Hyperspectral Remote Sensing Image Classification Based on Rotation Forest

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

***Abstract*—In this letter, an ensemble learning approach, Rotation Forest, has been applied to hyperspectral remote sensing image classification for the first time. The framework of Rotation Forest is to project the original data into a new feature space using transformation methods for each base classifier (decision tree), then the base classifier can train in different new spaces for the purpose of encouraging both individual accuracy and diversity within the ensemble simultaneously. Principal component analysis (PCA), maximum noise fraction, independent component analysis, and local Fisher discriminant analysis are introduced as feature transformation algorithms in the original Rotation Forest. The performance of Rotation Forest was evaluated based on several criteria: different data sets, sensitivity to the number of training samples, ensemble size and the number of features in a subset. Experimental results revealed that Rotation Forest, especially with PCA transformation, could produce more accurate results than bagging, AdaBoost, and Random Forest. They indicate that Rotation Forests are promising approaches for generating classifier ensemble of hyperspectral remote sensing.**

***Index Terms*—Classification, decision tree, ensemble learning, hyperspectral remote sensing image, Rotation Forest.**

I. Introduction

HYPERSPECTRAL remote sensing image classification is a challenging problem because of its high dimensional inputs (hundreds of bands), many class outputs, and limited availability of reference data [1], [2]. Therefore, we require some powerful techniques to improve the accuracy of classification results. Since it is always difficult to select an optimal classifier, an attractive type of machine learning algorithm called the multiple classifier system (MCS) or classifier ensemble is rapidly developing and enjoying a lot of attentions due to their potential to improve the classification accuracy of hyperspectral remote sensing image significantly [1], [3].

MCS integrates the outputs of individual classifiers according to a certain combination approach (such as majority vote, Bayesian rule, etc.) or based on an iterative error minimization [3]–[7]. The outputs can be generated by the same classifier of different training samples, or the different classifiers of same/different training set. Previous studies demonstrated that a successful MCS should be one where the member classifiers are accurate as well as the diversities among them are obvious, because combining similar classification results would not further improve the accuracy [8]–[10]. Two popular approaches for creating strong ensembles are boosting and bagging [4], [11]. Boosting processes data with iterative retraining, and the weights of misclassified samples are increased to concentrate the learning algorithm on specific samples [4], [12]. In contrast, bagging can produce accurate ensemble by training many classifiers on boot-strapped samples from training set [11]. Diversity in bagging is provided with further randomization yielding Random Forest ensemble approach [13]. Random Forest adopts decision trees trained on bootstrap samples and the diversity is promoted with random choice of features at each node while constructing the trees. It can overcome the drawbacks of bagging and boosting algorithms (e.g., high computational cost and sensitivity to noise) [7]. In addition, limiting the number of variables in Random Forest used for a spilt, the computational complexity can be reduced and the correlation between the trees be decreased. This enables Random Forest to deal with high-dimensional datasets [9].

Rotation Forest, proposed by Rodriguez *et al.* [14] is based on the idea of Random Forest. The main idea of Rotation Forest is to encourage simultaneously both member diversities and individual accuracy within a classifier ensemble. In the framework of Rotation Forest, each classifier is independently constructed using decision tree method, and each tree is trained on the training samples in a rotated feature space derived from principal component analysis (PCA) transformation. One of the most important point of ensemble methods is to select the base classifier. Decision tree is always used for rotation task because of its sensitivity to rotation of the feature axes [15]. Though Rotation Forest performs much better than other ensemble methods (bagging, AdaBoost, Random Forest) on some benchmark classification from UCI repository [14], [16], the performance for classify hyperspectral remote sensing im-

Manuscript received October 29, 2012; revised February 21, 2013; accepted March 13, 2013. Date of publication May 16, 2013; date of current version November 8, 2013. This work was supported in part by the Jiangsu Provincial Natural Science Foundation under Grant BK2012018 and BK2010182, the Natural Science Foundation of China under Grant 41171323, and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

J. Xia is with Key Laboratory for Land Environment and Disaster Monitoring of State Bureau of Surveying and Mapping of China, China University of Mining and Technology, 221116 Xuzhou, China and also with the GIPSA-lab, Grenoble Institute of Technology, Grenoble 38400, France.

P. Du is with the Department of Geographical Information Science, Nanjing University, Nanjing 210093, China (e-mail: dupjrs@gmail.com).

