Dynamic Synthetic Minority Over-Sampling Technique-Based Rotation Forest for the Classification of Imbalanced Hyperspectral Data

Wei Feng, Gabriel Dauphin, Wenjiang Huang, Yinghui Quan, Wenxing Bao, Mingquan Wu, and Qiang Li

***Abstract*—Rotation forest (RoF) is a powerful ensemble classifier and has attracted substantial attention due to its performance in hyperspectral data classification. Multi-class imbalance learning is one of the biggest challenges in machine learning and remote sensing. The standard technique for constructing RoF ensemble tends to increase the overall accuracy; RoF has difficulty to sufficiently recognize the minority class. This paper proposes a novel dynamic SMOTE (synthetic minority oversampling technique)-based RoF algorithm for the multi-class imbalance problem. The main idea of the proposed method is to dynamically balance the class distribution before building each rotation decision tree. A resampling rate is set in each iteration (ranging from 10% in the first iteration to 100% in the last) and this ratio defines the number of minority class instances randomly resampled (with replacement) from the original dataset in each iteration. The rest of the minority class instances are generated by the SMOTE method. The reported results on three real hyperspectral datasets show that the proposed method can get better performance than random forest, RoF, and some popular data sampling methods.**

***Index Terms*—Ensemble learning, hyperspectral image classification, imbalance learning, rotation forest (RoF).**

Manuscript received September 15, 2018; revised April 12, 2019 and May 20, 2019; accepted May 31, 2019. This work was supported in part by the National Key R&D Program of China under Grant 2016YFB0501501, in part by Hainan Provincial Key R&D Program of China under Grant ZDYF2018073, and in part by the National Natural Science Foundation of China under Grants 41601466, 41871339, and 61461003. *(Corresponding author: Wei Feng.)*

W. Feng is with the Key laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China, and also with the School of Electronic Engineering, Xidian University, Xian 710071, China (e-mail: fengwei@radi.ac.cn).

G. Dauphin is with the Laboratory of Information Processing and Transmission, L2TI, Institut Galilée, University Paris XIII, 93430 Villetaneuse, France (e-mail: gabriel.dauphin@univ-paris13.fr).

W. Huang is with the Key laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China (e-mail: huangwj@radi.ac.cn).

Y. Quan is with the Key Laboratory for Radar Signal Processing, Xidian University, Shaanxi 710071, China (e-mail: yhquan@mail.xidian.edu.cn).

W. Bao is with the School of Computer Science and Engineering, Beifang University of Nationalities, Yinchuan 750021, China (e-mail: bwx71@163.com).

M. Wu is with the State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China (e-mail: wumq@radi.ac.cn).

Q. Li is with the Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, China (e-mail: liruo@itp.ac.cn).

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/JSTARS.2019.2922297

I. Introduction

HYPERSPECTRAL image data is widely used in land cover mapping, environmental modeling and monitoring, and updating geographical databases [1]. Class distribution, i.e., the proportion of instances belonging to each class in a dataset, plays a key role in remote sensing research [2]. However, many hyperspectral data classification tasks suffer from the class imbalanced problem, in which some classes are highly underrepresented as compared to other classes [3], [4]. This skewed distribution makes many classic classification algorithms less effective [3]. In addition, dealing with imbalanced multi-class tasks, which always happen in remote sensing data, is harder than dealing with two-class ones [5]–[8]. Consequently, how to classify imbalanced data of multi-classes effectively has emerged as one of the biggest challenges in machine learning and remote sensing.

The objective of imbalance learning can be generally described as *obtaining a classifier that will provide high accuracy for the minority class without severely jeopardizing the accuracy of the majority class* [2]. Many efforts have been devoted to imbalanced learning problems, such as cost-sensitive methods [9], kernel-based methods [10], and active learning methods [11]. However, cost-sensitive methods need to define misclassification costs, which are not usually available in the datasets. The kernel-based methods and active learning methods have large computation costs, especially for large datasets [11], [12]. The sampling approach is popular in dealing with the class imbalance problem of remote sensing [13]–[15]. The method avoids the modification of the learning algorithm and tries to decrease the effect of imbalanced data with a preprocessing step. Random under-sampling (RUS), random over-sampling (ROS) and synthetic minority oversampling technique (SMOTE) are the most popular data sampling methods. Ensemble learning, which fuses the predictions made by the single classifiers into global predictions, has been successfully applied to hyperspectral image classification [13], [16]–[18]. The ensemble-based method is a new category in imbalanced domains [12]. Chawla *et al.* proposed the SMOTEBoost method by combining SMOTE with the AdaBoost.M2 [19] ensemble algorithm [20]. Similar boosting-based ensemble methods are RUSBoost (Random UnderSampling Boosting) [21], EUSBoost (Evolutionary UnderSampling Boosting), [22] and cost-sensitive boosting [23]. However, these methods pay more attention to the binary imbalanced problem

[24]. Moreover, most boosting-based methods are very sensitive to noise [2]. Bagging-based methods have some advantages such as easy operation, robustness, and high scalability of bagging, hence, they are more popular in imbalance learning [25]. Examples in this regard include SMOTEBagging [5] and UnderBagging [26], which could be adopted in dealing with both binary and multi-class imbalance issues [2]. Random forest (RF) is a successful version of bagging. Hence, RF should deal with imbalanced data by being extended from all bagging-based imbalance learning methods and may outperform these methods, especially for multi-class tasks [2]. Two well-known RF-based imbalance learning methods are balanced random forests (BRF) [27] and weighted random forest (WRF) [27].

Rotation forest (RoF) is a powerful ensemble classifier that draws upon the idea of RF but has attracted more attention due to its performance in the hyperspectral remote sensing domain [13], [17], [28]–[31]. Several approaches have been proposed to further improve the performance of RoF [30]–[32]. However, the performance of these methods are evaluated only in the case of balanced data distribution. There is relatively less investigation of RoF when dealing with multi-class imbalanced data. Moreover, as the standard technique for constructing the RoF ensemble only tends to increase the overall accuracy, RoF still has difficulty to sufficiently recognize the minority class [33]. Hence, the RoF ensemble methods have to be designed specifically to effectively handle the class imbalance problem [12].

The major contribution of this paper is to propose a novel dynamic synthetic minority oversampling technique combined RoF (DSRoF) algorithm for the multi-class imbalance problem. The main idea of the proposed method is to dynamically balance the class distribution by SMOTE before building each rotation decision tree. A resampling rate $\alpha\%$ is set in each iteration (ranging from 10% in the first iteration to 100% in the last) and this ratio defines the number of minority class instances randomly resampled (with replacement) from the original dataset in each iteration. The rest of the minority class instances are generated by the SMOTE method. The effectiveness of the proposed approach is assessed on three real hyperspectral datasets. The reported results show that the proposed method can get better performance than RF, RoF, and data sampling methods (RUS, ROS, and SMOTE).

The remaining part of this paper is organized as follows. Section II presents an overview of the related work. Section III describes in detail the proposed methodology. Then, Section IV presents the results and discussion. Finally, the conclusions are given in Section V.

II. Related Work

When RF and RoF are employed for accurate hyperspectral image classification, the data sampling methods, such as RUS, ROS, and SMOTE, are always adopted in the data preprocessing step to alleviate the negative effect of the class imbalance. The objective of this paper is to find an internal oversampling-based ensemble method to overcome the multi-class imbalance problem. This section provides an introduction to the background on the data sampling methods (RUS, ROS, and SMOTE), RF, and RoF. A training set with N samples and L classes is denoted as $S = [X, Y] = \{x_i, y_i\}_{i=1}^{N}$, where X is the dataset containing the training objects, x_i is a D dimensional vector with feature values $\mathbb{F}$, Y is a vector with class labels for the data, and y_i takes a value from the set of class labels $1, ..., L$.

