Enhanced-Random-Feature-Subspace-Based Ensemble CNN for the Imbalanced Hyperspectral Image Classification

Qinzhe Lv, Wei Feng, Yinghui Quan, Gabriel Dauphin, Lianru Gao, *Senior Member, IEEE*, and Mengdao Xing, *Fellow, IEEE*

***Abstract*—Hyperspectral image (HSI) classification often faces the problem of multiclass imbalance, which is considered to be one of the major challenges in the field of remote sensing. In recent years, deep learning has been successfully applied to the HSI classification, a convolutional neural network (CNN) is one of the most representative of them. However, it is difficult to effectively improve the accuracy of minority classes under the problem of multiclass imbalance. In addition, ensemble learning has been successfully applied to solve multiclass imbalance, such as random forest (RF) This article proposes a novel enhanced-random-feature-subspace-based ensemble CNN algorithm for the multiclass imbalanced problem. The main idea is to perform random oversampling of training samples and multiple data enhancements based on random feature subspace, and then, construct an ensemble learning model combining random feature selection and CNN to the HSI classification. Experimental results on three public hyperspectral datasets show that the performance of the proposed method is better than the traditional CNN, RF, and deep learning ensemble methods.**

***Index Terms*—Convolutional neural network (CNN), enhanced random feature subspace (ERFS), ensemble learning, hyperspectral image (HSI) classification, multiclass imbalance.**

I. Introduction

HYPERSPECTRAL images (HSIs), which contain abundant information in hundreds of continuous spectral features, have been widely applied in land cover mapping, military field, environmental modeling and monitoring, and precision agriculture [1]–[4]. HSIs usually contain multiple classes, hence, one of the most important tasks in hyperspectral research is to design a supervised classification algorithm that assigns a class label to each pixel after proper training. However, the number of instances of each class tends to vary greatly, which is called class imbalance [5], [6]. The problem of class imbalance brings serious challenges to the HSI classification, which reduces the effectiveness of many existing algorithms [5]. Therefore, how to improve the accuracy of minority classes without damaging the accuracy of majority classes is a great challenge [7].

There has been a lot of effort devoted to the problem of imbalance data classification, such as cost-sensitive algorithms [8], kernel-based algorithms [9], and active learning algorithms [10]. However, the aforementioned approaches have large computation costs, especially for HSIs [11]. Some methods try to solve class imbalanced by generating artificial samples of the minority class, e.g., random undersampling (RUS), random oversampling (ROS) [12], the synthetic minority oversampling technique (SMOTE) [13], and the orthogonal complement subspace projection (OCSP) [14]. However, these methods have limitations, RUS and ROS lose valid information or increase the risk of overfitting, respectively. SMOTE generates artificial samples by a linear combination of real samples, hence, the new features generated are limited. The OCSP uses nonlinear artificial sample generation, which can produce rich new features, but the gradient-constraints-based algorithm is computationally intensive and time consuming [15].

As a powerful feature extraction tool to solve nonlinear problems, deep learning is widely used in several image processing tasks [16], [17]. Inspired by these successful applications, deep learning is also introduced into HSI classification [18], [19]. The convolutional neural network (CNN) is one of the deep learning models successfully applied to the HSI classification [20]. A well-designed CNN model can be used to extract the spectral [21] or spectral-spatial [22], [23] features of hyperspectral images and obtain the final classification results. Although the CNN-based method can effectively improve the classification effect of HSIs, it usually needs a large number of samples [24]. When facing the class imbalanced problem, the accuracy of minority class is severely impaired. Therefore, to effectively

Manuscript received November 28, 2020; revised January 11, 2021; accepted March 13, 2021. Date of publication March 26, 2021; date of current version April 21, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61772397 and Grant 12005169, in part by the National Key R&D Program of China under Grant 2016YFE0200400, in part by the Science and Technology Innovation Team of Shaanxi Province under Grant 2019TD-002, and in part by the Open Research Fund of Key Laboratory of Digital Earth Science under Grant 2019LDE005. *(Qinzhe Lv and Wei Feng contributed equally to this work.) (Corresponding authors: Wei Feng; Yinghui Quan.)*

Qinzhe Lv, Wei Feng, and Yinghui Quan are with the Department of remote sensing science, and technology, School of Electronic Engineering, Xidian University, Xi'an 710071, China (e-mail: qzlv@stu.xidian.edu.cn; wfeng@xidian.edu.cn; yhquan@mail.xidian.edu.cn).

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

Lianru Gao is with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China (e-mail: gaolr@aircas.ac.cn).

Mengdao Xing is with the Academy of Advanced Interdisciplinary Research, Xidian University, Xi'an 710071, China (e-mail: xmd@xidian.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2021.3069013

TABLE I
DATA INFORMATION

	Indian Pines AVRIS				University of Pavia ROSIS				Salinas			
No.	Class Name	Train.	Test	Total	Class Name	Train.	Test	Total	Class Name	Train.	Test	Total
1	Alfalfa	2	44	46	Asphalt	332	6299	6631	Brocoli_green_weeds_1	100	1909	2009
2	Corn-notill	71	1357	1428	Meadows	932	17717	18649	Brocoli_green_weeds_2	186	3540	3726
3	Corn-mintill	41	789	830	Gravel	105	1994	2099	Fallow	99	1877	1976
4	Corn	12	225	237	Trees	153	2911	3064	Fallow_rough_plow	70	1324	1394
5	Grass-pasture	24	459	483	Painted metal sheets	67	1278	1345	Fallow_smooth	134	2544	2678
6	Grass-trees	37	693	730	Bare Soil	251	4778	5029	Stubble	198	3761	3959
7	Grass-pasture-mowed	1	27	28	Bitumen	67	1263	1330	Celery	179	3400	3579
8	Hay-windrowed	24	454	478	Self-Blocking Bricks	184	3498	3682	Grapes_untrained	564	10707	11271
9	Oats	1	19	20	Shadows	47	900	947	Soil_vinyard_develop	310	5893	6203
10	Soybean-notill	49	923	972					Corn_senesced_green_weeds	164	3114	3278
11	Soybean-mintill	123	2332	2455					Lettuce_romaine_4wk	53	1015	1068
12	Soybean-clean	30	563	593					Lettuce_romaine_5wk	96	1831	1927
13	Wheat	10	195	205					Lettuce_romaine_6wk	46	870	916
14	Woods	63	1202	1265					Lettuce_romaine_7wk	54	1016	1070
15	Buildings-Grass-Trees-Drives	19	367	386					Vinyard_untrained	363	6905	7268
16	Stone-Steel-Towers	5	88	93					Vinyard_vertical_trellis	90	1717	1807
Total		512	9737	10249		2138	40638	42776		2706	51423	54129
IR	123				19.83				12.26			

TABLE II
ARCHITECTURE OF THE CNN DESIGN FOR INDIAN PINES AVRIS

Layer	Conv. Kernel	Activation	MaxPool	Dropout
Input	$25\times25\times n_f$			
Conv.1	4×4×32	ReLU	2×2	0.5
Conv.2	4×4×64	ReLU	2×2	0.25
Conv.3	3×3×64	ReLU	2×2	0.25
Flatten	-	-	-	-
Dense	-	Softmax	-	-
Output	$1\times L$			

TABLE III
ARCHITECTURE OF THE CNN DESIGN FOR UNIVERSITY OF PAVIA ROSIS, AND SALINAS

Layer	Conv. Kernel	Activation	MaxPool	Dropout
Input	$15\times15\times n_f$			
Conv.1	3×3×32	ReLU	2×2	0.5
Conv.2	3×3×64	ReLU	2×2	0.25
Conv.3	2×2×64	ReLU	-	0.25
Flatten	-	-	-	-
Dense	-	Softmax	-	-
Output	$1\times L$			

deal with the multiclass imbalanced problem, the CNN-based method must be improved, such as reducing network parameters or simplifying network architecture [25] or generating artificial samples for minority classes.

