

Article

# Kernel Supervised Ensemble Classifier for the Classification of Hyperspectral Data Using Few Labeled Samples

Jike Chen <sup>1</sup>, Junshi Xia <sup>2,3</sup>, Peijun Du <sup>1,\*</sup>, Jocelyn Chanussot <sup>4,5</sup>, Zhaohui Xue <sup>6</sup> and Xiangjian Xie <sup>1</sup>

<sup>1</sup> Key Laboratory for Satellite Mapping Technology and Applications of State Administration of Surveying, Mapping and Geoinformation of China, Nanjing University, 210093 Nanjing, China; Jikechen.rs@gmail.com (J.C.); xiexjrs@gmail.com (X.X.)

<sup>2</sup> Intégration, du Matériau au Système (IMS), Université de Bordeaux, UMR 5218, F-33405 Talence, France; xiajunshi@gmail.com

<sup>3</sup> Intégration, du Matériau au Système (IMS), Centre National de la Recherche Scientifique (CNRS), UMR 5218, F-33405 Talence, France

<sup>4</sup> Grenoble-Image-speech-Signal-Automatics Lab (GIPSA)-lab, Grenoble Institute of Technology, 38400 Grenoble, France; jocelyn.chanussot@gipsa-lab.grenoble-inp.fr

<sup>5</sup> Faculty of Electrical and Computer Engineering, University of Iceland, 101 Reykjavik, Iceland

<sup>6</sup> Department of Geomatics, Hohai University, 8 West of Focheng Road, 211100 Nanjing, China; zhaohui.xue@hhu.edu.cn

\* Correspondence: dupjrs@gmail.com; Tel.: +86-159-0515-9291

Academic Editors: András Jung, Magaly Koch and Prasad S. Thenkabail

Received: 25 February 2016; Accepted: 11 July 2016; Published: 15 July 2016

**Abstract:** Kernel-based methods and ensemble learning are two important paradigms for the classification of hyperspectral remote sensing images. However, they were developed in parallel with different principles. In this paper, we aim to combine the advantages of kernel and ensemble methods by proposing a kernel supervised ensemble classification method. In particular, the proposed method, namely RoF-KOPLS, combines the merits of ensemble feature learning (i.e., Rotation Forest (RoF)) and kernel supervised learning (i.e., Kernel Orthonormalized Partial Least Square (KOPLS)). In particular, the feature space is randomly split into  $K$  disjoint subspace and KOPLS is applied to each subspace to produce the new features set for the training of decision tree classifier. The final classification result is assigned to the corresponding class by the *majority voting* rule. Experimental results on two hyperspectral airborne images demonstrated that RoF-KOPLS with radial basis function (RBF) kernel yields the best classification accuracies due to the ability of improving the accuracies of base classifiers and the diversity within the ensemble, especially for the very limited training set. Furthermore, our proposed method is insensitive to the number of subsets.

**Keywords:** Rotation Forest; Kernel-based methods; Kernel Orthonormalized Partial Least Square; classification; hyperspectral data

## 1. Introduction

Hyperspectral remote sensing images, which can record hundreds of contiguous spectral bands in each pixel of the image, contains plenty of spectral information. The growing availability of hyperspectral imagery has opened up a new area for the investigation of the urbanization, land cover mapping, surface material analysis and target detection with improved accuracy [1–5]. The rich spectral information in hyperspectral images provides a great potential for generating more accurate classification maps compared to the ones produced by the multi-spectral images.

However, high dimensionality and relatively small size of training set pose the well-known Hughes phenomenon, which limits the performances of supervised classification methods [6]. In

order to alleviate this problem, many strategies have been proposed. As far as classification algorithms are concerned, ensemble learning or classifier ensemble has been shown to have the ability of alleviating the contradict of small training set and high-dimensionality. Furthermore, ensemble learning proved to provide better and more robust solutions in numerous remote sensing applications [7–9] in terms of the variety of available classification algorithms and the complexity of the hyperspectral data. The effectiveness of an ensemble method relies on the diversity and accuracy of the base classifiers [10,11]. Since the ensemble is typically more effective than a single classifier, many approaches have been developed and widely used in remote sensing applications [12–16]. For instance, [15] applied multiple classifiers (e.g., Bagging, Boosting and consensus theory) to multisource remote sensing data, and demonstrated that they outperformed several traditional classifiers in terms of accuracies. In [16] suggested that the Random Forest (RF) classifier performed equally to or better than the support vector machines (SVMs) for the classification of hyperspectral data. In particular, a special attention has been paid to the Rotation Forest (RoF), which is a relatively new classifier ensemble that can improve the accuracy of individual classifiers and diversity within the ensemble simultaneously [17]. The authors of [18–20] adapted RoF to classify hyperspectral images and found that it achieved better performances than traditional ensemble methods, e.g., Bagging, AdaBoost and RF. The authors of [21] proposed to apply RoF and RF for fully polarized SAR image classification using polarimetric and spatial features, and demonstrated that RoF can get better accuracy than SVM and RF.

Although RoF has demonstrated great performances in the classification of hyperspectral data, feature extraction methods used in RoF are limited to unsupervised ones in the previous studies, e.g., principle component analysis (PCA). RoF builds classifier ensembles based on independent decision tree by using feature extraction and random subspace so that each tree is trained on the training samples in a rotated feature space. It must be pointed out that, in the context of RoF, all the components derived from the feature extraction are kept, the discriminatory information is preserved even though it lies with the component responsible for the least variance [17]. According to the available prior class information, feature extraction as a pre-processing step of hyperspectral image analysis can be categorized into unsupervised and supervised ones [22,23].

In terms of feature reduction, PCA is one of the most popular unsupervised feature extraction methods in remote sensing community [24,25]. In contrast, supervised methods take into account prior class information to increase the separability of classes. A number of supervised feature extraction approaches, e.g., Fisher's linear discriminant analysis (FLDA) [26], partial least square regression (PLS) [27] and orthonormalized partial least square regression (OPLS) [28], have been developed. In remote sensing community, a modified FLDA was presented for the dimensionality reduction of hyperspectral remote sensing imagery, and the desired class information was well preserved and separated in the low-dimensional space [29]. The authors of [30] have found that PLS was superior to PCA when achieving the goal of discrimination and dimensionality reduction. OPLS is a variant of PLS, which is applicable to supervised problems, with certain optimality conditions regarding PLS. Moreover, considering that OPLS projections are obtained to predict the output labels, in consequence much more discriminative projection vectors are extracted compared to LDA, PLS [31,32].

A critical shortcoming of supervised feature extraction methods mentioned above is that they are based on the linear relation between the input and output spaces, which does not reflect the real data behavior [31,33,34]. In order to alleviate this problem, kernel methods have been developed and applied to the feature selection and feature reduction in hyperspectral image [35,36]. Moreover, as far as OPLS is concerned, the estimation of required parameter in OPLS is inaccurate without sufficient training set [37]. In order to circumvent these limitations, a non-linear version of OPLS, i.e., kernel OPLS (KOPLS), has been developed [38]. It is a very powerful feature extractor due to its appealing property of obtaining the non-linear projections by using kernel functions. In [31], experimental results revealed that KOPLS largely outperformed the traditional (linear) PLS algorithm especially in the context of nonlinear feature extraction.

In view of the above-mentioned facts, in this paper, we propose a novel kernel supervised feature learning classification scheme, namely RoF-KOPLS, which succeeds in taking advantages of the merits of KOPLS and RoF simultaneously. In the training step, the feature space is randomly split into  $K$  disjoint subspace and KOPLS is applied to each subspace to generate the kernel matrix and the transformation matrix. Then all the extracted features are retained to reformulate the new feature set for the training of decision tree (DT) classifier. In the prediction step, the new feature set of test samples is obtained by the kernel matrix and the transformation matrix, and then used to predict the class labels. The final classification result is assigned to the corresponding class which gets the maximum number of votes. We would like to emphasize that in this work we focus on pixel-wise classification, although RoF can be combined with spatial information, such as Markov random fields [20]. In order to examine the effectiveness of the proposed classification algorithm, experiments were conducted on two different hyperspectral airborne images: an AVIRIS image acquired over the Northwestern Indiana's Indian Pines site and a ROSIS image of the University of Pavia, Italy.

The remainder of this paper is organized as follows. In Section 2, Rotation Forest and OPLS are introduced. In Section 3, the proposed classification scheme is described based on the introduction of OPLS, KOPLS and RoF. Experimental results obtained on two different hyperspectral images are presented in Section 4. In Section 5, experimental results are discussed. Finally, we conclude this paper with some conclusions and future lines.

