IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 22, 2020, accepted October 3, 2020, date of publication October 13, 2020, date of current version October 27, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3030649

Hyperspectral Image Classification Using Comprehensive Evaluation Model of Extreme Learning Machine Based on Cumulative Variation Weights

YUPING YIN[1] **AND LIN WEI**[2]

[1]Faculty of Electrical and Control Engineering, Liaoning Technical University, Huludao 125105, China
[2]Department of Basic Teaching, Liaoning Technical University, Huludao 125105, China

Corresponding author: Lin Wei (29766164@qq.com)

This work was supported in part by the Science and Technology Research Project of the Education Department of Liaoning under Grant LJ2017QL021 and Grant LJ2020QNL013, and in part by the Doctoral Initiation Foundation of Liaoning Technical University under Grant 19-1026.

ABSTRACT In order to improve the classification of hyperspectral image(HSI), we propose a novel hyperspectral image classification method based on the comprehensive evaluation model of extreme learning machine(ELM) with the cumulative variation weights(CVW), referred to as ELM with the cumulative variation weights and comprehensive evaluation (CVW-CEELM). To be specific, the cumulative variation value is proposed as a new metric. The inefficient bands are eliminated by the cumulative variation quotient values based on the cumulative variation values. The cumulative variation weights based on the cumulative variation values are used to determine the contribution of each weak ELM classifier to the hyperspectral image classification algorithm. The remaining effective bands are divided by grouping strategy. In each group of the effective bands, the different numbers of bands are selected to reduce the dimension of the hyperspectral image dataset by the weighted random-selecting-based method. After dimensionality reduction, the spatial-spectral features of each pixel are extracted and multiple weak ELM classifiers are trained by the training samples. Then, the results of several weak classifiers are synthetically evaluated by the cumulative variation weights to get the final classification results. Experimental results on the typical hyperspectral image datasets illustrate that the proposed CVW-CEELM has few adjustable parameters to make the operation simple, and outperforms a variety of the image classification counterparts in terms of the calculation cost and classification accuracy.

INDEX TERMS Hyperspectral image, extreme learning machine, cumulative variation weights, comprehensive evaluation.

I. INTRODUCTION

In recent years, hyperspectral image classification has attracted much research from the remote sensing community [1]. The spectrum information of a hyperspectral image dataset is very abundant, and it enhances the ability to identify target areas in the aspect of spatial information, spectral information and radiation information [2]. Due to the rich information and characteristics, hyperspectral images have been widely used in many fields,e.g., environmental monitoring [3], rock mineral identification [4], precision agriculture [5] and military target monitoring [6]. In addition to the state-of-the-art neural networks, support vector machines and other methods can also be applied to hyperspectral image classification. However, the hyperspectral data sets have the high-dimensional data structures and bring some challenges for hyperspectral image classification. In particular, the unbalance between the limited number of training samples and the high dimensionality of the data can cause Hughes phenomenon.

With a purpose of improving the accuracy of the hyperspectral image classification, previous researchers suggested

The associate editor coordinating the review of this manuscript and approving it for publication was Joewono Widjaja.

a lot of excellent methods, e.g. principal component analysis(PCA) [7], segment autoencoder(SA) [8], singular spectrum analysis(SSA) [9], deep learning(DL) [10]. These methods not only avoids Hughes, but also improves the classification results of hyperspectral image datasets. On the other hand, support vector machines [11], kernel-based algorithms [12] and extreme learning machine (ELM) [13], [14] have been proved to be very useful for the hyper-spectral image classification. Multiple composite features and the composite features have been used to improved the classification accuracy [15]. Also, some sampling query strategies have been proposed to address the limited availability of training samples, such as semi-supervised and active learning methods [16]. Among these algorithms, a new machine learning approach that is termed the extreme learning machine has attracted a lot of attention. It is a single hidden layer feedforward neural network(SLFN). During the training process of traditional SLFNs, all the weights and biases need to be tuned iteratively, which is usually solved by gradient-based iterative techniques, e.g.,back-propagation (BP) algorithm [17]. Compared with the traditional BP neural network and support vector machine, the advantages of ELM are fast computation, few parameters, better recognition efficiency and generalization ability [15].

With these salient advantages, extreme learning machine (ELM) has attracted the attention of a large number of researchers in the area of hyperspectral image classifification and analysis over the past several years. e.g., an extreme learning machine approach is presented in the reference [18]. ELM has been successfully applied to many areas of classification in the reference [19]. Nonetheless, the existing ELM methods only use spectral information, hyperspectral image classification accuracy is not high. In order to achieve better search performance, researchers have improved their methods. Li *et al.* have proposed a method for the parameters of the extreme learning machine, which is based on differential evolution algorithm [20]. Liu *et al.* have proposed a hyperspectral image classification method that combines spatial-spectral and ELM [21]. Extreme learning machine with composite kernels for hyperspectral image classification have been proposed in the reference [22]. This improved ELM algorithm can not capture accurate spatial information, classification accuracy is still not ideal. Cao *et al.* have proposed linear vs nonlinear extreme learning machine for spectral-spatial classification of hyperspectral image in the reference [23], it works effectively, but its high computation complexity interferes with its application on HSI containing large scenes. In the reference [24], ensemble extreme learning machines for hyperspectral image classification have been proposed, howerer, E^2LMs have used upsampling to train weak learning classifier and the classification results are not optimal with limited number of training samples. A classification method based on combination of spatial-spectral features and ensemble extreme learning machines(SS-EELM) is proposed in the reference [25], several rounds of sampling are carried out based on ensemble learning theory, and several weak classifiers with poor classification accuracy are trained and then combined to build a strong classifier using majority vote method, this method has low computational complexity and it is easy to implement. Howerer, SS-EELM does not consider the contribution of each band, the algorithm does not optimize sample selection, its classification results can still be further improved.

In order to reduce the complexity of classification algorithm and construct the band grouping strategy which is more suitable for the classification of many different types of small size homogeneous areas, we propose a novel hyperspectral image classification algorithm based on the comprehensive evaluation model of extreme learning machine with the cumulative variation weights(CVW-CEELM). At present, a set of optimization criteria and algorithms have better effects in bands selection but also higher calculation complexity. Considering the retention of the original image information and the strong correlation between neighboring bands, the effective bands can be equally divided into multiple groups, in which the subsets of bands are selected randomly according to certain weights to reduce the complexity in bands selection. The proposed method, which is a suboptimal bands selection method, is used to reduce the dimension of the hyperspectral image dataset, namely the weighted random-selecting-based method. Especially, the classification results of small samples is improved. First, the cumulative variation value is proposed, and the inefficient bands can be eliminated based on the cumulative variation quotient values determined by the cumulative variation values. Second, we propose the improved grouping strategy for the remaining effective bands to increase the feature vectors of the hyperspectral image dataset, and the different numbers of bands are selected by the weighted random-selecting-based method to reduce the feature dimension. Third, after dimensionality reduction, the spatial-spectral features of each pixel are extracted and multiple weak ELM classifiers based on the supervised classification framework are trained by the training samples. Finally, the results of several weak classifiers are synthetically evaluated using the cumulative variation weights determined by the cumulative variation values to get the final classification results. Experimental results obtained from two benchmark hyperspectral datasets confirm the attractive properties of the proposed method in terms of classification accuracy and computation time.

The rest of this paper is organized as follows: Section 2 describes the ELM algorithm. Section 3 describes the CVW-CEELM algorithm. Section 4 presents the experimental results, and Section 5 provides concluding remarks.