Ensemble learning combines multiple learners to achieve more generalization than a single learner and has been successfully applied to the HSI processing, such as rotation forest (RoF) and its improvement [26]–[29], but most of them are implemented on balanced datasets. In recent years, ensemble learning has been used to solve the class imbalanced problem [30], such as dynamic synthetic minority oversampling technique-based RoF [7]. Given the superior performance of ensemble learning compared with the subclassifier and the strong feature extraction ability of the CNN, this article will explore a new method combining ensemble learning with the CNN to improve the imbalanced HSI classification performance.

The contribution of this article is twofold. First, we propose a novel enhanced random feature subspace (ERFS)-based ensemble CNN (EECNN) algorithm for the multiclass imbalanced problem. Second, the proposed ERFS algorithm can be used alone to generate a balanced training set and improve the classification accuracy of the RF and CNN. The proposed method makes ROS of a dynamic sampling rate for each training sample of minority classes and divides each sample obtained by ROS into different random feature subspaces (RFSs). We use a variety of data augmentation methods such as rotation, folding, etc., to randomly enhance the data on each RFS and obtain the ERFS. Finally, ensemble learning, which combines a well-designed CNN and random feature selection is used to obtain HSI classification results. The experimental results on three common datasets validate the effectiveness of the aforementioned two contributions.

The rest of this article is organized as follows. Section II provides an overview of the related work. Section III describes the proposed methodology in detail. Section IV evaluates the performance of the proposed approach. Finally, Section V concludes this article.

II. RELATED WORK

When faced with the multiclass imbalanced problem, the CNN model needs to oversample the training samples before classifying to obtain a relatively balanced dataset. Ensemble learning can improve the generalization ability of the overall model to a certain extent, thus improving the overall classification effect [31]. The purpose of this article is to find an ensemble CNN based on the ERFS to overcome the multiclass imbalanced problem. This section introduces the background knowledge of the sampling methods (ROS, RUS), ERFS, CNN, and ensemble learning.

A. Random Sampling Methods (ROS, RUS)

The random sampling method is usually used to balance the distribution of multiclass data [32]. Simple random sampling method mainly includes ROS and RUS. ROS obtains balanced

TABLE IV
CLASSIFICATION RESULTS OF EECNN IN INDIAN PINES AVRIS

Ensemble Size (T)	Index	Feature Subset Size (F)				
		10	20	30	40	50
5	OA (%)	97.10±0.27	96.86±0.24	96.85±0.39	97.02±0.16	96.91±0.19
	AA (%)	95.80±0.44	95.23±0.54	94.43±1.13	95.36±0.42	94.62±0.80
	F1 (%)	96.35±0.37	95.81±0.37	95.40±0.82	95.99±0.41	95.59±0.39
	Recall (%)	97.40±0.33	97.03±0.43	97.15±0.39	97.17±0.38	97.26±0.59
	K (%)	96.69±0.30	96.42±0.27	96.46±0.44	96.60±0.18	96.48±0.21
	Time (s)	142.37	178.38	216.55	258.95	308.29
10	OA (%)	97.32±0.16	97.21±0.12	97.34±0.04	97.29±0.20	97.25±0.23
	AA (%)	95.84±0.21	95.48±0.63	95.87±0.21	95.71±0.37	95.67±0.19
	F1 (%)	96.45±0.19	96.25±0.41	96.50±0.10	96.36±0.29	96.35±0.10
	Recall (%)	97.58±0.21	97.59±0.26	97.68±0.18	97.52±0.20	97.59±0.26
	K (%)	96.95±0.19	96.82±0.14	96.95±0.04	96.91±0.23	96.86±0.27
	Time (s)	287.07	381.54	440.84	531.18	641.23
20	OA (%)	97.44±0.10	**97.57±0.07**	97.44±0.09	96.41±0.13	97.43±0.12
	AA (%)	95.91±0.13	**96.23±0.20**	95.89±0.13	95.93±0.23	95.90±0.35
	F1 (%)	96.51±0.13	**96.82±0.19**	96.55±0.12	96.53±0.16	96.61±0.17
	Recall (%)	97.67±0.15	97.89±0.16	97.77±0.16	97.73±0.17	97.86±0.20
	K (%)	97.08±0.12	**97.23±0.08**	97.09±0.11	97.05±0.15	97.07±0.13
	Time (s)	585.86	811.40	942.24	1088.94	1343.67
30	OA (%)	97.46±0.07	97.42±0.10	97.41±0.07	96.53±0.13	97.41±0.14
	AA (%)	95.96±0.16	96.10±0.13	95.77±0.23	96.66±0.20	96.01±0.18
	F1 (%)	96.58±0.14	96.63±0.14	96.47±0.17	96.66±0.17	96.66±0.15
	Recall (%)	97.74±0.16	97.70±0.20	97.76±0.14	**98.01±0.15**	97.89±0.13
	K (%)	97.10±0.07	97.06±0.11	97.05±0.08	97.18±0.09	97.05±0.15
	Time (s)	865.39	1197.33	1525.33	1816.62	2304.97

data by randomly extracting and replicating a certain number of minority samples, to obtain balanced training samples, the original sample size of each minority class is N_{i}, and the maximum class sample size is $N_{\max}$, the ROS times of each minority class are equal to $N_{\max} - N_{\mathrm{i}}$. On the contrary, RUS obtains balanced data by randomly eliminating samples of majority classes, to obtain balanced training samples, the original training sample size of majority classes is N_{m}, the original training sample size of the minimum class is $N_{\min}$, and the RUS times of each majority class are equal to $N_{\mathrm{m}} - N_{\min}$. RUS loses samples of majority classes and leads to the loss of effective information, hence, this article chooses the ROS to generate balanced data.

B. Enhanced Random Feature Subspace (ERFS)

After ROS processing for imbalanced training samples, there are a large number of duplicate samples in the balanced sample set, especially for those with few original samples. However, a large number of repeated samples increase the risk of overfitting of the classification model, which is unfavorable for correct classification. Random subspace ensembles can be used to alleviate the curse of dimensionality and the high feature-to-instance ratio [33]. The ERFS is constructed by RFS and data augmentation. First, a 3-D sample x_i generated by ROS, which containing spectral-spatial feature is selected and randomly divided into *K* RFSs in the band dimension [31]. Second, k kinds of data augmentation methods, such as horizontal flip, vertical flip, diagonal flip, and rotation with different angles, are performed on RFSs. Finally, when all samples generated by ROS have completed the aforementioned processing, a new training set is obtained that not only balances the data distribution but also avoids the risk of overfitting caused by repeated data. Compared with the random data augment of each minority sample, the ERFS can generate more diverse training samples, which is conducive to HSI classification.

C. Convolutional Neural Network (CNN)

A well-designed CNN has been proved to be more accurate for HSI classification [34], [35]. The classical CNN structure includes convolution layers, nonlinearity mapping layers, and pooling operation layers.

The convolution layer is the most characteristic part of the CNN, for example, when the input of the CNN is a 3-D matrix X with spectral-spatial information, its shape is $m \times m \times d$, where $m \times m$ represents the size of the spatial information and d is the number of bands, suppose the first convolution layer contains k filters, the output of the jth filter can be represented as follows [18]:

$$\mathbf{y}_j = \sum_{i=1}^{d} f\left(\mathbf{x}_i * \mathbf{w}_j + b_j\right), j = 1, 2, \ldots, k \tag{1}$$

where $\mathbf{x}_i$ is the ith feature map of $\mathbf{X}$; $\mathbf{w}_j$ and $\mathbf{b}_j$ represents the weight and offset of the jth filter respectively; $*$ represents