A. Data Random Sampling Methods

RUS balances class distribution through the random elimination of majority class examples, i.e., $|N'_{maj}| = |N_{min}|$, where $|N'_{maj}|$ is the size of the majority class of the balanced dataset and $|N_{min}|$ is the size of the minority class of the original dataset.

ROS gets balanced data by randomly replicating minority class instances. In ROS, a new minority training set is sampled (with replacement) from the original minority class training instances such that $|N'_{min}| = |N_{maj}|$, where $|N'_{min}|$ is the size of the minority class of the final balanced dataset and $|N_{maj}|$ is the size of the majority class of the original dataset.

B. Synthetic Minority Over-Sampling Technique (SMOTE)

SMOTE, proposed by Chawla *et al.*, is the most popular oversampling method and it can avoid the overfitting problem [34]–[36]. Its main idea is to create new minority class examples by interpolating several minority class instances that lie together. Different from the ROS, the new minority class instances of SMOTE are synthesized in the feature space rather than in the data space. The new synthetic samples are created by specifying two parameters: the number of nearest neighbors (k) and the over-sampling rate. In SMOTE, a minority class sample x_1 is first chosen. Then, new synthetic samples x_{new} for continuous features are generated through the following three steps 1): calculate the distance between x_1 and one of its k nearest neighbors x_2, $x_1 - x_2$; 2) multiply the distance obtained in the first step by a random number δ between 0 and 1, $\delta(x_1 - x_2)$; and 3) add the value obtained from second step to the feature value of the original feature vector, $x_{new} = x_1 + \delta(x_1 - x_2)$. The over-sampling rate decides how many synthetic instances of the minority class will be generated.

C. Random Forest (RF)

RF is an improved version of bagging. But it uses two randomization principles: random instance selection and random feature selection. A RF could be described as a classifier consisting of a collection of tree-structured classifiers $h(x, \Theta_k), k = 1, ..., L$ where Θ_k are independent and identically distributed random vectors and each tree $h(x, \Theta_k)$ casts a unit vote for the most popular class at input x [37]. The base classifier in RF is Classification and Regression Trees (CART). However, there are two main differences between the trees in RF and traditional CART: 1) in the growing step, at each node, the best split is calculated only among a fixed number of input variables which are randomly chosen; and 2) the tree in RF grows to the maximum depth with no pruning.

D. Rotation Forest (RoF)

The RoF method [28] draws upon the idea of RF. It has more diversity by combining several randomization techniques to build a sub-problem dataset which is projected into a new feature space. An RoF model of T size is built by carrying out the following steps.

1) RoF uses bootstrap technology to get several diverse training sets.
2) The feature space $\mathbb{F}$ of a training set is split randomly into G subspaces which are disjoint.
3) Then the principal component analysis (PCA) [38] is applied separately on each subspace $\mathbb{F}_{t,g}$ $(g \leq G, t \leq T)$ to obtain the coefficients $c_{t,g}$ $(g \leq G, t \leq T)$. The obtained vectors with coefficients are organized in a sparse "rotation" matrix R_t $(t \leq T)$.

$$R_t = \begin{bmatrix} c_{t,1}^{(1)} \cdots c_{t,1}^{(M_1)} & 0 & \cdots & 0 \\ 0 & c_{t,2}^{(1)} \cdots c_{t,2}^{(M_2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c_{t,g}^{(1)} \cdots c_{t,g}^{(M_G)} \end{bmatrix}$$

where $M = \lfloor D/G \rfloor$ is the feature number in each subset.

4) A rotation matrix R_t' $(t \leq T)$ is constructed by rearranging the columns of the matrix R_t to match the order of original features $\mathbb{F}$.
5) A new training dataset $S_t' = [S_t \cdot R_t', Y_t]$ is produced, and then used to train an individual classifier h_t $(t \leq T)$.
6) The results of a series of individual classifiers, generated by repeating the aforementioned process on all diverse training sets, are fused by majority vote rule.

III. METHOD

The process of the novel DSRoF method is detailed in Algorithm 1. DSRoF is inspired by SMOTEBagging [5]. This method has a lower risk of losing information than using RUS in the preprocessing step then training a RoF (RUS-RoF). In addition, DSRoF could result in more ensemble diversity than using ROS or SMOTE in the data preprocessing step before building a RoF model (ROS-ROF and SMOTE-RoF). Moreover, the DSRoF framework could be used in dealing with the multi-class imbalanced data problem, one of the biggest challenges in hyperspectral data classification.

Let us suppose N_i is the number of training instances of the ith class. Those classes are sorted in descending order according to their number of instances. Then, N_1 is the training size of the largest class 1 and N_L is the training size of the smallest class L. A resampling rate $\alpha\%$ is adopted to define the number of minority class instances which will be selected from the original dataset in each iteration. By considering the case of N_i $(1 \leq i \leq L) < \alpha\% \cdot N_1$, we use random resample with replacement to obtain the $\alpha\% \cdot N_1$ minority class samples from the original dataset. The $(1 - \alpha\%) \cdot N_1$ minority class instances are generated by the SMOTE algorithm. All the instances of the largest class are kept. A new balanced dataset is produced by combining the samples of the largest class, bootstrapped samples, and generated instances of all the other classes. In the next phase, the feature set $\mathbb{F}$ of the balanced training set is split randomly into G disjoint subsets, then principal component analysis (PCA) is applied separately on each subset. Finally, a decision tree is trained on the new training data which is produced by concatenating the linear extracted features contained in each subset as in the traditional RoF method.

Algorithm 1: Dynamic SMOTE-Based Rotation Forest (DSRoF).

Training phase
Input: $S = [X, Y] = \{x_i, y_i\}_{i=1}^{N}$: training set; $\alpha\%$: resampling rate; L: number of classes; N_i: the number of training instances of ith class; $\mathbb{F}$: feature set; G: number of feature subsets; T: number of classifiers; h: Ensemble creation algorithm; $E = \varnothing$: a rotation forest ensemble.
Process:
Sort the classes of the imbalanced data in descending order according to their number of instances.
for t = 1:T **do**
  Keep all the N_1 instances of the largest class 1
  for c = 2:L **do**
    Obtain the dataset S_{ct} $(c = 2, ..., L)$ by randomly resampling $\alpha\% \cdot N_1$ instances from the original dataset with replacement
    Obtain a balanced dataset S_{ct}' $(c = 2, ..., L)$ by combining S_{ct} $(c = 2, ..., L)$ and $(1 - \alpha\%) \cdot N_1$ minority class instances generated by the SMOTE algorithm
  end for
  Construct a new balanced dataset S_t $(c = 1, ..., L)$ by combining the N_1 largest class training instances with S_{ct}' $(c = 2, ..., L)$.
  Randomly split the feature set $\mathbb{F}$ of S_t into G subsets $\mathbb{F}_{t,g}$
  for g = 1:G **do**
    Compose the data $S_{t,g}$ for the features in $\mathbb{F}_{t,g}$ frame and apply PCA on $S_{t,g}$ to get the coefficients $c_{t,k}$
  end for
  Construct a rotation matrix R_t' by rearranging the columns of the matrix R_t composed of the coefficients $c_{t,g}$ to match the order of original features $\mathbb{F}$.
  Obtain the final balanced training set $S_t' = [S_t \cdot R_t', Y_t]$
  Train an classifier h_t using S_t'
  $E \leftarrow E \cup h_t$
  Change percentage $\alpha\%$.
end for
Output: The ensemble E

