \operatorname{argmax}_{c \in \{1, 2, \dots, C\}} \sum_{t: L_t(\mathbf{x}^*)=c} 1$
-

output ensemble and the class with the maximum number of votes is chosen as the label for \mathbf{x}^* .

2) *Boosting random forest (BoostRF)*: Like bagging, boosting is another ensemble scheme for improving the performance of a weak learner. Adaptive boosting (AdaBoost) processes data with weights, and the weights of misclassified samples are increased to concentrate the learning algorithm on specific samples [20]. Indeed, AdaBoost provides the most informative training samples by changing their distributions dynamically for each base learner. In this case, AdaBoost can reduce both the variance and bias of the classification, leading to reduction in the smaller upper bound of the testing error. For this reason, a boosting random forest (BoostRF), which is described in Algorithm 2, is used in this paper. The classification of a new sample \mathbf{x}^* in BoostRF is performed on the output weights obtained from the probabilistic distribution in each RF classifier.

Algorithm 2 BoostRF

Input: $\{\mathbf{X}, \mathbf{Y}\} = \{\mathbf{x}_i, y_i\}_{i=1}^n$: training samples, T : number of RF classifiers, L : RF classifier.

Output: The ensemble \mathcal{L}

- 1: initialize the same weights to each sample $\omega_0 = 1/n$
- 2: **for** $t = 1 : T$ **do**
- 3: fit a classifier L_t to the training data using ω_0
- 4: the error $\epsilon_t = \sum_{i=1}^n I(L_t(\mathbf{x}_i) \neq y_i) \omega_t(\mathbf{x}_i)$, where $I(a = b)$ equals 1 when a equals b , otherwise equals 0.
- 5: if $\epsilon_t > 0.5$ or $\epsilon_t = 0$, abort the loop.
- 6: $\beta_t = \frac{\epsilon_t}{1 - \epsilon_t}$.
- 7: **for** $i = 1 : n$ **do**
- 8: update the weights $\omega_t(\mathbf{x}_i) = \omega_t(\mathbf{x}_i) \times \beta_t$.
- 9: **end for**
- 10: normalize weights to make the total weight of ω_t equal to 1.
- 11: add the classifier to the current ensemble, $\mathcal{L} = \mathcal{L} \cup L_t$.
- 12: **end for**

Prediction phase

Input: The ensemble $\mathcal{L} = \{L_t\}_{t=1}^T$. A new sample \mathbf{x}^* . The weights β_t

Output: class label y^*

- 1: $y^* = \operatorname{argmax}_{c \in \{1, 2, \dots, C\}} \sum_{t: L_t(\mathbf{x}^*)=c} \frac{1}{\beta_t}$
-

Algorithm 3 RSRF

Training phase

Input: $\{\mathbf{X}, \mathbf{Y}\} = \{\mathbf{x}_i, y_i\}_{i=1}^n$: training samples, T : number of RF classifiers, L : RF classifier. \mathbb{F} : Feature set.

Output: The ensemble \mathcal{L} .

- 1: **for** $t = 1 : T$ **do**
- 2: randomly select a subset from \mathbb{F} without replacement to form a new training set composed of M features ($M < D$).
- 3: train L_t using a new training set.
- 4: add the classifier to the current ensemble, $\mathcal{L} = \mathcal{L} \cup L_t$.
- 5: **end for**

Prediction phase

Input: The ensemble $\mathcal{L} = \{L_t\}_{t=1}^T$. A new sample \mathbf{x}^* .

Output: class label y^*

- 1: $y^* = \operatorname{argmax}_{c \in \{1, 2, \dots, C\}} \sum_{t: L_t(\mathbf{x}^*)=c} 1$
-

3) *Random subspace random forest (RSRF)*: In the random subspace approach the training set is also modified, as it is in bagging. However, this modification is performed in the feature space [21]. Algorithm 3 shows the main steps of random subspace random forest (RSRF). For the classification phase, a new sample \mathbf{x}^* will be run on the output ensemble, and the class with the maximum number of votes is chosen as the label for \mathbf{x}^* .

4) *Rotation random forest (RoRF)*: Rotation random forest (RoRF) uses data transformation and random feature selection to construct RF ensembles [48]. The main training and prediction steps of RoRF are presented in Algorithm 4.

In the training phase, the feature space is first divided into K disjoint subspaces. PCA is performed on each subspace with the bootstrapped samples from 75% of the original training set. A transformed training set is generated by rotating the original training set with a sparse matrix \mathbf{R}_t^a . An individual classifier is trained on this rotated training set. In the prediction phase, a new sample \mathbf{x}^* is rotated by \mathbf{R}_t^a , where \mathbf{R}_t^a is a rearranged

Algorithm 4 RoRF**Training phase**

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

Output: The ensemble \mathcal{L}

- 1: **for** $t = 1 : T$ **do**
- 2: randomly split the features \mathbb{F} into K subsets \mathbb{F}_k^t
- 3: **for** $k = 1 : K$ **do**
- 4: extract from \mathbf{X} the new training set $\mathbf{X}_{t,k}$ with the corresponding features \mathbb{F}_k^t
- 5: generate a subset $\hat{\mathbf{X}}_{t,k}$ by selecting with the bootstrap algorithm, the 75% of the initial training samples in $\mathbf{X}_{t,k}$
- 6: transform $\hat{\mathbf{X}}_{t,k}$ to get the coefficients $v_{t,k}^{(1)}, \dots, v_{t,k}^{(M_k)}$ using PCA
- 7: **end for**
- 8: sparse matrix \mathbf{R}_t is composed of the above coefficients

$$\mathbf{R}_t = \begin{bmatrix} v_{t,1}^{(1)}, \dots, v_{t,1}^{(M_1)} & 0 & \dots & 0 \\ 0 & v_{t,2}^{(1)}, \dots, v_{t,2}^{(M_2)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & v_{t,k}^{(1)}, \dots, v_{t,k}^{(M_K)} \end{bmatrix}$$

- 9: with respect to the original feature set, rearrange \mathbf{R}_t to \mathbf{R}_t^a
- 10: obtain the new training samples $\{\mathbf{X}\mathbf{R}_t^a, \mathbf{Y}\}$
- 11: build RF classifier L_t using $\{\mathbf{X}\mathbf{R}_t^a, \mathbf{Y}\}$
- 12: add the classifier to the current ensemble, $\mathcal{L} = \mathcal{L} \cup L_t$.
- 13: **end for**

Prediction phase

Input: The ensemble $\mathcal{L} = \{L_t\}_{t=1}^T$. A new sample \mathbf{x}^* . Rotation matrix: \mathbf{R}_t^a .

Output: class label y^*

- 1: get the output ensemble with $\mathbf{x}^* \mathbf{R}_t^a$.
- 2: $y^* = \operatorname{argmax}_{c \in \{1,2,\dots,C\}} \sum_{t: L_t(\mathbf{x}^* \mathbf{R}_t^a) = c} 1$

matrix of the rotation matrix \mathbf{R}_t with respect to the order of the original features and t is the index of the ensemble. Then, the transformed set, i.e., $\mathbf{x}^* \mathbf{R}_t^a$, is classified by the ensemble, and the class with the maximum number of votes is chosen as the final class. It is important to note Step 5 in Algorithm 4, in which 75% of the original number of training samples are selected to avoid obtaining the same coefficients of the transformed components if the same features are selected, thus enhancing diversity among the member classifiers.

