This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Imbalanced Hyperspectral Image Classification With an Adaptive Ensemble Method Based on SMOTE and Rotation Forest With Differentiated Sampling Rates

Wei Feng, Wenjiang Huang, and Wenxing Bao

***Abstract*— Rotation forest (RoF) is a powerful ensemble classifier and has been demonstrated the outstanding performance in hyperspectral data classification. However, the classification task suffers from the class imbalanced problem which has been considered to be one of the most important challenges. The traditional construction method of RoF biases classifying the majority classes and ignores recognizing the minority classes samples. This letter proposes a novel adaptive ensemble method based on SMOTE and RoF with differentiated sampling rates (AdaSRoF) for the multiclass imbalance problem. The proposed method adaptively generates several balanced data sets with more diversity and less noise by using SMOTE and a dynamic data sampling ratio for base classifiers. The obtained results on two publicly available hyperspectral images show that the proposed method can get more diversity and better performance than support vector machine (SVM), random forest (RF), and RoF in multiclass imbalance learning.**

***Index Terms*— Classification, ensemble learning, hyperspectral image, multiclass imbalance learning, SMOTE.**

I. Introduction

HYPERSPECTRAL image classification has been a vibrant area of research in recent years [1]. However, the classification task suffers from the class imbalanced problem, which has been considered to be one of the most important challenges [2], [3]. The class imbalance occurs when the number of instances belonging to one of the classes is much larger than the numbers in other class [1], [3]. Classification of imbalanced data is difficult because most of the canonical classifiers are driven by overall accuracy (OA), hence fail in the identification of the minority class, besides generally all classifiers present some performance loss when the data are imbalanced [4], [5]. In imbalance learning, there are two scenarios, binary data and multiclass data. Some difficulties, such as the uneven distribution of examples among classes, the small sample size, and the class overlapping, make multiclass imbalance learning tasks much harder than binary ones [5]–[8].

The ideal objective of multiclass imbalance learning can be described as how to design a classifier that could result in high accuracies for the minority classes without sacrificing the accuracies of the majority classes [9]. Many methods have been devoted to imbalanced learning problems [9], [10]. However, most of the reported solutions focus on the binary data but not applicable to multiclass cases. The resampling approaches are the popular imbalance learning methods [11], [12]. Since these methods balance the data distribution in the preprocessing step and do not need modifying the learning model, they are easy to be developed for multiclass task [9]. Among these approaches, random undersampling (RUS) and random oversampling (ROS) are two of the most popular resampling methods [2]. However, RUS discards examples from the majority classes randomly, thus could miss important information. ROS could avoid information loss but has a risk of overfitting. Synthetic minority oversampling technique SMOTE [13] is a very famous oversampling method. It could avoid overfitting, but the disadvantage of this method is to produce additional noise [7]. Consequently, how to properly implement SMOTE to balance the complex multiclass data set is a great challenge.

Ensemble learning has got a successful application to classify hyperspectral images [14]–[16]. Rotation forest (RoF), a powerful ensemble classifier, has attracted lots of attentions [14]–[18]. It obtains outstanding performance by combining several randomization techniques such as data bootstrap, random feature selection, and feature space rotation. Some improved versions of the RoF have been proposed and verified better performances [18], [19]. However, most of these methods are implemented depending on a balanced training set. At present, there is less investigation of RoF for the multiclass imbalanced problem, especially in hyperspectral

Manuscript received November 12, 2018; revised December 18, 2018; accepted April 5, 2019. This work was supported in part by the Chinese Academy of Sciences through the Strategic Priority Research Program under Grant XDA19080304, in part by the Hainan Provincial Key Research and Development Program of China under Grant ZDYF2018073, and in part by the National Natural Science Foundation of China under Grant 41601466, Grant 41871339, Grant 61461003, and Grant 41701411. *(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, Xi'an 710071, China (e-mail: fengwei@radi.ac.cn; fw.enp@qq.com).

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).