II. EXTREME LEARNING MACHINE

The ELM is a learning algorithm for single-hidden-layer feedforward neural network. As long as the activation function can be infinitely differentiable in any real number range, the input weight and bias in ELM are randomly generated and independent of the training samples [26]. Compared with traditional BP neural network, ELM does not require bias of

the output layer, and ELM only needs to determine the weight matrix linking the hidden neurons and the output neurons. For α arbitrary different training samples (P_i, Y_i), $i = 1, 2, \cdots \alpha$, where $P_i = [x_{i1}, x_{i2}, \cdots x_{im}]^T$ is the feature matrix, and $Y_i = [t_{i1}, t_{i2}, \cdots t_{in}]^T$ is the label information corresponding to P_i. The relationship between input and output of ELM can be expressed as [27]

$$\sum_{j=1}^{n} \beta_i g\left(\omega_{ij} \cdot P_i + b_j\right) = o_i \quad i = 1, 2, \cdots, m; j = 1, 2, \cdots, n \tag{1}$$

where β_j is the weight vector linking the jth hidden neuron and all the output neurons. b_j is the jth bias. o_i is the ith sample output vector. ω_{ij} is the weight value linking the jth hidden neuron and the ith input neuron. $g(\cdot)$ is the sigmoid function. $\beta = (\beta_1, \beta_2, \cdots, \beta_n)^T$ is the weight matrix linking the hidden neurons and the output neurons. $Y = (Y_1, Y_2, \cdots, Y_m)$ is the label matrix. The optimization objective function of ELM is as follows.

$$\underset{\beta}{Min} \|H\beta - Y\| \tag{2}$$

where the hidden-layer output matrix is expressed by H, as shown at the bottom of the page.

In order to determine the optimal parameters of the objective function, the matrix H can be derived once the parameter pair (ω_j, b_j) is fixed, and then training ELM means finding the least-squares solution $\hat{\beta}$ of the $H\beta = Y$, i.e., $\left\|H\hat{\beta} - Y\right\| = Min \ \|H\beta - Y\|$. $\hat{\beta}$ can be expressed as

$$\hat{\beta} = H^{+}Y \tag{3}$$

where H^+ is the Moore-Penrose generalized inverse of matrix H. When the parameters are determined, formula(1)can be used to calculate the output function of ELM.

III. PROPOSED CVW-CEELM

Considering the characteristics of the hyperspectral image dataset, the effective dimensionality reduction method is proposed, and then the corresponding classification method is studied for HSI dataset. The proposed CVW-CEELM is described as follows.

A. NORMALIZATION

Due to a lot of noise information in HSI dataset, normalization of HSI dataset should be adopted to reduce the high signal-to-noise ratio of HSI. The normalization of HSI dataset is expressed as

$$Nor(x_{pq}^i) = 255 \times \frac{x_{pq}^i - \min(x^i)}{\max(x^i) - \min(x^i)} \tag{4}$$

where x_{pq}^i is the ith band gray value in row p, column q. x^i is the ith band gray value matrix.

B. BAND CUMULATIVE VARLATION FUNCTION

The information variation value can reflect the data dispersion degree when analyzing the data. Compared with the deviation of data set, the measurement scale has little influence on the information variation value. Based on the above advantage, the information variation value is improved for dimensionality reduction of HSI dataset and comprehensive evaluation in this paper. We propose a new metric named cumulative variation value with the intention of reflecting the data dispersion degree. The corresponding concepts and operations of the cumulative variation weights based on cumulative variation value are described as follows.

The cumulative variation value of the intraclass is expressed as

$$CN_{ik}(t) = \frac{\sum_{j=1}^{t-1} |C_{ik}(j+1) - C_{ik}(j)|}{\bar{C}_{ik}(t)}, 2 \le t \le T_{ik}, \quad 1 \le k \le CLN \tag{5}$$

$$\bar{C}_{ik}(t) = \frac{1}{t} \sum_{j=1}^{t} C_{ik}(j) \tag{6}$$

where $CN_{ik}(t)$ is the cumulative variation value of the kth category of the ith band. $\bar{C}_{ik}(t)$ is the sample mean of the kth category of the ith band. $C_{ik}(j)$ is the jth sample of the kth category of the ith band. T_{ik} is the sample size of the kth category of the ith band. CLN is the number of categories.

If the cumulative variation value of the intraclass is close to zero, it shows that there is no significant difference in gray values of the same category of a band, and the classification effect is better by using this band. If the cumulative variation value of the intraclass is away from zero, it shows that the gray values of the same category of a band differ significantly, and the classification effect is poor by using this band. The cumulative variation value of the interclass is expressed as

$$CZ_i(t) = \frac{\sum_{j=1}^{t-1} |C_i(j+1) - C_i(j)|}{\bar{C}_i(t)}, 2 \le t \le T_i \tag{7}$$

$$\bar{C}_i(t) = \frac{1}{t} \sum_{j=1}^{t} C_i(j) \tag{8}$$

$$H = \begin{bmatrix} g(\omega_1 \cdot P_1 + b_1) & g(\omega_2 \cdot P_1 + b_2) & \cdots & g(\omega_n \cdot P_1 + b_n) \\ g(\omega_1 \cdot P_2 + b_1) & g(\omega_2 \cdot P_2 + b_2) & \cdots & g(\omega_n \cdot P_2 + b_n) \\ \vdots & \vdots & & \vdots \\ g(\omega_1 \cdot P_m + b_1) & g(\omega_2 \cdot P_m + b_2) & \cdots & g(\omega_n \cdot P_m + b_n) \end{bmatrix}_{m \times n}$$

where $CZ_i(t)$ is the cumulative variation value of the ith band. $\bar{C}_i(t)$ is the sample mean of the ith band. $C_i(j)$ is the jth sample of the ith band. T_i is the sample size of the ith band. $T_i = \sum_{k=1}^{CLN} T_{ik}$.

The cumulative variation value of the interclass can represent the difference between different classes in the same band. If the cumulative variation value of the interclass is close to zero, it shows that there is no significant difference in gray values of the different classes in the same band, and the classification effect is poor by using this band. If the cumulative variation value of the interclass is away from zero, it shows that the gray values of the different classes in the same band differ significantly, and the classification effect is improved by using this band.

C. DIMENSIONALITY REDUCTION

Actually, the gray values of the intraclass are quite different because some pixels may contain multiple ground objects. Therefore, the related variation information of the interclass and the related variation information of the intraclass should be comprehensively considered by the cumulative variation quotient. Meanwhile, the cumulative variation weights can be constructed by the cumulative variation quotient.

The cumulative variation quotient of a band is expressed as

$$F(i) = \frac{\|CZ_i(t)\|_2}{\sum_{k=1}^{CLN} \|CN_{ik}(t)\|_2}, 1 \le i \le BN \tag{9}$$

where $F(i)$ is the cumulative variation quotient of the ith band. BN is the number of bands.

The norms used above are expressed as

$$\|CZ_i(t)\|_2 = \left(\int_1^{T_i} (CZ_i(t))^2 \, dt \right)^{1/2} \tag{10}$$

$$\|CN_{ik}(t)\|_2 = \left(\int_1^{T_{ik}} (CN_{ik}(t))^2 \, dt \right)^{1/2} \tag{11}$$

When there is significant difference between the interclass and the intraclass, the cumulative variation quotient is away from 1, and the classification effect is better by using this band. Otherwise, the cumulative variation quotient is close from 1, and the classification effect is poor by using this inefficient band.

An important point needs to be added as follows. The arrangement order of training samples may change. This change may affect the distribution of the cumulative variation values of the intraclass, but the cumulative variation values of the interclass also change. The cumulative variation quotient value expresses the difference between them as a whole. Therefore, there is no illconditioned problem in the cumulative variation quotient value about the arrangement order of training samples.

Due to the strong correlation between neighboring bands, the weighted random-selecting-based method is proposed to reduce the dimension of the hyperspectral image dataset. First, the inefficient bands can be eliminated and average group is generated for the remaining effective bands. When eliminating inefficient bands, three aspects need to be noted as follows: If the cumulative variation quotient value of a band is close from 1, this band can be eliminated; Loss of original information should be minimized; The number of inefficient bands is determined by the needs of the grouping strategy. Second, the cumulative variation quotient is used to calculate the weight of each group. Finally, the weights are used to randomly select the bands in each group.

The band weights of each group are expressed as

$$\rho_L = \frac{\sum F_L(i)}{\sum F(i)}, 1 \le L \le GN \tag{12}$$

where ρ_L is the weight of Lth group. $\sum F_L(i)$ is the sum of cumulative variation quotient of the Lth group. $\sum F(i)$ is the sum of the cumulative variation quotient of all bands. GN is the number of groups.