RoRF can improve the performance of random forest by introducing further diversity through performing a data transformation within the ensemble, which is beneficial for the ensemble.

5) *Boosted rotation random forest (BRoRF)*: Both empirical and theoretical studies have demonstrated that the individual classifiers are accurate, but disagree on some different parts of the input space as much as possible at the same time [49]. The implication of this finding is that the member classifiers should be diverse. Performance would not be improved if the individual classification results were identical or similar. Boosting and rotation-based methods employ different perturbation techniques to promote the diversity in the ensemble. By combining the two perturbation techniques, we expect to enhance the diversity in order to further improve

Algorithm 5 BRoRF**Training phase**

Input: $\{\mathbf{X}, \mathbf{Y}\} = \{\mathbf{x}_i, y_i\}_{i=1}^n$: training samples, T : number of RF classifiers, K : number of subsets (M : number of features in each subset), L : RF classifier. J : iterations in Boosting.

Output: The ensemble \mathcal{L}

- 1: **for** $t = 1 : T$ **do**
- 2: randomly split the features \mathbb{F} into K subsets \mathbb{F}_k^t
- 3: **for** $k = 1 : K$ **do**
- 4: extract from \mathbf{X} the new training set $\mathbf{X}_{t,k}$ with the corresponding features \mathbb{F}_k^t
- 5: generate a subset $\hat{\mathbf{X}}_{t,k}$ by selecting with the bootstrap algorithm, the 75% of the initial training samples in $\mathbf{X}_{t,k}$
- 6: transform $\hat{\mathbf{X}}_{t,k}$ to get the coefficients $v_{t,k}^{(1)}, \dots, v_{t,k}^{(M_k)}$ using PCA
- 7: **end for**
- 8: sparse matrix \mathbf{R}_t is composed of the above coefficients

$$\mathbf{R}_t = \begin{bmatrix} v_{t,1}^{(1)}, \dots, v_{t,1}^{(M_1)} & 0 & \dots & 0 \\ 0 & v_{t,2}^{(1)}, \dots, v_{t,2}^{(M_2)} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & v_{t,k}^{(1)}, \dots, v_{t,k}^{(M_K)} \end{bmatrix}$$

- 9: with respect to the original feature set, rearrange \mathbf{R}_t to \mathbf{R}_t^a
- 10: obtain the new training samples $\{\mathbf{X}\mathbf{R}_t^a, \mathbf{Y}\}$
- 11: initial the same weights to each sample $\omega_0 = \frac{1}{n}$.
- 12: **for** $j = 1 : J$ **do**
- 13: the error $\epsilon_j = \sum_{i=1}^n I(L_j(\mathbf{x}_i \mathbf{R}_t^a) \neq y_i) \omega_j(\mathbf{x}_i)$.
- 14: If $\epsilon_j > 0.5$ or $\epsilon_j = 0$, abort the loop.
- 15: $\beta_j = \frac{\epsilon_j}{1 - \epsilon_j}$.
- 16: **for** $i = 1 : n$ **do**
- 17: update the weights $\omega_j(\mathbf{x}_i) = \omega_j(\mathbf{x}_i) \times \beta_j$.
- 18: **end for**
- 19: normalize weights to make the total weights of ω_t equal to 1.
- 20: **end for**
- 21: add $L_t^* = \operatorname{argmax}_{j=1}^J \log \frac{1}{\beta_j} I(C_j(\mathbf{x} \mathbf{R}_t^a) = y)$ to the current ensemble \mathcal{L} .
- 22: **end for**

Prediction phase

Input: The ensemble $\mathcal{L} = \{L_t^*\}_{t=1}^T$. A new sample \mathbf{x}^* .

Output: class label y^*

- 1: $y^* = \operatorname{argmax}_{c \in \{1,2,\dots,C\}} \sum_{t: L_t^*(\mathbf{x}^*) = c} 1$

the classification performance [50].

Under this assumption, we propose the boosted rotation random forest (BRoRF) method. Compared to RotBoost in [50], we use random forests instead of decision trees as the base classifier in BRoRF. Algorithm 5 illustrates the main procedures. In BRoRF, the new training set is obtained by procedures used in RoRF (steps 1 to 9); then AdaBoost (steps 10 to 21) is used to manipulate the obtained training set to generate the ensemble. It should be noted that the individual classification results achieved by AdaBoost are combined to generate the final result in BRoRF, whereas the base classification results in BagRF, BoostRF, RSRF, and RoRF are obtained from random forest. Therefore, BRoRF promotes diversity without decreasing the accuracy of the member classifiers, which could be expected to improve the classification performance.

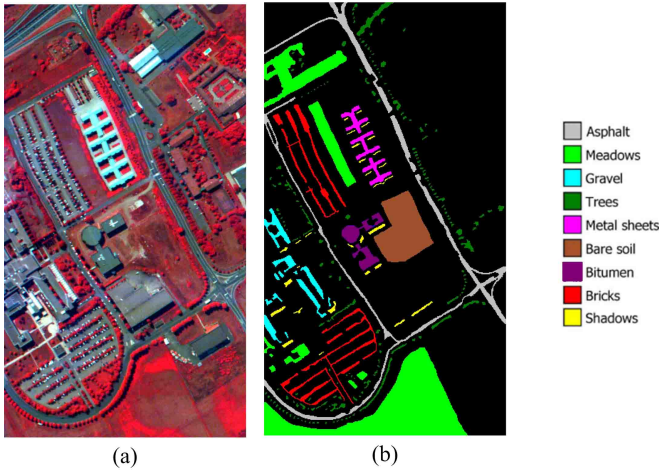


Fig. 3: (a) Three-band color composite image of ROSIS data. (b) Reference map.

III. EXPERIMENTAL SETTINGS AND RESULTS

A. Hyperspectral Datasets

Two well-known public hyperspectral datasets were used to evaluate the performance of the proposed method.

1) *Indian Pines AVIRIS*: The first dataset was captured by Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over northwestern Indiana in June 1992, with 220 spectral bands (wavelength range from 400 nm to 2500 nm) and a spatial resolution of 20 m/pixel. The whole scene (145 × 145) consists of 16 classes, ranging in size from 20 to 2468 pixels (shown in Table I). After removing some noisy bands, 200 bands were kept.

2) *University of Pavia ROSIS*: The second dataset was acquired over a university area in the city of Pavia, Italy by the Reflective Optics System Imaging Spectrometer (ROSIS), with 115 bands (wavelength range from 430 nm to 860 nm) and a very high spatial resolution of 1.3 m/pixel. It consists of 610 × 340 pixels. As with Indian Pines, 12 noisy bands were removed, and the remaining 103 bands were used in the classification. The reference data with nine classes of interest is composed of 42,776 pixels (shown in Table I). Fig. 3 shows the three-band color composite image and the reference map of the University of Pavia data.

B. Experimental settings

In this paper, two input datasets are used in the experiments: (1) with spectral bands, and (2) with EMEPs. The dimensionality of spectral features are 200 and 103 for Indian Pines AVIRIS and University of Pavia ROSIS images, respectively. The first three components extracted by ICA are used to construct the EMEPs with different extinction profiles, including area, diagonal of the bounding box, volume, height, and standard deviation, which consists of 213 features. The results reported are achieved by the mean of five Monte Carlo runs.