X. He and J. Chanussot are with the GIPSA-lab, Grenoble Institute of Technology, Grenoble 38400, France.

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

Digital Object Identifier 10.1109/LGRS.2013.2254108

age has not been investigated. The objective of this letter is to adapt Rotation Forest to classify hyperspectral remote sensing image and compare it with other approaches (SVM, bagging, AdaBoost, and Random Forest) to evaluate its performance and applicability. Furthermore, the capability of different feature transformation algorithms, including PCA, independent component analysis (ICA), maximum noise fraction (MNF), and local Fisher discriminant analysis (LFDA) is also explored.

The remainder of this letter is arranged as follows. In Section II, we briefly introduce the Rotation Forest approach. Experimental results and discussion are given in Section III. Finally, Section IV draws the conclusions and addresses the future work.

II. Rotation Forest

Let $\mathbf{x} = [x_1, ..., x_n]^T$ be a training sample characterized by n features and X represents the training set of an $N \times n$ matrix. $Y = [y_1, ..., y_N]^T$ is defined as the class labels $\{1, ..., c\}$, where c is the total number of classes. Denote by $\Gamma_1, ..., \Gamma_L$ the classifiers in the ensemble, and F is the feature set.

The steps for training classifier Γ_i, $i = 1, ..., L$ are handled in the following [14]–[16].

1) F is split into K feature sets and each subset contains $M = n/K$ number of features.
2) Let $F_{i,j}$ be the jth, $j = 1, .., K$ subset of features for Γ_i, and $X_{i,j}$ be the features in $F_{i,j}$ from X. $X'_{i,j}$ is denoted as a new training set which is selected from $X_{i,j}$ randomly with the 75% size using bootstrap algorithm. Then, we transform X'_{ij} to get the coefficients $a_{i,j}^{(1)}, ..., a_{i,j}^{(M_j)}$, the size of $a'_{i,j}$ is $M \times 1$.
3) A sparse rotation matrix R_i is organized with the above coefficients

$$R_i = \begin{bmatrix} a_{i,1}^{(1)}, ..., a_{i,1}^{(M_1)} & 0 & \cdots & 0 \\ 0 & a_{i,2}^{(1)}, ..., a_{i,2}^{(M_2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{i,K}^{(1)}, ..., a_{i,K}^{(M_K)} \end{bmatrix}$$

The columns of R_i is rearranged to R_i^a with respect to the original feature set. Then, the training set will become XR_i^a. In this case, all classifiers will be trained in parallel style.

For a given test sample χ, the confidence is calculated for each class by the average combination method

$$\mu_k(\chi) = \frac{1}{L}\sum_{i=1}^{L} \gamma_{i,k}\left(\chi R_i^a\right), \quad k = 1, ..., c \tag{1}$$

where, $\gamma_{i,k}\left(\chi R_i^a\right)$ is the probability generated by the classifier Γ_i to the hypothesis that χ belongs to class k.

Finally, χ will be assigned to the class with the largest confidence. It is very important to note 2), selecting of the sample size $X'_{i,j}$ smaller than $X_{i,j}$ aims at two aspects: one is to avoid obtaining the same coefficients of the transformed components if the same features are chosen, and the other is to enhance the diversity among the generated ensemble member classifiers.

The success of Rotation Forest relies on the base classifier and the rotation matrix created by the transformation methods. Decision tree is always adopted for Rotation Forest because it is sensitive to rotation of the axes. Here, we selected classification and regression tree (CART) as the base classifier [17].

CART is based on the *Gini* index, which is treated as node impurity criterion [9], [17]

$$Gini(t) = \sum_{i=1}^{c} P_{\omega_i}(1 - P_{\omega_i}) \tag{2}$$

where c is the number of classes and P_{ω_i} is the probability of class ω_i at node t. P_{ω_i} is defined as

$$P_{\omega_i} = \frac{n_{\omega_i}}{N} \tag{3}$$

where n_{ω_i} is the number of samples of class ω_i and N is the total number of training samples.

The summed *Gini* index selects the split that maximizes the decrease in impurity. By employing this rule, CART generates a sequence of subtrees by growing a large tree and pruning it back until only the root node is left. Then it uses cross-validation to estimate the misclassification cost of each subtree and chooses the one with the lowest estimated cost [18].