The result of multiplying ρ_L with the required number is the number of randomly selected bands in each group. In practice, the number of selected bands should not exceed the total number of bands in the group, and all selected bands should retain at least 50% of the valid original information according to the experimental experience in reference [25].

D. SPATIAL-SPECTRAL FEATURE

Considering the strong correlation between neighboring bands [28], [29], the spatial-spectral features are used to classify the HSI dataset after dimensionality reduction based on the cumulative variation quotient. x_{pq}^{spe} is the spectral feature of a given pixel, and x_{pq}^{spa} is the spatial feature of a given pixel. The spatial feature of a given pixel is expressed as

$$x_{pq}^{spa} = \frac{1}{QN} \sum_{(p,q) \in Q(p,q)} x_{pq}^{spe} \tag{13}$$

where $Q(p, q)$ is the squared neighborhood of a given pixel. QN is the number of samples in the squared neighborhood $Q(p, q)$.

The spatial-spectral feature of a given pixel is expressed as

$$x_{pq} = \gamma x_{pq}^{spe} + (1 - \gamma) x_{pq}^{spa}, \gamma \in [0, 1] \tag{14}$$

where x_{pq} is the spatial-spectral feature of a given pixel. γ can be determined through numerical experiments.

E. COMPREHENSIVE EVALUATION MODEL OF ELM BASED ON CVW

After dimensionality reduction, the spatial-spectral features of each pixel is extracted and multiple weak ELM classifiers are trained by the training samples. Finally, the several weak classifiers with the cumulative variation weights are synthetically evaluated to obtain the final classification results.

The following formula is used to calculate the cumulative variation weight of each classifier.

$$\varphi_{LS_w} = \sum_{i \in LS_w} F(i) \tag{15}$$

$$\eta_w = \frac{\varphi_{LS_w}}{\sum \varphi_{LS_w}}, 1 \le w \le CN \tag{16}$$

where w is the classifier label. CN is the number of weak classifiers. LS_w is the selected band set of the wth weak classifier. φ_{LS_w} is the sum of cumulative variation quotient of the wth weak classifier. η_w is the cumulative variation weight of the wth weak classifier.

The comprehensive evaluation of weak classifiers can be expressed as

$$result_w = [o_w(1), o_w(2), \cdots, o_w(CLN)]^T \tag{17}$$

$$\text{Re}sult = \arg\max \left\{ \sum_{w=1}^{CN} \eta_w \cdot result_w \right\} \tag{18}$$

where $result_w$ is the output vector of the wth weak classifier. $o_w(i)$ is the ith component of the output vector. Re$sult$ is the final comprehensive evaluation result.

F. ALGORITHM PROCEDURES

The complete algorithm procedures are described as follows. The corresponding flowchart of CVW-CEELM is shown in Fig. 1.

IV. EXPERIMENTAL RESULTS

In this section, two well-known HSI datasets (Indian Pines and Pavia University scene) are used to verify the effectiveness of the proposed method. SVM [30],ELM [31],GELM [18],KELM,GELM-CK,KELM-CK [21],SS-EELM [25] serve as the benchmark models to verify the proposed CVW-CEELM. In addition, the general Gaussian radial basis function kernel for SVM,KELM and KELM-CK is adopted as $K(x_i, x_j) = \exp(-\|x_i - x_j\|^2/2\sigma^2)$, where $\sigma = 2^q$, $q \in$ {-4, -3, …, 4}, and the optimal penalty parameter C and kennel width σ are determined through the grid search algorithm [21]. Eight independent experiments are conducted, and class accuracy(CA), average accuracy(AA), the average values of overall accuracy (OA), kappa coefficient and calculation cost are used to evaluate the classification results.

A. INDIAN PINES

This dataset was acquired in 1992 over the Indian Pines test site in Northwestern Indiana by the AVIRIS sensor. This dataset consists of 16 land cover classes. Indian Pines compose of 145 × 145 pixels and 220 spectral bands, the wavelength range is from $0.4\mu m$ to $2.5\mu m$. For Indian Pines images, 10% of the labeled samples are randomly selected for training, and the remaining samples are used for testing. All models use the same training samples. Table 1 shows the number of the training samples, the number of testing samples and classification results of all comparison models. Fig. 2 shows the false color image and the ground-truth map of the Indian Pines.

To describe the role of cumulative variation quotient values, bands 173 and 220 are chosen as representatives of all bands. The cumulative variation curve of band 173 is shown in Fig. 3. Compared with the cumulative variation values of the interclass,the fluctuation amplitude of the cumulative variation values of the intraclass are obviously larger, and the cumulative variation quotient value of band 173 is 3.4559. This result indicates that the differences among the 16 classes based on band 173 are significant, and band 173 is favorable for HSI classification. The band 173 should be retained for HSI classification. The cumulative variation curve of band 220 is shown in Fig. 4. The fluctuation amplitude of the cumulative variation values of the intraclass are close to that of the cumulative variation values of the interclass, and the

Algorithm 1 CVW-CEELM

Input: Training set $X_{trs} = \{(P_i, Y_i) \,|i = 1, \cdots, \alpha\}$; S: All samples; Testing set $X_{tes} = S - X_{trs}$; CN: The number of weak classifiers.

Output: Classification result

Step 1) Normalization for HSI according to (4).

Step 2) Calculate the cumulative variation quotient values for all bands.

Step 2.1) Calculate the cumulative variation values of the intraclass according to (5)-(6). ($CN_{ik}(t)$)

Step 2.2) Calculate the cumulative variation values of the interclass according to (7)-(8). ($CZ_i(t)$).

Step 2.3) Calculate the $F(i)$, $1 \le i \le BN$ for all bands according to (9)-(11).

Step 3) Eliminate inefficient bands base on $F(i)$.

Step 4) Reduce the dimension for each ELM classifier.

Step 4.1) The effective bands are equally divided into GN groups.

Step 4.2) Calculate the weight of each group according to (12). (ρ_L, $1 \le L \le GN$).

Step 4.3) In each group, the different numbers of bands can be determined to reduce the dimension of HSI by ρ_L, $1 \le L \le GN$.

Step 5) for i=1 to CN. do the following:

Step 5.1) Generate random label information of bands in each group based on step 4.3) to form a new band label vector.

Step 5.2) Generate the spatial-spectral features based on the new band label vector according to (13)-(14).

Step 5.3) Feed the spatial-spectral features of X_{trs} to ELM to train a classifer: $\hat{\beta} = H^+ Y$ according to (3).

end for

Step 6) Establish the comprehensive evaluation model of ELM

Step 6.1) Calculate the cumulative variation weights η_w, $1 \le w \le CN$ for all ELM classifiers according to (15)-(16).

Step 6.2) Generate the classification result from CN classification results of each test sample according to(17)-(18).

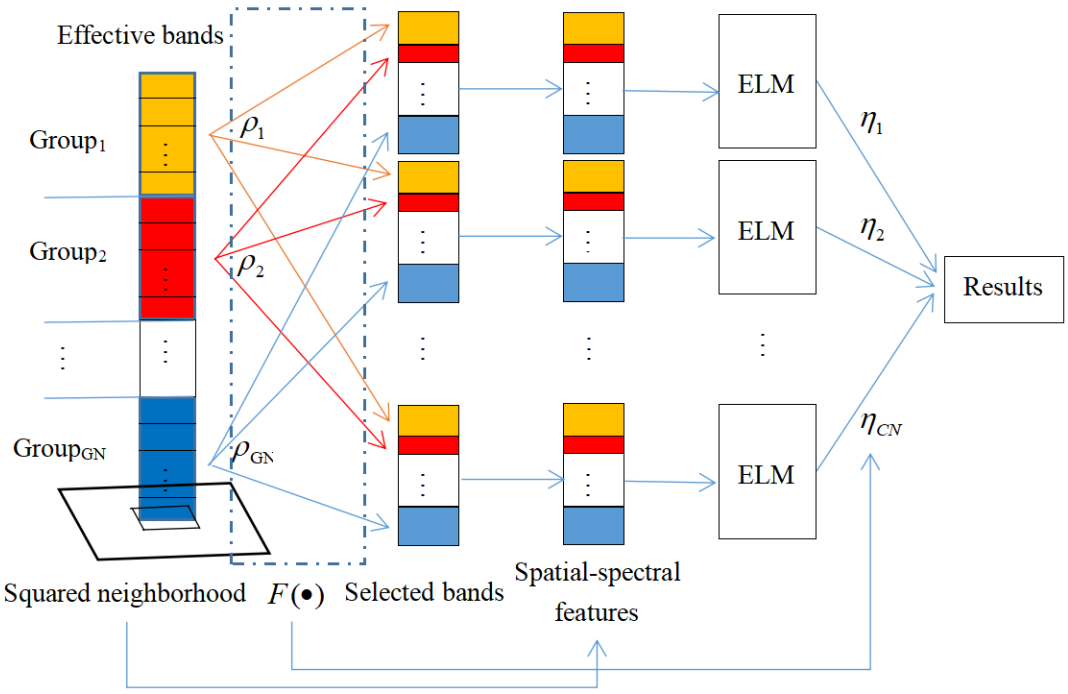

FIGURE 1. The overall flowchart of CVW-CEELM.