TABLE V
CLASSIFICATION RESULTS OF EECNN IN UNIVERSITY OF PAVIA ROSIS

Ensemble Size (T)	Index	Feature Subset Size (F)				
		10	20	30	40	50
5	OA (%)	95.28±0.17	97.07±0.35	96.85±0.29	96.38±0.43	97.45±0.23
	AA (%)	96.58±0.51	97.31±0.32	97.08±0.35	97.30±0.49	97.40±0.23
	F1 (%)	95.91±0.16	97.16±0.21	97.15±0.33	96.83±0.37	97.39±0.24
	Recall (%)	95.45±0.55	97.05±0.19	97.27±0.34	96.52±0.60	97.41±0.27
	K (%)	93.71±0.23	96.11±0.46	95.83±0.38	95.18±0.58	96.62±0.30
	Time (s)	284.34	317.6	364.68	404.40	445.49
10	OA (%)	95.93±0.28	97.42±0.38	97.87±0.09	97.84±0.24	97.95±0.28
	AA (%)	97.45±0.16	97.96±0.15	98.09±0.24	98.14±0.08	97.99±0.19
	F1 (%)	96.38±0.17	97.43±0.27	97.90±0.11	97.77±0.16	97.75±0.23
	Recall (%)	95.60±0.26	97.01±0.41	97.73±0.20	97.46±0.23	97.55±0.28
	K (%)	94.55±0.39	96.56±0.51	97.17±0.12	97.13±0.33	97.27±0.38
	Time (s)	576.20	642.96	741.96	793.02	891.30
20	OA (%)	96.28±0.15	97.30±0.11	98.12±0.13	**98.48±0.03**	98.23±0.11
	AA (%)	97.86±0.05	97.92±0.11	98.21±0.10	**98.37±0.03**	98.22±0.03
	F1 (%)	96.83±0.12	97.48±0.11	97.99±0.09	**98.21±0.05**	97.99±0.08
	Recall (%)	96.10±0.17	97.12±0.12	97.80±0.13	**98.08±0.08**	97.79±0.14
	K (%)	95.02±0.21	96.41±0.16	97.50±0.17	**97.98±0.04**	97.65±0.15
	Time (s)	1166.93	1287.16	1469.60	1796.53	1807.47
30	OA (%)	96.22±0.39	96.52±0.33	97.43±0.20	98.13±0.14	98.38±0.11
	AA (%)	97.56±0.19	97.94±0.09	98.10±0.04	98.17±0.12	98.25±0.12
	F1 (%)	96.66±0.31	96.98±0.28	97.57±0.14	97.98±0.10	98.07±0.12
	Recall (%)	96.00±0.43	96.29±0.39	97.16±0.23	97.83±0.20	97.92±0.14
	K (%)	94.94±0.53	95.34±0.46	96.57±0.27	97.52±0.19	97.84±0.14
	Time (s)	1853.80	1942.33	2389.19	2863.23	2780.83

the convolution operation; and $f(\cdot)$ is an activation function used to enhance the nonlinear representation of models. ReLU is one of the activation functions that researchers used most frequently [16], [36]. The ReLU function can be expressed as follows:

$$\sigma(\mathbf{x}) = \max(0, \mathbf{x}). \tag{2}$$

Pooling, also known as undersampling, is mainly used to reduce the size of feature maps, compress the number of parameters and reduce overfitting. It mainly includes max pooling and average pooling.

Compared with other fully connected networks, the unique design of the CNN lies in its local connections and shared weights, which greatly reduces the number of parameters, speeds up the learning rate, and reduces the possibility of overfitting to a certain extent [37].

D. Ensemble Learning

The ensemble method is a metaalgorithm that combines several machine learning techniques into a prediction model to reduce variance, bias, or improve prediction. The classical ensemble learning algorithms can be divided into three categories: bagging [38], boosting [39], and stacking [40].

In bagging, the original dataset is randomly extracted (with replacement) part of the samples to form a new dataset, which is used to train subclassifier, and each subclassifier can be generated in parallel because of its independence. The prediction results of all subclassifiers are voted by the majority or weighted to get the overall prediction results of the ensemble model. The RF is an upgrade based on Bagging and adopts two randomization ideas: random instance selection and random feature selection [7].

Boosting, which is also known as adaptive resampling and combining (ARCing) [11]. AdaBoost is the most famous algorithm in this family and uses a complete dataset to train each subclassifier, which is characterized by adjusting weights after each iteration to improve the classifier's attention to misclassified samples [41].

Stacking, also known as stacked generalization, is different from the two methods described previously that use deterministic algorithms to realize ensemble, and stacking uses a secondary model to ensemble primary models. In short, the stacking method first trains multiple primary learners with the original training set, then trains the secondary model with the outputs of the primary models, and resulting in a complete model.

III. PROPOSED APPROACH

In this article, a novel EECNN is proposed, which is inspired by DSRoF [7]. First, the data preprocessing is completed through ERFS, and then, the ensemble learning model with the CNN as subclassifiers is used for classification, in which random feature selection provides a subset of data for each CNN. Algorithm 1

TABLE VI
CLASSIFICATION RESULTS OF EECNN IN SALINAS

Ensemble Size (T)	Index	Feature Subset Size (F)				
		10	20	30	40	50
5	OA (%)	96.92±0.31	97.91±0.33	99.10±0.10	99.08±0.13	98.73±0.29
	AA (%)	97.32±0.21	97.94±0.26	98.99±0.16	99.06±0.11	98.81±0.15
	F1 (%)	96.92±0.31	97.49±0.30	98.84±0.17	99.05±0.10	98.68±0.16
	Recall (%)	96.81±0.31	97.32±0.30	98.75±0.17	99.07±0.08	98.62±0.16
	K (%)	96.57±0.35	97.67±0.36	98.99±0.12	98.98±0.14	98.59±0.32
	Time (s)	315.86	331.18	493.50	616.51	726.63
10	OA (%)	96.89±0.44	98.71±0.30	98.49±0.17	99.01±0.13	99.03±0.04
	AA (%)	97.17±0.41	98.71±0.25	98.31±0.10	99.05±0.11	98.84±0.08
	F1 (%)	96.32±0.85	98.48±0.29	98.01±0.12	99.01±0.11	98.74±0.09
	Recall (%)	96.15±0.79	98.36±0.30	97.87±0.13	99.00±0.10	98.70±0.09
	K (%)	96.54±0.49	98.57±0.34	98.32±0.19	98.89±0.14	98.92±0.05
	Time (s)	546.73	635.34	930.61	1285.90	1464.95
20	OA (%)	96.97±0.40	98.13±0.30	97.24±0.18	**99.34±0.06**	99.32±0.04
	AA (%)	97.69±0.36	98.11±0.32	98.40±0.07	**99.30±0.04**	99.25±0.02
	F1 (%)	97.21±0.50	97.83±0.43	98.29±0.07	**99.27±0.04**	99.22±0.02
	Recall (%)	97.08±0.52	97.74±0.44	98.27±0.08	**99.26±0.04**	99.21±0.02
	K (%)	96.63±0.44	97.92±0.34	96.93±0.20	**99.27±0.07**	99.24±0.04
	Time (s)	1204.15	1361.28	2178.19	2612.63	3025.55
30	OA (%)	96.60±0.32	98.86±0.10	99.07±0.06	97.71±0.05	97.91±0.09
	AA (%)	97.20±0.14	98.79±0.06	98.99±0.07	98.73±0.03	98.78±0.06
	F1 (%)	96.55±0.19	98.60±0.05	98.87±0.09	98.68±0.06	98.70±0.05
	Recall (%)	96.40±0.17	98.49±0.05	98.80±0.10	98.64±0.09	98.65±0.06
	K (%)	96.22±0.35	98.73±0.11	98.97±0.07	97.46±0.05	97.67±0.10
	Time (s)	1770.20	2059.43	2541.97	3134.72	3789.66

TABLE VII
CLASSIFICATION RESULTS (%) OF THE INDIAN PINES AVRIS, RESPECTIVELY, OBTAINED BY RF, CNN, ECNN, ERFS+RF, ERFS+CNN, AND THE PROPOSED EECNN