For the standard RF, \hat{T} is set to be 10, and the number of features in a subset is set to be the default value (square root

TABLE I: Indian Pines AVIRIS and University of Pavia ROSIS images: class name and number of train and test samples

No.	AVIRIS			ROSIS		
	Name	Train	Test	Name	Train	Test
1	Alfalfa	15	39	Bricks	514	3682
2	Corn-no till	50	1384	Shadows	231	947
3	Corn-min till	50	784	Metal Sheets	256	1345
4	Bldg-Grass-Tree-Drives	50	184	Bare Soil	532	5029
5	Grass/pasture	50	447	Trees	524	3064
6	Grass/trees	50	697	Meadows	540	18649
7	Grass/pasture-mowed	15	11	Gravel	392	2099
8	Corn	50	439	Asphalt	548	6631
9	Oats	15	5	Bitumen	375	1330
10	Soybeans-no till	50	918			
11	Soybeans-min till	50	2418			
12	Soybeans-clean till	50	564			
13	Wheat	50	162			
14	Woods	50	1244			
15	Hay-windrowed	50	330			
16	Stone-steel towers	50	45			

of the number of the features used). For all the ensembles, the number of RFs (T) is fixed to be 10.

The number of features in a subset (M) in RSRF is set to be one-half of the features used. Following the studies in [31] and [24], the number of features in a subset (M) in RoRF and BRoRF is set as follows. For the AVIRIS, M is set to be 100 and 3 for spectral features and EMEPs, respectively. For the ROSIS, M is set to be 10 and 3 for spectral features and EMEPs, respectively.

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

- Overall accuracy (OA): The percentage of correctly classified samples.
- Average accuracy (AA): Average percentage of correctly classified samples for an individual class.
- Kappa coefficient (κ): The percentage agreement corrected by the level of agreement that could be expected by chance alone.
- Average of OA (AOA): The percentage average overall accuracies of the individual member classifiers. The member classifier is the decision tree and RF for RF and RF ensembles, respectively.
- *Coincident Failure Diversity (CFD)* [51]: The measure of difference among the member classifiers. A higher value of *CFD* represents stronger diversity.
- CPU time (seconds): All methods were implemented in MATLAB on a computer with Intel Xeon 2 CPUs, 3.2 GHz, and 16 GB of memory. RF implementation was downloaded from the website <http://code.google.com/p/randomforest-matlab/>. Please note that the CPU processing time of EMEPs construction in RF ensemble is not included in this manuscript.

In addition, we present McNemar's test (Z) to compare the classification results. McNemar's test is calculated by

$$Z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}}, \quad (6)$$

where f_{12} means the number of samples correctly classified by classifier 1 and the number incorrectly classified by classifier 2. If $Z > 0$, then classifier 1 is more accurate than classifier 2. The difference between classifiers 1 and 2 is statistically

significant if $|Z| > 1.96$.

C. Experimental Results

1) *Indian Pines AVIRIS image*: The low spatial resolution and the presence of highly mixed pixels in this scene make the classification of this dataset a challenging task.

Table II provides the OAs, AAs, κ with standard deviations, class-specific accuracies, and computational time for RF and its ensembles.

The OA of RF with spectral information is only 62.38%. This reveals that the RF classifier is not sufficient to discriminate between different classes when using hundreds of spectral bands as the input. All the RF ensemble classifiers yield better performance than the individual RF classifier. In particular, the proposed BRoRF shows the best results, where the OA and AA are 73.60% and 81.45%, which are 11.22 and 8.15 percentage points higher, respectively, than those of RF.

The accuracies in Table II involving the EMEPs are significantly better than those achieved by only the spectral information, demonstrating the effectiveness of EMEPs. The OA of RF is 90.31%. BagRF and BoostRF cannot improve on this performance. However, their AAs are better than that of RF. The proposed BRoRF gains the highest accuracies. However, the difference in OA between BRoRF and RF is reduced to 1.93 percentage points. Boosting techniques and data transformation in the rotation-based ensemble took a longer computation time than those of other algorithms. The computational time of BRoRF is at least 10 times longer than other RF ensembles. However, the CPU processing time is acceptable (about 48s).

Generally, it is time-consuming and expensive to prepare training samples for remote sensing data. Therefore, it is of great interest for users to see generalization performance of classifiers using a limited training set. In order to assess the effectiveness of the RF ensemble for a limited training set, we randomly extracted 15 samples per class from the training set. The results are shown in Table III. The proposed BRoRF still shows the best performance.

With respect to the analysis of AOA and CFD diversities in Tables II and III, it can be seen that the main reason for BRoRF's superior performance is that it provides higher values of AOA and CFD diversities. We also provide McNemar's test between the BRoRF and RF or other RF ensembles in Table IV. As can be seen, the BRoRF method shows significant improvement compared to the other methods, except in the case of BRoRF and RoRF with EMEPs using the standard training set. Another observation from Tables II and III is that BRoRF shows stable results (with smaller standard deviations) than other RF ensembles and the RF classifier.

2) *University of Pavia ROSIS image*: Table V provides information related to the classification accuracies of the RF and its ensembles, as well as the CPU processing time. The corresponding classification maps can be found in Fig. 4. It can be seen from Table V that spectral classification with RoRF and BRoRF improves the OA of RF by about 8 and 9 percentage points, respectively. Specifically, the class accuracies of *bare soil* and *meadows* are significantly improved.

By using spectral classification (see the top of Fig. 4), many samples of *meadows* and *bare soil* are misclassified as one another. The class *gravel* is incorrectly classified and confused with the similar classes *bricks* and *asphalt*. When the EMEPs are included in the classification process, the classification results of these classes are greatly improved, as shown at the bottom of Fig. 4. The findings are similar to those in the AVIRIS image. RoRF and BRoRF share the top positions. The proposed BRoRF is slightly better than RoRF. Also, the major accuracy difference between RoRF and BRoRF occurs for class *gravel*. The computation time of RoRF and BRoRF is about 20s and 350s. The use of other ensemble methods, such as BagRF, BoostRF, and RSRF, also leads to high accuracies.

The experimental results shown in Table VI on very limited training samples (5% samples from the training set) demonstrate the superior performance of the proposed methods. As shown in Tables V and VI, the proposed BRoRF produces higher values of AOAs and diversities than other methods, leading to better classification results. Also, the BRoRF have shown significantly better and more stable results than other RF ensembles (as seen in Table VII).

From the above results, the pros and cons of the RF ensembles chosen for this study are as follows:

- 1) BagRF slightly improves the performance over RF and the computational time is short.
- 2) BoostRF provides better results than BagRF. However, it requires longer computational time than BagRF. BagRF and BoostRF only define the number of RFs (T).
- 3) RSRF achieves results that are comparable to BoostRF's results; the computational time is shorter than that of BagRF.
- 4) RoRF gains better results than BagRF, BoostRF, and RSRF. The computational time is slightly longer than that of BoostRF.
- 5) BRoRF yields the best results among the RF ensembles. However, the computational time is about 17 times longer than that of RoRF. In addition to T , we should set the value of M in RSRF, RoRF, and BRoRF.