Prediction phase
Inputs: the ensemble $E = \{h_t\}_{t=1}^{T}$; a new sample x^*.
Output: Class label

$$y^* = argmax \sum_{t}^{T} {}_{(h_t(x^*)=c,\ c \in \{1,2,...L\})} 1.$$

TABLE I
DATA INFORMATION

	Indian Pines AVRIS			University of Pavia ROSIS			Salinas		
	Class No.	Train.	Test		Train.	Test		Train.	Test
1	Alfalfa	23	23	Asphalt	331	6300	Brocoli_green_weeds_1	100	1909
2	Corn-notill	428	1000	Meadows	932	17717	Brocoli_green_weeds_2	186	3540
3	Corn-mintill	249	581	Gravel	104	1995	Fallow	98	1878
4	Corn	71	166	Trees	153	2911	Fallow_rough_plow	69	1325
5	Grass-pasture	144	339	Painted metal sheets	67	1278	Fallow_smooth	133	2545
6	Grass-trees	219	511	Bare Soil	251	4778	Stubble	197	3762
7	Grass-pasture-mowed	14	14	Bitumen	66	1264	Celery	178	3401
8	Hay-windrowed	143	335	Self-Blocking Bricks	184	3498	Grapes_untrained	563	10708
9	Oats	10	10	Shadows	47	900	Soil_vinyard_develop	310	5893
10	Soybean-notill	291	681				Corn_senesced _green_weeds	163	3115
11	Soybean-mintill	736	1719				Lettuce_romaine_4wk	53	1015
12	Soybean-clean	177	416				Lettuce_romaine_5wk	96	1831
13	Wheat	61	144				Lettuce_romaine_6wk	45	871
14	Woods	379	886				Lettuce_romaine_7wk	53	1017
15	Buildings-Grass -Trees-Drives	115	271				Vinyard_untrained	363	6905
16	Stone-Steel-Towers	46	47				Vinyard_vertical_trellis	90	1717
Total		3106	7143		2135	40641		2697	51432

TABLE II
CLASSIFICATION RESULTS (%) OF THE INDIAN PINES AVRIS IMAGE, RESPECTIVELY OBTAINED BY RF, ROF, RUS-ROF, ROS-ROF, SMOTE-ROF, AND THE PROPOSED DSROF IN THE CASE OF THE IMBALANCE RATIO OF 73.6

IR:73.6	RF	RoF	RUS-RoF	ROS-RoF	SMOTE-RoF	DSRoF
1	59.13±6.22	0.00±0.00	**87.61±13.70**	80.43±5.89	79.13±3.43	76.96±5.04
2	71.72±1.73	71.97±2.47	55.98±9.49	63.35±2.26	62.59±2.25	**84.31±0.65**
3	59.14±1.86	51.03±1.16	50.10±7.75	56.75±3.94	57.49±1.18	**75.39±1.34**
4	47.59±2.95	3.49±6.68	74.88±10.82	87.47±1.77	**87.53±1.5**	87.05±2.68
5	89.20±1.49	90.44±1.03	77.14±6.95	88.73±2.05	89.12±1.89	**95.96±0.64**
6	95.23±1.35	97.69±0.58	88.66±5.48	94.01±0.87	93.97±0.83	**98.40±0.56**
7	52.14±5.88	0.00±0.00	**86.07±6.34**	85.00±2.26	85.71±0.00	84.29±3.01
8	96.42±1.03	**99.94±0.13**	86.21±9.52	94.12±1.75	94.36±1.78	99.70±0.01
9	40.00±11.55	0.00±0.00	**92.00±11.52**	79.00±11.01	87.00±10.59	73.00±10.59
10	74.70±1.76	75.10±1.39	71.7±10.07	80.18±1.09	80.26±0.82	**90.47±0.63**
11	84.18±1.01	**89.59±1.26**	45.25±10.65	52.96±1.92	54.89±2.83	87.23±0.34
12	57.84±3.14	31.13±12.23	63.82±8.74	76.66±4.25	73.10±4.66	**91.9±1.50**
13	89.72±1.38	**97.92±0.46**	94.55±3.38	95.76±0.83	96.67±1.08	97.5±0.88
14	95.69±0.56	**97.86±0.20**	83.65±6.39	91.29±1.78	90.85±1.20	96.85±0.37
15	50.18±2.82	44.21±2.56	44.69±2.64	57.90±3.66	56.68±1.89	**75.57±1.32**
16	90.64±2.69	94.26±1.44	93.09±4.52	97.45±2.42	96.81±2.07	**100±0.00**
AA (%)	72.09±0.85	59.04±1.04	74.71±1.86	80.07±1.11	80.38±0.76	**88.41±0.76**
OA (%)	78.80±0.32	77.22±0.97	64.31±2.73	72.35±0.68	72.52±0.53	**89.21±0.22**
F-measure (%)	75.03±0.70	61.74±1.52	67.13±1.87	75.49±1.06	75.52±0.91	**89.19±0.57**
Gmean (%)	69.13±1.63	0.00±0.00	71.86±2.14	78.55±1.17	78.89±0.68	**87.85±0.95**
Recall (%)	40.00±11.55	0.00±0.00	44.69±2.64	52.96±1.92	54.89±2.83	**73.00±10.59**
Time (s)	1.34	222.64	21.39	350.12	348.68	206.45

An example is used to illustrate the definition of $\alpha\%$ and of the yielded subsets. When the range of the sampling ratio $\alpha\%$ is set from 10% to 100%. For the building of the first tree, $10\% \cdot N_1$ samples will be selected from the original minority class dataset. SMOTE will generate the rest $(1 - 10\%) \cdot N_1$ minority class instances for the final balanced dataset. Then the $a\%$ will be updated to 20% for the second tree, 30% for the third tree, and 100% for the tenth tree. If we build T = 30 classifiers as ensemble members, every ten classifiers will be built with different resampling rates ranging from 10% to 100%.

TABLE III
CLASSIFICATION RESULTS (%) OF THE UNIVERSITY OF PAVIA ROSIS IMAGE, RESPECTIVELY OBTAINED BY RF, ROF, RUS-ROF, ROS-ROF, SMOTE-ROF, AND THE PROPOSED DSROF IN THE CASE OF THE IMBALANCE RATIO OF 19.83

IR:19.83	RF	RoF	RUS-RoF	ROS-RoF	SMOTE-RoF	DSRoF
1	90.63±0.75	93.63±0.39	72.99±5.10	68.20±0.84	69.31±1.71	**94.24±0.46**
2	95.66±0.44	**97.27±0.24**	76.36±5.34	74.90±3.08	77.75±2.33	95.77±0.29
3	57.29±2.74	22.66±4.04	75.91±5.15	78.07±4.06	**81.68±1.27**	76.77±0.90
4	85.66±1.11	78.13±0.93	94.96±2.89	**96.91±0.47**	96.46±0.24	93.82±0.50
5	98.47±0.43	99.39±0.10	99.16±0.59	99.01±0.45	99.53±0.15	**99.70±0.09**
6	56.75±1.83	63.51±3.71	86.46±4.32	**88.19±2.42**	87.79±2.24	87.04±0.79
7	71.48±2.11	0.00±0.00	89.84±3.16	89.89±0.64	**90.06±0.56**	82.14±1.15
8	81.67±0.88	**92.11±1.31**	75.63±6.61	81.47±2.57	77.95±1.24	87.02±0.51
9	99.13±0.20	99.80±0.18	**99.95±0.10**	99.82±0.18	99.89±0.19	99.93±0.09
AA (%)	81.86±0.38	71.83±0.56	85.69±0.82	86.27±0.34	86.71±0.51	**90.71±0.24**
OA (%)	85.91±0.30	84.36±0.45	79.93±1.89	79.50±1.11	80.73±1.05	**92.47±0.27**
F-measure (%)	83.78±0.32	74.76±0.39	81.62±0.92	81.60±0.45	82.25±0.62	**91.22±0.26**
Gmean (%)	80.20±0.54	0.00±0.00	84.97±0.91	85.57±0.41	86.10±0.57	**90.39±0.26**
Recall (%)	56.75±1.83	0.00±0.00	72.99±5.10	68.20±0.84	69.31±1.71	**76.77±0.90**
Time (s)	0.89	62.66	23.37	96.38	95.35	85.32