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

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

Digital Object Identifier 10.1109/LGRS.2019.2913387

study. In addition, the traditional construction method of RoF biases classifying the majority classes and ignores recognizing the minority classes samples [5]. Therefore, some special developments are necessary for the RoF ensemble to hand such a difficulty problem. Diversity is essential in order to build an accurate ensemble of classifiers in imbalance learning [5]. Hence, how to properly increase the diversity of RoF to make the method more suited to the imbalanced data classification is an interesting research direction.

The major contribution of this letter is to propose a novel adaptive ensemble method based on SMOTE and RoF with differentiated sampling rates (AdaSRoF) for the multiclass imbalance problem of hyperspectral image classification. The proposed method adaptively generates several balanced data sets with more diversity and less noise by using SMOTE and a dynamic data sampling ratio for base classifiers. The remaining part of this letter is organized as follows. Section II describes in detail the proposed methodology. Section III shows the results and discussion. The conclusions are given in Section IV.

II. Method

A. Differentiated Sampling Rates

The proposed method adopts the differentiated sampling rates for different base classifiers. The instances of all minority classes are sampled randomly with different sampling rates $\beta\%$ before implementing a SMOTE. The differentiated sampling strategy is useful for the diversity increase in the ensemble system and then improves the system's performance. The following example is used to illustrate the definition of $\beta\%$ and of the yielded training set. When the range of the sampling ratio $\beta\%$ is set from 10% to 100% and N_1 is assumed as the number of training instances of the largest class, for the first classifier, $\beta\%$ equals 10%, then $\beta\% \cdot N_1$ instances will be sampled with replacement from the original minority class data set. To match the size of the largest class, $(1-\beta\%) \cdot N_1$ minority class instances will be generated by SMOTE and combined with the previous sampled instances to construct the first balanced data set. Then, the $\beta\%$ will be updated to 20% to yield a diversity training set for the second base classifier. When $\beta\%$ equals to 100%, it is completely ROS. If $T = 30$ base classifiers are built, then every ten classifiers will have different $\beta\%$ values, which range from 10% to 100% [9].

B. Adaptive Weight Function

The proposed method uses an adaptive weight function to mitigate SMOTE's risk of producing artificial noise. The adaptive weight function is introduced in (1). Let us denote a training set as $S = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$, where the sample size is N and the class label of the instance x_i is y_i. The significance of a training sample $(\mathbf{x}_i, y_i)$ is assessed by

$$\begin{aligned} W(x_i) &= 1 - \frac{1}{\sum_{c=1}^{L}(\upsilon_c)} |\upsilon(\mathbf{x}_i, y_i) - \max_{c \neq y_i} \upsilon(\mathbf{x}_i, c)| \\ &= 1 - \frac{1}{T} |\upsilon(\mathbf{x}_i, y_i) - \max_{c \neq y_i} \upsilon(\mathbf{x}_i, c)| \end{aligned} \tag{1}$$

where $\upsilon(\mathbf{x}_i, y_i)$ is the vote number of the true class y and $\upsilon(\mathbf{x}_i, c)$ is the vote number of any other class c. The high weight $W(x)$ of a sample means that this sample is closed to the classification decision boundary in the current forest and will be selected with high probability to produce the synthetic instances for minority classes by combining SMOTE for the next classifier. Hence, with the growth of the ensemble size, the weights of the boundary instances are increased and higher than those of other types of instances.

Algorithm 1 Adaptive Ensemble Method Based on SMOTE and RoF With Differentiated Sampling Rates (AdaSRoF)