D. Study of Effects on Parameter Selection

In this work, the number of trees in RF (\hat{T}), the number of RFs in the ensemble (T), and the number of features in a subset (M) are the essential parameters of the proposed method. The effects of \hat{T} and T on the EMEPs are shown in Figs. 5 and 6. Fig. 5 shows that OA increases as \hat{T} increases, although good performance is already achieved with $\hat{T} = 10$. Note that, in the previous subsection, we adopted $\hat{T} = 10$ to reduce the computational burden.

From Fig. 6, we observe that the proposed methods are not sensitive to values of \hat{T} when it is greater than 10. Fig. 7 presents the influence of M on the classification performance. In all cases, RSRF tends to have better performance with a larger value of M . For the AVIRIS image, RoRF (BRoRF) with spectral information yields better performance with larger values of M . The optimal value is 100 in the studied range. For the ROSIS image, RoRF (BRoRF) with spectral information first increases the OA and then decreases it when the value of

TABLE II: Classification accuracies obtained for the Indian Pines AVIRIS image using different ensemble classifiers.

Class	Spectral						EMEPs					
	RF	BagRF	BoostRF	RSRF	RoRF	BRORF	RF	BagRF	BoostRF	RSRF	RoRF	BRORF
1	70.97	83.87	80.65	70.97	80.65	80.65	82.37	93.55	93.55	93.55	93.55	93.55
2	46.52	45.72	37.45	47.17	61.39	59.87	85.13	78.66	76.71	82.58	85.27	82.87
3	51.15	50.38	52.31	53.85	58.46	57.05	95.72	90.38	82.95	86.92	88.21	89.36
4	67.38	68.98	69.52	67.91	87.17	88.24	96.30	97.33	97.33	97.33	96.79	96.79
5	82.68	84.99	86.37	85.91	83.14	87.07	89.85	97.00	96.07	96.54	96.54	96.30
6	80.44	86.18	86.32	87.35	87.79	88.53	84.62	89.56	91.62	91.47	92.06	92.94
7	84.62	76.92	84.62	84.62	84.62	84.62	84.62	92.31	92.31	92.31	92.31	92.31
8	90.42	89.49	89.95	92.52	95.09	95.79	100.00	100.00	100.00	100.00	100.00	100.00
9	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
10	61.39	67.14	58.57	73.43	74.62	72.78	82.75	80.26	82.43	83.08	82.00	82.00
11	48.52	62.25	61.00	62.49	61.95	64.49	89.52	90.02	91.60	91.52	92.77	93.93
12	56.54	60.04	64.09	65.56	84.53	81.40	91.71	93.92	93.92	93.37	93.92	95.40
13	90.97	94.19	93.55	92.90	94.19	94.84	99.35	99.35	99.35	99.35	99.35	99.35
14	87.08	86.91	88.97	88.07	90.86	92.84	99.18	99.18	99.34	99.18	99.18	99.18
15	54.17	47.92	52.98	57.44	56.85	55.06	99.40	99.11	99.70	99.70	100.00	99.70
16	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
OA	62.38	66.76	65.34	68.90	73.17	73.60	90.31	90.27	90.13	91.30	92.08	92.24
	± 2.95	± 2.02	± 1.45	± 1.21	± 0.74	± 0.34	± 0.91	± 0.82	± 0.76	± 0.75	± 0.74	± 0.51
AA	73.30	75.31	75.40	76.89	81.33	81.45	92.89	93.79	93.55	94.18	94.50	94.61
	± 2.34	± 2.11	± 2.07	± 1.96	± 0.88	± 0.62	± 0.74	± 0.89	± 0.64	± 0.63	± 0.51	± 0.45
κ	57.70	62.32	60.78	64.80	69.67	70.09	88.91	88.87	88.71	90.05	90.93	91.12
	± 2.87	± 1.96	± 1.47	± 1.24	± 0.81	± 0.37	± 0.86	± 0.83	± 0.71	± 0.78	± 0.65	± 0.47
AOA	54.24	57.47	55.34	61.31	64.38	64.42	87.28	86.25	87.12	88.28	89.37	89.54
CFD	0.44	0.48	0.46	0.50	0.56	0.57	0.67	0.64	0.66	0.68	0.70	0.71
Time (s)	0.39	2.48	4.23	1.84	5.14	48.92	0.31	2.24	3.80	1.61	4.89	47.48

TABLE III: Classification accuracies obtained for the Indian Pines AVIRIS image using different ensemble algorithms with a very limited training set.

Class	Spectral						EMEPs					
	RF	BagRF	BoostRF	RSRF	RoRF	BRORF	RF	BagRF	BoostRF	RSRF	RoRF	BRORF
OA	51.17	55.64	56.82	56.67	60.87	62.79	83.78	84.62	85.17	84.22	86.09	88.11
	± 1.78	± 1.75	± 1.46	± 1.28	± 1.74	± 1.42	± 1.84	± 1.65	± 1.57	± 1.48	± 1.14	± 1.01
AA	63.73	67.73	68.14	68.02	73.04	74.28	88.78	89.15	89.24	89.01	90.15	90.72
	± 1.63	± 1.12	± 1.96	± 1.81	± 1.15	± 1.12	± 1.78	± 1.62	± 1.54	± 1.64	± 1.12	± 1.01
κ	45.26	50.98	51.02	52.11	54.28	56.17	78.11	79.14	79.67	79.01	80.18	81.54
	± 1.84	± 1.54	± 1.14	± 1.01	± 1.76	± 1.04	± 1.51	± 1.63	± 1.52	± 1.24	± 1.01	± 1.04
AOA	43.64	48.18	48.21	49.98	52.17	53.11	79.14	80.89	81.74	81.27	83.11	83.52
CFD	0.40	0.44	0.44	0.44	0.46	0.48	0.55	0.57	0.58	0.57	0.61	0.63

TABLE IV: Statistic of McNemar's test for Indian Pines AVIRIS image using the RF ensembles.

Z	Standard training set		Limited training set	
	Spectral	EMEPs	Spectral	EMEPs
BRORF vs. RF	17.28	5.36	20.84	10.15
BRORF vs. BagRF	13.85	5.98	19.65	8.27
BRORF vs. BoostRF	14.98	60.21	10.25	6.12
BRORF vs. RSRF	7.64	4.23	9.32	8.25
BRORF vs. RoRF	1.98	1.54	4.11	3.02

M increases. Thus, the optimal value is 10. However, when we include EMEPs in the classification process, a small value of M is preferred for the RoRF (BRORF). Indeed, the better performance is achieved by the above-mentioned parameters (i.e., T , \hat{T} and M) with higher diversity. The influences of parameters on the performance can be deeply analyzed by using the evolution of CFD diversities, as we did in [52]. Finally, we can conclude the general recommendation of M

for the new dataset:

- 1) for the RF classifier, the default value of M is recommended;
- 2) for the RSRF, M is recommended to be set to one-half of the number of features used;
- 3) for the RoRF (BRORF) with spectral information, one-half of the number of the features used or a small number (e.g., 10) is recommended to configure M ;
- 4) for the RoRF (BRORF) with EMEPs, a very small value (e.g., 3) of M is recommended.