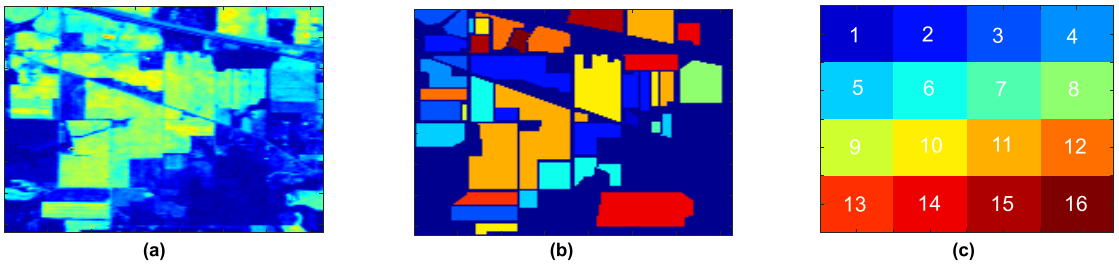

FIGURE 2. False color image, ground-truth map and corresponding colors based on band 173 of Indian Pines. ((a) false color image; (b) ground-truth map; (c) corresponding colors).

cumulative variation quotient value of band 220 is 1.1592 which is close to 1. This result indicates that the differences among the 16 classes based on band 220 are not significant, and band 220 is not favorable for HSI classification. Therefore band 220 should be eliminated. The cumulative variation quotient values of all 220 bands are also calculated and shown in Fig. 5.

The twenty-four bands whose cumulative variation quotient values are close to 1 should be eliminated. The twenty-four eliminated bands are clearly marked in red, as shown in Fig. 5. Table 2 lists the cumulative variation quotient values of the twenty-four eliminated bands. From the experimental results, the twenty-four eliminated bands (bands 1, 88, 104-109, 150-163, 219-220) include exactly twenty bands (bands 104-108, 150-163, 220) affected by water vapor noise. This result suggests that the proposed method based on the cumulative variation quotient is effective. The remaining 196 bands are effective bands for HSI classification.

In order to avoid missing some effective bands in the process of dimensionality reduction, the effective bands are

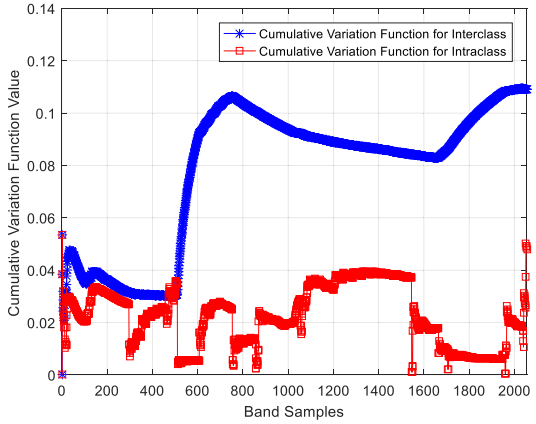

FIGURE 3. Cumulative variation curve of band 173.

divided into fourteen groups of fourteen bands on average. The number of weak ELM classifiers is set as 14. The average grouping is shown in Fig. 6. Grouping strategy is performed for the remaining effective bands according to step 3), and

TABLE 1. Classification results obtained for the Indian pines dataset.

/%

NO.	Class	Train	Test	SVM	ELM	GELM	KELM	GELM_CK	KELM_CK	SS-EELM	CVW-CEELM
1	Alfalfa	5	41	72.9±8.69	45.5±12.8	26.9±17.9	71.8±10.9	96.2±4.37	95.5±12.0	96.0±3.50	**97.5±3.51**
2	Corn-notill	143	1285	84.9±1.47	74.8±3.57	80.8±1.16	84.1±1.81	97.2±1.37	97.1±0.66	97.2±0.65	**98.3±0.66**
3	Corn-mintill	83	747	77.1±4.99	59.1±2.38	58.2±3.25	71.7±4.16	97.5±1.59	98.8±0.60	98.2±1.45	**99.1±1.43**
4	Corn	24	213	77.1±6.70	41.3±5.79	38.9±7.44	67.2±6.13	96.7±2.58	95.8±2.91	96.3±0.86	**98.6±0.87**
5	Grass-pasture	48	435	93.1±1.25	85.4±4.28	88.8±1.96	92.2±1.98	97.3±1.41	**97.7±1.54**	97.4±1.64	96.8±1.65
6	Grass-trees	73	657	96.4±1.28	94.2±1.37	96.7±0.99	96.6±1.43	**99.5±0.37**	99.3±0.27	99.3±0.37	99.3±0.36
7	Grass-pasture-mowed	3	25	79.9±9.68	11.2±3.44	6.09±8.01	64.9±16.2	86.9±9.58	94.5±11.8	93.1±8.96	**94.8±8.88**
8	Hay-windrowed	48	430	98.1±0.88	98.5±0.26	99.6±0.18	98.9±0.45	99.6±0.18	99.7±0.27	99.6±0.11	**99.8±0.12**
9	Oats	2	18	73.2±20.3	10.1±6.97	3.81±5.34	57.1±20.2	57.1±26.3	49.5±20.8	54.9±16.3	**77.5±16.2**
10	Soybean-notill	97	875	78.8±4.96	65.6±4.58	65.4±2.58	80.1±3.09	96.3±1.49	96.4±0.79	96.4±1.52	**97.8±1.52**
11	Soybean-mintill	246	2209	86.2±1.67	76.9±1.50	80.4v2.28	86.2±2.43	98.7±0.41	98.7±0.48	98.6±0.37	**99.3±0.38**
12	Soybean-clean	59	534	83.5±2.59	64.7±3.61	75.6±4.62	81.7±3.09	97.4±1.11	97.2±1.42	97.3±0.68	**99.1±0.69**
13	Wheat	21	184	98.7±0.91	98.2±0.88	99.5±0.26	99.1±0.52	**99.8±0.12**	99.8±0.13	99.1±0.60	98.8±0.59
14	Woods	127	1138	96.0±1.69	91.8±1.90	95.7±1.34	96.3±1.15	99.3±0.47	99.7±0.27	99.4±0.09	**99.7±0.09**
15	Buildings-Grass-Trees-Drives	39	347	58.3±4.04	59.4±6.10	62.9±3.61	64.0±4.74	95.7±2.61	98.7±2.56	96.5±0.26	**99.6±0.25**
16	Stone-Steel-Towers	9	84	90.1±4.85	39.8±11.2	68.8±9.98	80.5±4.81	93.3±3.62	**94.8±10.3**	93.4±5.06	93.5±5.06
	OA			86.3±0.58	76.2±0.79	79.4±0.65	85.5±0.33	97.9±0.15	98.1±0.27	98.0±0.25	**98.5±0.25**
	AA			84.0±1.64	63.5±1.14	65.5±1.82	80.8±2.01	94.3±1.36	94.6±1.22	94.5±1.23	**96.9±1.23**
	k			84.5±0.81	72.6±0.87	76.5±0.79	83.4±0.36	97.5±0.19	97.8±0.33	97.6±0.27	**97.9±0.28**
	Time(S)			51.6±1.49	0.23±0.01	0.90±0.06	4.98±0.07	2.47±0.05	16.0±0.13	36.5±0.03	15.4±0.02

the best metric values are in bold

TABLE 2. Cumulative variation quotient values of eliminated bands and corresponding bands.