## 2. Related Works

### 2.1. Rotation Forest

Rotation Forest is a novel ensemble classifier for building independent decision trees built on the different sets of extracted features [17]. The main steps of RoF are summarized as follows: (1) the feature space is randomly split into  $K$  disjoint subsets and each subset contains  $M$  features; (2) PCA is applied to each feature set with a bootstrapped samples of 75% size of the original training set; (3) a sparse rotation matrix  $R_i$  is constructed by concatenating the coefficients of the principal components in each subset; (4) an individual DT classifier is trained with the new training samples formed by concatenating  $M$  linear extracted features in each subset; (5) by repeating the above steps several times, multiple classifiers were generated, and the final result is achieved by combining the outputs of all classifiers. The main training and prediction steps of RoF are shown in Algorithm 1. Classification and regression tree (CART) is adopted as the base classifier in this paper because of its sensitiveness to the rotations of axes [39]. The Gini index, is used to select the best split in the construction process of DT.

### 2.2. Orthonormalized Partial Least Square (OPLS)

OPLS is a multivariate analysis method for feature extraction, which exploits the correlation between the features and the target data by combining the merits of canonical variate analysis and PLS [28,31,32]. Given a set of training samples  $\{\mathbf{X}, \mathbf{Y}\} = \{\mathbf{x}_i, y_i\}_{i=1}^n$ , where  $\mathbf{x}_i \in \mathbb{R}^D$  and  $y_i \in \mathbb{R}$ .  $n$  and  $D$  represent the number of training samples and the dimensionality, respectively. Let  $\mathbf{X}$  and  $\mathbf{Y}$  represent  $\mathbf{X} = [x_1, \dots, x_n]^\top$  and  $\mathbf{Y} = [y_1, \dots, y_n]^\top$ , respectively. Here, we denote by  $\tilde{\mathbf{X}}$  and  $\tilde{\mathbf{Y}}$  the columnwise-centered version of  $\mathbf{X}$  and  $\mathbf{Y}$ , and denote by  $d$  the number of extracted features from the original data. Let  $\mathbf{C}_{\mathbf{X}\mathbf{Y}} = \frac{1}{n} \tilde{\mathbf{X}}^\top \tilde{\mathbf{Y}}$  represent the covariance between  $\mathbf{X}$  and  $\mathbf{Y}$ , whereas the covariance matrix of  $\mathbf{X}$  is given by  $\mathbf{C}_{\mathbf{X}\mathbf{X}} = \frac{1}{n} \tilde{\mathbf{X}}^\top \tilde{\mathbf{X}}$ .  $\mathbf{U} \in \mathbb{R}^{D \times d}$  is referred as the projection matrix, thus the extracted features can be formulated by  $\tilde{\mathbf{X}}' = \tilde{\mathbf{X}}\mathbf{U}$ .

The objective of OPLS is formulated as Equation (1)

$$\begin{aligned} \text{OPLS:} \quad & \text{maximize: } \text{Tr} \left\{ \mathbf{U}^\top \mathbf{C}_{\mathbf{X}\mathbf{Y}} \mathbf{C}_{\mathbf{X}\mathbf{Y}}^\top \mathbf{U} \right\} \\ & \text{subject to: } \mathbf{U}^\top \mathbf{C}_{\mathbf{X}\mathbf{X}} \mathbf{U} = \mathbf{I} \end{aligned} \quad (1)$$

OPLS is optimal (i.e., in the sense of mean-square-error) for performing linear multiregression on a given number of features extracted from the input data [40].

---

**Algorithm 1** Rotation Forest

---

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

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

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

- 9:   rearrange  $\mathbf{R}_i$  to  $\mathbf{R}_i^a$  so as to correspond to the original feature set
  - 10:   build an DT classifier  $L_i$  using  $\{\mathbf{X}\mathbf{R}_i^a, \mathbf{Y}\}$
  - 11:   add the classifier to the current ensemble,  $\mathcal{L} = \mathcal{L} \cup L_i$ .
  - 12: **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: calculate the confidence  $\mathbf{x}^*$  for each class,  $y_j$ , by average combination method:  $p(y_i|\mathbf{x}^*) = \frac{1}{T} \sum_{i=1}^T p(y_i|\mathbf{x}^* \mathbf{R}_i^a)$ . As a result,  $\mathbf{x}^*$  is assigned to the class with the largest confidence.
- 

### 3. Proposed Classification Scheme

#### 3.1. Kernel Orthonormalized Partial Least Square (KOPLS)

OPLS assumes that there exists the linear relation between the input features and the label. It might not be applicable when the linearity assumption is not hold. Kernel methods have been developed to alleviate this problem and demonstrated to be effective in many application domains [41,42]. In kernel methods, the original input data is mapped into a high or even infinite dimensional feature space by a non-linear function. The core of kernel methods lies in the implicit non-linear mapping since only the inner products are needed in the transformation [38,43].

Let us consider the function  $\phi : \mathbb{R}^D \rightarrow \mathcal{H}$  that maps the input data into a Reproducing Kernel Hilbert feature space  $\mathcal{H}$  of very high-dimension or even infinite dimension. Thus, the input variables  $\{\mathbf{x}_i, y_i\}_{i=1}^n$  is mapped to  $\{\phi(\mathbf{x}_i), y_i\}_{i=1}^n$ , where  $\Phi \in \mathbb{R}^{n \times \dim(\mathbb{H})}$  is the non-linear mapping with  $i$ -th row of vector  $\phi(\mathbf{x}_i)$ . The extracted features can be given by  $\Phi' = \Phi \mathbf{U}$ .

The kernel version of OPLS can be expressed as follows:

$$\begin{aligned} \text{KOPLS:} \quad & \text{maximize: } \text{Tr} \left\{ \mathbf{U}^\top \tilde{\Phi}^\top \tilde{\mathbf{Y}} \tilde{\mathbf{Y}}^\top \tilde{\Phi} \mathbf{U} \right\} \\ & \text{subject to: } \mathbf{U}^\top \tilde{\Phi}^\top \tilde{\Phi} \mathbf{U} = \mathbf{I} \end{aligned} \tag{2}$$

where  $\tilde{\Phi}$  is the centered version of  $\Phi$ .

According to the Representer Theorem [41], each projection vector in  $\mathbf{U}$  can be written as the linear combination of the training data, such as  $\mathbf{U} = \tilde{\Phi}^\top \mathbf{A}$ , where matrix  $\mathbf{A} = [\alpha_1, \dots, \alpha_d]$  and  $\alpha_i$  is the column vector containing the coefficients for the  $i$ -th projection vector [31], which is a new argument for the maximization optimization problem. KOPLS method can be reformulated as follows:

$$\begin{aligned} \text{KOPLS:} \quad & \text{maximize: } \text{Tr} \left\{ \mathbf{A}^\top \mathbf{K}_X \tilde{\mathbf{Y}} \tilde{\mathbf{Y}}^\top \mathbf{K}_X \mathbf{A} \right\} \\ & \text{subject to: } \mathbf{A}^\top \mathbf{K}_X \mathbf{K}_X \mathbf{A} = \mathbf{I} \end{aligned} \quad (3)$$

where, the kernel matrix is defined as  $K_x = \tilde{\Phi}^\top \tilde{\Phi}$ . In this paper, three kernels are used:

- *Linear Kernel:*

$$[\text{leftmargin} = *, \text{labelsep} = 5\text{mm}] \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j \quad (4)$$

- *Polynomial Kernel:*

$$\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j + 1)^c, \quad c \in \mathbb{Z}^+ \quad (5)$$

- *Radial Basis Function Kernel:*

$$\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right), \quad \sigma \in \mathbb{R}^+ \quad (6)$$

### 3.2. Rotation Forest with OPLS

Rotation Forest with OPLS (RoF-OPLS) is a variant of RoF. The major difference between RoF and RoF-OPLS is that OPLS is used to extract features for RoF-OPLS, while the feature extraction of RoF is based on PCA. The main steps of RoF-OPLS are: firstly, divide the feature space into  $K$  disjoint subspaces; then, OPLS is applied to each subspace with the bootstrapped samples of 75% of the training set; in the next step, the new training set obtained by rotating the original training set is treated as input to the individual classifier; finally, by repeating the above steps several times, the final result is generated by combining the outputs of all classifiers.

### 3.3. Rotation Forest with KOPLS

The success of MCSs (Multiple Classifier Systems) depend on not only the choice of base classifier, but also the diversity within the ensemble [12,44]. Aiming at improving both the diversity and classification accuracies of the DT classifiers within the ensemble, we propose a novel ensemble method, i.e., Rotation Forest with KOPLS (RoF-KOPLS), which aims at combining the advantages of KOPLS and RoF together. The proposed method can be summarized with the following steps (see Algorithm 2 and Figure 1). In the training phase, the feature space is randomly split into  $K$  disjoint subspace. For each subset, the initial training samples with 75% are drawn from the training data by using a bootstrap sampling method. KOPLS is applied to each subspace to get the coefficients  $R_k$ . In the next step, the kernel matrices of  $\hat{\mathbf{X}}_{i,j}$  are calculated, and an individual classifier is trained on the extracted features  $\mathbb{F}_i^{\text{new}}$ . In the prediction phase, the kernel matrices between  $\hat{\mathbf{X}}_{i,j}$  and a new sample  $\mathbf{x}^*$  is generated firstly. Then, the new transformed dataset  $\mathbb{F}_i^{\text{test}}$  is classified by the ensemble, and the final result will be assigned to the corresponding class by the *majority voting* rule. We expect that RoF-KOPLS can improve the performance of RoF-OPLS by introducing further diversity by performing a kernel feature extraction within the ensemble. The base classifiers in RoF-KOPLS are expected to be more diverse compared to these in RoF-OPLS, thus yielding more powerful ensemble. Furthermore, depending on the types of kernel function, RoF-KOPLS can be more specific, i.e., RoF with linear kernel (RoF-KOPLS-Linear), RoF with polynomial kernel (RoF-KOPLS-Polynomial), and RoF with RBF kernel (RoF-KOPLS-RBF).