Training phase
Input: $S = [X, Y] = (x_1, y_1), (x_2, y_2,), \cdots, (x_N, y_N)$: training set; $\mathbb{F}$: feature set; L: number of classes; N_i: the sample size of ith class; $\beta\%$: resampling rate; K: number of feature subsets; ζ: base classifier; T: ensemble size; $E = \varnothing$: an ensemble.
Process:
Initialize the weight distribution for all $x_i \in S$ as $W_1(x_i) = 1/N$, and sort the classes c of the imbalanced data S in descending order according to their sample sizes.
for t=1:T **do**
  Compose a set $S_{1,t}$ by sampling N_1 instances (with replacement) from the largest class 1 according to the instance weights $W_t(x_i)$
  for c=2:L **do**
    Obtain $S_{c,t}$ by resampling (with replacement) $\beta\% \cdot N_1$ instances according to the adaptive weight function W_t from the original class c
    Product $S'_{c,t}$ with the sample size of $(1-\beta\%) \cdot N_1$ by SMOTE algorithm
  end for
  Construct a new balanced data set S_t by combining the sampled data set $S_{c,t}$ $(c = 1, ..., L)$ with the artificially generated data set $S'_{c,t}$ $(c = 2, ..., L)$.
  Carry out a random assignment on the feature set $\mathbb{F}$ of S_t for K disjoint subsets $\mathbb{F}_{t,k}$
  for k=1:K **do**
    Compose the data $S_{t,k}$ for the features in $\mathbb{F}_{t,k}$ frame and apply PCA on $S_{t,k}$ to get the coefficients $c_{t,k}$
  end for
  Construct a rotation matrix M'_t by rearranging the columns of the matrix M_t composed of $c_{t,k}$ to match the order of original features $\mathbb{F}$, then compose the diversity training set $S'_t = [S_t \cdot M'_t, Y_t]$
  Train a classifier ζ_t on S'_t
  $E \leftarrow E \cup \zeta_t$
  Change the sampling ratio $\beta\%$.
  for all $x_i \in S$ **do**
    Update $W_t(x_i)$ according to the equation 1
    $W_{t+1}(x_i) \leftarrow W_t(x_i)$
  end for
end for
Output: The ensemble E

C. Adaptive Ensemble Method Based on SMOTE and Rotation Forest With Differentiated Sampling Rates

The process of the proposed adaptive ensemble method based on SMOTE and RoF with differentiated sampling rates (AdaSRoF) is detailed in Algorithm 1. AdaSRoF is an

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE I

DESCRIPTION OF THE EXPERIMENTAL DATA SETS

No.	Pavia University (IR:19.96)			Indian Pines						
				Class name	Case 1 (IR:24.50)		Case 2 (IR:49.10)		Case 3 (IR:73.60)	
	Class name	Tra.	Test		Tra.	Test	Tra.	Test	Tra.	Test
1	Asphalt	198	6433	Alfalfa	23	23	23	23	23	23
2	Meadows	559	18090	Corn-notill	142	1286	285	1143	428	1000
3	Gravel	62	2037	Corn-mintill	83	747	166	664	249	581
4	Trees	91	2973	Corn	23	214	47	190	71	166
5	Painted metal sheets	40	1305	Grass-pasture	48	435	96	387	144	339
6	Bare Soil	150	4879	Grass-trees	73	657	146	584	219	511
7	Bitumen	39	1291	Grass-pasture-mowed	14	14	14	14	14	14
8	Self-Blocking Bricks	110	3572	Hay-windrowed	47	431	95	383	143	335
9	Shadows	28	919	Oats	10	10	10	10	10	10
10				Soybean-notill	97	875	194	778	291	681
11				Soybean-mintill	245	2210	491	1964	736	1719
12				Soybean-clean	59	534	118	475	177	416
13				Wheat	20	185	41	164	61	144
14				Woods	126	1139	253	1012	379	886
15				Bldg-Grass-Tree-Drives	38	348	77	309	115	271
16				Stone-Steel-Towers	46	47	46	47	46	47
Total		1277	41499		1094	9155	2102	8147	3106	7143