In [19], the authors compared the performance of different transformation algorithms (e.g., PCA, NDA, and RP) and found that PCA produced the best results. In this letter, we will further examine the efficiency of common transformation algorithms applied to hyperspectral remote sensing image classification, such as PCA [20], maximum noise fraction (MNF) [21], [22], independent component analysis (ICA) [23], [24], and LFDA. PCA, MNF, and ICA are all unsupervised feature extraction methods, while LFDA is supervised. PCA and MNF maximize the amount of data variance and signal-to-noise ratio (SNR), respectively, yielding a transformed data set in a new uncorrelated coordinate system, while ICA transforms the data into maximally independent components [20]–[22], [24]. However, PCA, MNF, and ICA all maximize the information contained in the first transformed components, relegating variations of less significant size to low-order components [25]. LFDA effectively combines the ideas of Fisher discriminant analysis (FDA) and locality-preserving projection (LPP) [26]. That makes LFDA can both maximize between-class separability and preserves with-class local structure. More details about LFDA can be seen in [26]. In order to preserve the variability information in the images, all components using the above three transformation methods are retained.

III. Experiments and Result Analysis

In order to assess the performance of Rotation Forest algorithm, we conduct the experiments with three widely used hyperspectral images obtained from NASA's Airborne Visible Infra-Red Imaging Spectrometer (AVIRIS), Reflective Optics System Spectrographic Imaging System (ROSIS), and Digital Airborne Imaging Spectrometer (DAIS) owned by the German Aerospace Center (DLR). AVIRIS dataset is captured over a vegetation area of Indian Pines, Indiana, USA. The

TABLE I
TRAINING AND TEST SAMPLES OF THE THREE HYPERSPECTRAL IMAGES

AVIRIS			ROSIS			DAIS		
Class name	Training set	Test set	Class name	Training set	Test set	Class name	Training set	Test set
Soybeans-min till	493	2468	Asphalt	548	6641	Water	202	4281
Grass/pasture	99	497	Meadows	540	18649	Trees	205	2424
Hay-windrowed	98	489	Gravel	392	2099	Meadows	206	1251
Soybeans-clean till	123	614	Trees	524	3064	Bricks	315	2237
Corn-no till	287	1434	Metal Sheets	265	1345	Soil	205	1475
Grass/tree	149	747	Soil	532	5029	Asphalt	204	1704
Soybeans-no till	194	968	Bitumen	375	1330	Bitumen	202	685
Woods	259	1294	Bricks	514	3682	Parking lot	201	287
Corn-min till	167	834	Shadows	231	947	Shadows	119	241

TABLE II
OVERALL ACCURACIES (%) FOR THE INDIAN PINES *AVIRIS* IMAGE USING DIFFERENT NUMBER OF TRAINING SAMPLES

Number of Training sample	CART	Bagging	AdaBoost	RF	RoF (PCA)	RoF (ICA)	RoF (MNF)	RoF (LFDA)	SVM	LORSAL
Case 1	57.25	66.5	66.98	71.38	79.65	76.1	76.78	71.66	76.82	**84.3**
Case 2	62.26	73.12	73.3	75.82	84.87	84.52	82.03	77.78	82.02	**87.46**
Case 3	67.74	76.86	77.6	80.31	87.51	86.58	84.39	81.26	84.57	**89.13**
Case 4	68.57	80.76	80.35	83.96	88.6	88.36	86.59	84.01	87.06	**90.01**

image contains 145 × 145 pixels, with 200 spectral bands after removing 20 water absorption bands (104–108, 150–163, and 220). The spatial resolution is 20 m/pixel. ROSIS image with 115 spectral channels is acquired over the University of Pavia, Italy. The image size is 610 × 340 with the spatial resolution of 1.3 m. Twelve noisy channels were removed and the remaining 103 bands with a spectral range from 0.43 to 0.86 μm were used for the experiments. The DAIS image was collected at 1500 m flight altitude over the city of Pavia, Italy, with ground resolution of 5 m and size of 400 × 400 pixels with 80 spectral bands. Training and test samples are detailed in Table I.

In all cases, the performance achieved by Rotation Forest is illustrated using the following designs.

1) Number of features in each subset: $M = 10$.
2) Number of classifiers in the ensemble: $L = 10$.
3) Feature extraction method: PCA [20], ICA [23], MNF [21], and LFDA [26].
4) Base classifier: CART.