IV. EXPERIMENTAL RESULTS

A. Experiment Settings

To evaluate the performance of DSRoF, described in the previous section, RF, traditional RoF, data preprocessing involving RUS, ROS, and SMOTE combined rotation forest (RUS-RoF, ROS-RoF, and SMOTE-RoF) are utilized in the comparative analysis. Classification and Regression Trees (CART) are used as base classifiers in the experiments. All ensembles are implemented with 30 trees. Other parameters of RF and CART are kept to their default values in R-project packages, "randomForest" and "rpart" (https://cran.r-project.org/web/packages/randomForest/index.html, https://cran.r-project.org/web/packages/rpart/). For the proposed DSRoF and other RoF-based methods, the number of feature subsets G is set to 30. The range of sampling parameter $\alpha\%$ is set to [10%, 20%, 30%...100%]. All the presented results are averaged over ten independent runs of the algorithm.

B. Evaluation Methods

Average accuracy (AA) and overall accuracy (OA), which respectively represent the average of the accuracies for each class and the percentage of correctly predicted instances, are first employed as evaluation methods. F-measure, Gmean, and minimum Recall [2], which are the most popular to evaluate multi-class imbalance learning methods, are also used in the experiments. The computing times of all the algorithms are given. The significance of the differences in the classification performance between DSRoF and other models is assessed by McNemar's test (Z). McNemar's test is calculated by $Z = (f_{12} - f_{21})/\sqrt{f_{12} + f_{21}}$, where f_{12} is the number of samples correctly classified by classifier 1 and misclassified by classifier 2. Diversity [39], [40], considered as a very important concept in ensemble learning, is adopted for the analysis of each model in this study. It can be calculated by $Diversity = -\frac{1}{NT^2}\sum_{i=1}^{N} t(x_i)(T - t(x_i))$, where T is the ensemble size, $t(x_i)$ is the number of classifiers that correctly recognize sample x_i, and N represents the number of samples. For an intuitive comparison of all the algorithms, we use an ensemble margin-based categorization map method, in which the prediction labels of the unknown instances are shown only when those instances' unsupervised margin values are over one [36]. Generally, the instance with margin value of 1 means the data has high probability to be classified correctly [36].

C. Datasets

The DSRoF is evaluated on three standard hyperspectral images: *Indian Pines AVRIS*; *University of Pavia ROSIS*; and *Salinas*.

1) *Indian Pines AVRIS* is highly imbalanced and composed of 145 * 145 pixels, with a spatial resolution of 20 m/pixel and 200 spectral bands. The reference data with 16 classes are composed of 10 249 samples.
2) *University of Pavia ROSIS* consists of 610 * 340 pixels with a spatial resolution of 1.3 m/pixel, and 103 spectral bands. The reference data with nine classes are composed of 42 776 instances.
3) *Salinas* consists of 512 * 217 pixels with a spatial resolution of 3.7 m/pixel, and 224 spectral bands. The reference data with 16 classes are composed of 54 129 instances.

In order to objectively evaluate the performance of a classifier, the training set and the test set should be independent. Hence, the data sampling (without replacement) technology is used to divide the reference dataset into two nonoverlapping

TABLE IV
CLASSIFICATION RESULTS (%) OF THE SALINAS IMAGE, RESPECTIVELY OBTAINED BY RF, ROF, RUS-ROF, ROS-ROF, SMOTE-ROF, AND THE PROPOSED DSROF IN THE CASE OF THE IMBALANCE RATIO OF 12.51

IR:12.51	RF	RoF	RUS-RoF	ROS-RoF	SMOTE-RoF	DSRoF
1	99.46±0.12	99.41±0.23	98.91±0.88	99.70±0.06	99.69±0.09	**99.79±0.01**
2	99.58±0.17	99.78±0.14	99.37±0.68	99.85±0.08	99.79±0.13	**99.87±0.13**
3	94.37±1.05	94.55±2.42	94.03±4.02	97.31±1.54	98.36±0.62	**98.62±0.65**
4	99.38±0.31	99.63±0.23	99.42±0.43	99.55±0.45	99.70±0.01	**99.70±0.01**
5	96.90±0.21	96.39±0.93	96.28±1.84	96.77±0.34	97.39±0.28	**98.14±0.13**
6	99.70±0.02	99.71±0.04	99.39±1.08	99.81±0.03	99.79±0.02	**99.81±0.06**
7	99.06±0.18	96.85±0.26	99.10±1.03	99.49±0.08	99.58±0.09	**99.75±0.02**
8	81.98±0.65	89.33±0.66	69.93±9.20	79.61±3.55	80.33±2.08	**90.04±0.52**
9	98.92±0.15	98.84±0.11	98.94±0.20	99.04±0.16	99.12±0.11	**99.56±0.18**
10	88.89±0.50	90.51±0.35	89.34±2.13	90.94±0.56	91.36±0.47	**95.06±0.15**
11	91.15±1.61	90.95±0.63	93.17±1.53	93.56±0.41	93.59±0.39	**95.60±0.11**
12	98.62±0.27	99.22±0.09	99.54±0.35	99.09±0.23	99.21±0.09	**99.82±0.11**
13	95.52±1.13	96.67±0.28	96.49±1.67	96.68±0.20	**96.76±0.23**	95.98±0.11
14	96.42±0.47	95.99±0.34	95.47±1.05	95.71±0.21	95.78±0.12	**97.97±0.25**
15	60.97±1.11	51.89±1.42	**66.90±5.78**	62.15±3.81	62.44±2.04	62.82±0.76
16	97.28±0.37	97.38±0.51	98.14±0.78	97.97±0.38	98.81±0.13	**99.28±0.18**
AA (%)	93.64±0.15	93.57±0.23	93.40±0.37	94.20±0.15	94.48±0.10	**95.74±0.05**
OA (%)	89.24±0.15	89.53±0.20	87.56±1.26	89.31±0.42	89.64±0.22	**92.16±0.06**
F-measure (%)	93.44±0.11	93.05±0.24	92.50±0.44	93.54±0.15	93.87±0.14	**95.70±0.06**
Gmean	93.03±0.17	92.63±0.26	92.71±0.48	93.58±0.23	93.88±0.10	**95.22±0.03**
Recall	60.97±1.11	51.89±1.42	66.90±5.78	62.15±3.81	62.44±2.04	**62.82±0.76**
Time (s)	1.9	291.23	144.07	402.02	397.15	256.35

TABLE V
STATISTICAL SIGNIFICANCE OF DIFFERENCE AMONG THE CLASSIFIERS

	Indian Pines AVRIS	University of Pavia ROSIS	Salinas
DSRoF vs.RF	17.33	31.55	20.58
DSRoF vs.RoF	24.92	42.58	25.68
DSRoF vs.RUS-RoF	37.48	51.42	25.00
DSRoF vs.ROS-RoF	30.97	56.97	20.07
DSRoF vs.SMOTE-RoF	31.40	66.68	23.01

parts: training set and test set. In addition, to better investigate the effectiveness of the proposed DSRoF for the classification of the imbalanced hyperspectral data, the data with a high imbalance ratio (IR) is considered in our experiment. For data *Indian Pines AVRIS*, 30% of original reference data are sampled randomly without replacement to construct training sets. We note that for the smallest classes *Grass-pasture-mowed* and *Oats* of the data *Indian Pines AVRIS*, only half of their instances are selected randomly to construct training sets. For data *University of Pavia ROSIS* and *Salinas*, 5% of their instances are selected randomly from original reference data to construct training sets. All the unselected instances compose corresponding test sets. More details about the data information could be found in the Table I. The IR is calculated by dividing the number of the smallest class by the number of samples of the largest class. The IR of the training sets of the three datasets are 73.6, 19.83, and 12.51, respectively.