IR=123	RF	CNN	ECNN	ERFS+RF	ERFS+CNN	EECNN
1	**100.00±0.00**	99.11±1.78	**100.00±0.00**	97.37±2.63	99.76±0.73	**100.00±0.00**
2	67.61±1.59	94.42±3.22	95.91±0.66	64.16±2.43	91.64±4.66	**99.14±0.41**
3	73.69±3.60	96.22±1.84	97.41±0.44	66.57±2.39	94.47±2.77	**97.43±0.28**
4	89.88±3.98	**96.34±4.03**	96.09±0.90	83.94±1.59	95.72±2.24	95.91±1.03
5	78.43±5.50	92.54±1.72	92.78±1.38	70.09±0.18	95.55±2.74	**98.91±0.09**
6	72.43±2.63	92.83±2.99	97.31±2.91	67.74±3.47	96.61±4.03	**99.48±0.26**
7	20.00±40.00	40.00±48.99	**100.00±0.00**	50.00±50.00	76.97±17.74	**100.00±0.00**
8	91.38±2.04	98.66±1.05	**100.00±0.00**	92.39±0.33	99.93±0.14	**100.00±0.00**
9	50.00±50.00	77.29±22.66	**100.00±0.00**	**100.00±0.00**	87.74±18.04	99.00±2.00
10	76.50±2.82	94.19±3.04	96.88±1.01	74.57±0.93	89.02±8.26	**98.23±0.15**
11	72.01±1.66	95.67±1.39	96.58±0.11	87.06±0.03	97.07±2.04	**97.80±0.09**
12	81.22±3.44	87.12±7.04	90.51±0.42	79.22±1.30	92.92±5.27	89.27±1.19
13	98.49±0.91	96.39±2.24	99.49±0.00	99.18±0.82	96.88±3.35	**100.00±0.00**
14	84.22±1.54	97.73±3.07	97.99±0.36	88.49±0.39	97.93±1.37	**99.36±0.20**
15	72.52±3.29	92.77±4.70	98.39±0.22	75.73±3.36	96.70±3.07	**98.82±0.59**
16	**95.34±5.00**	61.48±7.75	66.05±1.82	89.17±1.20	67.42±5.30	66.41±1.77
OA (%)	75.68±0.88	94.21±0.77	96.16±0.05	77.55±0.65	94.66±0.90	**97.57±0.07**
AA (%)	76.48±2.65	88.30±3.02	95.34±0.10	80.35±3.63	92.27±1.72	**96.23±0.20**
F1 (%)	59.37±1.51	86.79±1.35	89.71±0.69	65.59±2.07	90.82± 1.63	**96.82±0.19**
Recall (%)	55.72±1.27	87.44±1.74	89.30±0.69	63.11±1.34	91.01±2.15	**97.89±0.16**
K (%)	71.79±1.05	93.40±0.88	95.62±0.05	74.36±0.72	93.73±1.59	**97.23±0.08**
Time (s)	3.92	34.29	208.90	6.17	82.98	811.40

TABLE VIII
CLASSIFICATION RESULTS (%) OF THE UNIVERSITY OF PAVIA ROSIS, RESPECTIVELY, OBTAINED BY RF, CNN, ECNN, ERFS+RF, ERFS+CNN, AND THE PROPOSED EECNN

IR=19.83	RF	CNN	ECNN	ERFS+RF	ERFS+CNN	EECNN
1	86.43±0.75	98.39±0.16	98.99±0.12	90.06±0.43	98.65±0.19	**99.19±0.04**
2	87.60±0.07	96.29±1.15	94.15±0.79	88.78±0.26	97.14±0.66	**98.54±0.02**
3	81.70±2.17	91.25±0.78	97.58±0.27	83.69±0.69	98.55±1.07	**98.18±0.20**
4	97.38±0.31	95.73±0.15	97.93±0.28	**98.36±0.28**	94.02±3.62	97.76±0.15
5	97.50±0.78	99.42±0.04	99.92±0.00	97.66±0.47	99.79±0.13	**99.97±0.04**
6	68.65±0.17	91.70±5.62	99.23±0.10	71.15±0.71	96.72±1.27	**99.35±0.13**
7	84.75±3.11	93.17±5.45	**98.09±0.17**	80.48±1.29	96.26±1.91	97.55±0.28
8	87.80±0.90	95.70±0.33	**97.90±0.27**	86.24±0.35	89.71±3.10	96.30±0.37
9	99.94±0.06	96.40±2.70	99.27±0.20	**99.98±0.04**	94.42±2.03	98.46±0.40
OA (%)	86.67±0.11	95.65±0.32	96.44±0.40	87.81±0.15	96.43±0.35	**98.48±0.03**
AA (%)	87.97±0.02	95.34±1.25	98.12±0.12	88.49±0.20	96.14±0.83	**98.37±3.07**
F1 (%)	82.72±0.16	95.41±0.43	96.98±0.29	85.24±0.35	95.82±0.47	**98.21±0.05**
Recall (%)	79.95±0.13	95.64±0.34	96.18±0.37	83.19±0.39	95.70±0.16	**98.08±0.08**
K (%)	81.90±0.14	94.23±0.39	95.36±0.54	83.59±0.21	95.25±0.45	**97.98±0.04**
Time (s)	3.90	30.72	614.92	7.55	118.09	1796.53

TABLE IX
CLASSIFICATION RESULTS (%) OF THE SALINAS, RESPECTIVELY, OBTAINED BY RF, CNN, ECNN, ERFS+RF, ERFS+CNN, AND THE PROPOSED EECNN

IR=12.26	RF	CNN	ECNN	ERFS+RF	ERFS+CNN	EECNN
1	100.00±0.00	99.47±0.75	100.00±0.00	99.98±0.02	100.00±0.00	**100.00±0.00**
2	99.78±0.07	81.52±14.63	99.82±0.04	99.85±0.15	99.92±0.12	**99.93±0.04**
3	82.90±0.55	97.28±3.04	94.56±1.13	81.83±1.25	99.10±0.76	**99.97±0.06**
4	98.24±0.44	97.83±2.31	97.68±0.18	**99.15±0.51**	97.20±1.23	97.56±0.37
5	99.83±0.38	99.76±0.17	99.92±0.08	99.67±0.11	99.97±0.02	**100.00±0.00**
6	99.87±0.06	99.94±0.05	99.99±0.01	99.93±0.10	99.80±0.20	**100.00±0.00**
7	99.93±0.04	99.53±0.58	99.17±0.07	99.84±0.05	100.00±0.00	**100.00±0.00**
8	91.90±0.23	94.88±0.66	99.59±0.03	78.26±0.48	97.10±1.64	**99.85±0.02**
9	99.77±0.07	**99.98±0.02**	99.81±0.02	99.93±0.03	99.89±0.12	99.88±0.02
10	90.18±0.18	99.65±0.24	99.35±0.07	92.74±0.30	**99.86±0.11**	99.28±0.08
11	98.69±0.54	97.52±2.06	97.74±0.52	**98.80±0.94**	97.48±1.88	97.65±0.11
12	85.61±0.37	96.03±1.81	96.21±1.39	91.89±0.52	97.63±0.63	**97.67±0.21**
13	90.41±1.32	96.70±1.09	99.26±0.07	98.03±0.50	96.84±1.33	**99.76±0.15**
14	97.84±0.40	99.06±0.59	99.75±0.15	99.39±0.14	**99.90±0.14**	99.82±0.07
15	67.20±0.34	90.29±2.26	94.84±0.33	77.60±0.65	84.00±8.55	**97.38±0.49**
16	99.88±0.10	99.08±1.31	100.00±0.00	99.51±0.66	99.88±0.16	**100.00±0.00**
OA (%)	90.63±0.13	95.23±1.20	98.58±0.04	90.84±0.04	96.40±1.23	**99.34±0.06**
AA (%)	93.88±0.16	96.78±0.76	98.61±0.01	94.77±0.16	98.04±0.35	**99.30±0.04**
F1 (%)	93.36±0.14	94.57±2.96	98.42±0.01	93.98±0.17	98.01±0.38	**99.27±0.04**
Recall (%)	93.32±0.10	94.91±2.42	98.33±0.01	93.45±0.20	98.17±0.26	**99.26±0.04**
K (%)	89.61±0.14	94.68±1.34	98.43±0.04	89.78±0.04	96.00±1.36	**99.27±0.07**
Time (s)	17.03	56.48	626.28	35.65	180.77	2612.63

introduces the process of this method in detail. This method has a lower risk of information loss than using RUS in the preprocessing step, and the feature extraction capability of the CNN as subclassifiers is more powerful. Most importantly, the EECNN framework can be used to solve the problem of multiclass imbalance, which is regarded as one of the biggest challenges in the HSI classification.