Band	Cumulative variation quotient values	Band	Cumulative variation quotient values
160	0.9794	161	1.0401
158	0.9907	150	1.0422
159	1.0072	106	1.0449
157	1.0074	105	1.0512
156	1.0185	107	1.0620
104	1.0187	153	1.0700
108	1.0200	163	1.0768
155	1.0252	220	1.1592
151	1.0267	1	1.1597
162	1.0314	219	1.1862
154	1.0316	109	1.1969
152	1.0353	88	1.2135

a certain number of bands are selected by the weighted random-selecting-based method to reduce the dimension of hyperspectral image dataset according to step 4). All selected information should retain at least 50% of the valid original information. At least ninety-eight bands are selected as the new band label vector. The spatial-spectral features are generated based on the new band label vector according to (13)-(14); Meanwhile the spatial-spectral features of training set are fed to ELM to train a classifer. All classifiers can be established according to step 5). The number of bands selected for each group is shown in Table 3. According to Table 3 and Fig.6, if the weight is larger based on the cumulative variation quotient values, this group of bands is conducive to HSI classification, and the more bands are selected in this group. On the other hand, the weight is smaller based on the cumulative variation quotient values, fewer bands are selected in this group. This grouping strategy increases the feature vectors of each sample to enrich the description of small samples, as shown in Fig. 1.

In order to extract spatial-spectral features, the width of neighborhood window is set to 3. Next, we discuss the impact of the number of hidden nodes and γ on the HSI classification accuracy based on CVW-CEELM. We investigate the impact of the number of hidden nodes on the classification performance of CVW-CEELM. FIG. 7 depicts OAs curves

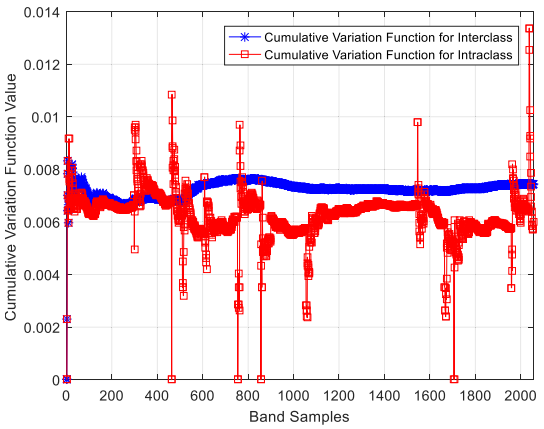

FIGURE 4. Cumulative variation curve of band 220.

versus the number of hidden nodes of CVW-CEELM, where the parameter is $\gamma = 0.1$. When the number of hidden nodes of CVW-CEELM equals 400, the best OA can be obtained for Indian Pines dataset. On the other hand, the number of hidden nodes of SS-EELM is set to 500 and other parameters are set according to reference [25]. The number of hidden nodes of GELM-CK and KELM-CK is set to 1000, and other parameters are set according to reference [21].

TABLE 3. The number of bands selected by the weighted random-selecting-based method for Indian pines.

NO.	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Weight	0.0677	0.0740	0.0570	0.0572	0.0518	0.0438	0.0406	0.0771	0.0836	0.0841	0.0857	0.0983	0.0953	0.0837
Number	7	7	6	6	5	4	4	8	8	8	8	10	9	8

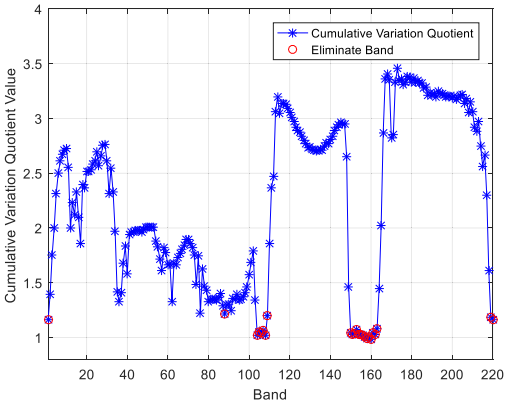

FIGURE 5. Cumulative variation quotient curve of Indian pines.

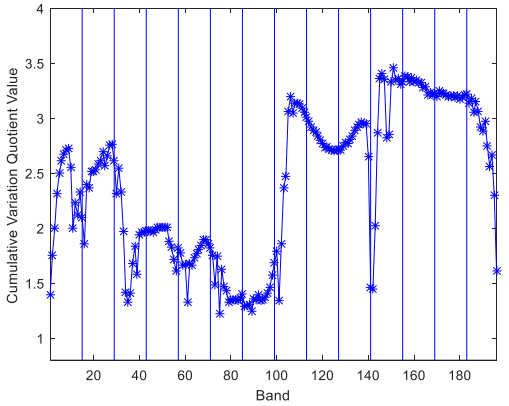

FIGURE 6. Average grouping of Indian pines.

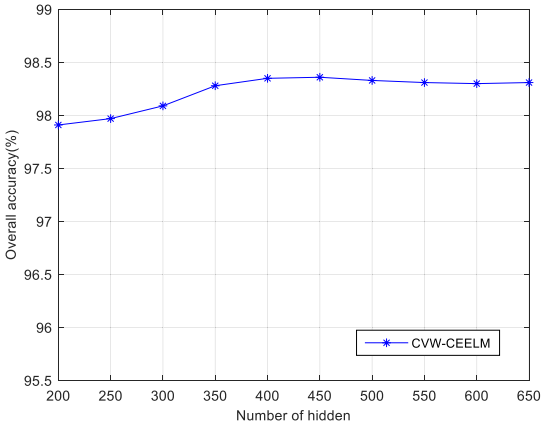

FIGURE 7. The relation between the OAs of Indian Pines and the number of hidden nodes.

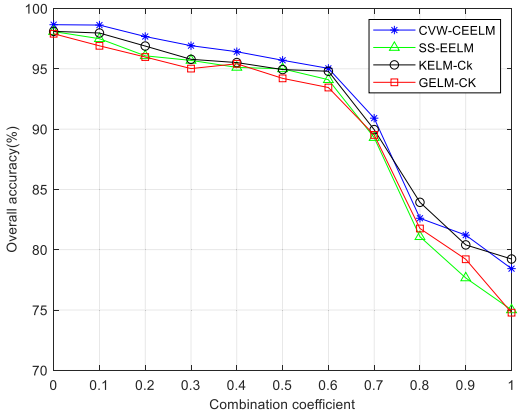

FIGURE 8. The relation between the OAs of Indian Pines and the combination coefficient γ.

In the following experiments, we explore the impact of γ on the classification performance of CVW-CEELM. FIG. 8 shows the OAs curves versus γ for the GELM-CK, KELM-CK, SS-EELM and CVW-CEELM, while other parameters have been determined in the previous experiment. Referring to this figure, the proposed CVW-CEELM outperforms the GELM-CK, KELM-CK and SS-EELM when γ is sufficiently large. For example, when $1 - \gamma = 0.9$, the CVW-CEELM achieves an OA of 98.53% while the GELM-CK, KELM-CK and SS-EELM accomplish the OAs of 96.69%, 97.91% and 97.10% respectively. For the convenience of experimental analysis, γ of the above algorithms is assumed to be 0.1 unless otherwise stated. The mean of the five performance metrics corresponding to the above eight algorithms are listed in Table 1. Moreover, the classification maps of the SVM, ELM, GELM, KELM, GELM-CK, KELM-CK, SS-EELM and CVW-CEELM for Indian Pines corresponding to Table 1 are shown in FIG. 9.

Under the conditions of adopting the same effective bands and training samples, the analysis results for Indian Pines dataset are as follows.