**Algorithm 2** Rotation Forest with KOPLS

**Training phase**

**Input:**  $\{X, Y\} = \{x_i, y_i\}_{i=1}^l$ : training samples,  $T$ : number of classifiers,  $K$ : number of subsets,  $M$ : number of features in a subset,  $L$ : base classifier. The ensemble  $\mathcal{L} = \emptyset$ .  $\mathbb{F}$ : Feature set

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

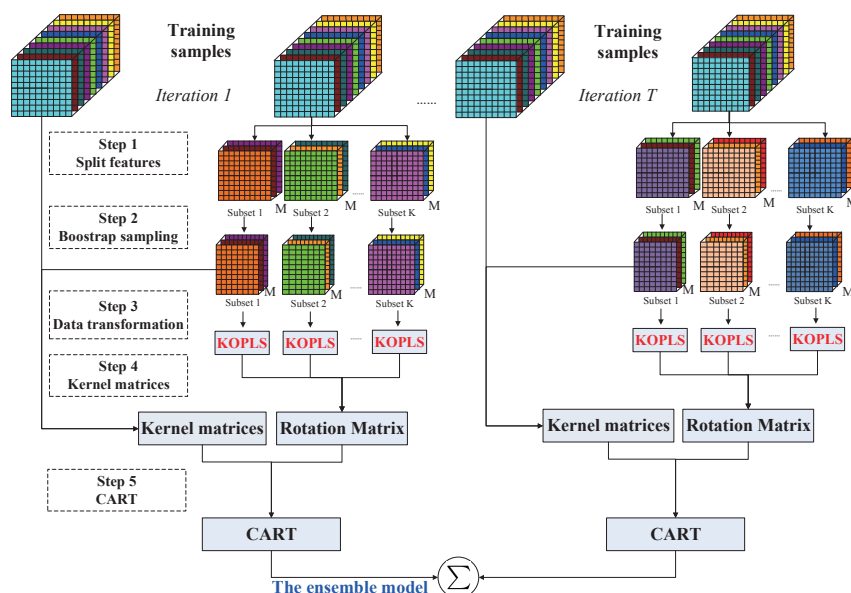
- 1: **for**  $i = 1 : T$  **do**
- 2:   randomly split the features  $\mathbb{F}$  into  $K$  subsets  $\mathbb{F}_j^i$
- 3:   **for**  $j = 1 : K$  **do**
- 4:     form the new training set  $X_{i,j}$  with  $\mathbb{F}_j^i$
- 5:     randomly select the 75% of the initial training samples to generate  $\hat{X}_{i,j}$
- 6:     using KOPLS to transform  $\hat{X}_{i,j}$  with the aim of getting the coefficients  $R_{i,j} = [\alpha_{i,j}^1, \dots, \alpha_{i,j}^M]$
- 7:     calculate the kernel matrices by  $\hat{X}_{i,j}$ ,  $K_{train,i,j} = K(\hat{X}_{i,j}, \hat{X}_{i,j})$
- 8:   **end for**
- 9:   the features extracted will be given by:  $\mathbb{F}_i^{new} = [K_{train,i,1}^\top R_{i,1}, \dots, K_{train,i,K}^\top R_{i,K}]$
- 10:   train a DT classifier  $L_i$  using  $\{\mathbb{F}_i^{new}, Y\}$
- 11:   add the classifier to the current ensemble,  $\mathcal{L} = \mathcal{L} \cup L_i$ .
- 12: **end for**

**Prediction phase**

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

**Output:** class label  $y^*$

- 1: **for**  $i = 1 : T$  **do**
- 2:   **for**  $j = 1 : K$  **do**
- 3:     generate the kernel matrices between  $\hat{X}_{i,j}$  and  $x^*$ ,  $K_{test,k} = K(\hat{X}_{i,j}, x_{i,j}^*)$
- 4:     generate the test features of  $x^*$ ,  $\mathbb{F}_i^{test} = [K_{test,i,1}^\top R_{i,1}, \dots, K_{test,i,K}^\top R_{i,K}]$
- 5:   **end for**
- 6:   run the classifier  $L_i$  using  $\mathbb{F}_i^{test}$  as input
- 7: **end for**
- 8: calculate the confidence  $x^*$  for each class and assign the class label  $p(y_i|x^*) = \frac{1}{T} \sum_{i=1}^T p(y_i|\mathbb{F}_i^{test})$  to the class with the largest confidence.



For pixel  $x^*$ , the output class probability  $p(y_i|x^*) = \frac{1}{T} \sum_{i=1}^T p(y_i|V_i)$ .

**Figure 1.** Illustration of the RoF-KOPLS.

#### 4. Experimental Results

Two popular hyperspectral airborne images were used for experiments. More detailed descriptions of the two data sets and the corresponding results are discussed in the next two subsections.

The following measures were used to evaluate the performances of different classification approaches:

- *Overall accuracy* (OA) is the percentage of correctly classified pixels.
- *Average accuracy* (AA) is the average of percentages of classified pixels for individual class.
- *Kappa coefficient* ( $\kappa$ ) is the percentage of agreement corrected by the level of agreement that would be expected by casually [23].

For the purpose of analysing the ensemble clearly, we adopted the following measures to estimate its performance.

- *Average of OA* (AOA) is the average of OAs of individual classifiers within the ensemble.
- *Diversity in classifier ensemble*. Diversity has been regarded as a very significant characteristic in classifier ensemble [45]. In this paper, coincident failure diversity (CFD) is used as the diversity measure [10]. The higher the value of CFD, the more diverse the ensemble.

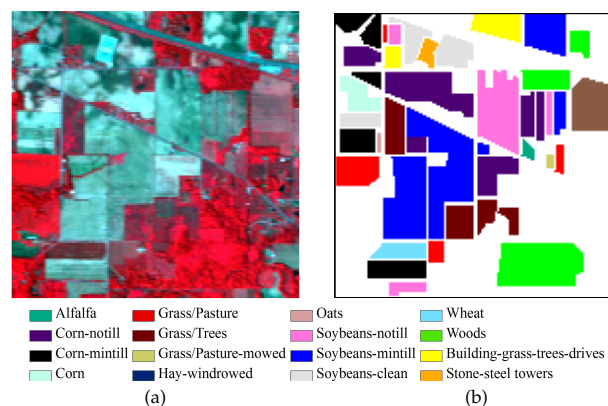
##### 4.1. Results of the AVIRIS Indian Pines Image

The Indian Pines image was acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensors over the Indian Pines test site in Northwestern Indiana of an agricultural area. The image is  $145 \times 145$  pixels, with a spatial resolution of 20 m per pixel. In order to evaluate the performance of our proposed methods, all full spectral bands, including 20 noisy and water absorption bands, was used for experiment. It is composed of 220 spectral channels in the wavelength ranging from 0.4 to 2.5  $\mu\text{m}$ . Sixteen classes of interest are reported in Table 1. Figure 2 depicts the three-band false color composite image and the reference data of this image.