As described in Algorithm 1, N_l is the number of training instances of the lth class, with a total of L classes. We sort in descending order based on the number of instances of each class, and N_1 is the training size of the largest class 1. We achieve $N_l\ (1 \le l \le L) = N_1$ by ROS training samples from each of the minority class with the oversampling rate $\alpha\% = (N1 - N_l)/N_l \cdot 100\%$, and the obtained dataset is defined as $S_c\ (2 \le l \le L)$. We designed A methods of data augmentation such as horizontal flip, vertical flip, diagonal flip, and rotation with different angles. Randomly split the feature set $\mathbb{F}$ of S_c into A subsets $\mathbb{F}_{c,a}\ (a = 1, \ldots, A)$, and different data augmentation

Algorithm 1: ERFS-Based Ensemble CNN.

Training phase
Input: $S = [X, Y] = \{x_i, y_i\}_{i=1}^{N}$: training set, where sample x_i represents a 3-D pixel block containing feature set $\mathbb{F}$ and y_i represents its label; F: the size of $\mathbb{F}$; L: number of classes; N_l: the number of training instances of lth class; $\alpha\%$: resampling rate; A: number of data augmentation methods; M_t $(t = 2, \ldots, T)$: tth subclassifier; T: number of classifiers; $E = \varnothing$: the ensemble CNN model.
ERFS:
Sort the classes of the imbalanced data in descending order according to their number of instances and keep all the N_1 instances of the largest class 1
for l=2:L **do**
  $\alpha\% = (N1 - N_l)/N_l \cdot 100\%$
  Obtain the dataset S_c $(c = 2, \ldots, L)$ by randomly resampling $\alpha\% \cdot N_l$ instances from the original dataset with replacement.
  Randomly split the feature set $\mathbb{F}$ of S_c into A subsets $\mathbb{F}_{c,a}$ $(a = 1, \ldots, A)$, and different data augmentation methods were used for each $\mathbb{F}_{c,a}$ to obtain S_{ca}.
end for
Build a new balanced dataset S_{balanced} $(c = 1, \ldots, L)$ by mixing the S_{ca} with original training set S.
Ensemble CNN:
The size of feature subset $\mathbb{F}_f$ $(0 < f \leq F)$ is given as f.
for i=1:T **do**
  The feature subset $\mathbb{F}_f$ is extracted randomly from the feature space $\mathbb{F}$ of S_{balanced} by random feature selection.
  The training set S_f with feature subset $\mathbb{F}_f$ is used to train a sub-classifier CNN M_t.
  Save the trained model M_t.
end for
$E = \{M_t\}_{t=1}^{T}$
Output: The trained E.

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

$$y^* = \operatorname{argmax} \sum_{1(M(x^*)=c, c\in\{1,2,\ldots L\})}^{T}$$

methods were used for each $\mathbb{F}_{c,a}$ to obtain S_{ca}. A new balanced dataset S_{balanced} $(c = 1, \ldots, L)$ is built by mixing the S_{ca} with original training set S.

In the next phase, the feature subset $\mathbb{F}_f$ is extracted randomly from the feature space $\mathbb{F}$ of S_{balanced} by random feature selection, and the training set S_f is used to train subclassifier M_t. The overall prediction results of the ensemble model are obtained by majority voting on the prediction results of all subclassifiers. Finally, we get a fully trained model $E = \{M_t\}_{t=1}^{T}$ that can be used to classify multiclass imbalanced HSIs.

IV. Experimental Results

A. Datasets

To evaluate the proposed ERFS-based Emsemble CNN algorithm, three standard hyperspectral datasets were used in the experiment: *Indian Pines AVRIS*;*University of Pavia ROSIS*; and *Salinas.*

1) *Indian Pines AVRIS* is extremely imbalanced and consisted of 145 * 145 pixels, and 200 spectral bands after removing 20 water absorption bands. The dataset was gathered by the AVIRIS sensor over the Indian Pines test site in North-western Indiana with 16 classes and 10 249 instances.

2) *University of Pavia ROSIS* was acquired by the ROSIS sensor during a flight campaign over Pavia, northern Italy, consists of 610 * 340 pixels, and 103 spectral bands. The dataset consists of nine land cover classes and 42 776 instances.

3) *Salinas* was collected by the 224-band AVIRIS sensor over Salinas Valley, CA, USA, consists of 512 * 217 pixels, and 204 spectral bands after removing 20 water absorption bands. The dataset consists of 16 land cover classes and 54 129 instances.

The aforementioned three datasets are divided into independent training sets and test sets, respectively, to verify the effect of various algorithms in the experiment. Randomly selected (without replacement) 5% labeled samples of each dataset as training samples and the rest as test samples. Besides, we give the class imbalance ratio (IR) of the dataset that can be obtained by dividing the number of the smallest class by the number of samples of the largest class to better verify the effectiveness of the proposed EECNN. The IR of the original training set sampled from three hyperspectral datasets is 123, 19.83, and 12.26, respectively. More detailed information on the aforementioned three experimental datasets is shown in Table I.

B. Experiment Settings

As mentioned previously, the *Indian Pines AVRIS* dataset is the most imbalanced, and the other two datasets are larger and have similar IR. Considering the differences in IR and experimental conditions, we set the spatial feature size $m \times m$ of *Indian Pines AVRIS* dataset with small data volume and large IR as 25 × 25, which for *University of Pavia ROSIS* dataset and *Salinas* dataset with larger data volume but small IR is 15×15. Therefore, two different structures of CNNs as subclassifiers are designed to accommodate different sample sizes, and they are all convoluted by three convolutions (Conv.) layer, three ReLU layers, and three pooling layers. The detailed structural information of the two CNN models is given in Tables II and III, respectively. In the proposed EECNN algorithm, the optimal values of empirical parameters F (the size of feature subset) and T (the number of sub-classifiers) need to be obtained from experiments, so we decided $F =$ (10, 20, 30, 40, 50) and $T =$ (5, 10, 20, 30).