E. Comparisons with State-of-the-Art Methods

In this section, we compare the proposed method with some state-of-the-art algorithms in terms of OAs and AAs in order to demonstrate the capability of the combination of RF ensembles (especially the RoRF and BRORF) and EMEPs. Tables VIII and IX report the OAs and AAs for state-of-the-art methods, such as SVM+WH [53], SVMMSF [54], SVMRF-E [29],

TABLE V: Classification accuracies obtained for University of Pavia ROSIS image using different ensemble classification algorithms.

Class	Spectral						EMEPs					
	RF	BagRF	BoostRF	RSRF	RoRF	BRoRF	RF	BagRF	BoostRF	RSRF	RoRF	BRoRF
1	89.57	91.55	90.66	90.39	93.07	93.86	99.08	99.51	99.38	99.57	99.38	99.54
2	97.15	98.63	98.52	98.31	99.58	99.79	100.00	100.00	100.00	99.89	100.00	100.00
3	97.92	99.03	98.81	98.96	99.48	99.48	99.93	99.93	99.93	99.93	99.93	99.93
4	78.25	76.62	76.32	79.08	95.11	94.97	90.28	96.92	98.25	98.01	98.53	98.87
5	98.73	98.69	98.53	98.60	98.56	98.60	96.51	99.09	99.12	98.83	99.58	99.15
6	56.13	55.12	57.96	54.81	64.86	67.01	91.44	93.77	94.83	93.88	94.73	94.18
7	53.26	50.88	52.17	51.55	56.69	57.22	87.37	83.52	85.37	85.28	85.33	92.81
8	78.21	79.31	79.69	80.12	85.24	84.30	95.84	96.37	96.38	96.55	96.38	96.76
9	84.36	85.11	87.44	84.44	89.62	90.53	99.92	99.85	99.85	99.92	99.85	99.85
OA	70.72 ±0.98	71.04 ±0.72	72.02 ±0.63	70.90 ±0.64	78.64 ±0.52	79.55 ±0.41	93.52 ±0.45	95.43 ±0.41	96.14 ±0.38	95.71 ±0.37	96.16 ±0.31	96.36 ±0.29
AA	81.51 ±0.82	81.66 ±0.74	82.23 ±0.71	81.81 ±0.65	86.91 ±0.52	87.31 ±0.48	95.60 ±0.64	96.55 ±0.52	97.01 ±0.53	96.87 ±0.48	97.08 ±0.41	97.90 ±0.40
κ	64.02 ±0.94	64.40 ±0.75	65.45 ±0.66	64.30 ±0.62	73.50 ±0.50	74.53 ±0.43	91.54 ±0.41	94.02 ±0.43	94.93 ±0.37	94.38 ±0.39	94.96 ±0.33	95.24 ±0.30
AOA	64.89	65.78	66.17	65.01	72.89	73.08	90.87	92.27	93.64	93.01	94.12	94.28
CFD	0.51	0.52	0.53	0.51	0.60	0.61	0.73	0.75	0.76	0.75	0.76	0.77
Time (s)	1.18	11.15	27.52	9.02	17.24	300.35	1.59	15.12	32.34	11.45	20.56	354.15

TABLE VI: Classification accuracies obtained for University of Pavia ROSIS image using different ensemble algorithms with a very limited training set.

Class	Spectral						EMEPs					
	RF	BagRF	BoostRF	RSRF	RoRF	BRoRF	RF	BagRF	BoostRF	RSRF	RoRF	BRoRF
OA	60.25 ±1.28	61.75 ±1.27	62.59 ±1.18	61.27 ±1.19	69.71 ±1.07	70.82 ±1.05	85.14 ±0.98	85.99 ±0.95	86.28 ±0.93	86.29 ±0.92	89.96 ±0.90	91.02 ±0.88
AA	72.52 ±1.15	73.05 ±1.16	73.82 ±1.11	73.08 ±1.07	80.24 ±1.07	81.78 ±1.02	89.17 ±0.82	90.25 ±0.81	90.38 ±0.82	90.26 ±0.79	92.17 ±0.64	93.24 ±0.61
κ	55.14 ±1.45	56.73 ±1.33	56.99 ±1.37	56.01 ±1.29	63.49 ±0.89	64.28 ±1.01	81.69 ±0.99	82.12 ±0.97	82.73 ±0.96	82.78 ±0.95	85.01 ±0.88	86.72 ±0.89
AOA	53.21	54.04	54.78	55.12	63.78	64.25	81.21	82.01	81.99	82.07	83.82	84.17
CFD	0.41	0.44	0.44	0.45	0.48	0.50	0.63	0.65	0.66	0.66	0.68	0.70

TABLE VII: Statistics of McNemar's test of University of Pavia ROSIS image using the RF ensembles.

Z	Standard training set		Limited training set	
	Spectral	EMEPs	Spectral	EMEPs
BRoRF vs. RF	37.25	10.24	32.16	20.54
BRoRF vs. BagRF	32.67	7.51	29.62	18.62
BRoRF vs. BoostRF	28.64	3.26	25.31	15.64
BRoRF vs. RSRF	35.64	5.92	27.53	17.28
BRoRF vs. RoRF	7.24	2.81	6.84	10.12

TABLE VIII: Comparisons with state-of-the-art approaches for Indian Pines AVIRIS image.

Methods	Standard training set		Limited training set	
	OA	AA	OA	AA
RoRF+EMEPs	92.08	94.50	86.09	90.15
BRoRF+EMEPs	92.24	94.61	88.11	90.72
SVM+WH [53]	86.63	91.61	71.24	75.67
SVMMSF [54]	91.80	94.28	72.78	77.38
SVMRF-E [29]	91.83	95.69	73.12	75.83
RoF-MRF [31]	90.45	94.17	74.18	82.42
RoRF+EMPs [55]	87.32	91.38	80.11	84.56
RoRF+EMAPs [56]	91.34	93.25	85.14	86.21

RoF-MRF [31], RoRF+EMPs [55], and RoRF+EMAPs [56], in both standard and limited training set cases. Please refer to the original references for a better understanding of the compared methods. It can be seen from the table that the proposed methods provide competitive classification results compared to the other investigated methods.

Since there is no need to initialize the parameters in the construction of EMEPs, users might select a very small value (e.g., 3) of M , as in our case, to achieve better performance because higher diversity is obtained.

IV. CONCLUSIONS

In this paper, we developed a framework to classify hyper-spectral images. Our framework uses random forest ensembles for the classification with spatial information modeled by EMEPs. The proposed methods were tested on two benchmark hyperspectral datasets: Indian Pines AVIRIS and University of Pavia ROSIS images. Different strategies were used to construct RF ensembles, and the results compared in terms of classification accuracies and CPU time.