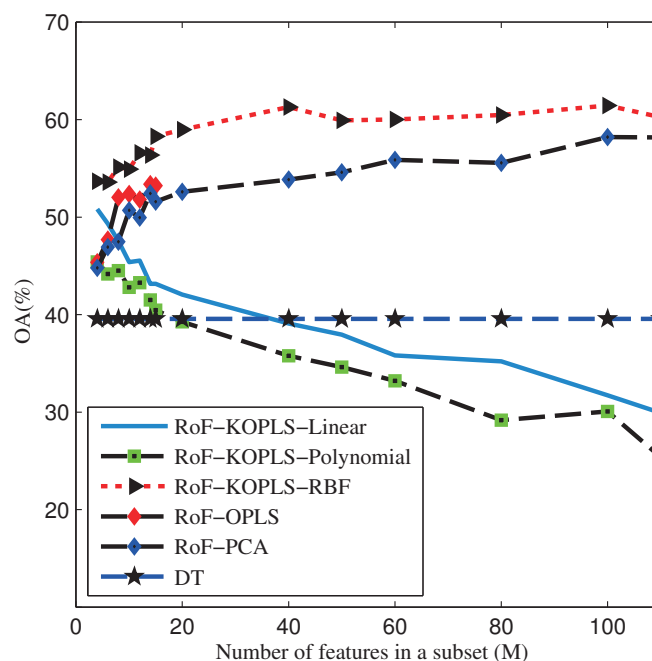
In order to evaluate the performance of the proposed classification techniques, some methods including support vector machine (SVMs), DT, RotBoost [46,47], DT with KOPLS (DT-KOPLS), and RoF-PCA were implemented for comparison. The reason why we select SVMs and DT in comparison to the proposed methods is that they are two of the leading classification techniques of hyperspectral data. As far as SVM is concerned, the radial basis function kernel is chosen for classification, which include two parameters (i.e., penalty term  $C$  and the width of the exponential  $\sigma$ ). Furthermore, in our experiments, fivefold cross-validation was used to select the best combination of parameters under the condition that  $C$  and  $\sigma$  were set to  $[2^{-4}, 2^{12}]$  and  $[2^{-10}, 2^5]$ , respectively. Furthermore, DT-KOPLS is the variant of DT. In terms of RoF-PCA, it is an ensemble method using independent DT built on a different set of extracted features. It is worth noting that the feature extraction for RoF-PCA is based on PCA. For DT-KOPLS, KOPLS is used for feature extraction prior to DT classifier. The range of extracted components is from 2 to 30. In this paper, three kernels, e.g., linear, RBF and polynomial, are used in KOPLS feature extraction prior to DT classifier. Only the best results are reported in this paper. The kernel width  $\sigma$  in RBF kernel was computed by the median of all pairwise distances between the samples [48], and  $c$  in polynomial kernel was set to 2. The reported results were achieved by averaging the results obtained from ten Monte Carlo runs. According to our previous studies [19,20],  $T$  is set to be 10 in the ensembles.

The number of features in a subset ( $M$ ) is a crucial parameter for the Rotation Forest ensembles. In order to investigate the impact of  $M$  on the performance of different classification scheme, we randomly select a very limited training set, i.e., 10 samples per class. The evaluation of OA with the increase of  $M$  is depicted in Figure 3. It should be noted that, the value of  $M$  should be less than the number of classes for RoF-OPLS. For other methods,  $M$  ranges from 2 to 110. The results presented in Figure 3 obviously show that there is no consistent pattern of the relationship between  $M$  and OAs,

which is in accordance with the conclusions obtained in our studies [20,49]. The OAs obtained by RoF-KOPLS-Linear and the RoF-KOPLS-Polynomial decrease as the increase of  $M$ . In particular, it is worth noting that the RoF-KOPLS-RBF can obtain the best OAs in all cases. Furthermore, RoF-KOPLS-RBF is insensitive to  $M$  in comparison with other classification methods when the value of  $M$  is greater the number of classes (i.e., 16). Another observation is that, the optimal values of  $M$  for different classification methods are various. For instance, RoF-KOPLS-RBF achieves the best classification result when  $M = 100$ . To ensure a fair comparison, the optimal values of  $M$  are independently selected for specific methods. Thus, the optimal values of  $M$  for RoF-OPLS, RoF-PCA, RoF-KOPLS-RBF, RoF-KOPLS-Linear, RoF-KOPLS-Polynomial were set to be 14, 100, 100, 4 and 4, respectively. Figure 4 plots the classification maps obtained by the individual and ensemble learning methods (only one Monte Carlo run).



**Figure 2.** AVIRIS Indian Pines data set. (a) Three-band color composite (bands 57, 27, 17); (b) Ground-truth map containing 16 mutually exclusive land-cover classes. The legend of this scene is shown at the bottom.

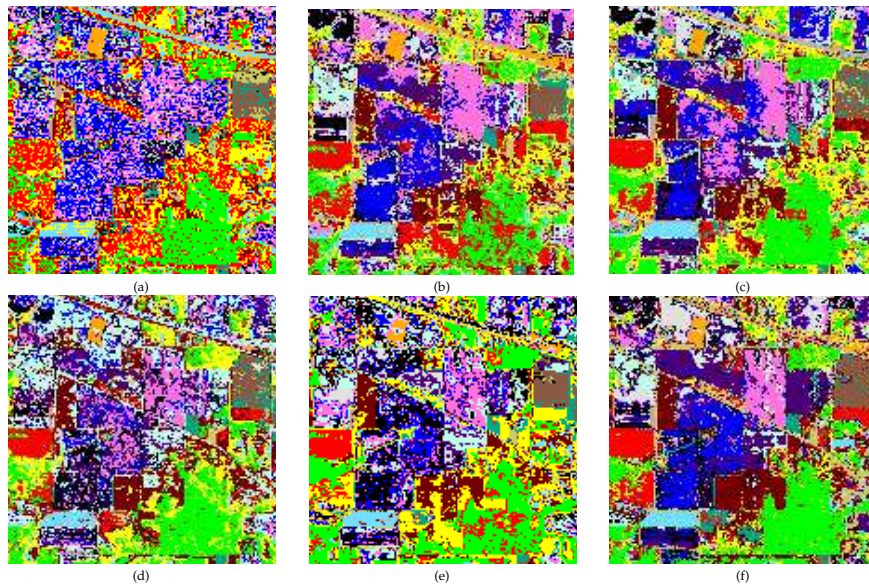


**Figure 3.** Indian Pines AVIRIS Image. OAs obtained by DT, RoF-PCA, RoF-OPLS, RoF-KOPLS-Linear, RoF-KOPLS-Polynomial, RoF-KOPLS-RBF with different number of  $M$ .



**Table 1.** Overall, Average and Class-specific Accuracies for the Indian Pines AVIRIS image.

Class	Train	Test	SVM	DT	RotBoost	DT-KOPLS	RoF-PCA	RoF-OPLS	RoF-KOPLS		
									RBF	Linear	Polynomial
Alfalfa	10	54	76.30	74.81	82.50	42.41	85.91	86.11	<b>89.81</b>	81.85	73.52
Corn-no till	10	1434	27.33	29.87	<b>56.64</b>	11.05	52.01	46.69	53.18	39.52	32.36
Corn-min till	10	834	33.39	26.62	<b>50.85</b>	16.94	50.69	45.30	47.28	<b>44.44</b>	36.02
Bldg-Grass-Tree-Drives	10	234	56.37	26.79	<b>75.00</b>	8.550	66.16	73.55	67.31	49.15	45.68
Grass/pasture	10	497	53.76	57.24	76.18	34.35	71.17	72.72	<b>78.17</b>	69.72	69.72
Grass/trees	10	747	60.83	40.13	83.88	26.05	81.38	74.66	<b>88.59</b>	69.65	64.79
Grass/pasture-mowed	10	26	90.77	82.69	90.63	68.08	91.87	92.31	<b>95.00</b>	91.54	87.31
Corn	10	489	51.76	49.28	82.15	25.01	78.04	64.34	<b>87.83</b>	67.71	62.35
Oats	10	20	94.00	83.50	96.00	50.50	95.00	95.00	<b>100.0</b>	89.50	87.50
Soybeans-no till	10	968	45.61	31.24	<b>67.12</b>	17.07	62.21	54.32	55.51	52.36	40.19
Soybeans-min till	10	2468	34.89	30.06	<b>43.00</b>	17.32	41.17	29.11	41.17	34.85	31.67
Soybeans-clean till	10	614	32.98	24.92	48.66	14.66	45.15	40.54	<b>56.81</b>	31.89	23.21
Wheat	10	212	93.54	84.95	96.63	50.09	94.70	95.61	<b>98.49</b>	89.25	87.36
Woods	10	1294	67.67	68.63	80.02	37.33	73.75	80.02	<b>83.79</b>	73.22	70.83
Hay-windrowed	10	380	29.76	35.03	38.08	11.34	43.38	45.18	<b>52.50</b>	38.53	30.82
Stone-steel towers	10	95	88.00	89.68	<b>97.41</b>	64.42	95.29	92.21	90.84	91.58	92.63
OA			44.73	39.56	<b>61.50</b>	21.55	58.29	53.38	61.44	50.83	45.40
AA			58.56	52.22	72.80	30.95	70.49	67.98	<b>74.14</b>	63.42	58.50
$\kappa$			38.65	33.17	<b>57.03</b>	14.62	53.52	48.21	56.98	45.38	39.53