Furthermore, the popular methods, including bagging [11], AdaBoost [4], Random Forest (RF) [13], support vector machine (SVM) [27] and logistic regression via variable splitting and augmented Lagrangian (LORSAL) [28] are added to be compared with Rotation Forest. The required parameters of SVM with RBF are the penalty factor C and kernel width γ. We use a fivefold cross-validation grid search method to find the best combination of C and γ within the set $C \in \{2^1, 2^2, 2^3, 2^4, 2^5, 2^6, 2^7, 2^8\}$ and $\gamma \in \{2^{-3}, 2^{-2}, 2^{-1}, 2^0, 2^1, 2^2, 2^3\}$. The number of neighbors in LFDA is chosen from $\{1, 3, 5, 7\}$ using fivefold cross validation. In the following experiments, we employed RoF (PCA, ICA, MNF, and LFDA) as the abbreviations of Rotation Forest with PCA, ICA, MNF, and LFDA transformations. Overall accuracy (OA) is used as the quantitative index. In order to investigate the impact of the labeled samples on the classification accuracy, we randomly select the labeled samples with replacement from the original training samples whose sizes correspond to four cases: 25% (case 1), 50% (case 2), 75% (case 3), 100% (case 4) rate of original size of training samples. In order to increase the statistical significance of the results, each value of OA reported in this letter is obtained as the average of 10 Monte Carlo (MC) runs.

A. *Results of Indian Pines AVIRIS Image*

Table II shows the classification accuracies (OA%) obtained by the Rotation Forest approaches as well as other algorithms using different training samples. We highlight the highest OA of each case in bold font. It can be seen that RoF (PCA, ICA) achieve good results than other ensemble approaches (bagging, Adaboost, and Random Forest), where the OA is always increased as the number of training samples is increased. For instance, in case 1, CART, bagging, Adaboost, and RF acquired an OA of 57.25%, 66.5%, 66.98%, and 71.38, respectively. RoF (PCA) and RoF (ICA), respectively, increased the OA to 79.65% and 76.78%, while the OA of RoF (MNF) and RoF (LFDA) were improved to 76.78% and 71.66%. LORSAL yielded the highest OA and RoF (PCA) gave the better performance than SVM in all cases. In addition, we have compared the computation time of these methods on an Intel(R) Xeon(R) CPU X5660 @ 2.80 GHz 2.79 GHz, two processors, 12 GB memory. In case 2, the computation time of RoF (PCA), RoF (MNF), RoF (ICA), RoF (LFDA), SVM, and LORSAL was 8.53 s, 9.28 s, 9.16 s, 36.84 s, 42.65 s, and 3.18 s, respectively. The computation time of Rotation Forest approaches is longer than the one of bagging, Adaboost, and Random Forest. For the approaches of SVM and RoF (LFDA), the computational time also included the time consumed on the parameter determination. So RoF (PCA), RoF (MNF), and RoF (ICA) are more efficient than RoF (LFDA) and SVM. Among them, LORSAL method is the fastest.

TABLE III
OVERALL ACCURACIES (%) FOR THE UNIVERSITY OF PAVIA *ROSIS* IMAGE USING DIFFERENT NUMBER OF TRAINING SAMPLES

Number of Training sample	CART	Bagging	AdaBoost	RF	RoF (PCA)	RoF (ICA)	RoF (MNF)	RoF (LFDA)	SVM	LORSAL
Case 1	59.33	67.38	66.83	66.77	**78.38**	71.92	70.45	73.02	76.42	71.28
Case 2	62.82	68.26	67.8	68.9	**80.71**	76.37	71.48	75.2	77.35	76.97
Case 3	63.39	69.64	70.13	69.9	**82.89**	75.91	72.59	75.73	77.86	78.08
Case 4	64.93	70.11	70.3	71.11	**83.14**	78.04	73.28	75.57	79.98	80.09

TABLE IV
OVERALL ACCURACIES (%) FOR THE CENTER OF PAVIA *DAIS* IMAGE USING DIFFERENT NUMBER OF TRAINING SAMPLES

Number of Training sample	CART	Bagging	AdaBoost	RF	RoF (PCA)	RoF (ICA)	RoF (MNF)	RoF (LFDA)	SVM	LORSAL
Case 1	87.95	90.89	91.83	93.12	**95.64**	95.2	95.06	95.52	93.95	93.76
Case 2	90.51	91.49	92.45	93.93	**95.76**	95.36	94.91	95.6	94.17	94.45
Case 3	91.25	92.09	92.22	93.95	**95.78**	95.64	95.15	95.57	94.67	94.47
Case 4	91.57	92.17	92.61	94.8	95.81	95.48	95.28	**95.92**	95.1	94.8