The first four spectral-feature-based models(SVM,ELM, GELM and KELM) only use the spectral features. Although, SVM and KELM increase the calculation cost due to the introduction of the kernel function, SVM and KELM have better classification effect than ELM and GELM. As observed from the Table 1, the last four spatial-spectral-feature-based models (GELM-CK,KELM-CK,SS-EELM and CVW-CEELM) show better classification performance than the first four models. Through the further analysis of the last four models, we find that the classification accuracy of GELM-CK is not as good as KELM-CK,SS-EELM and CVW-CEELM. KELM-CK model consumes time cost up to 16.0 seconds. Although SS-EELM model based on the ensemble extreme learning machines does not introduce the kernel functions, SS-EELM also has large calculation cost, whereas it is equal to 36.5 seconds. The proposed CVW-CEELM, which adopts the suboptimal bands selection method, only need 15.4 seconds to do so. OA of the proposed CVW-CEELM is 98.5%;

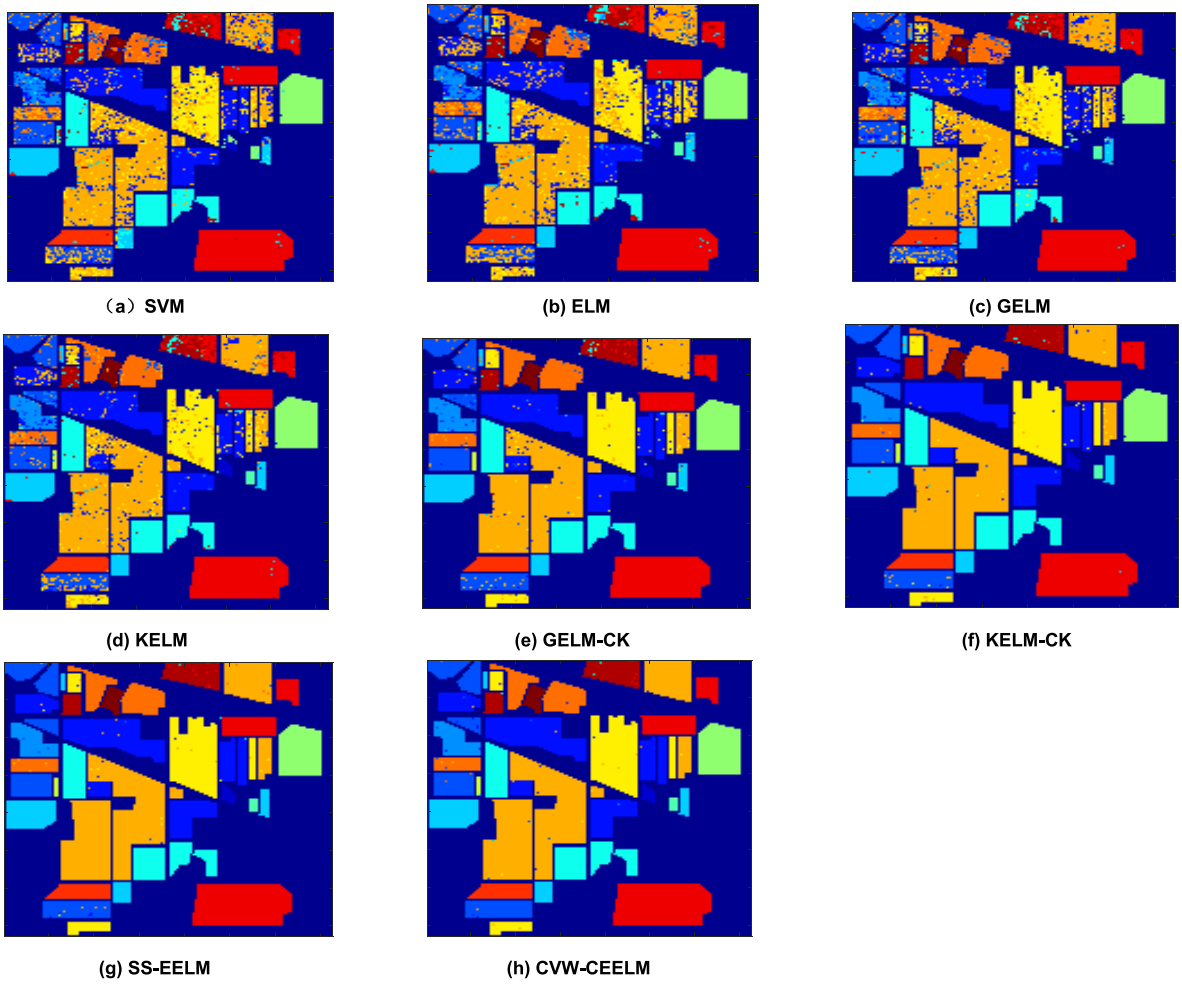

FIGURE 9. Classification maps of the SVM,ELM,GELM,KELM,GELM-CK,KELM-CK,SS-EELM and CVW-CEELM for the Indian Pines.

OA of KELM-CK is 98.1%; OA of SS-EELM is 98.0%; OA of GELM-CK is 97.9%. Especially for many different types of small-size homogeneous areas in Indian Pines dataset, the proposed CVW-CEELM outperforms other classification models. As shown in Table 1 and FIG. 9, the classification accuracies of Alfalfa, Grass-pasture-mowed and Oats are the highest using CVW-CEELM compared with other models. Although, the classification accuracy of Stone-Steel-Towers is the highest using KELM-CK, but the time cost of KELM-CK is higher than CVW-CEELM. Through the analysis above, the proposed CVW-CEELM can effectively improve the classification performance from two aspects of the time cost and classification accuracy.

B. PAVIA UNIVERSITY

The Pavia University scene images were acquired by Reflective Optics System Imaging Spectrometer optical sensor. Each band of the Pavia university has the size of 610 × 340, the wavelength range is from $0.43\mu m$ to $0.86\mu m$. The 12 noisy and water absorption bands are removed, and the remaining 103 bands are used for classification. This dataset consists of 9 land cover classes. Fig. 10 shows the false color image and the ground-truth map of the Pavia University scene. The Pavia University scene has a large number of samples of various types, on the assumption that ensure accuracy of classification, 4% of the labeled samples are randomly selected for training to reduce operation time, and the remaining samples are used for testing. All models use the same training samples. Table 4 shows the number of the training samples, the number of testing samples and classification results of all comparison models.

The Pavia University dataset is processed in the same way as experiment 1. The cumulative variation quotient values of all bands are shown in Fig. 11. The three bands whose cumulative variation quotient values are close to 1 are eliminated. The three eliminated bands(bands 68-70) are clearly marked in red, as shown in Fig. 11. The remaining 100 bands are effective bands for HSI classification. The average grouping is shown in Fig. 12. Grouping strategy is performed for the remaining effective bands, and a certain number of bands are selected by the weighted random-selecting-based

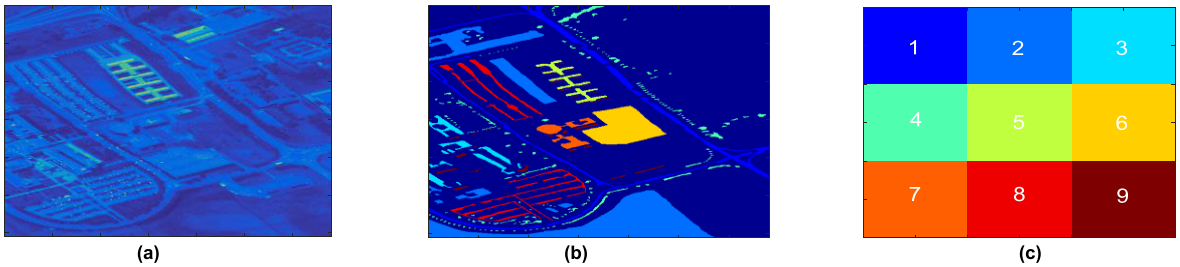

FIGURE 10. False color image based on band 11, ground-truth map and corresponding colors for the Pavia University ((a) false color image; (b) ground-truth map; (c) corresponding colors).