internal oversampling-based ensemble method, which reduces the imbalance ratio (IR) in each iteration and generates artificial instances during each subset construction. Suppose there are L classes in the data set S and the sample size of the ith class is N_i. The classes of S are sorted in the descending order according to their sample sizes N_i. Then, N_L is assumed as the number of the smallest class L. The weight value W for each instance of S is initialized as $1/N$. In each iteration t, the first set is produced by performing a sampling (with replacement) according to W on each class from the original data set, and the second set is generated by the SMOTE algorithm. During the data construction, for the class c $(2 \leqslant c \leqslant L)$, a resampling rate $\beta\%$ is adopted to define the sample sizes of both sampled instances and artificial instances as we aforementioned in Section II-A. A diversity balanced data set S_t is produced by combining all the sampled instances and the artificially generated instances. In the next phase, the feature space rotation is performed on the balanced data set S_t, as in the traditional RoF method, to obtain the final training data set S_t'. Then, a classifier ζ_t is trained on the data S_t'. The weight value W of each training instance is then recalculated and the resampling rate $\beta\%$ is also updated. These processes are repeated until the maximum iteration number is reached. The results output from a series of classifiers are finally fused by a majority vote method.

III. EXPERIMENTAL STUDY

A. Experiment Settings

In order to demonstrate the advantages of the proposed AdaSRoF, three popular methods, support vector machine (SVM), random forest (RF) [20], and traditional RoF [17], are employed as comparisons. This experiment utilizes the classification and regression tree (CART) as the baseline learner and all ensembles consist of 30 decision tree classifiers. The parameter K is set to 30 for both the traditional RoF and the proposed method. The range of sampling parameter a is set from 10% to 100%. All the presented results are averaged over ten independent runs of the algorithm.

B. Evaluation Methods

The experiments are first carried out using five different comparative measures including the average accuracy (AA), OA, F-measure, and Gmean and minimum Recall. The ensemble diversity, presented in [9], and an important nonparametric pair-wise test named McNemar's test [21] are also used to compare the performances of AdaSRoF and other models in terms of the diversity improvement and the statistical significance. The computing times of all the algorithms are given.

C. Data Sets

To assess the effectiveness of the proposed AdaSRoF, the experiments are performed on two publicly available hyperspectral images. The first image is Pavia University, whose spatial dimensions is $610 * 340$ pixels with 103 spectral channels and a spatial resolution of 1.3 m per pixel. Its ground reference data consist of nine classes. The second image is Indian Pines, which is with the size of $145 * 145$ pixels, 200 spectral channels, and a spatial resolution of 20 m per pixel. Its ground reference data contain 16 mutually exclusive classes.

For a better investigation of the behavior of the proposed AdaSRoF for the multiclass imbalanced hyperspectral image classification, the data sets of different IRs are utilized in the experimental study. For data Pavia University, 3% of the reference data are selected randomly to construct the training set and the last 97% instances construct the test set. For data Indian Pines, 10% (case 1), 20% (case 2), and 30% (case 3) of original reference data are sampled randomly without replacement to construct training sets, respectively. All the unselected instances compose the corresponding test sets. We note that for the two smallest classes of the data Indian Pines, only half of their instances are sampled randomly without replacement from the original data to construct the training sets. The IR of the training set is calculated by N_1/N_L. Hence, the IRs of the four data sets are 19.96, 24.50, 49.10, and 73.60, respectively. More details are given in Table I.