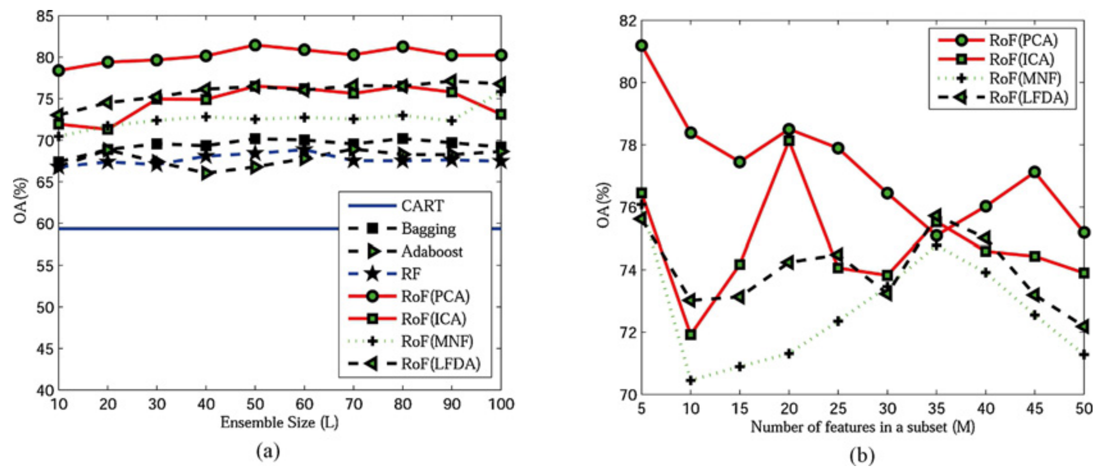

Fig. 1. Impact of OA using different number of L and M obtained from ROSIS image (case 1).

B. Results of University of Pavia ROSIS Image

Table III reports the OA for each approaches using different number of training samples. From Table III, it can be observed that RoF (PCA) outperform other algorithms in all cases. RoF (LFDA) gave the better performance than RoF (MNF). In case 4, the corresponding OA of RoF (PCA) achieved on the test set was 83.14%, higher than the one of SVM (79.98%) and LORSAL(80.09%). More details can be seen in Table III. The computation complexity is similar with the previous AVIRIS experiment. The computation complexity of RoF (PCA), RoF (MNF), and RoF (ICA) are much less than the one of SVM and RoF (LFDA). And LORSAL method is more effective than RoF (PCA) algorithm.

C. Results of Center of Pavia DAIS Image

The global accuracies of different method using different number of samples are reported in Table IV. The Pavia Center *DAIS* data set was easy to classify since even the CART acquires extraordinarily high classification accuracy. Regarding the global accuracies, Rotation Forest with different transformation algorithms are all superior to other compared approaches. In case 1–3, RoF (PCA) achieved the best global accuracies with the OA (95.58%). And RoF (LFDA) yielded the highest OA (95.92%) in case 4.

D. Sensitivity of Parameters

Ensemble size (L) and the number of features in a subset (M) are the key parameters of Rotation Forest, also known as an indicator of the operating complexity. In order to investigate the impacts of these parameters, we have performed the classification results using different ensemble size when the number of subset M is fixed to 10, different number of features in a subset when ensemble size L is fixed to 10.

Fig. 1 shows the OA (%) using different number of L and M obtained from ROSIS image (case 1). With the increment of L, the classification results are significantly improved. Different approaches achieved the best classification result at different number of L and M. For instance, when M was fixed, RoF (PCA) obtained the best OA when L equals to 50, RoF (MNF) obtained the best result when L equals to 80. Furthermore, we also conducted the similar experiments on AVIRIS and DAIS images.

IV. CONCLUSION

A method for generating ensemble of classifiers, Rotation Forest, was introduced into hyperspectral remote sensing image classification. It consists in splitting the feature set into K subsets, running transformation algorithms separately on each subset and then reassembling a new extracted feature set while keeping all the components. CART decision tree classifier is used as the base classifier. Different splits of the feature set lead to different rotations. Thus diverse classifiers are obtained. Thus, we target diversity and accuracy together. We have applied Rotation Forest using different transformation approaches, including PCA, MNF, ICA, and LFDA to classify hyperspectral remote sensing image and compared with bagging, AdaBoost, Random Forest, and other advance classifiers. Experimental results have shown that RoF (PCA) outperformed other methods in terms of accuracies. The key parameters of Rotation Forest are also explored in this letter. Future studies will be focused on the integration of Rotation Forest with other ensemble approaches, the selection of an optimized decision tree model, and the use of other effective feature extraction algorithms.