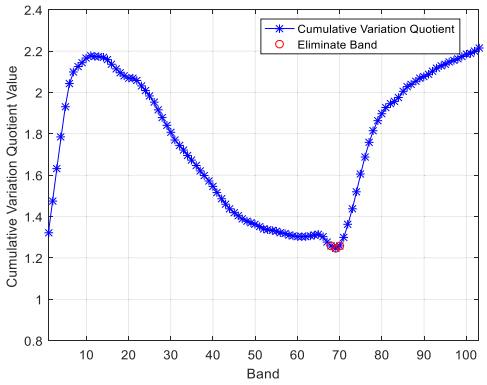

FIGURE 11. Cumulative variation quotient curve of Pavia University.

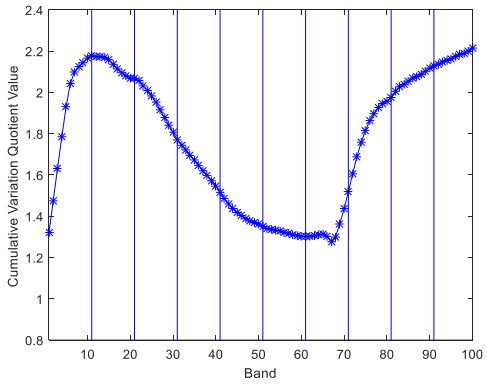

FIGURE 12. Average grouping of Pavia University.

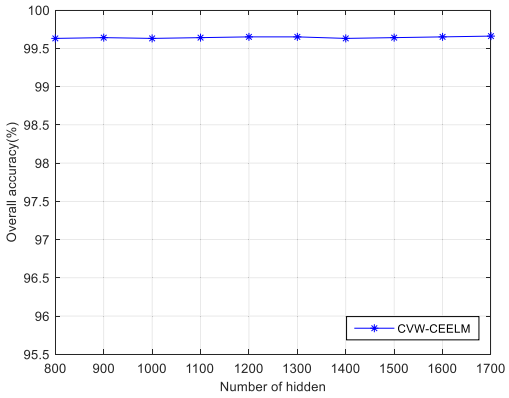

FIGURE 13. The relation between the OAs of Pavia University and the number of hidden nodes.

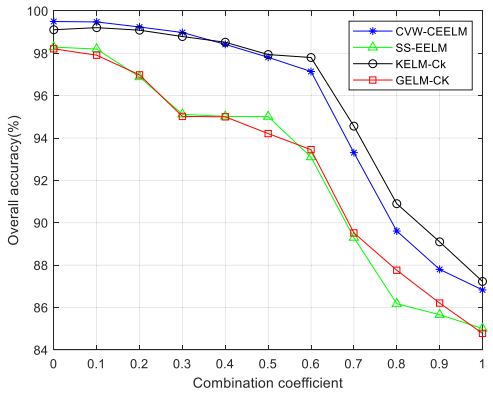

FIGURE 14. The relation between the OAs of Pavia University and the combination coefficient γ.

method to reduce the dimension of the Pavia University dataset. All selected information should retain at least 50% of the valid original information, and at least fifty bands are selected as the new band label vector according to step 4). The spatial-spectral features are generated based on the new band label vector according to (13)-(14); Meanwhile the spatial-spectral features of training set are fed to ELM to train a classifer. All classifiers can be established according to step 5). The number of bands selected for each group is shown in Table 5. The effective bands are divided into ten groups of ten bands on average and the number of weak ELM classifiers is set as 10.

In order to extract spatial-spectral features, the width of neighborhood window is set to 3. We investigate the impact of the number of hidden nodes on the classification performance of CVW-CEELM. FIG. 13 depicts OAs curves versus the number of hidden nodes of CVW-CEELM, where the parameter is $\gamma = 0.1$. When the number of hidden nodes of CVW-CEELM equals 900, the best OA can be obtained for the Pavia University dataset. For this reason, the number of hidden nodes of CVW-CEELM is set as 900. The parameters of other algorithms are similar to experiment 1.

After the number of hidden nodes is determined, we explore the impact of γ on the classification performance of CVW-CEELM. FIG. 14 shows the OAs curves versus γ for the GELM-CK, KELM-CK, SS-EELM and CVW-CEELM. When $1 - \gamma = 0.9$, the CVW-CEELM achieves an OA of 99.63% while the GELM-CK, KELM-CK and SS-EELM accomplish the OAs of 98.10%, 99.21% and 97.91% respectively. Referring to FIG. 14, the proposed

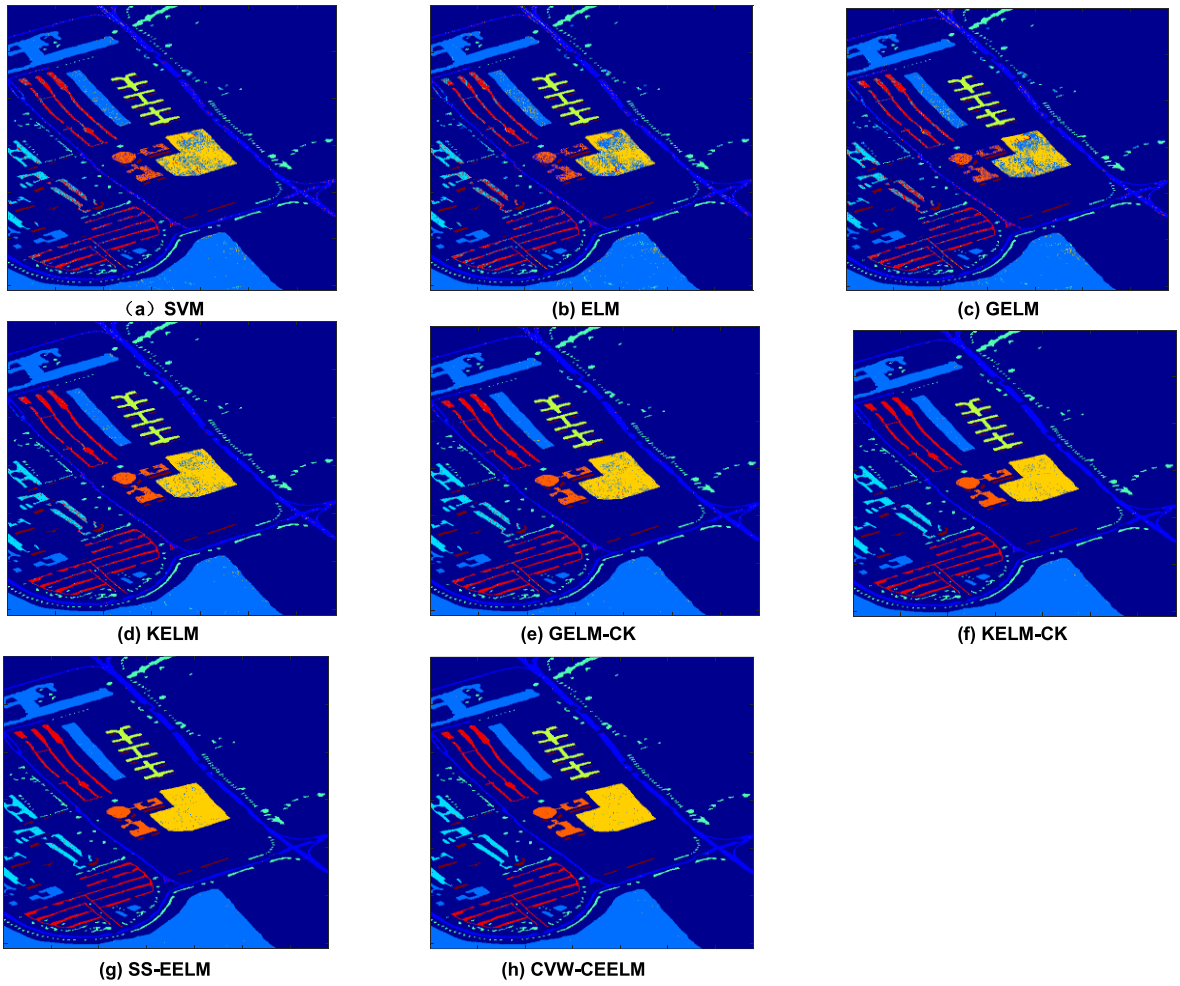

FIGURE 15. Classification maps of the SVM,ELM,GELM,KELM,GELM-CK,KELM-CK,KELM-CK,SS-EELM and CVW-CEELM for the Pavia University.