D. Results and Analysis

Tables II, III, and IV present the AA, OA, F-measure, Gmean, and minimum Recall of the RF, traditional RoF, three data sampling combined RoF methods, and the proposed DSRoF, on the hyperspectral images *Indian Pines AVRIS*, *University of Pavia ROSIS* and *Salinas,* respectively. The best results for each data are highlighted in bold font. The experimental results in those tables show that traditional RoF biases of the classification of

TABLE VI
ENSEMBLE DIVERSITIES OF THE CLASSIFIERS

	Indian Pines AVRIS	University of Pavia ROSIS	Salinas
RF	0.1499	0.1109	0.0752
RoF	0.1094	0.0683	0.0610
RUS-RoF	0.1814	0.1164	0.0721
ROS-RoF	0.1403	0.0942	0.0597
SMOTE-RoF	0.1393	0.0956	0.0574
DSRoF	0.1546	0.1110	0.0753

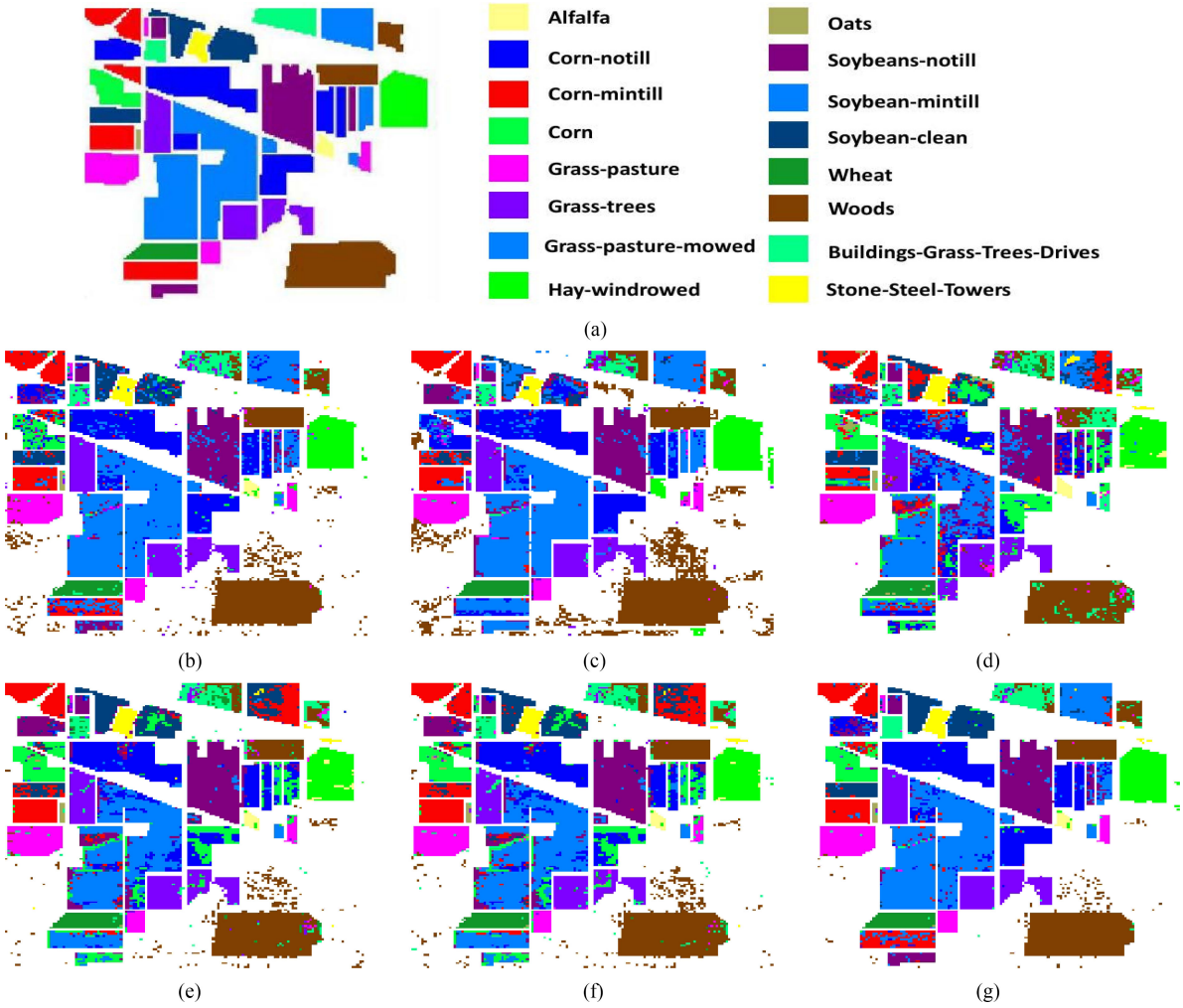

Fig. 1. Ground truth (GT) and classification maps of random forest (RF), rotation forest (RoF), random under-sampling (RUS) combined, random oversampling (ROS) combined, SMOTE combined rotation forest (RUS-RoF, ROS-RoF and SMOTE-RoF), as well as the proposed dynamic SMOTE-based rotation forest method DSRoF on the hyperspectral data *Indian Pines AVRIS*. (a) GT. (b) RF. (c) RoF. (d) RUS-RoF. (e) ROS-RoF. (f) SMOTE-RoF. (g) DSRoF.

most major classes and RF presents better performance in the imbalance case. In addition, all the imbalance learning algorithms lead to an improved classification of the smallest class.

Oversampling combined methods obviously outperform undersampling-based ensemble classifiers. The major drawback of undersampling is that it can discard potentially useful data, which could be important for the induction process. SMOTE-based schemes obtain better results when compared with ROS-based RoF. Although the SMOTE technique could produce additional noise during the data balance process, the RoF has good noise robustness and moderate noise could even improve the performance of RoF. The proposed DSRoF always achieves the best performance. It can increase not only the average accuracy but also the overall accuracy especially in dealing with the data of higher imbalance ratio (Table II). On the dataset *Indian Pines AVRIS*, with respect to RF, RoF, RUS-RoF,