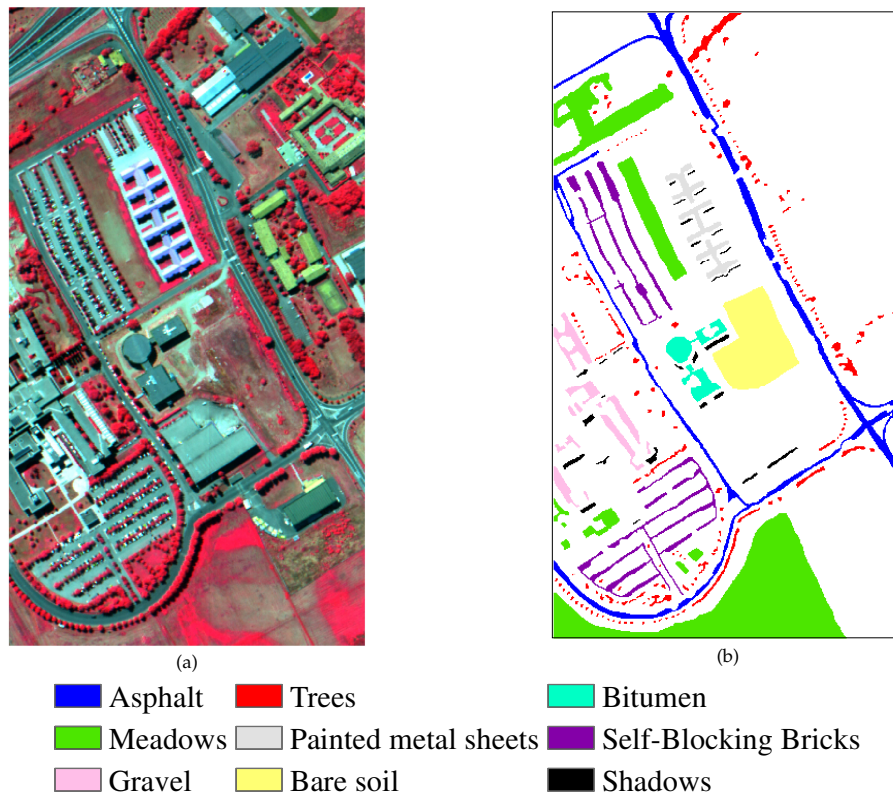


**Figure 4.** Classification maps of the Indian Pines AVIRIS image (only one Monte Carlo run). OAs of the classifiers are presented as follows: (a) DT (40.20%); (b) RoF-PCA (57.39%); (c) RoF-OPLS (54.97%); (d) RoF-KOPLS-Linear (45.39%); (e) RoF-KOPLS-Polynomial (42.80%); (f) RoF-KOPLS-RBF (64.25%)

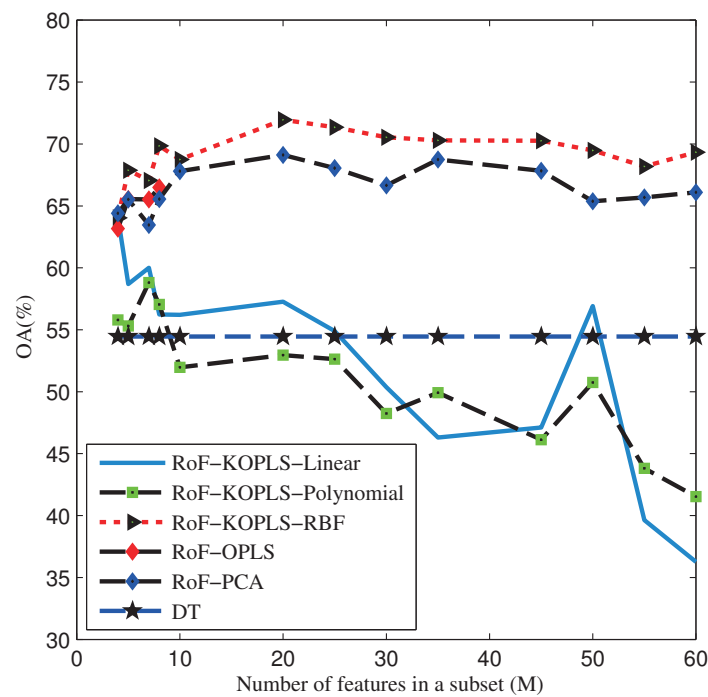
#### 4.2. Results of the University of Pavia ROSIS Image

In the second study, the proposed scheme was tested on the ROSIS image, which is collected from an university area with a spatial resolution of 1.3 m and 103 bands. The original recorded image has a spatial dimension of  $610 \times 340$  pixels, with 103 channels left for experiments by removing 12 noisy bands. Nine classes of interest are contained in the reference data with a total number of 42776 labeled samples. A false color composite image and the reference data are shown in Figure 5. For this experiment, we randomly select only 10 samples per class as training samples, which represents a very limited training set. In order to ensure a fair comparison, we conducted ten independent runs for each experiment in terms of training samples selection and classification.

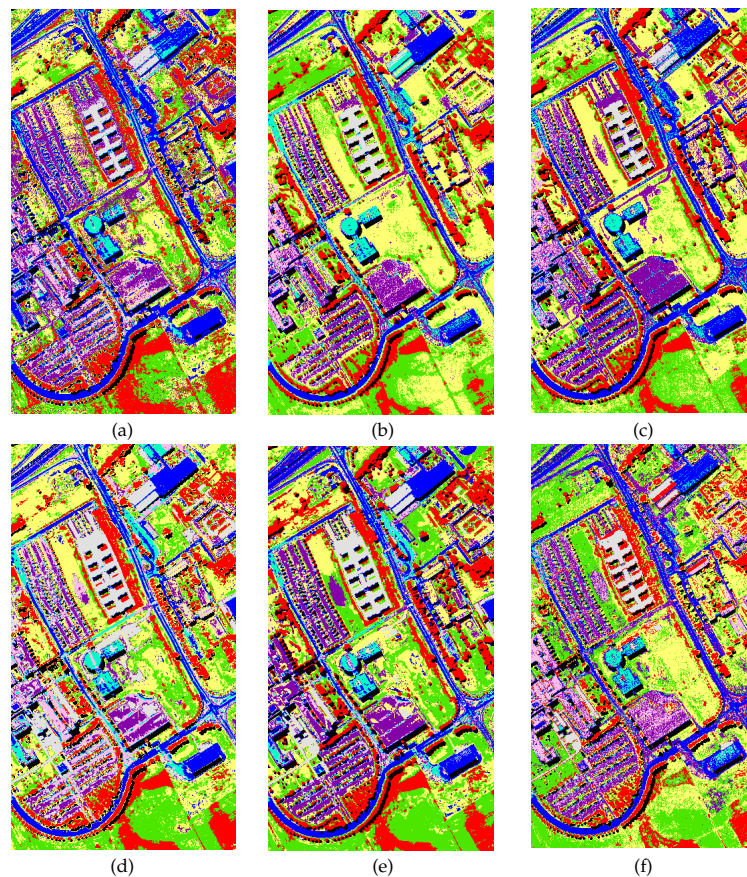
In the first experiment, the impacts of  $M$  on the global accuracies obtained by all classification approaches were investigated. For the RoF-OPLS algorithm, the value of  $M$  should be less than the number of classes. Hence, the values of  $M$  were set to be 4, 5, 7 and 8. However, this limitation is not necessary for RoF-KOPLS and RoF-PCA methods. In order to clearly examine the effect of  $M$  on the OAs obtained by RoF-KOPLS and RoF-PCA methods, the value of  $M$  was set to the range from 4 to 60. Figure 6 shows the OAs obtained by different methods as a function of different values of  $M$ . Similar conclusion can be drawn with the former experiments. First, the performances of the RoF methods rely on the values of  $M$ . It should be noted that the RoF-KOPLS-RBF is insensitive to the value of  $M$  compared to other classification techniques when  $M$  is greater than 9 (i.e., the number of classes). Second, the impact of  $M$  on OA seems to be irregular. Third, the overall accuracies obtained by RoF-KOPLS-RBF are more accurate than those achieved by all other methods. Finally, the overall accuracies obtained by the RoF-KOPLS-Linear method and the RoF-KOPLS-Polynomial method exhibit larger variations as the increase of  $M$ . Nevertheless, the overall accuracies achieved by the presented RoF-KOPLS-RBF method tend to be stable with the increase of the value of  $M$ . In order to make fair comparisons, the value of  $M$  should be selected as the one achieving the best accuracy for each classification algorithm. In consequence, the values of RoF-OPLS, RoF-PCA, RoF-KOPLS-RBF, RoF-KOPLS-Linear, RoF-KOPLS-Polynomial were set to 8, 20, 20, 4 and 7, respectively. Figure 7 depicts the classification maps obtained by all the considered methods.



**Figure 5.** ROSIS University of Pavia data set. (a) Three-band color composite (bands 102, 56, 31); (b) Reference map containing 9 mutually exclusive land-cover classes. The legend of this scene is shown at the bottom.



**Figure 6.** OAs obtained by DT, RoF-PCA, RoF-OPLS, RoF-KOPLS-Linear, RoF-KOPLS-Polynomial, RoF-KOPLS-RBF with different number of  $M$  from the University of Pavia ROSIS Image.