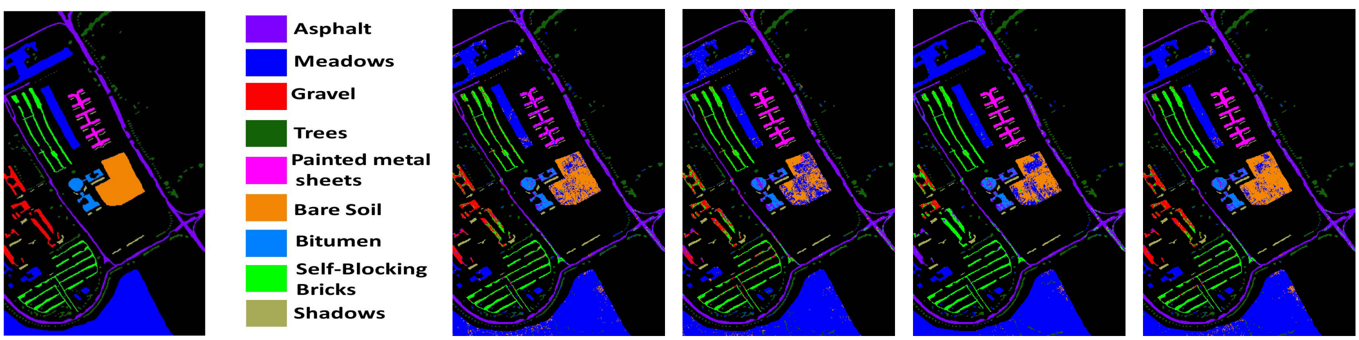
D. Results and Analysis

The results obtained in the experiments according to AA, OA, F-measure, and Gmean and minimum Recall values are

TABLE II

CLASSIFICATION RESULTS OF THE PAVIA UNIVERSITY IMAGE, AND INDIAN PINES IMAGE WITH DIFFERENT IRs, RESPECTIVELY, OBTAINED BY SVM, RF, RoF, AND THE PROPOSED AdaSRoF METHODS

Class	Pavia University (IR:19.96)				Indian Pines											
					Case 1 (IR:24.50)				Case 2 (IR:49.10)				Case 3 (IR:73.60)			
	SVM	RF	RoF	AdaSRoF	SVM	RF	RoF	AdaSRoF	SVM	RF	RoF	AdaSRoF	SVM	RF	RoF	AdaSRoF
1	88.35	88.06	89.25	**92.64**	73.91	60.43	82.61	**90.87**	86.96	84.78	0.00	**96.52**	43.48	61.30	0.00	**73.91**
2	96.63	96.77	**98.40**	96.71	76.84	63.17	79.18	**82.32**	78.99	69.63	69.82	**83.89**	**87.14**	74.59	74.28	86.14
3	70.04	62.65	30.89	**75.94**	65.52	54.03	57.80	**69.89**	74.46	56.02	48.01	**77.41**	77.28	60.19	49.98	**78.47**
4	81.82	85.64	82.54	**90.04**	51.21	31.68	3.08	**76.40**	66.21	43.63	0.00	**90.05**	72.29	52.95	0.90	**86.87**
5	82.99	97.33	97.98	**98.41**	87.40	82.69	87.10	**89.70**	89.43	85.58	90.21	**93.54**	**96.46**	90.65	90.56	96.19
6	75.90	54.74	66.06	**83.85**	95.89	95.98	**97.11**	97.03	97.53	94.13	97.96	**97.84**	97.26	96.48	97.69	**98.88**
7	**75.16**	63.17	60.39	70.92	**85.71**	69.29	71.43	**85.71**	92.86	53.57	0.00	**95.00**	78.57	50.00	0.00	**83.57**
8	86.40	84.22	**93.73**	88.20	93.97	95.85	98.05	**99.19**	97.36	97.81	99.24	**99.61**	97.61	98.09	**99.97**	99.70
9	97.61	99.66	99.64	**99.95**	**100.00**	65.00	0.00	**100.00**	80.00	64.00	0.00	**98.00**	70.00	49.00	0.00	**81.00**
10					72.78	67.97	72.91	**84.75**	75.99	68.86	73.02	**85.82**	84.73	78.38	75.49	**91.03**
11					82.64	86.27	**89.99**	84.75	86.91	88.20	**89.24**	86.89	86.24	88.50	**89.80**	87.53
12					62.66	45.88	33.88	**80.58**	74.84	55.07	37.62	**88.53**	80.67	65.91	32.21	**92.84**
13					92.97	91.73	96.97	**98.86**	96.95	97.20	98.41	**99.51**	95.14	89.58	**97.64**	97.08
14					96.07	94.52	96.21	**96.65**	94.07	94.62	**96.06**	95.65	96.09	96.15	97.86	**98.10**
15					47.99	41.06	47.84	**63.07**	66.15	48.54	49.35	**73.24**	73.36	53.62	44.54	**73.87**
16					95.74	98.72	100.00	**100.00**	**100.00**	99.36	97.45	**100.00**	95.74	91.91	94.26	**99.79**
AA	83.88	81.36	79.88	**88.52**	80.08	71.52	69.64	**87.49**	84.92	75.06	59.15	**91.34**	83.25	74.83	59.07	**89.06**
OA	88.59	85.96	87.16	**91.66**	79.93	75.36	78.88	**85.61**	84.31	78.35	76.71	**88.42**	87.45	81.72	77.56	**90.03**
F-measure	86.55	83.66	84.17	**90.04**	77.20	70.60	74.21	**85.19**	82.51	75.64	60.33	**90.17**	84.82	79.55	61.63	**90.10**
Gmean	83.40	79.65	75.69	**87.98**	78.28	67.86	0.00	**86.74**	84.18	71.97	0.00	**90.96**	81.82	72.45	0.00	**88.56**
Recall	70.04	54.74	30.89	**70.92**	47.99	31.68	0.00	**63.07**	66.15	43.63	0.00	**73.24**	43.48	49.00	0.00	**73.87**
Time (s)	134.95	0.67	43.58	53.90	435.52	0.58	86.53	78.50	1577.41	0.92	146.42	128.45	3448.62	1.31	209.35	172.83