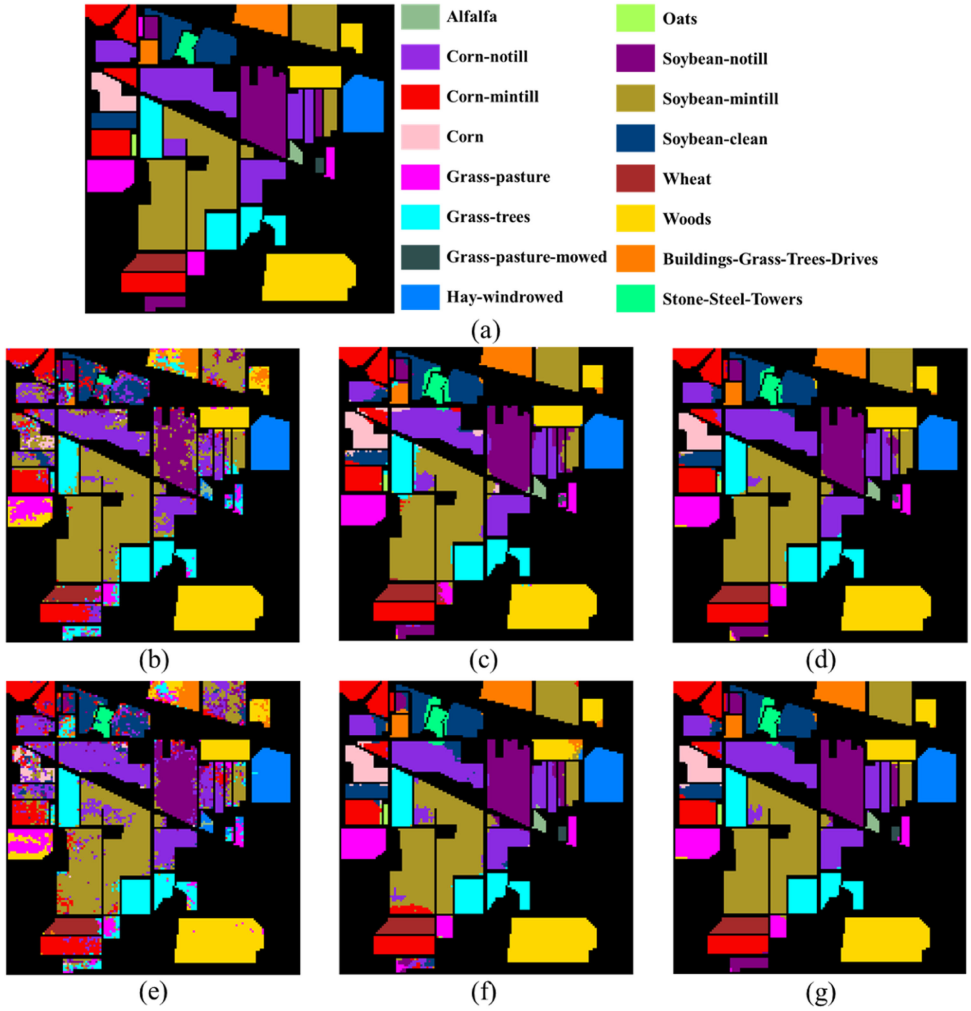


Fig. 1. Classification maps of the Indian Pines AVRIS (IR=123) obtained by (a) ground truth, (b) RF (OA = 75.68%, AA = 76.48%), (c) CNN (OA = 94.21%, AA = 88.30%), (d) ECNN (OA = 96.16%, AA = 95.34%), (e) ERFS+RF (OA = 77.55%, AA = 80.35%), (f) ERFS+CNN (OA = 94.66%, AA = 92.27%), and (g) proposed EECNN (OA = **97.57**%, AA = **96.23**%), respectively.

To evaluate the performance of the EECNN, described in the previous section, RF, traditional CNN, Ensemble CNN, ERFS-based RF, and ERFS-based CNN (recorded as RF, CNN, ECNN, and ERFS+RF, ERFS+CNN) are utilized in the comparative experiment. In this article, all experiments are performed on a computer with 8-core 3.0 GHz CPU and 16.0-GB RAM using Python 3.7 (64-bit). All the results are the average of the algorithms run independently over ten times.

C. Evaluation Metrics

In the experiment, we choose overall accuracy (OA), average accuracy (AA), F1 Score (F1), Recall, and Kappa (K) as evaluation indexes to measure the effect of various algorithms in a multiclass imbalanced hyperspectral classification clearly. AA and OA represent the average of the accuracies for each class and the percentage of correctly predicted instances, respectively, which are the most common indicators in classification tasks. Recall is the percentage of instances correctly classified in each class, which is considered as an important index to measure the performance of algorithms in a multiclass imbalanced problem [42]. F1 score, also known as balanced F score, is defined as the harmonic average of precision rate and recall rate. Kappa is used for the consistency test. In the multiclass imbalanced classification, the training model is biased, which is more conducive to the classification of majority classes of samples, and Kappa of the model with strong bias is low.

D. Results and Analysis

Tables IV–VI show the experimental results of the proposed EECNN algorithm with different ensemble size T and feature subset size F on *Indian Pines AVRIS*, *University of Pavia ROSIS*, and *Salinas*, including OA, AA, F1, Recall, Kappa, and running time. The best experimental results are shown in bold. In the experiment of *Indian Pines AVRIS*, when the ensemble size $T = 20$ and the feature subset size $F = 20$, the proposed EECNN algorithm has the highest OA, AA, F1, and Kappa,

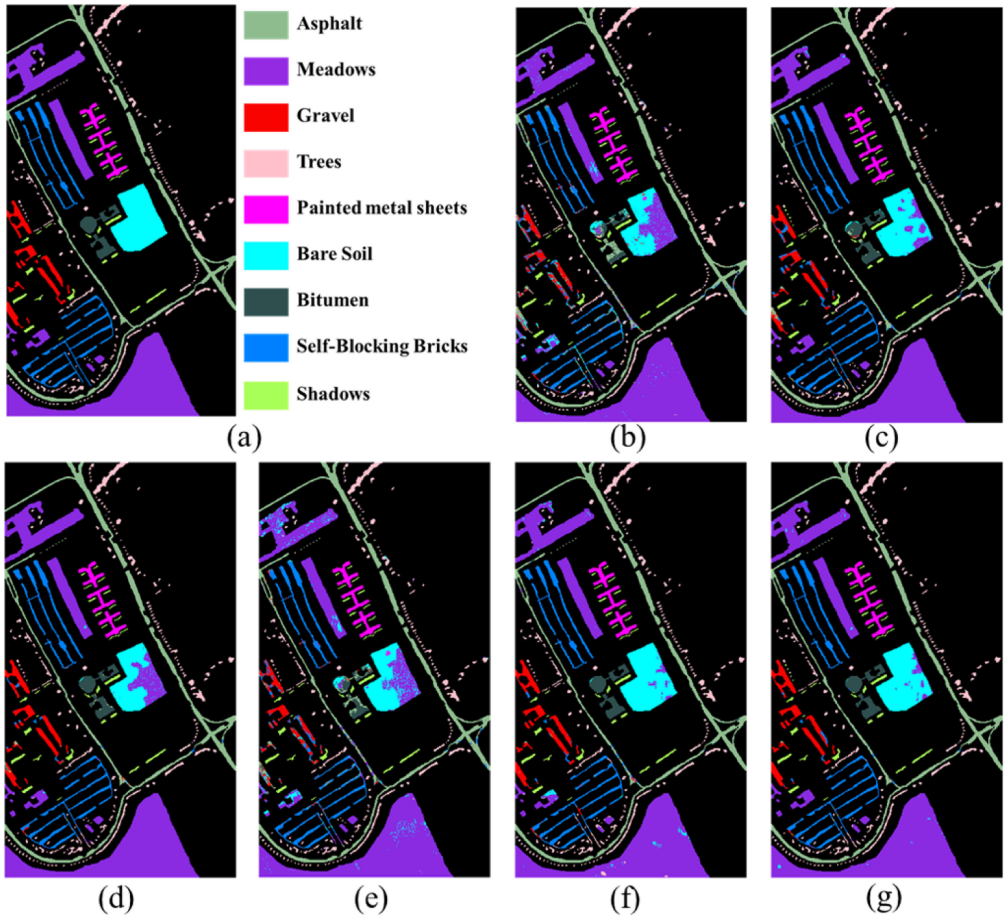


Fig. 2. Classification maps of the University of Pavia ROSIS (IR=19.83) obtained by (a) ground truth, (b) RF (OA = 86.67%, AA = 87.97%), (c) CNN (OA = 95.65%, AA= 95.34%), (d) ECNN (OA = 96.44%, AA = 98.12%), (e) ERFS+RF (OA = 87.81%, AA = 88.49%), (f) ERFS+CNN (OA = 96.43%, AA = 96.14%), and (g) proposed EECNN (OA = **98.48**%, AA = **98.37**%), respectively.

which are 97.57%, 96.23%, 96.82%, and 97.23%, respectively, which preliminarily proves the effectiveness and robustness of the algorithm. For *University of Pavia ROSIS* and *Salinas*, the proposed EECNN algorithm has the highest OA, AA, F1, Recall, and Kappa when the ensemble size $T = 20$ and the feature subset size $F = 40$, which are 98.48%, 98.37%, 98.21%, 98.08%, and 97.89%, respectively, in *University of Pavia ROSIS*, and 99.34%, 99.30%, 99.27%, 99.26%, and 99.27% in *Salinas*. The experimental results of the three datasets show that although the *Indian Pines AVRIS* dataset adopts a larger spatial feature size of 25 (15 for others), the classification effect is the lowest due to the largest IR (IR=123), and the other two classification results are also inversely proportional to the IR value.