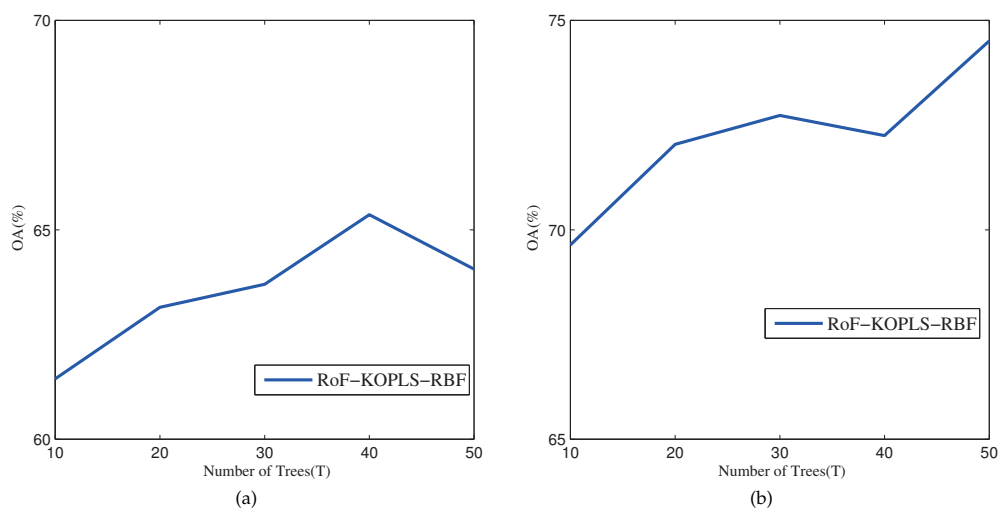
**Figure 7.** Classification maps of the University of Pavia ROSIS image (only one Monte Carlo run). OAs of the classifiers are presented as follows: (a) DT (54.06%); (b) RoF-PCA (66.0%); (c) RoF-OPLS (65.24%); (d) RoF-KOPLS-Linear (57.26%); (e) RoF-KOPLS-Polynomial (60.57%); (f) RoF-KOPLS-RBF (70.65%).

## 5. Discussion

### 5.1. Discussion on the AVIRIS Indian Pines Image

The overall and class-specific accuracies of different classification algorithms are presented in Table 1. The results reveal that the classifier ensembles can yield more accurate accuracies compared to single classifiers. It is apparent that the proposed RoF-KOPLS-RBF method provides good results roughly equivalent to the recently proposed method, Rotboost, which is followed by RoF-PCA and RoF-OPLS. Furthermore, it should be noted that the proposed RoF-KOPLS-RBF method can achieve considerable increases in most class-specific accuracies, which significantly outperforms others. The McNemar's test revealed that the difference between RoF-KOPLS-RBF and RoF-OPLS are statistically significant ( $|z| > 1.96$ ) [50]. The kernel-based method improves the accuracies by 8.06% in OA and 6.16% in AA. Furthermore, as we can see from the Figure 4, the Rotation Forest ensembles can improve the classification accuracies and produce more smooth classification maps. These results validate the good performance of the proposed RoF-KOPLS-RBF by combining KOPLS and RoF.

The number of classifiers ( $T$ ) and training samples are the key parameters for the proposed method. In order to investigate the influence of  $T$  on the classification accuracies, we have performed the classification results when the number of feature in a subset  $M$  is set to 100. As we can see from the Figure 8a, the classification accuracies are improved with the increase of  $T$ .



**Figure 8.** Sensitivity to the change of the number of trees. (a) Indian Pines AVIRIS image; (b) University of Pavia ROSIS image.

Table 2 presents classification accuracies obtained by individual classifier using different numbers of training samples. As reported in the table, the proposed RoF-KOPLS-RBF, RoF-KOPLS-Linear, RoF-KOPLS-Polynomial, and RoF-OPLS methods are superior to DT and DT-KOPLS. RoF-KOPLS-RBF, RoF-OPLS, and RoF-PCA achieve better classification accuracies when compared to SVM. It can be found that the proposed RoF-KOPLS-RBF method gains the best classification results under most of training scenarios as compared to other classification techniques. As we can see from the Table 2, when we compare the proposed method with the recently new classification method RotBoost, our proposed method is equivalent or superior to the RotBoost approach. Therefore, it can be concluded that RoF-KOPLS-RBF works more efficiently with relatively low number of labeled training samples.

Table 3 provides the OAs, AOAs, and diversities obtained by different RoF ensembles using 10 samples for each class. The accuracy of individual classifier and diversity are two important properties for a classifier ensemble as higher values of AOA and diversity always give rise to better performance. The results in this table show that the proposed RoF-KOPLS-RBF method acquires the highest AOA and diversity, leading to the best classification accuracies. Furthermore, it is worth noting that the effect of kernel functions on the classification accuracies are significant. RoF-KOPLS-RBF method obtains better classification results in comparison to RoF-KOPLS-Linear and RoF-KOPLS-Polynomial methods. This can be attributed to RoF-KOPLS-RBF's higher values of AOA and diversity.

**Table 2.** OAs and AAs (in Parentheses) Obtained for Different Classification Methods When Applied to the Indian Pines AVIRIS image.

Samples Per Class	SVM	DT	RotBoost	DT-KOPLS	RoF-PCA	RoF-OPLS	RoF-KOPLS		
							RBF	Linear	Polynomial
10	44.73 (58.56)	39.56 (52.22)	<b>61.50</b> (72.80)	21.55 (30.95)	58.29 (70.49)	53.38 (67.98)	61.44 ( <b>74.14</b> )	50.83 (63.42)	45.40 (58.50)
20	55.45 (68.76)	44.48 (58.01)	<b>68.34</b> (77.97)	22.74 (32.89)	65.32 (77.01)	61.28 (74.67)	67.40 ( <b>79.38</b> )	59.44 (71.25)	53.31 (66.80)
30	60.81 (73.23)	49.39 (61.94)	71.58 (80.62)	26.38 (32.49)	69.06 (78.67)	65.81 (77.20)	<b>71.88 (82.52)</b>	63.74 (75.35)	59.31 (71.40)
50	65.69 (77.39)	53.81 (65.11)	<b>75.83</b> (83.40)	54.33 (64.49)	73.54 (82.88)	69.65 (80.24)	75.55 ( <b>85.86</b> )	67.84 (78.21)	63.98 (74.97)
60	69.53 (79.64)	55.61 (66.13)	<b>77.24</b> (83.39)	58.62 (68.30)	75.46 (82.91)	71.17 (80.97)	76.99 ( <b>86.66</b> )	70.37 (79.56)	66.36 (76.58)
80	72.58 (80.81)	58.11 (68.27)	78.83 (84.76)	66.43 (74.52)	77.02 (83.34)	74.05 (82.66)	<b>79.70 (88.27)</b>	73.49 (81.26)	70.32 (78.57)
100	73.50 (79.48)	60.67 (69.70)	79.82 (84.71)	67.90 (74.97)	78.12 (84.00)	75.72 (83.48)	<b>82.56 (89.51)</b>	74.36 (81.49)	71.51 (79.79)
120	78.04 (85.35)	62.95 (70.77)	81.00 (85.36)	71.01 (77.23)	79.48 (84.99)	76.93 (83.76)	<b>83.97 (90.39)</b>	75.98 (82.93)	74.56 (81.16)

**Table 3.** OAs (in Percent), AOAs (in Percent), and Diversities Obtained for Different Rotation Forest Ensembles When Applied to the Indian Pines AVIRIS Image.

Classifiers	RoF-PCA	RoF-OPLS	RoF-KOPLS		
			RBF	Linear	Polynomial
OA	58.29	53.38	<b>61.44</b>	50.83	45.40
AOA	45.76	42.75	<b>48.16</b>	41.13	40.01
Diversity	47.76	44.19	<b>48.84</b>	40.95	37.75

### 5.2. Discussion on the University of Pavia ROSIS Image

The classification accuracies of all the classification techniques are summarized in Table 4. From this table, the best OA, kappa coefficient, and class-specific accuracies for most classes are achieved by the presented RoF-KOPLS-RBF method, which is followed by the RotBoost, RoF-PCA and RoF-OPLS approach. In this case, the OA of the RoF-OPLS approach is improved by 5.46% compared to the RoF-KOPLS-RBF. According to the results of McNemar's test, the RoF-KOPLS-RBF classification map is significantly more accurate compared to those achieved by other methods except the RotBoost approach with a confident level of 5%. We can conclude that the proposed RoF-KOPLS-RBF method inherits the good merits of KOPLS and RoF, thus leading to improved classification result.

As like in the first experiment, the impacts of  $T$  and training samples on the classification results have also been explored. When investigating the influence of  $T$  on the classification accuracies, the number of feature in a subset  $M$  is set to 20 achieving the best accuracy for the proposed method. Figure 8b shows the OA (%) using different number of  $T$ . With the increase of  $T$ , the classification results are significantly improved. Table 5 gives the OAs and AAs (in parentheses) obtained by different classification approaches when using different number of training samples. As expected, the classification accuracies obtained by all methods becomes higher with the increase of the training set size. Analogous to the first experiment, the proposed RoF-KOPLS-RBF method demonstrates relatively higher performance with a very limited number of training samples in terms of OAs and AAs, as compared to the other classification approaches. Moreover, from Figure 7, we can draw that the Rotation Forest ensembles generate more accurate classification maps with reduced data noise in comparison with the individual classifiers.