Experimental results indicate good generalization perfor-

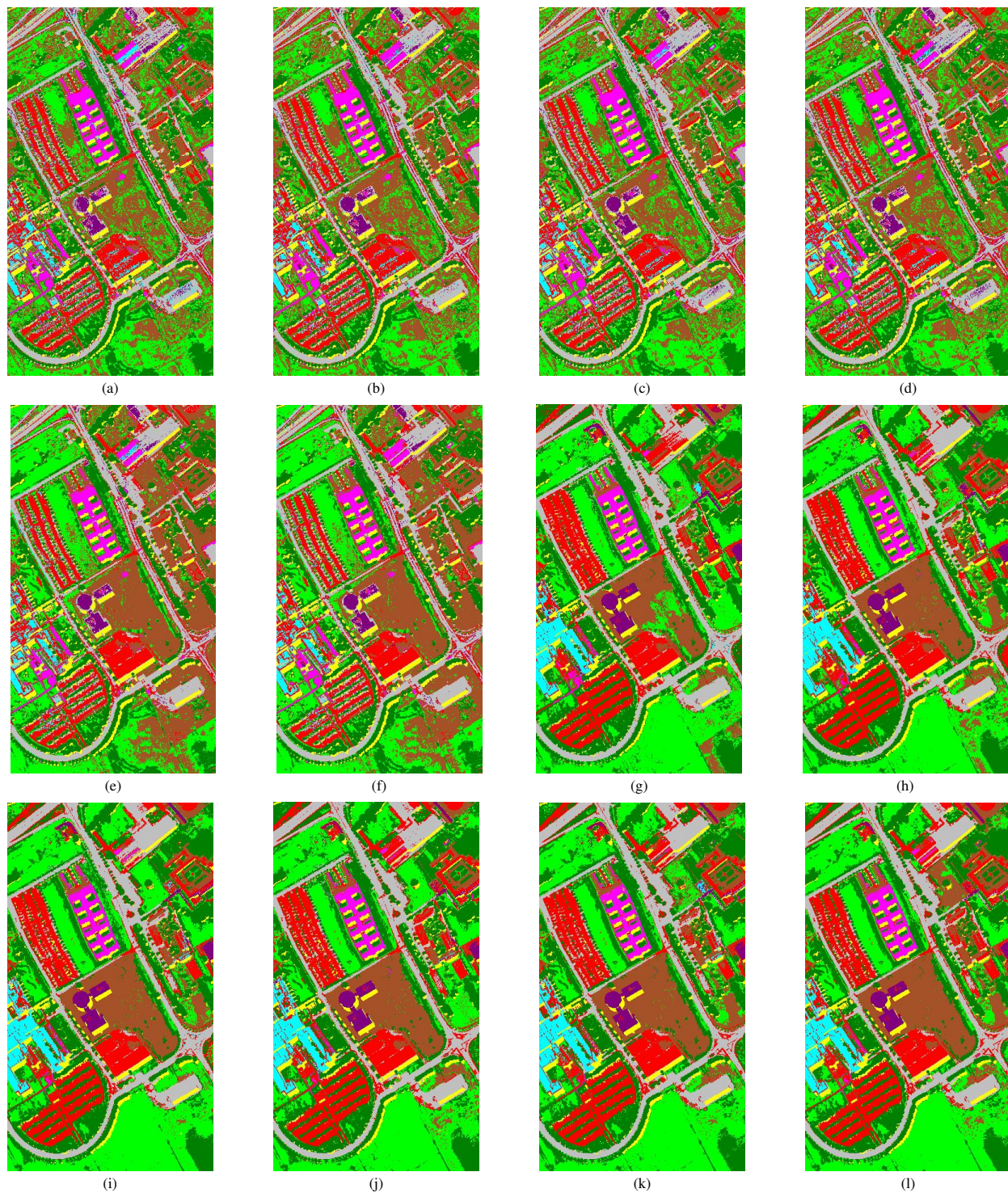


Fig. 4: Classification maps of University of Pavia ROSIS image using RF, BagRF, BoostRF, RSRF, RoRF and BRoRF with spectral features (a)-(f) and EMEPs (g)-(l).

mance of RF ensembles on EMEPs features. Particular attention should be paid to the BRoRF classifier, which shows the best results due to its property of generating the ensemble with high accuracy of member classifiers as well as diversity. Moreover, the proposed methods achieve better performance than other spectral-spatial classifiers with accepted CPU process time.

In our future studies, we will apply the proposed methodology to other types of data (e.g., multispectral remote sensing data) and extend our framework for the classification of multi-sensor datasets, i.e., hyperspectral and LiDAR datasets. In addition, parallel implementations will be introduced in our framework to further decrease CPU processing times.

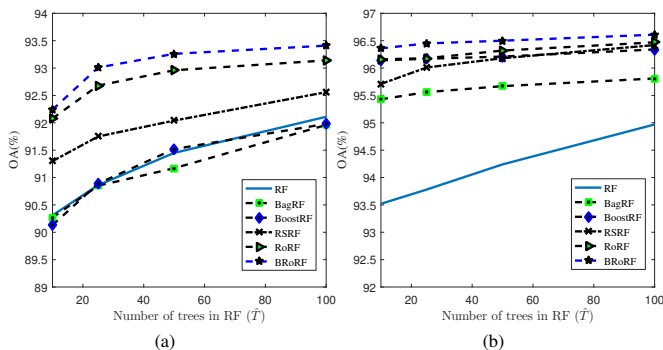


Fig. 5: Sensitivity to the change of number of trees in the RF (T) for (a) Indian Pines AVIRIS and (b) University of Pavia data with EMEPs.

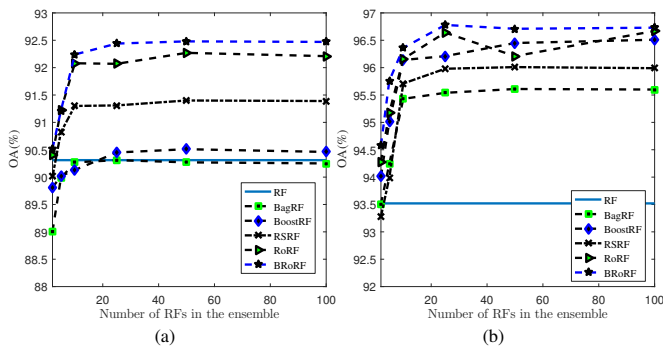


Fig. 6: Sensitivity to the change of number of RFs in the ensemble (T) for (a) Indian Pines AVIRIS and (b) University of Pavia data with EMEPs.

TABLE IX: Comparisons with state-of-the-art approaches for University of Pavia ROSIS image.

Methods	Standard training set		Limited training set	
	OA	AA	OA	AA
RoRF+EMEPs	96.16	97.08	89.95	92.17
BRoRF+EMEPs	96.36	97.90	91.02	93.24
SVM+WH [53]	85.42	91.31	72.48	82.59
SVMMSF [54]	91.08	94.76	75.96	83.67
SVMRF-E [29]	87.63	93.41	76.21	84.52
RoF-MRF [31]	92.15	93.06	80.29	87.68
RoRF+EMPs [55]	90.20	92.76	84.21	88.96
RoRF+EMAPs [56]	95.47	97.02	88.24	91.24

ACKNOWLEDGMENTS

The authors would like to thank Prof. P. Gamba and Prof. D. A. Landgrebe for kindly providing the hyperspectral remote sensing images. The authors gratefully appreciate the work of the Associate Editor and the Anonymous Reviewers who evaluated the manuscript and provided very relevant and constructive comments and suggestions.