Tables VII–IX show the experimental results of RF, CNN, ECNN, ERFS + RF, ERFS + CNN, and the proposed EECNN algorithm on the aforementioned three datasets, including OA, AA, F1, Recall, Kappa, and running time. In the experiments of three datasets, the proposed EECNN algorithm always has the best classification effect. Figs. 1–3 show the ground truth (gt) of *Indian Pines AVRIS*, *University of Pavia ROSIS,* and *Salinas*, and their classification effect maps obtained by different algorithms. In the *Indian Pines AVRIS* dataset, the OA and AA of the proposed EECNN algorithm are 21.89% and 19.75% higher than RF, 3.36% and 7.93% higher than the CNN, 1.41% and 0.89% higher than the ECNN, respectively. In the *University of Pavia ROSIS* and *Salinas* datasets with lower IR, the classification effect of all algorithms are better than their performance in *Indian Pines AVRIS*, and the classification effect of the proposed EECNN algorithm is better than other algorithms. In addition, the EECNN algorithm is more effective than the traditional algorithm in the dataset with larger IR, and all the classification effects of the traditional algorithms combined with the ERFS are improved. Moreover, the results of F1, Recall, and Kappa in all datasets also show that the proposed EECNN has better classification performance than other methods.

Figs. 4 and 5, respectively, show the change curves of OA and AA of the aforementioned algorithms when the spatial feature size $m \times m$ changes. In *Indian Pines AVRIS*, set $T = 20$, $E = 20$, $m = (25, 27, 29, 31, 33)$, and in *University of Pavia ROSIS* and *Salinas*, set $T = 20$, $E = 40$, $m = (15, 17, 19, 21, 23)$. From Figs. 4 and 5, it can be seen that the performance of all algorithms remains stable or increases slightly with the increase of the spatial feature size m, and the proposed EECNN algorithm always works best.

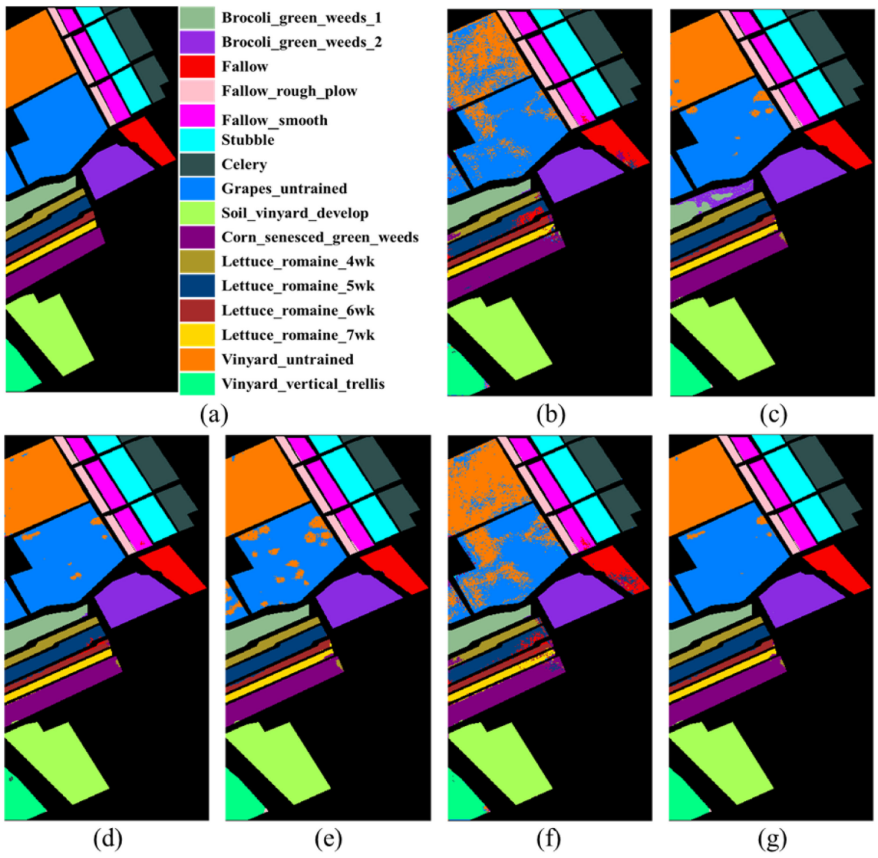


Fig. 3. Classification maps of the Salinas (IR=12.26) obtained by (a) ground truth, (b) RF (OA = 90.63%, AA = 93.88%), (c) CNN (OA = 95.23%, AA = 96.78%), (d) ECNN (OA = 98.58%, AA = 98.61%), (e) ERFS+RF (OA = 90.84%, AA = 94.77%), (f) ERFS+CNN (OA = 96.40%, AA = 98.04%), and (g) proposed EECNN (OA = **99.34**%, AA = **99.30**%), respectively.

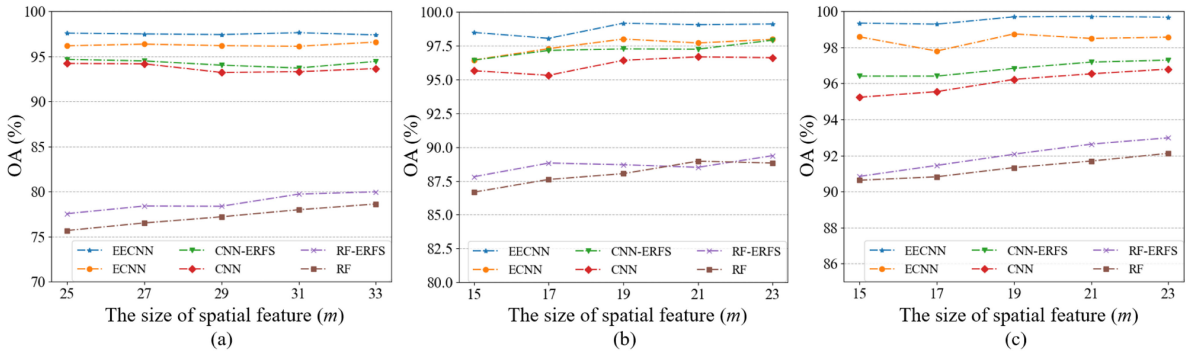


Fig. 4. Evaluation of the overall accuracy according to the size of spatial feature (m). (a) Indian Pines AVRIS. (b) University of Pavia ROSIS. (c) Salinas.

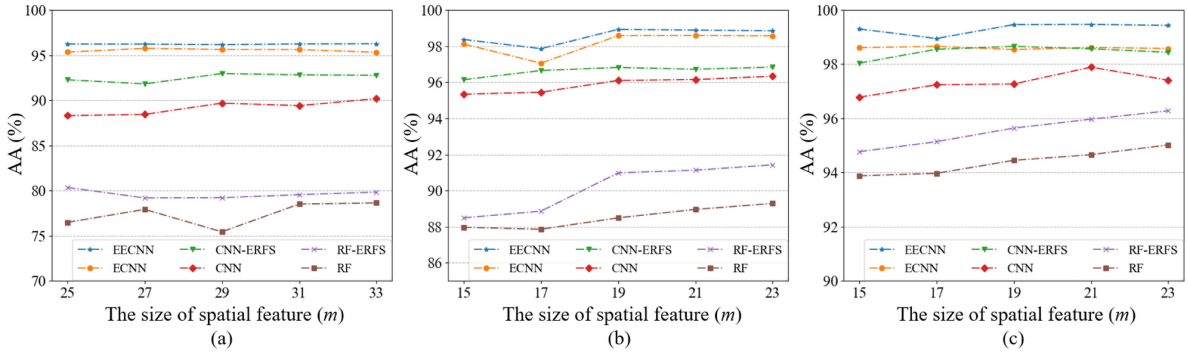


Fig. 5. Evaluation of the average accuracy according to the size of spatial feature (m). (a) Indian Pines AVRIS. (b) University of Pavia ROSIS. (c) Salinas.

V. Conclusion

In this article, a novel ERFS-based ensemble CNN (EECNN) algorithm is proposed to solve the multiclass imbalanced problem in the HSI classification. The EECNN mainly generates training samples by ROS of dynamic sampling rate, and enhances the data in the RFS to get the balanced training set, and constructs ensemble learning model with the CNN as subclassifiers for classification. The proposed method is compared with other similar algorithms (RF, CNN, ECNN, ERFS+CNN, ERFS+CNN) on three HSI datasets. The performance of this method is evaluated by several methods. The experimental results show that the proposed method provides an effective solution to the multiclass imbalanced problem of HSIs, and it is more effective than the existing algorithms.