The OAs, AOAs, and diversities obtained by Rotation Forest ensembles are reported in Table 6 to evaluate the ensemble clearly. It can be noted that the proposed RoF-KOPLS-RBF approach gives the highest AOA and diversity, when compared to other classification approaches. RoF-KOPLS-RBF gains the best overall accuracy due to the fact that higher AOA and diversity lead to better ensemble performance, which confirms the validity of combining the merits of KOPLS and Rotation Forest. As can be seen from the table, we can conclude that the kernel function can give rise to significant impact on the classification accuracies, which is similar to the first experiment. RoF-KOPLS-RBF achieves the higher values of AOA and diversity when compared to RoF-KOPLS-Linear and RoF-KOPLS-Polynomial, leading to better classification results.

In addition, it should be noted that although the proposed method has shown good performance in the classification of hyperspectral data, it is confronted with some common cons for Rotation Forest, e.g., the relative low computational efficiency and sensitivity to the number of features in a subset [21]. Moreover, the proposed method only consider the spectral information so that it obtains suboptimal classification results when compared to the method taking advantage of the spatial and spectral information simultaneously [20].

**Table 4.** Overall, Average and Class-specific Accuracies for the Pavia ROSIS image.

Class	Train	Test	SVM	DT	RotBoost	DT-KOPLS	RoF-PCA	RoF-OPLS	RoF-KOPLS		
									RBF	Linear	Polynomial
Bricks	10	3682	74.40	55.89	69.16	33.58	66.55	67.47	<b>71.94</b>	69.70	65.17
Shadows	10	947	99.97	94.19	<b>99.98</b>	84.09	99.54	99.95	99.88	99.86	99.80
Metal Sheets	10	1345	99.20	96.88	<b>99.70</b>	56.27	99.40	99.30	98.70	96.60	95.97
Bare Soil	10	5029	69.70	49.81	71.32	22.81	71.94	<b>73.81</b>	67.69	61.44	48.88
Trees	10	3064	88.18	72.11	<b>94.38</b>	42.28	90.42	90.16	89.67	86.06	72.40
Meadows	10	18649	62.26	46.63	61.65	35.81	63.05	56.47	<b>68.44</b>	54.60	52.70
Gravel	10	2099	63.60	37.81	<b>68.64</b>	37.63	61.02	54.82	66.83	48.99	37.85
Asphalt	10	6631	64.90	58.93	63.43	38.95	64.83	67.92	67.35	<b>70.68</b>	63.83
Bitumen	10	1330	86.66	70.75	<b>90.48</b>	57.97	81.63	76.90	80.58	74.34	74.41
OA			69.27	54.46	69.34	37.51	69.06	66.49	<b>71.95</b>	64.11	58.81
AA			78.76	64.78	<b>79.86</b>	45.49	77.60	76.31	79.01	73.59	67.89
$\kappa$			61.76	44.63	62.12	26.42	61.55	58.81	<b>64.69</b>	55.72	49.36

**Table 5.** OAs and AAs (in Parentheses) Obtained for Different Classification Methods Using Different Numbers of Training Samples When Applied to the Pavia ROSIS Image.

Samples Per Class	SVM	DT	RotBoost	DT-KOPLS	RoF-PCA	RoF-OPLS	RoF-KOPLS		
							RBF	Linear	Polynomial
10	69.27 (78.76)	54.46 (64.78)	69.34 ( <b>79.86</b> )	37.51 (45.49)	69.06 (77.60)	66.49 (76.31)	<b>71.95</b> (79.01)	64.11 (73.59)	58.81 (67.89)
30	78.30 (84.06)	62.88 (72.96)	79.22 (85.31)	61.56 (67.88)	75.75 (82.68)	78.92 (83.91)	<b>80.25 (86.28)</b>	70.04 (79.33)	61.85 (74.01)
40	81.69 (86.50)	64.03 (73.45)	81.40 ( <b>87.21</b> )	65.61 (72.69)	79.68 (84.63)	80.47 (85.03)	<b>81.96</b> (87.10)	71.74 (81.39)	64.62 (75.97)
50	83.36 (87.84)	64.71 (74.04)	<b>83.71</b> (88.13)	73.08 (77.40)	81.71 (86.45)	80.97 (85.87)	83.56 ( <b>88.35</b> )	73.52 (83.06)	66.91 (77.59)
60	84.22 (88.39)	66.64 (75.15)	84.61 (88.89)	72.07 (79.04)	82.48 (87.31)	81.58 (86.52)	<b>84.47 (89.17)</b>	74.51 (82.91)	68.05 (77.99)
80	85.65 (89.39)	68.58 (76.87)	85.06 (89.42)	73.54 (78.37)	83.66 (87.83)	82.62 (87.33)	<b>86.20 (90.22)</b>	76.47 (84.64)	69.96 (79.47)
100	87.28 (90.17)	69.77 (77.56)	86.05 (90.37)	80.05 (83.56)	85.56 (89.55)	83.38 (88.05)	<b>87.33 (90.93)</b>	77.59 (85.33)	71.49 (81.0)



**Table 6.** OAs (in Percent), AOAs (in Percent), and Diversities Obtained for Different Rotation Forest Ensembles When Applied to the Pavia ROSIS image.

Classifiers	RoF-PCA	RoF-OPLS	RoF-KOPLS		
			RBF	Linear	Polynomial
OA	69.06	66.49	<b>71.95</b>	64.11	58.81
AOA	57.48	57.16	<b>58.09</b>	56.42	56.81
Diversity	55.78	57.86	<b>59.00</b>	53.56	46.99

## 6. Conclusions

In this paper, a new classification approach is presented by combining the advantages of kernel-based feature extraction, i.e., KOPLS, and ensemble method, i.e., Rotation Forest. The performance of the proposed methods was evaluated by several experiments based on two popular hyperspectral images. Experimental results demonstrated that the proposed RoF-KOPLS methodology can inherit the merits of RoF and KOPLS to achieve more accurate classification results.

The following conclusions can be drawn according to the experimental results:

- RoF-KOPLS with RBF kernel yields the best accuracies against the comparative methods above-mentioned due to the ability of improving the accuracy of base classifiers and the diversity within the ensemble, especially for the very limited training set.
- In RoF-KOPLS, the kernel functions can give rise to significant influences on the classification results. Experimental results have shown that RoF-KOPLS with RBF kernel obtained the best performances.
- RoF-KOPLS with RBF kernel is insensitive to the number of features in a subset when compared to other methods.

In the future, we will further explore the integration of Rotation Forest and kernel methods in classifier ensemble for real application of the hyperspectral images. On the one hand, we will attempt to combine the proposed method with Adaboost or Bagging [51]. On the other hand, given the important role of spatial features in the classification of hyperspectral image [52], spatial information will be incorporated to improve the performances of the proposed classification scheme in the following work.

**Acknowledgments:** This work is partially supported by the Natural Science Foundation of China (No. 41171323), Jiangsu Provincial Natural Science Foundation (No. BK2012018) and National Key Scientific Instrument and Equipment Development Program (No. 012YQ050250). The authors would like to thank D. Landgrebe from Purdue University for providing the AVIRIS hyperspectral data and P. Gamba for providing the University of Pavia ROSIS Image, along with the training and test data sets.

**Author Contributions:** Jike Chen and Junshi Xia conceived and designed the experiments; Jike Chen performed the experiments, analyzed the data and wrote the paper. Peijun Du, Jocelyn Chanussot, Zhaohui Xue and Xiangjian Xie gave comments, suggestions to the manuscript and checked the writing.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

PCA	Principle Component Analysis
RBF	Radial Basis Function
FLDA	Fisher's Linear Discriminant Analysis
PLS	Partial Least Square Regression
OPLS	Orthonormalized Partial Least Square Regression
KOPLS	Kernel Orthonormalized Partial Least Square Regression
RF	Random Forest
SVMs	Support Vector Machines
RoF	Rotation Forest
DT	Decision Trees

CART	Classification and Regression Tree
RoF-OPLS	Rotation Forest with OPLS
RoF-KOPLS	Rotation Forest with KOPLS
OA	Overall Accuracy
AA	Average Accuracy
AOA	Average of OA
$\kappa$	Kappa coefficient
CFD	Coincident Failure Diversity
RotBoost	Rotation Forest with Adaboost
DT-KOPLS	DT with KOPLS