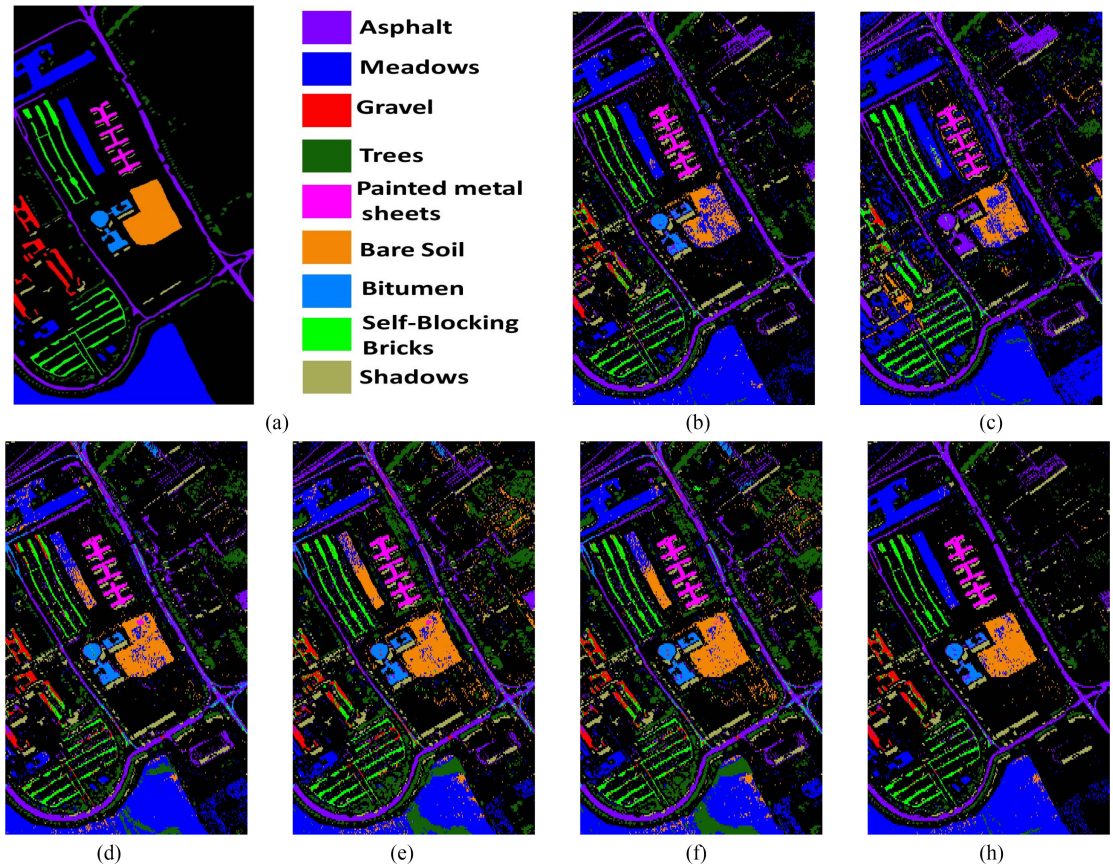

Fig. 2. Ground truth (GT) and classification maps of random forest (RF), rotation forest (RoF), random undersampling combined (RUS), random oversampling (ROS) combined, SMOTE combined rotation forest (RUS-RoF, ROS-RoF and SMOTE-RoF), as well as the proposed dynamic SMOTE-based rotation forest method DSRoF on the hyperspectral data *University of Pavia ROSIS*. (a) GT. (b) RF. (c) RoF. (d) RUS-RoF. (e) ROS-RoF. (f) SMOTE-RoF. (g) DSRoF.

ROS-RoF, and SMOTE-RoF, the best increases in AA are over 16%, 29%, 14%, 8%, and 8%, respectively, and the best increases in OA are over 10%, 12% 25%, 16%, and 16%, respectively. Moreover, the results of F-measure and Gmean also demonstrate the better performance of the proposed method for the classification of the hyperspectral images when compared with other methods.

Table V shows the results of McNemars test for statistical comparisons of the classification performance of the proposed DSRoF to that of RF, RoF, RUS-RoF, ROS-RoF, and SMOTE-RoF, respectively on the three datasets. The numbers in the table are the Z values. $Z > 0$ means that classifier 1 is more accurate than classifier 2. Moreover, the difference between the two classifiers is to be statistically significant if $|Z|$ is greater than 1.96. Table V shows that all of the Z values are over 1.96. Therefore, according to McNemars test, DSRoF significantly outperforms RF, RoF, and other data sampling combined RoF methods.

Diversity is an essential definition of ensemble learning. The diversities of all the methods are shown in Table VI. This table shows that the traditional RoF statistically obtains the lowest diversity values. Although the RUS method could lead to higher diversity than other methods, it does not benefit the overall accuracy increase. The proposed DSRoF method has the best results among the three oversampling combined methods. Hence, the dynamic internal imbalance sampling-based ensemble approach works better than the methods consisting of first preprocessing data and then using a standard ensemble on balanced data.

For a more intuitive comparison of the proposed method and other reference methods, the categorization map is adopted. Figs. 1–3 exhibit the classification maps obtained by different classification methods for *Indian Pines AVRIS*, *University of Pavia ROSIS,* and *Salinas* images. DSRoF could result in more accurate cartography with respect to RF, traditional RoF, and other improved RoF methods.

In order to study the influence of subspace size on RoF construction, we present in Figs. 4 and 5 the evaluation of the average accuracy and overall accuracy for all the RoF methods on the three datasets. In this experiment, the size T of the ensemble is still set to 30 and the tested number of subspace G is set as (5, 10, 20, 30, 40, 50). Fig. 4 shows that almost all the improved methods have better results with respect to the traditional RoF. The best increase in average accuracy is obtained by the proposed method. Fig. 5 exhibits that RoF results in higher overall accuracy than UnderRoF, OverRoF, and SMOTERoF. But the best increase results are still obtained by DSRoF. Moreover, the proposed method DSRoF is not very sensitive to the subspace size.

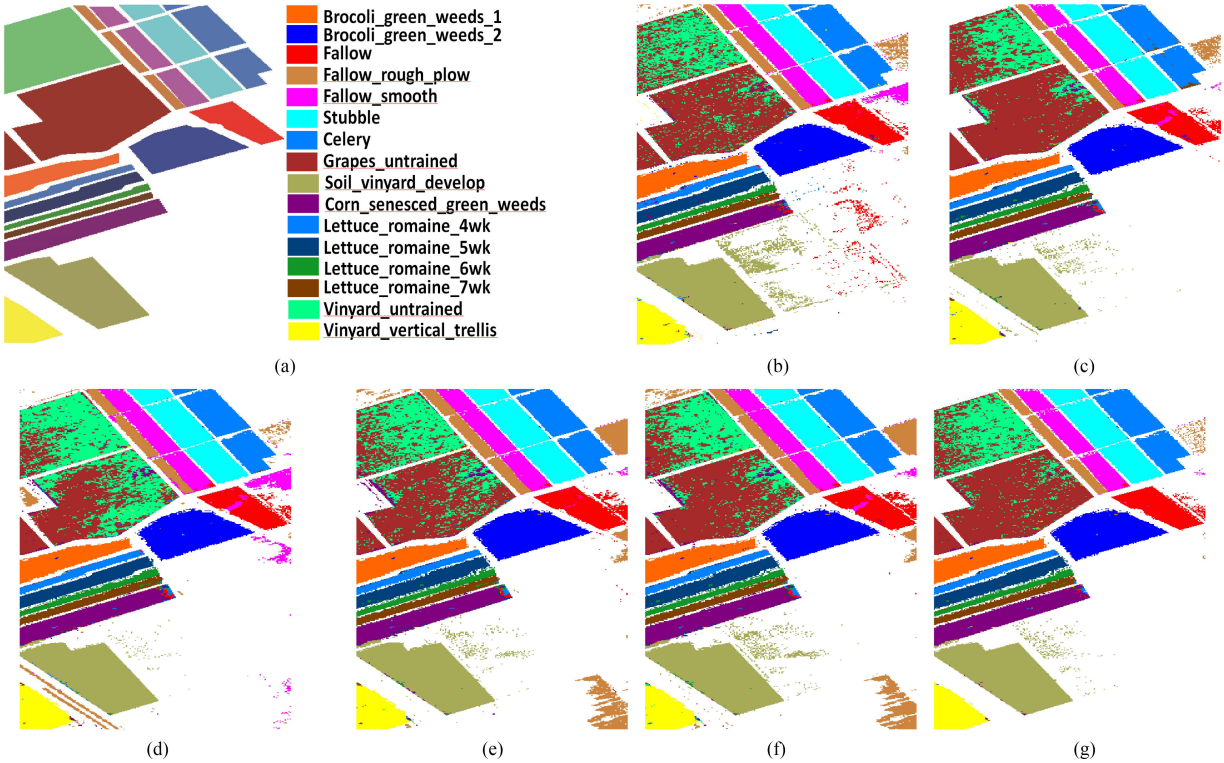

Fig. 3. Ground truth (GT) and classification maps of random forest (RF), rotation forest (RoF), random undersampling(RUS) combined, random oversampling (RS) combined, SMOTE combined rotation forest (RUS-RoF, ROS-RoF and SMOTE-RoF), as well as the proposed dynamic SMOTE-based rotation forest method DSRoF on the hyperspectral data *Salinas*. (a) GT. (b) RF. (c) RoF. (d) RUS-RoF. (e) ROS-RoF. (f) SMOTE-RoF. (g) DSRoF.