References

[1] X. Shang and L. A. Chisholm, "Classification of australian native forest species using hyperspectral remote sensing and machine-learning classification algorithms," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2481–2489, Jun. 2014.

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

[3] K. Bakos, G. Lisini, G. Trianni, and P. Gamba, "A novel framework for urban mapping from multispectral and hyperspectral data," *Int. J. Remote Sens.*, vol. 34, no. 3, pp. 759–770, 2013.

[4] N. C. Sims, D. Culvenor, G. Newnham, and N. C. Coops, "Towards the operational use of satellite hyperspectral image data for mapping nutrient status and fertilizer requirements in Australian plantation forests," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 320–328, Apr. 2013.

[5] S. García, Z. L. 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.

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

[7] W. Feng *et al.*, "Dynamic synthetic minority over-sampling technique-based rotation forest for the classification of imbalanced hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 7, pp. 2159–2169, Jul. 2019.

[8] 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.

[9] S. Ding, B. Mirza, Z. Lin, J. Cao, and J. Sepulveda, "Kernel based online learning for imbalance multiclass classification," *Neurocomputing*, vol. 277, pp. 139–148, 2018.

[10] 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.*, 2007, pp. 127–136.

[11] 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.

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

[13] J. Mathew, M. Luo, C. K. Pang, and H. L. Chan, "Kernel-based smote for SVM classification of imbalanced datasets," in *Proc. 41st Annu. Conf. IEEE Ind. Electron. Soc.*, 2015, pp. 001127–001132.

[14] J. Li, Q. Du, Y. Li, and W. Li, "Hyperspectral image classification with imbalanced data based on orthogonal complement subspace projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 3838–3851, Jul. 2018.

[15] M. Jiang, Y. Fang, Y. Su, G. Cai, and G. Han, "Random subspace ensemble with enhanced feature for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 8, pp. 1373–1377, Aug. 2020.

[16] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, Jun., 2017, doi: 10.1145/3065386.

[17] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.

[18] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, Sep. 2019.

[19] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Comput. Sci.*, 2014, arXiv:1409.1556.

[20] V. Singhal, H. K. Aggarwal, S. Tariyal, and A. Majumdar, "Discriminative robust deep dictionary learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.,* vol. 55, no. 9, pp. 5274–5283, Sep. 2017.

[21] H. Wei, H. Yangyu, W. Li, Z. Fan, and L. Hengchao, "Deep convolutional neural networks for hyperspectral image classification," *J. Sensors*, vol. 2015, pp. 1–12, 2015.

[22] L. Ying, Z. Haokui, and S. Qiang, "Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network," *Remote Sens.*, vol. 9, no. 1, 2017, Art. no, 67, doi: 10.3390/rs9010067.

[23] J. Xia, J. Chanussot, P. Du, and X. He, "Spectral-spatial classification for hyperspectral data using rotation forests with local feature extraction and Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2532–2546, May 2015.

[24] Y. Kong, X. Wang, and Y. Cheng, "Spectral-spatial feature extraction for HSI classification based on supervised hypergraph and sample expanded CNN," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 4128–4140, Nov. 2018.

[25] A. Santara *et al.*, "BASS Net: Band-adaptive spectral-spatial feature learning neural network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5293–5301, Sep. 2017.

[26] 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.

[27] 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. Photogrammetry Remote Sens.*, vol. 105, pp. 38–53, 2015.

[28] W. Feng, G. Dauphin, W. Huang, Y. Quan, and W. Liao, "New margin-based subsampling iterative technique in modified random forests for classification," *Knowl.-Based Syst.*, vol. 182, 2019, Art. no. 104845 2019, ISSN: 0950- 482 7051, doi: 10.1016/j.knosys.2019.07.016.

[29] 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.

[30] W. Feng, W. Huang, and W. Bao, "Imbalanced hyperspectral image classification with an adaptive ensemble method based on smote and rotation forest with differentiated sampling rates," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 12, pp. 1879–1883, Dec. 2019.

[31] J. Xia, M. D. Mura, J. Chanussot, P. Du, and X. He, "Random subspace ensembles for hyperspectral image classification with extended morphological attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 4768–4786, Sep. 2015.

[32] M. P. Paing, C. Pintavirooj, S. Tungjitkusolmun, S. Choomchuay, and K. Hamamoto, "Comparison of sampling methods for imbalanced data classification in random forest," in *Proc. 11th Biomed. Eng. Int. Conf.*, 2018, pp. 1–5.

[33] J. Xia, M. Dalla Mura, J. Chanussot, P. Du, and X. He, "Random subspace ensembles for hyperspectral image classification with extended morphological attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 4768–4786, Sep. 2015.

[34] Y. Chen, L. Zhu, P. Ghamisi, X. Jia, G. Li, and L. Tang, "Hyperspectral images classification with Gabor filtering and convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2355–2359, Dec. 2017.

[35] G. Qishuo, L. Samsung, and J. Xiuping, "Hyperspectral image classification using convolutional neural networks and multiple feature learning," *Remote Sens.*, vol. 10, no. 2, 2018, Art. no. 299.

[36] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.

[37] Y. Chen, Y. Wang, Y. Gu, X. He, P. Ghamisi, and X. Jia, "Deep learning ensemble for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 6, pp. 1882–1897, Jun. 2019.

[38] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, pp. 123–140, 1996.

[39] R. E. Schapire, "The strength of weak learnability," in *Proc. 2nd Annu. Workshop Comput. Learn. Theory*, vol. 5, no. 2, 1989, pp. 197–227.
[40] D. H. Wolpert, "Stacked generalization," *Neural Netw.*, vol. 5, no. 2, pp. 241–259, 1992.
[41] A. Shahraki, M. Abbasi, and Y. Haugen, "Boosting algorithms for network intrusion detection: A comparative evaluation of real AdaBoost, gentle AdaBoost and modest AdaBoost," *Eng. Appl. Artif. Intell.*, vol. 94, 2020, Art. no. 103770.
[42] 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.

Qinzhe Lv is currently working toward the master's degree with the Department of Remote Sensing Science and Technology, School of Electronic Engineering, Xidian University, Xi'an, China.

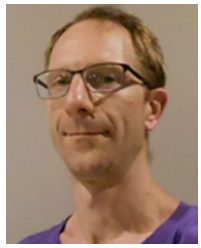
His research interests include radar signal processing, remote sensing, and machine learning

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 worked as a Postdoctoral Researcher with the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China, from 2017 to 2019. She is currently a Lecturer with the Department of Remote Sensing Science and Technology, School of Electronic Engineering, Xidian University, Xi'an, China. Her current research interests include remote sensing, machine learning, and image processing

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 Department of Remote Sensing Science and Technology, School of electronic engineering, Xidian University. His research interests include radar imaging, radar signal processing, and radar microsystem.



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

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

Lianru Gao (Senior Member, IEEE) received the B.S. degree in civil engineering from Tsinghua University, Beijing, China, in 2002, and the Ph.D. degree in cartography and geographic information system from the Institute of Remote Sensing Applications, Chinese Academy of Sciences (CAS), Beijing, China, in 2007.

He is currently a Professor with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, CAS. He has authored/coauthored more than 160 peer-reviewed papers, and there are more than 80 journal papers included by Science Citation Index. His research interests include hyperspectral image processing and information extraction.

Mengdao Xing (Fellow, IEEE) was born in Zhejiang, China, in November 1975. He received the B.S. and Ph.D. degrees in electrical engineering from Xidian University, Xi'an, China, in 1997 and 2002, respectively.

He is currently a Full Professor with the Academy of Advanced Interdisciplinary Research, Xidian University. He is also with the National Key Laboratory of Microwave Imaging Technology, Institute of Electronics, Chinese Academy of Sciences, Beijing, China. He is the author or coauthor of two books and more than 200 papers. His research interests include synthetic aperture radar (SAR), inverse SAR, and sparse signal processing.