## References

1. Plaza, A.; Benediktsson, J.A.; Boardman, J.W.; Brazile, J.; Bruzzone, L.; Camps-Valls, G.; Chanussot, J.; Fauvel, M.; Gamba, P.; Gualtieri, A.; et al. Recent advances in techniques for hyperspectral image processing. *Remote Sens. Environ.* **2009**, *113*, S110–S122.
2. Shang, X.; Chisholm, L.A. Classification of Australian native forest species using hyperspectral remote sensing and machine-learning classification algorithms. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2481–2489.
3. Bioucas-Dias, J.M.; Plaza, A.; Camps-Valls, G.; Scheunders, P.; Nasrabadi, N.M.; Chanussot, J. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–36.
4. Manolakis, D.; Marden, D.; Shaw, G.A. Hyperspectral image processing for automatic target detection applications. *Lincoln Lab. J.* **2003**, *14*, 79–116.
5. Dong, Y.; Zhang, L.; Zhang, L.; Du, B. Maximum margin metric learning based target detection for hyperspectral images. *ISPRS J. Photogramm. Remote Sens.* **2015**, *108*, 138–150.
6. Hughes, G. On the mean accuracy of statistical pattern recognizers. *IEEE Trans. Inf. Theory* **1968**, *14*, 55–63.
7. Oza, N.C.; Tumer, K. Classifier ensembles: Select real-world applications. *Inf. Fusion* **2008**, *9*, 4–20.
8. Benediktsson, J.A.; Chanussot, J.; Fauvel, M. Multiple classifier systems in remote sensing: From basics to recent developments. In *Multiple Classifier Systems*; Springer: Berlin, Germany, 2007; pp. 501–512.
9. Du, P.; Xia, J.; Zhang, W.; Tan, K.; Liu, Y.; Liu, S. Multiple classifier system for remote sensing image classification: A review. *Sensors* **2012**, *12*, 4764–4792.
10. Kuncheva, L.I.; Whitaker, C.J. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Mach. Learn.* **2003**, *51*, 181–207.
11. Shipp, C.A.; Kuncheva, L.I. Relationships between combination methods and measures of diversity in combining classifiers. *Inf. Fusion* **2002**, *3*, 135–148.
12. Kuncheva, L.I. *Combining Pattern Classifiers: Methods and Algorithms*; John Wiley & Sons: Hoboken, NJ, USA, 2004.
13. Rokach, L. *Pattern Classification Using Ensemble Methods*; World Scientific: Singapore, Singapore, 2009; Volume 75.
14. Waske, B.; Braun, M. Classifier ensembles for land cover mapping using multitemporal SAR imagery. *ISPRS J. Photogramm. Remote Sens.* **2009**, *64*, 450–457.
15. Briem, G.J.; Benediktsson, J.A.; Sveinsson, J.R. Multiple classifiers applied to multisource remote sensing data. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 2291–2299.
16. Pal, M. Random forest classifier for remote sensing classification. *Int. J. Remote Sens.* **2005**, *26*, 217–222.
17. Rodriguez, J.J.; Kuncheva, L.I.; Alonso, C.J. Rotation forest: A new classifier ensemble method. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 1619–1630.
18. Xia, J.; Chanussot, J.; Du, P.; He, X. Rotation-Based Ensemble Classifiers for High-Dimensional Data. In *Fusion in Computer Vision*; Springer: Berlin, Germany, 2014; pp. 135–160.
19. Xia, J.; Du, P.; He, X.; Chanussot, J. Hyperspectral remote sensing image classification based on rotation forest. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 239–243.
20. Xia, J.; Chanussot, J.; Du, P.; He, X. Spectral–Spatial Classification for Hyperspectral Data Using Rotation Forests with Local Feature Extraction and Markov Random Fields. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2532–2546.

21. Du, P.; Samat, A.; Waske, B.; Liu, S.; Li, Z. Random Forest and Rotation Forest for fully polarized SAR image classification using polarimetric and spatial features. *J. Photogramm. Remote Sens.* **2015**, *105*, 38–53.
22. Hsu, P.H. Feature extraction of hyperspectral images using wavelet and matching pursuit. *ISPRS J. Photogramm. Remote Sens.* **2007**, *62*, 78–92.
23. Richards, J.A. *Remote Sensing Digital Image Analysis*; Springer: Berlin, Germany, 1999; Volume 3,
24. Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **1933**, *24*, 417.
25. Plaza, A.; Martinez, P.; Plaza, J.; Perez, R. Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 466–479.
26. Fukunaga, K. *Introduction to Statistical Pattern Recognition*; Academic Press: New York, NY, USA, 2013.
27. Wold, S.; Albano, C.; Dunn, W.J., III; Edlund, U.; Esbensen, K.; Geladi, P.; Hellberg, S.; Johansson, E.; Lindberg, W.; Sjöström, M. Multivariate data analysis in chemistry. In *Chemometrics*; Springer: Berlin, Germany, 1984; pp. 17–95.
28. Worsley, K.J.; Poline, J.B.; Friston, K.J.; Evans, A. Characterizing the response of PET and fMRI data using multivariate linear models. *NeuroImage* **1997**, *6*, 305–319.
29. Du, Q. Modified Fisher's linear discriminant analysis for hyperspectral imagery. *IEEE Geosci. Remote Sens. Lett.* **2007**, *4*, 503–507.
30. Barker, M.; Rayens, W. Partial least squares for discrimination. *J. Chemom.* **2003**, *17*, 166–173.
31. Arenas-García, J.; Camps-Valls, G. Efficient kernel orthonormalized PLS for remote sensing applications. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 2872–2881.
32. Arenas-García, J.; Petersen, K.; Camps-Valls, G.; Hansen, L.K. Kernel multivariate analysis framework for supervised subspace learning: A tutorial on linear and kernel multivariate methods. *J. Educ. Psychol.* **2013**, *30*, 16–29.
33. Leiva-Murillo, J.M.; Artés-Rodríguez, A. Maximization of mutual information for supervised linear feature extraction. *IEEE Trans. Neural Netw.* **2007**, *18*, 1433–1441.
34. Arenas-García, J.; Camps-Valls, G. Feature extraction from remote sensing data using Kernel Orthonormalized PLS. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2007), Barcelona, Spain, 23–28 July 2007; pp. 258–261.
35. Persello, C.; Bruzzone, L. Kernel-Based Domain-Invariant Feature Selection in Hyperspectral Images for Transfer Learning. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 2615–2626.
36. Camps-Valls, G.; Mooij, J.; Schölkopf, B. Remote sensing feature selection by kernel dependence measures. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 587–591.
37. Jiménez-Rodríguez, L.O.; Arzuaga-Cruz, E.; Vélez-Reyes, M. Unsupervised linear feature-extraction methods and their effects in the classification of high-dimensional data. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 469–483.
38. Arenas-García, J.; Petersen, K.B.; Hansen, L.K. Sparse kernel orthonormalized PLS for feature extraction in large data sets. *Adv. Neural Inf. Process. Syst.* **2007**, *19*, 33–40.
39. Breiman, L.; Friedman, J.; Stone, C.J.; Olshen, R.A. *Classification and Regression Trees*; CRC Press: Boca Raton, FL, USA, 1984.
40. Roweis, S.; Brody, C. *Linear Heteroencoders*; Gatsby Computational Neuroscience Unit, Alexandra House: London, UK, 1999.
41. Shawe-Taylor, J.; Cristianini, N. *Kernel Methods for Pattern Analysis*; Cambridge University Press: Cambridge, UK, 2004.
42. Camps-Valls, G. *Kernel Methods in Bioengineering, Signal and Image Processing*; Igi Global: Hershey, PA, USA, 2006.
43. Rosipal, R.; Trejo, L.J. Kernel partial least squares regression in reproducing kernel hilbert space. *J. Mach. Learn. Res.* **2002**, *2*, 97–123.
44. Ranawana, R.; Palade, V. Multi-Classifer Systems: Review and a roadmap for developers. *Inf. Fusion* **2006**, *3*, 1–41.
45. Cunningham, P.; Carney, J. Diversity versus quality in classification ensembles based on feature selection. In *Machine Learning: ECML 2000*; Springer: Berlin, Germany, 2000; pp. 109–116.

46. Zhang, C.X.; Zhang, J.S. RotBoost: A technique for combining Rotation Forest and AdaBoost. *Pattern Recog. Lett.* **2008**, *29*, 1524–1536.
47. Li, F.; Xu, L.; Siva, P.; Wong, A.; Clausi, D.A. Hyperspectral Image Classification With Limited Labeled Training Samples Using Enhanced Ensemble Learning and Conditional Random Fields. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 2427–2438.
48. Blaschko, M.B.; Shelton, J.A.; Bartels, A.; Lampert, C.H.; Gretton, A. Semi-supervised kernel canonical correlation analysis with application to human fMRI. *Inf. Fusion* **2011**, *32*, 1572–1583.
49. Xia, J.; Mura, M.D.; Chanussot, J.; Du, P.; He, X. Random Subspace Ensembles for Hyperspectral Image Classification with Extended Morphological Attribute Profiles. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 4768–4786.
50. Foody, G.M. Thematic map comparison. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 627–633.
51. Li, F.; Wong, A.; Clausi, D.A. Combining rotation forests and adaboost for hyperspectral imagery classification using few labeled samples. In Proceedings of the 2014 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Quebec City, QC, Canada, 13–18 July 2014 ; pp. 4660–4663.
52. Fauvel, M.; Tarabalka, Y.; Benediktsson, J.A.; Chanussot, J.; Tilton, J.C. Advances in spectral-spatial classification of hyperspectral images. *Proc. IEEE* **2013**, *101*, 652–675.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).