ACKNOWLEDGMENT

The authors would like to thank Prof. D. Landgrebe from Purdue University, West Lafayette, IN, USA, for providing the publicly available AVIRIS data and Prof. P. Gamba from the University of Pavia, Italy, for providing ROSIS and DAIS images.

REFERENCES

[1] J. C.-W. Chan and D. Paelinckx, “Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery,” *Remote Sens. Environ.*, vol. 112, no. 6, pp. 2999–3011, Jun. 2008.

[2] Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, “Multiple spectral-spatial classification approach for hyperspectral data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4122–4132, Nov. 2010.

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

[4] Y. Freund and R. E. Schapire, “Experiments with a new boosting algorithm,” in *Proc. 13th Int. Conf. Mach. Learning*, Jul. 1996, pp. 148–156.

[5] P. Du, J. Xia, W. Zhang, K. Tan, Y. Liu, and S. Liu, “Multiple classifier system for remote sensing image classification: A review,” *Sensors*, vol. 12, no. 4, pp. 4764–4792, 2012.

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

[7] J. A. Benediktsson, J. Chanussot, and M. Fauvel, “Multiple classifier systems in remote sensing: From basics to recent developments,” in *Proc 7th Int. Workshop Multiple Classifier Syst.*, May 2007, pp. 501–512.

[8] X. Ceamanos, B. Waske, J. A. Benediktsson, J. Chanussot, M. Fauvel, and J. R. Sveinsson, “A classifier ensemble based on fusion of support vector machines for classifying hyperspectral data,” *Int. J. Image Data.*, vol. 1, no. 4, pp. 293–307, Dec. 2010.

[9] B. Waske, J. A. Benediktsson, K. Arnason, and J. R. Sveinsson, “Mapping of hyperspectral AVIRIS data using machine learning algorithms,” *Can. J. Remote Sens.*, vol. 35, no. S1, pp. 106–116, 2009.

[10] L. I. Kuncheva and C. J. Whitaker, “Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy,” *Mach. Learn.*, vol. 51, no. 2, pp. 181–207, 2003.

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

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

[13] L. Breiman, “Random Forest,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[14] 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.

[15] A. Ozcift and A. Gulten, “Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms,” *Comput. Meth. Prog. Bio.*, vol. 104, no. 3, pp. 443–451, Dec. 2011.

[16] C. Zhang and J. Zhang, “RotBoost: A technique for combining rotation forest and Adaboost,” *Pattern Recogn. Lett.*, vol. 29, no. 10, pp. 1524–1536, Jul. 2008.

[17] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*. Boca Raton, FL: CRC Press, 1984.

[18] W.-Y. Loh, “Classification and regression tree methods,” in *Encyclopedia of Statistics in Quality and Reliability*, F. Ruggeri, R. Kenett, and F. W. Faltin, Eds. New York: Wiley, 2008, pp. 315–323.

[19] L. I. Kuncheva and J. J. Rodriguez, “An experimental study on rotation forest ensembles,” in *Proc. 7th Int. Workshop Multiple Classifier Syst.*, May 2007, pp. 459–468.

[20] J. Richards and X. Jia, *Remote Sensing Digitial Image Analysis*. 4th ed. Berlin, Heidelberg: Springer-Verlag, 2006.

[21] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, “A transformation for ordering multispectral data in terms of image quality with implications for noise removal,” *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 1, pp. 65–74, Jan. 1988.

[22] C.-I. Chang and Q. Du, “Interference and noise adjusted principal components analysis,” *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2387–2396, Sep. 1999.

[23] A. Hyvarinen and E. Oja, “Independent component analysis: Algorithms and applications,” *Neural Netw.*, vol. 13, no. 4–5, pp. 411–430, Jun. 2000.

[24] J. Wang and C.-I. Chang, “Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis,” *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1586–1600, Jun. 2006.

[25] I. Dopido, M. Zortea, A. Villa, A. Plaza, and P. Gamba, “Unmixing prior to supervised classification of remotely sensed hyperspectral images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 760–764, Jul. 2011.

[26] M. Sugiyama, “Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis,” *J. Mach. Learn. Res.*, vol. 27, no. 8, pp. 1021–1064, 2007.

[27] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. Tilton, and G. Trianni, “Recent advances in techniques for hyperspectral image processing,” *Remote Sens. Environ.*, vol. 113, no. S1, pp. S110–S122, Sep. 2009.

[28] J. Li, J. Bioucas-Dias, and A. Plaza, “Spectral-spatial classification of hyperspectral data using loopy belief propagation and active learning,” *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 844–856, Feb. 2013.