REFERENCES

- [1] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proceedings of the IEEE*, vol. 101, no. 3, pp. 652–675, 2013.
- [2] C. I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. Plenum Publishing Co., 2003.
- [3] —, *Hyperspectral Data Exploitation: Theory and Applications*. Wiley-Interscience, Hoboken, NJ, 2007.

- [4] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inform. Theory*, vol. 14, no. 1, pp. 55–63, 1968.
- [5] 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, 2004.
- [6] 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, 2005.
- [7] W. Liao, A. Pizurica, P. Scheunders, W. Philips, and Y. Pi, "Semisupervised local discriminant analysis for feature extraction in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 184–198, 2013.
- [8] M. Pal and G. M. Foody, "Feature selection for classification of hyperspectral data by svm," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2297–2307, 2010.
- [9] V. N. Vapnik, *The Nature of Statistical Learning Theory*. Springer, 1995.
- [10] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geos. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct 2016.
- [11] P. Ghamisi, Y. Chen, and X. X. Zhu, "A self-improving convolution neural network for the classification of hyperspectral data," *IEEE Geosci. Remote Sensing Lett.*, vol. 13, no. 10, pp. 1537–1541, Oct 2016.
- [12] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [13] J. Ham, Y. Chen, M. M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 492–501, 2005.
- [14] M. Immitzer, C. Atzberger, and T. Koukal, "Tree species classification with Random Forest using very high spatial resolution 8-band worldview-2 satellite data," *Remote Sensing*, vol. 4, no. 9, pp. 2661–2693, 2012.
- [15] L. Guo, N. Chehata, C. Mallet, and S. Boukira, "Relevance of airborne lidar and multispectral image data for urban scene classification using random forests," *ISPRS J. Photogramm.*, vol. 66, no. 1, pp. 56–66, Jan. 2011.
- [16] C. Pelletier, S. Valero, J. Inglada, N. Champion, and G. Dedieu, "Assessing the robustness of random forests to map land cover with high resolution satellite image time series over large areas," *Remote Sensing of Environment*, vol. 187, pp. 156 – 168, 2016.
- [17] E. Tuv, A. Borisov, G. C. Runger, and K. Torkkola, "Feature selection with ensembles, artificial variables, and redundancy elimination," *Journal of Machine Learning Research*, vol. 10, pp. 1341–1366, 2009.
- [18] S. Schuster, P. Wohlhart, C. Leistner, A. Saffari, P. M. Roth, and H. Bischof, "Alternating decision forests," in *IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, June 23-28, 2013*, pp. 508–515.
- [19] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, Aug. 1996.
- [20] Y. Freund and R. E. Schapire, "Experiments with a new Boosting algorithm," in *International Conference on Machine Learning*, Bari, Italy, 1996, pp. 148–156.
- [21] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832–844, August 1998.
- [22] J. J. Rodriguez, L. I. Kuncheva, and C. J. Alonso, "Rotation forest: A new classifier ensemble method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1619–1630, Oct. 2006.
- [23] J. Xia, P. Du, X. He, and J. Chanussot, "Hyperspectral remote sensing image classification based on rotation forest," *IEEE Geosci. Remote Sensing Lett.*, vol. 11, no. 1, pp. 239 – 243, 2014.
- [24] J. Xia, M. Dalla Mura, J. Chanussot, P. Du, and X. He, "Random subspace ensembles for hyperspectral image classification with extended morphological attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 4768–4786, 2015.
- [25] J. Xia, N. Falco, J. A. Benediktsson, P. Du, and J. Chanussot, "Hyperspectral image classification with rotation random forest via kpc." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. in press, 2017.
- [26] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 36, no. 2, pp. 192–236, 1974.
- [27] Y. Boykov, O. Veksler, and Z. R., "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, 2001.

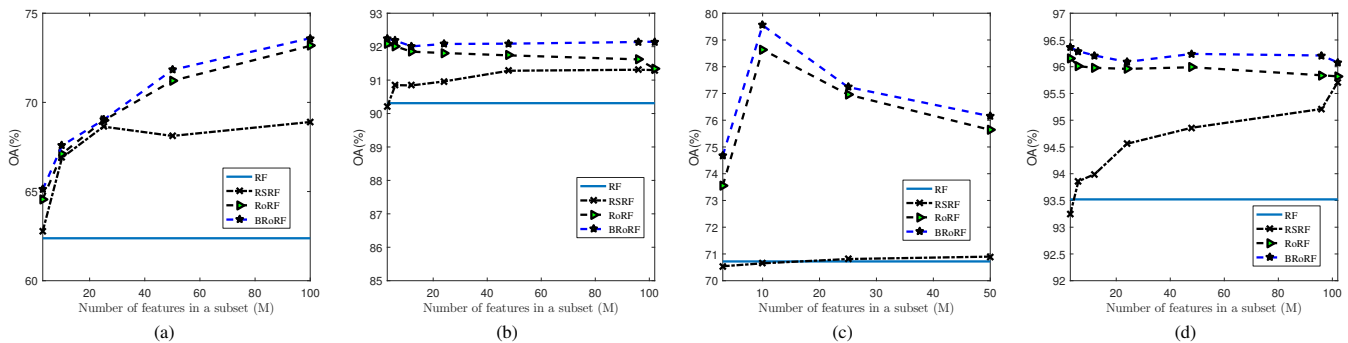


Fig. 7: Sensitivity to the number of features in a subset (M) for Indian Pines AVIRIS image using (a) only spectral information and (b) with EMEPs. Sensitivity to the number of features in a subset (M) for University of Pavia ROSIS image using (c) only spectral information and (d) with EMEPs.