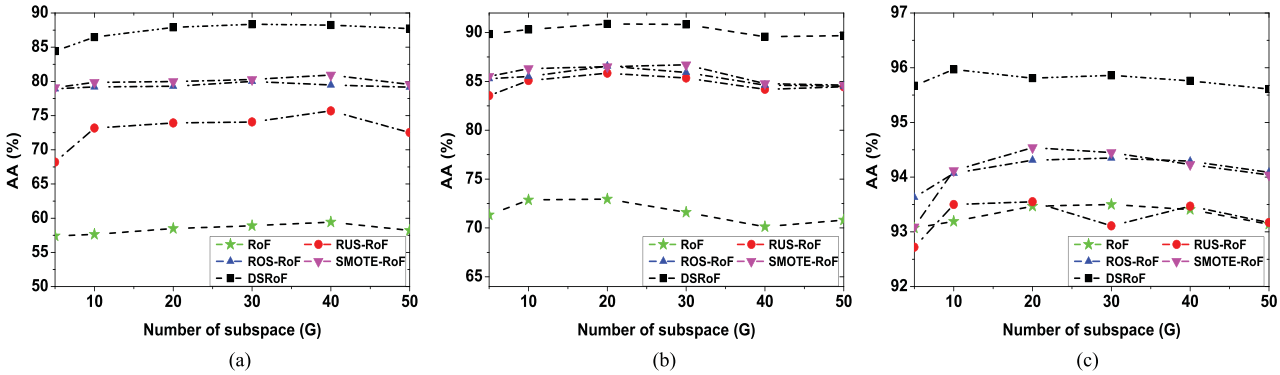

Fig. 4. Evolution of the average accuracy according to the subspace size (*G*). (a) *Indian Pines AVRIS*. (b) *University of Pavia ROSIS*. (c) *Salinas*.

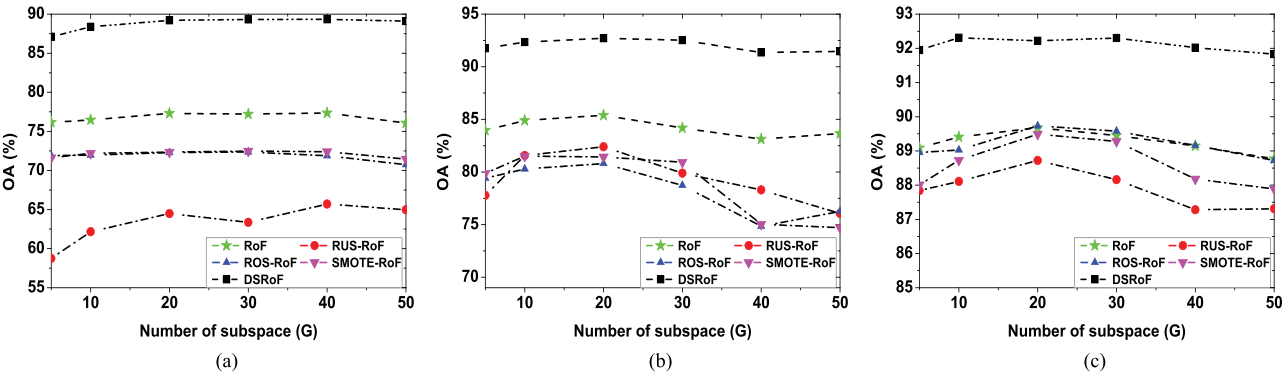

Fig. 5. Evolution of the overall accuracy according to the subspace size (*G*). (a) *Indian Pines AVRIS*. (b) *University of Pavia ROSIS*. (c) *Salinas*.

V. Conclusion

In this paper, we have proposed a novel dynamic SMOTE-based rotation forest (DSRoF) method for the imbalance learning of the multi-class hyperspectral data. DSRoF builds an ensemble of rotation decision trees by training each of them on the different balanced dataset generated dynamically by SMOTE and a resampling ratio. The proposed method is compared with other related algorithms (RF, RoF, RUS-RoF, ROS-RoF, and SMOTE-RoF) on three public hyperspectral image datasets. Several measures are used to evaluate the performance of the proposed method. Experimental results show that the proposed method can provide a competitive solution for imbalanced hyperspectral image classification. Moreover, the parameter analysis has also been operated, and the result shows that the proposed method has better robustness with respect to the reference methods.

References

[1] P. Mather and B. Tso, Eds., *Classification Methods for Remotely Sensed Data*, 2nd ed. Boca Raton, FL, USA: CRC Press, 2016.

[2] W. Feng, W. Huang, and J. Ren, "Class imbalance ensemble learning based on the margin theory," *Appl. Sci.*, vol. 8, no. 5, 2018.

[3] S. García, Z. Zhang, A. Altalhi, S. Alshomrani, and F. Herrera, "Dynamic ensemble selection for multi-class imbalanced datasets," *Inf. Sci.*, vol. 445–446, pp. 22–37, 2018.

[4] T. Sun, L. Jiao, J. Feng, F. Liu, and X. Zhang, "Imbalanced hyperspectral image classification based on maximum margin," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 522–526, Mar. 2015.

[5] S. Wang and X. Yao, "Diversity analysis on imbalanced data sets by using ensemble models," in *Proc. IEEE Symp. Comput. Intell. Data Mining.*, Nashville, TN, USA, Mar. 2009, pp. 324–331.

[6] B. Krawczyk, "Learning from imbalanced data: Open challenges and future directions," *Progress Artif. Intell.*, vol. 5, no. 4, pp. 221–232, 2016.

[7] J. A. Sáez, B. Krawczyk, and M. Woźniak, "Analyzing the oversampling of different classes and types of examples in multi-class imbalanced datasets," *Pattern Recognit.*, vol. 57, pp. 164–178, 2016.

[8] J. Bi and C. Zhang, "An empirical comparison on state-of-the-art multi-class imbalance learning algorithms and a new diversified ensemble learning scheme," *Knowl.-Based Syst.*, vol. 158, pp. 81–93, 2018.

[9] H. B. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009.

[10] D. Shuya *et al.*, "Kernel based online learning for imbalance multiclass classification," *Neurocomputing*, vol. 277, pp. 139–148, 2018.

[11] S. Ertekin, J. Huang, L. Bottou, and C. L. Giles, "Learning on the border: Active learning in imbalanced data classification." in *Proc. 16th ACM Conf. Inf. Knowl. Manage.*, Lisbon, Portugal, 2007, pp. 127–136.

[12] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: Bagging-, boosting, and hybrid-based approaches," *IEEE Trans. Syst. Man, Cybern. C, Appl. Rev.*, vol. 42, no. 4, pp. 463–484, Jul. 2012.

[13] W. Feng and W. Bao, "Weight-based rotation forest for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 2167–2171, Nov. 2017.

[14] A. Roy, R. M. Cruz, R. Sabourin, and G. D. Cavalcanti, "A study on combining dynamic selection and data preprocessing for imbalance learning," *Neurocomputing*, vol. 286, pp. 179–192, 2018.

[15] G. Douzas, F. Bacao, and F. Last, "Improving imbalanced learning through a heuristic oversampling method based on k-means and smote," *Inf. Sci.*, vol. 465, pp. 1–20, 2018.