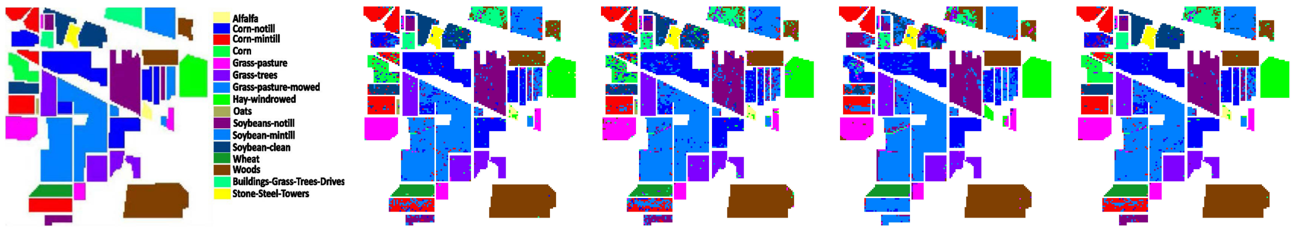


(a) (b) (c) (d) (e)

Fig. 1. Classification maps of the hyperspectral image Pavia University obtained by (a) ground truth, (b) SVM, (c) RF, (d) RoF, and the proposed (e) AdaSRoF, respectively.

exhibited in Table II for SVM, RF, traditional RoF, and the proposed AdaSRoF on the four imbalanced multiclass hyperspectral data sets. The best result in each data set for each performance measure is highlighted in bold font. The results show that the traditional RoF focuses on classifying the majority classes but sacrifices the accuracies of the minority classes. When the IR increases (cases 3 and 4), the recognition of the minority classes are ignored by RoF completely. In addition, different from the conclusion of the most existing literature, SVM and RF statistically outperform the traditional RoF for hyperspectral data classification in the imbalance case. The best statistical results are always achieved by AdaSRoF. The proposed method is effective for not only the classification of the minority classes but also the improvement of the OA. With respect to SVM, RF, and RoF, the proposed method gets about 7%, 16%, and 32% improvements in terms of AA, and over 5%, 10%, and 12% increases in terms of OA. Moreover, the results of F-measure, Gmean, and minimum Recall also prove that AdaSRoF significantly outperforms the reference methods for the classification of the multiclass imbalanced hyperspectral data sets.