- [28] P. Ghamisi, J. A. Benediktsson, and M. O. Ulfarsson, "Spectral-spatial classification of hyperspectral images based on hidden Markov random fields," *IEEE Trans. Remote Sens. Geos.*, vol. 52, no. 5, pp. 2565–2574, 2014.
- [29] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson, "SVM and MRF-based method for accurate classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens. Letters*, vol. 7, no. 4, pp. 736–740, 2010.
- [30] 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, 2012.
- [31] 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, 2015.
- [32] 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, 2001.
- [33] J. A. Benediktsson, S. Member, M. Pesaresi, and K. Arnason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 3, pp. 1940–1949, 2003.
- [34] 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, 2010.
- [35] —, "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.
- [36] P. Ghamisi, M. Dalla Mura, and J. A. Benediktsson, "A survey on spectral-spatial classification techniques based on attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2335–2353, 2015.
- [37] J. A. Benediktsson and P. Ghamisi, *Spectral-Spatial Classification of Hyperspectral Remote Sensing Images*. Chapters 5 and 6, Artech House Publishers, INC, Boston, USA, 2015.
- [38] P. Ghamisi, R. Souza, J. A. Benediktsson, X. X. Zhu, L. Rittner, and R. Lotufo, "Extinction profiles for the classification of remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 5631–5645, 2016.
- [39] P. Ghamisi, R. Souza, J. A. Benediktsson, L. Rittner, R. Lotufo, and X. X. Zhu, "Hyperspectral data classification using extended extinction profiles," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 11, pp. 1641–1645, 2016.
- [40] A. Hyvarinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13, no. 4-5, pp. 411–430, Jun. 2000.
- [41] P. Salembier, A. Oliveras, and L. Garrido, "Antiextensive connected operators for image and sequence processing," *IEEE Transactions on Image Processing*, vol. 7, no. 4, pp. 555–570, 1998.
- [42] C. Vachier, "Extinction value: a new measurement of persistence," in *IEEE Work. Nonlinear Sig. Image Proc.*, vol. I, 1995, pp. 254–257.
- [43] P. Soille, *Morphological Image Analysis: Principles and Applications*, 2nd ed. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2003.
- [44] Y. Xu, T. Géraud, and L. Najman, "Morphological filtering in shape spaces: Applications using tree-based image representations," in *2012 21st International Conference on Pattern Recognition (ICPR)*, 2012, pp. 485–488.
- [45] P. Monasse and F. Guichard, "Fast computation of a contrast-invariant image representation," *IEEE Trans. Image Process.*, vol. 9, no. 5, pp. 860–872, 2000.
- [46] R. Souza, L. Rittner, R. Lotufo, and R. Machado, "An array-based node-oriented max-tree representation," in *ICIP'15*, Sept 2015, pp. 3620–3624.
- [47] R. Souza, L. Rittner, R. Machado, and R. Lotufo, "A comparison between extinction filters and attribute filters," in *ISMM'15, Reykjavik, Iceland, May 27-29, 2015. Proceedings*, 2015, pp. 63–74.
- [48] G. Stiglic, J. J. Rodriguez, and P. Kokol, "Rotation of random forests for genomic and proteomic classification problems," *Software Tools and Algorithms for Biological Systems*, vol. 696, pp. 211–221, 2011.
- [49] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, 2004.
- [50] C. Zhang and J. Zhang, "Rotboost: A technique for combining rotation forest and adaboost," *Pattern Recognition Letters*, vol. 29, no. 10, pp. 1524–1536, 2008.
- [51] L. I. Kuncheva and C. J. Whitaker, "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy," *Machine Learning*, vol. 51, no. 2, pp. 181–207, 2003.
- [52] J. Xia, J. Chanussot, P. Du, and X. He, "Rotation-based support vector machines in classification of hyperspectral data with limited training samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1519–1531, 2016.
- [53] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using watershed transformation," *Pattern Recognition*, vol. 43, no. 7, pp. 2367–2379, 2010.
- [54] —, "Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers," *IEEE Trans. Systems, Man, and Cybernetics, Part B*, vol. 40, no. 5, pp. 1267–1279, 2010.
- [55] Y. Tarabalka, *Classification of Hyperspectral Data Using Spectral-Spatial Approaches*. PhD thesis, University of Iceland and Grenoble Institute of Technology, 2010.
- [56] 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 Sensing Lett.*, vol. 8, no. 3, pp. 542–546, 2011.



Junshi Xia (S'11–M'16) 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. He also received the Ph.D. degree in image processing with the Grenoble Images Speech Signals and Automatics Laboratory, Grenoble Institute of Technology, Grenoble, France, in 2014.

From December 2014 to April 2015, he was a Visiting Scientist with the Department of Geographic Information Sciences, Nanjing University, Nanjing, China. From May 2015 to April 2016, he was a Postdoctoral Research Fellow with the University of Bordeaux, Talence, France. From May 2016, he is the JSPS Postdoctoral Oversea Research Fellow with the University of Tokyo, Tokyo, Japan. His research interests include multiple classifier system in remote sensing, hyperspectral remote sensing image processing, and urban remote sensing.

In 2017, Dr. Xia won the first place in the IEEE Geoscience and Remote Sensing Society (GRSS) Data Fusion Contest organized by the Image Analysis and Data Fusion Technical Committee (IADF TC).



Pedram Ghamisi (S'12, M'15) graduated with a B.Sc. in civil (survey) engineering from the Tehran South Campus of Azad University. He obtained an M.Sc. degree with first class honors in remote sensing at K.N.Toosi University of Technology in 2012. In 2013/2014, he spent seven months at the school of Geography, Planning and Environmental Management, the University of Queensland, Australia. He received a Ph.D. in electrical and computer engineering at the University of Iceland, Reykjavik in 2015 and subsequently worked as a postdoctoral

research fellow at the University of Iceland. In 2015, Dr. Ghamisi won the prestigious Alexander von Humboldt Fellowship and started his work as a postdoctoral research fellow at Technical University of Munich (TUM) and Heidelberg University, Germany from October 2015. He has also been working as a researcher at German Aerospace Center (DLR), Remote Sensing Technology Institute (IMF), Germany since October 2015. His research interests are in remote sensing and image analysis, with a special focus on spectral and spatial techniques for hyperspectral image classification, multisensor data fusion, machine learning, and deep learning.

In the 2010-11 academic year, he received the Best Researcher Award for M.Sc. students in K. N. Toosi University of Technology. In July, 2013, he presented at the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Melbourne, and was awarded the IEEE Mikio Takagi Prize for winning the conference Student Paper Competition against almost 70 people. In 2016, he was selected as a talented international researcher by Iran's National Elites Foundation. In 2017, he won the Data Fusion Contest 2017 organized by the Image Analysis and Data Fusion Technical Committee (IADF) of the Geoscience and Remote Sensing Society (IEEE-GRSS). His model was the most accurate among more than 800 submissions. Additional information: <http://pedram-ghamisi.com/>



Naoto Yokoya (S10-M13) received the M.Eng. and Ph.D. degrees in aerospace engineering from the University of Tokyo, Tokyo, Japan, in 2010 and 2013, respectively. From 2012 to 2013, he was a Research Fellow with Japan Society for the Promotion of Science, Tokyo, Japan. Since 2013, he is an Assistant Professor with the University of Tokyo. Since 2015, he is also an Alexander von Humboldt Research Fellow with the German Aerospace Center (DLR), Oberpfaffenhofen, and Technical University of Munich (TUM), Munich, Germany. His research

interests include image analysis and data fusion in remote sensing.

Dr. Yokoya won the first place in the 2017 IEEE Geoscience and Remote Sensing Society (GRSS) Data Fusion Contest organized by the Image Analysis and Data Fusion Technical Committee (IADF TC) among more than 800 submissions. In 2016, he was a Guest Editor for the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (JSTARS) and in 2017 a Guest Editor for Remote Sensing. He was the Program Chair of the 7th IEEE GRSS Workshop on Hyperspectral Image and Signal Processing Evolution in Remote Sensing (WHISPERS) in 2015 and also on the technical committee of the 8th IEEE GRSS WHISPERS in 2016. Since 2017, he is a Co-chair of the IEEE GRSS IADF TC.



Akira Iwasaki received the M. Sc. degree in aerospace engineering from the University of Tokyo in 1987. He joined Electrotechnical Laboratory in 1987 where he engaged in research on space technology and remote sensing system. He received Doctoral degree of Engineering from the University of Tokyo in 1996. He is currently the professor in the University of Tokyo. He is also the project leader of the Japanese hyperspectral sensor, Hyperspectral Imager SUite (HISUI) that will be on the International Space Station.

In 2014, he won the best reviewers award of the IEEE Transactions of Geoscience and Remote Sensing. He was a Guest Editor for the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (JSTARS). He was the General Chair of the 7th IEEE GRSS Workshop on Hyperspectral Image and Signal Processing Evolution in Remote Sensing (WHISPERS) in 2015.