[16] W. Feng and S. Boukir, "Class noise removal and correction for image classification using ensemble margin," in *Proc. IEEE Int. Conf. Image Process.*, Quebec City, QC, Canada, Sep. 2015, pp. 4698–4702.

[17] J. Xia, N. Falco, J. A. Benediktsson, P. Du, and J. Chanussot, "Hyperspectral image classification with rotation random forest via kpca," *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 4, pp. 1601–1609, Apr. 2017.

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

[19] R. E. Schapire and Y. Singer, "Improved boosting algorithms using confidence-rated predictions," *Mach. Learn.*, vol. 37, no. 3, pp. 297–336, 1999.

[20] N. V. Chawla, A. Lazarevic, L. O. Hall, and K. W. Bowyer, "Smoteboost: Improving prediction of the minority class in boosting," in *Knowledge Discovery in Databases: PKDD 2003*, ser. Lecture Notes in Computer Science. Berlin Heidelberg, Germany: Springer, 2003, vol. 2838, pp. 107–119.

[21] C. Seiffert, T. M. Khoshgoftaar, J. V. Hulse, and A. Napolitano, "Rusboost: A hybrid approach to alleviating class imbalance," *IEEE Trans. Syst. Man Cybern. A, Syst. Humans*, vol. 40, no. 1, pp. 185–197, Jan. 2010.

[22] M. Galar, A. Fernández, E. Barrenechea, and F. Herrera, "Eusboost: Enhancing ensembles for highly imbalanced data-sets by evolutionary undersampling," *Pattern Recognit.*, vol. 46, no. 12, pp. 3460–3471, 2013.

[23] Y. Sun, M. S. Kamel, A. K. Wong, and Y. Wang, "Cost-sensitive boosting for classification of imbalanced data," *Pattern Recognit.*, vol. 40, no. 12, pp. 3358–3378, 2007.

[24] J. Díez-Pastor, J. Rodríguez, C. García-Osorio, and L. I. Kuncheva, "Random balance: Ensembles of variable priors classifiers for imbalanced data," *Knowl.-Based Syst.*, vol. 85, pp. 96–111, 2015.

[25] J. Sun, J. Lang, H. Fujita, and H. Li, "Imbalanced enterprise credit evaluation with dte-sbd: Decision tree ensemble based on smote and bagging with differentiated sampling rates," *Inf. Sci.*, vol. 425, pp. 76–91, 2018.

[26] R. Barandela, J. S. Sánchez, and R. M. Valdovinos, "New applications of ensembles of classifiers," *Pattern Anal. Appl.*, vol. 6, no. 3, pp. 245–256, 2003.

[27] C. Chen, A. Liaw, and L. Breiman, "Using random forest to learn imbalanced data," Dept. Statist., Univ. California, Berkeley, Tech. Rep. 666, 2004.

[28] 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.

[29] P. Du, A. Samat, B. Waske, S. Liu, and Z. Li, "Random forest and rotation forest for fully polarized sar image classification using polarimetric and spatial features," *ISPRS J. Photogramm. Remote Sens.*, vol. 105, pp. 38–53, 2015.

[30] F. Li, L. Xu, P. Siva, A. Wong, and D. A. Clausi, "Hyperspectral image classification with limited labeled training samples using enhanced ensemble learning and conditional random fields," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2427–2438, Jun. 2015.

[31] J. Xia, N. Falco, J. A. Benediktsson, J. Chanussot, and P. Du, "Class-separation-based rotation forest for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 4, pp. 584–588, Apr. 2016.

[32] H. Lu, L.Yang, K. Yan, Y. Xue, and Z. Gao, "A cost-sensitive rotation forest algorithm for gene expression data classification," *Neurocomputing*, vol. 228, pp. 270–276, 2017.

[33] J. Blaszczyński and J. Stefanowski, "Neighbourhood sampling in bagging for imbalanced data," *Neurocomputing*, vol. 150, no. Part B, pp. 529–542, 2015.

[34] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *J. Artif. Int. Res.*, vol. 16, no. 1, pp. 321–357, 2002.

[35] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *SIGKDD Explor. Newslett.*, vol. 6, no. 1, pp. 20–29, 2004.

[36] W. Feng, "Investigation of training data issues in ensemble classification based on margin concept. application to land cover mapping," Ph.D. dissertation, Univ. Bordeaux 3, France, 2017.

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

[38] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. New York, NY, USA: Springer-Verlag, 2002.

[39] R. Kohavi and D. Wolpert, "Bias plus variance decomposition for zero-one loss functions," in *Proc. 13th Int. Conf. Mach. Learn.*, Bari, Italy, 1996, pp. 275–283.

[40] M. Kapp, R. Sabourin, and P. Maupin, "An empirical study on diversity measures and margin theory for ensembles of classifiers," in *Proc. 10th Int. Conf. Inf. Fusion*, Quebec, Canada, Jul. 2007, pp. 1–8.

Wei Feng received the B.S. degree in computer science and technology from Northeast Agricultural University, Harbin, China, in 2009, the M.Sc. degree in computer applications technology from North Minzu University, Yinchuan, China, in 2013, and the Ph.D. degree in information science and technology from the Université Michel de Montaigne-Bordeaux 3, Bordeaux, France, in 2017.

She is a Postdoctoral Researcher with the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China. Her current research interests include the remote sensing, machine learning and image processing.

Gabriel Dauphin received the engineer's degree from Mines ParisTech, France, in 1996, and the Ph.D. degree in signal and image processing from Télécom Paris Tech University, France, in 2001.

Since 2002, he has been an Associate Professor at University Paris 13 in the Laboratory of Information Processing and Transmission (L2TI), Villetaneuse, France. His research interests include gate recognition, stereoscopic image compression, machine learning, and remote sensing.

Wenjiang Huang received the Ph.D. degree in physical geography from Beijing Normal University, Beijing, China, in 2005.

He is currently a Professor with the Key Laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China. His research interests include quantitative and hyperspectral remote sensing for vegetation, especially on crops, data fusion (multi-scale, multi-sensor, multi-temporal) for agricultural applications, and monitoring crop diseases using remote sensing technology.

Yinghui Quan received the B.S. and Ph.D. degrees in electrical engineering from Xidian University, Xi'an, China, in 2004 and 2012, respectively.

He is currently a Full Professor with the National Laboratory of Radar Signal Processing, Xidian University. His research interests include radar imaging, radar signal processing, and radar microsystem.

Wenxing Bao received the B.Eng. degree in industrial automation from Xidian University, Xi'an, China, in 1993, and the M.Sc. degree in electrical engineering and the Ph.D. degree in electronic science and technology from Xi'an Jiaotong University, Xi'an, China, in 2001 and 2006, respectively.

He is currently a Professor and the Vice President of North Minzu University, Yinchuan, China.

His current research interests include digital image processing and remote sensing image classification and fusion.

Mingquan Wu received the B.Sc. degree in surveying and mapping engineering from Chongqing Jiaotong University, Chongqing, China, and the M.S. degree from the Taiyuan University of Technology, Taiyuan, China, and the PhD. degree from the University of Chinese Academy of Sciences, Beijing, China, in 2013, both in cartography and geographic information system.

Since 2016, he has been an Associate Researcher of remote sensing with the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences. Since 2017, he has been a member of the Youth Innovation Promotion Association, Chinese Academy of Sciences. His research interests include remote sensing monitoring of infrastructure construction and agricultural remote sensing monitoring.

Qiang Li , received the B.S. degree in automation, and the M.S. and Ph. D degrees in physics from the Harbin Institute of Technology, Harbin, China, in 2011, 2013, and 2017, respectively. He has also studied with AMS collaboration in CERN (2013–2015).

Since August 2017, he has been a Postdoctoral Researcher with the Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing, China. His current research interests include the particle physics, statistical physics, and machine learning.