McNemar's test value over 1.96 ($p < 0.05$) means that there is a significant difference between two algorithms. Table III presents the McNemar's test results for the significant evaluation of the proposed AdaSRoF with respect to other methods on the four imbalanced data sets. All the values given in Table III are greater than 1.96, i.e., when compared to SVM, RF, and traditional RoF, the performance increase achieved by the proposed method is significant. The diversity results of all the methods are exhibited in Table IV. This table shows that the differentiated sampling strategy is effective to increase the diversity of the RoF ensemble system and improves its performance. Moreover, the analysis results here are consistent with those in Table II, i.e., the proposed AdaSRoF has the best results among the four methods. In addition, some feature extraction approaches, such as singular spectral analysis and sparse representation, have great performance in hyperspectral data classification. Hence, it is interesting to combine those feature level methods with the proposed methods to further increase the ensemble diversity in the future work. Graphical visualizations of the results are presented in Figs. 1 and 2. The two figures, respectively, exhibit the classification maps



(a) (b) (c) (d) (e)

Fig. 2. Classification maps of the hyperspectral data Indian Pines (IR of the training set: 73.6) obtained by (a) ground truth, (b) SVM, (c) RF, (d) RoF, and the proposed (e) AdaSRoF, respectively.

TABLE III

RESULTS OF MCNEMAR'S TEST FOR THE FOUR METHODS

	University	Indian Pines		
		Case 1	Case 2	Case 3
AdaSRoF vs.SVM	21.37	14.32	11.20	7.62
AdaSRoF vs.RF	34.43	22.92	22.69	18.31
AdaSRoF vs.RoF	32.45	17.55	25.39	26.72

TABLE IV

ENSEMBLE DIVERSITIES OF THE CLASSIFIERS

	University	Indian Pines		
		Case 1	Case 2	Case 3
SVM	-	-	-	-
RF	0.1124	0.1637	0.1573	0.1498
RoF	0.0777	0.1287	0.1092	0.1095
AdaSRoF	0.1242	0.1826	0.1750	0.1667

obtained by different classification methods for Pavia University and Indian Pines (case 3, IR:73.6) images. With respect to the SVM, RF, and traditional RoF, the proposed method not only obtains the significant improvement of the minority classes identification but also results in more accurate classification of the majority classes.

IV. CONCLUSION

In this letter, we have proposed an original adaptive ensemble method based on SMOTE and RoF with differentiated sampling rates (AdaSRoF) for the classification of the imbalanced multiclass hyperspectral data. The proposed AdaSRoF is an internal imbalance sampling-based ensemble approach. It increases the diversity of base classifiers via producing several different training sets with a dynamic sampling rate $\beta\%$ and decreases the risk of artificial noise in SMOTE by an adaptive weight function. Seven evaluation measures are adopted for model effectiveness assessment. Experimental results show that the proposed method significantly outperforms SVM, RF, and RoF and is effective for multiclass imbalanced data classification.

REFERENCES

[1] P. Ghamisi, J. Plaza, Y. Chen, J. Li, and A. J. Plaza, "Advanced spectral classifiers for hyperspectral images: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 1, pp. 8–32, Mar. 2017.

[2] J. Ren, "ANN vs. SVM: Which one performs better in classification of MCCs in mammogram imaging," *Knowl.-Based Syst.*, vol. 26, pp. 144–153, Feb. 2012.

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

[4] J. Ren, D. Wang, and J. Jiang, "Effective recognition of MCCs in mammograms using an improved neural classifier," *Eng. Appl. Artif. Intell.*, vol. 24, no. 4, pp. 638–645, Jun. 2011.

[5] J. Díez-Pastor, J. J. Rodríguez, C. I. García-Osorio, and L. I. Kuncheva, "Diversity techniques improve the performance of the best imbalance learning ensembles," *Inf. Sci.*, vol. 325, pp. 98–117, Dec. 2015.

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

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

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

[9] W. Feng, W. Huang, and J. Ren, "Class imbalance ensemble learning based on the margin theory," *Appl. Sci.*, vol. 8, no. 5, p. 815, 2018.

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

[11] 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, Apr. 2018.

[12] 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, Jun. 2018.

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

[14] 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.

[15] 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, Jul. 2015.

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

[17] 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.

[18] 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.

[19] 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, Mar. 2017.

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

[21] G. M. Foody, "Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 5, pp. 627–633, May 2004.