CVW-CEELM outperforms the GELM-CK, KELM-CK and SS-EELM when γ is sufficiently large. For the convenience of experimental analysis, γ of the above algorithms is assumed to be 0.1. The mean of the five performance metrics corresponding to the above eight algorithms are listed in Table 4. Moreover, the classification maps of the SVM, ELM, GELM, KELM, GELM-CK, KELM-CK, SS-EELM and CVW-CEELM for the Pavia University dataset corresponding to Table 4 are shown in FIG. 15.

Under the conditions of adopting the same effective bands and training samples, the analysis results for Pavia University dataset are as follows.

As observed from the Table 4, the last four spatial-spectral-feature-based models(GELM-CK, KELM-CK, SS-EELM and CVW-CEELM) show better classification performance than the first four models(SVM, ELM, GELM and KELM). Due to the large number of samples in the Pavia University dataset, the time cost of Pavia University dataset is higher than the time cost of Indian Pines dataset for all models. GELM-CK model consumes time cost up to 17.2 seconds; KELM-CK model consumes time cost up to 313 seconds; SS-EELM model consumes time cost up to 88.6 seconds; The proposed CVW-CEELM need 64.2 seconds. Due to the simple model structure of the proposed method, the time cost of the CVW-CEELM is smaller than that of KELM-CK and SS-EELM. Althouth GELM-CK model consumes the least time compared with KELM-CK, SS-EELM and CVW-CEELM, the classification accuracy of GELM-CK is not as good as KELM-CK, SS-EELM and CVW-CEELM. OA of the proposed CVW-CEELM is 99.4%; OA of KELM-CK is 99.1%; OA of SS-EELM is 99.1%; OA of GELM-CK is 98.9%. The classification accuracy of the proposed CVW-CEELM is the highest. In two typical HSI datasets, the number of training samples in Indian Pines dataset is less than that in Pavia University dataset. It is worth noting that the time deltas of KELM-CK, SS-EELM and CVW-CEELM are 297, 52.1 and 48.4 seconds, respectively. The time cost growth rate of CVW-CEELM is smaller than that of KELM-CK and SS-EELM with the increasing number of training samples. Therefore, the comprehensive time cost of CVW-CEELM is the smallest in the algorithms with higher classification accuracy.

TABLE 4. Classification results obtained for Pavia University dataset.

/%

NO.	Class	Train	Test	SVM	ELM	GELM	KELM	GELM_CK	KELM_CK	SS-EELM	CVW-CEELM
1	Asphalt	265	6366	89.9±1.02	83.7±0.75	83.6±0.65	87.1±0.73	97.8±0.41	98.3±0.47	98.0±0.17	**99.2±0.17**
2	Meadows	746	17903	94.5±0.82	93.1±0.54	91.7±0.74	94.8±0.42	99.5±0.14	99.8±0.17	99.6±0.15	**99.8±0.14**
3	Gravel	84	2015	83.1±1.04	75.4±1.56	75.5±1.57	84.5±1.57	96.1±1.08	98.1±0.68	97.4±0.67	**99.3±0.30**
4	Trees	123	2941	95.6±0.51	95.9±0.47	96.0±0.38	96.8±0.34	98.8±0.15	98.9±0.25	98.8±0.25	**99.0±0.77**
5	Painted metal sheets	54	1291	98.1±0.24	97.6±1.62	98.1±1.62	98.6±0.25	99.0±0.11	**99.6±0.24**	99.1±0.41	99.5±0.38
6	Bare Soil	201	4828	93.6±0.68	92.4±0.67	93.5±0.29	94.1±0.58	99.1±0.05	99.6±0.01	99.2±0.01	**99.7±0.02**
7	Bitumen	53	1277	92.4±0.97	92.0±0.66	92.0±0.66	92.8±0.83	99.2±0.12	99.3±0.18	99.2±0.31	**99.5±0.25**
8	Self-Blocking Bricks	147	3535	89.7±0.68	89.6±0.97	90.1±0.82	89.4±0.67	98.6±0.24	98.2±0.25	98.5±0.29	**99.0±0.24**
9	Shadows	38	909	97.5±0.01	98.2±0.21	98.4±0.16	98.6±0.18	99.1±0.16	99.0±0.39	99.2±0.44	**99.4±0.71**
OA				91.6±0.45	90.3±0.27	90.1±0.25	92.7±0.20	98.9±0.16	99.1±0.13	99.1±0.28	**99.4±0.31**
AA				92.7±0.16	90.9±0.21	91.0±0.17	93.0±0.23	98.6±0.15	99.0±0.14	98.8±0.25	**99.4±0.24**
k				89.7±0.68	87.1±0.40	87.0±0.29	90.4±0.26	96.5±0.17	99.1±0.12	97.5±0.23	**99.2±0.24**
Time(S)				91.3±1.81	1.47±0.21	9.82±0.62	25.7±0.49	17.2±0.41	313±0.88	88.6±1.01	64.2±1.03

the best metric values are in bold

TABLE 5. The number of bands selected by the weighted random-selecting-based method for Pavia University.

NO.	1	2	3	4	5	6	7	8	9	10
Weight	0.1057	0.1205	0.1103	0.0937	0.0803	0.0748	0.0746	0.1015	0.1160	0.1225
Number	5	6	6	5	4	4	4	5	6	6

In conclusion, in terms of the time cost and classification accuracy, the proposed CVW-CEELM has the best classification performance among the above hyperspectral image classification counterparts.

V. CONCLUSION

In this work, we have proposed a novel HSI classification algorithm based on the comprehensive evaluation model of extreme learning machine with the cumulative variation weights, which we referred to as CVW-CEELM. The cumulative variation value is proposed as a new metric used to improve the bands selection of grouping strategy and determine the weights of multiple weak classifiers. The CVW-CEELM proposes the bands selection of grouping strategy, which reduces the complexity of bands selection algorithm. Besides, The weighted random-selecting-based method improves the grouping strategy. To some extent, the improved grouping strategy increases the feature vectors describing a sample, which improves the classification of many different types of small-size homogeneous areas. The cumulative variation weights calculated based on the cumulative variation values weight all weak classifiers, which make the classification results more accurate. Our experiments, conducted using the benchmark HSI datasets, reveal that the proposed method can achive good classification performance.

In the future work, we will explore the impact of the spatial-spectral features based on the multi-scale neighborhood of a given pixel on the HSI classification effect of CVW-CEELM. The features selection will be further optimized, and a more efficient HSI classification model will be developed.

REFERENCES

[1] H. K. Zhang, Y. Li, and Y. N. Jiang, "Deep learning for hyperspectral imagery classification: The state of the art and prospects," *Acta Automatica Sinica*, vol. 44, no. 6, pp. 961–977, 2018, doi: 10.16383/j.aas.2018.c170190.

[2] W. Zhao and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, Aug. 2016, doi: 10.1109/TGRS.2016.2543748.

[3] K. Yao, X. D. Guo, and Y. Nan, "Research progress of hyperspectral remote sensing monitoring of vegetation biomass assessment," *Sci. Surveying Mapping*, vol. 41, no. 8, pp. 48–53, 2016, doi: 10.16251/j.cnki.1009-2307.2016.08.010.

[4] K. Zhang, B. Q. Hei, Z. Zh, and S. Y. Li, "CNN with coeffcient of variation-based dimensionality reduction for hyperspectral remote sensing images classification," *J. Remote Sens.*, vol. 22, no. 1, pp. 87–96, 2016, doi: 10.11834/jrs.20187075.

[5] C. M. Gevaert *et al.*, "Combining hyperspectral UAV and multispectral Formosat-2 imagery for precision agriculture applications," in *Proc. 6th Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, 2017, doi: 10.1109/WHISPERS.2014.8077607.

[6] Y. D. Tang, S. C. Huang, and A. J. Xue, "Sparse representation based band selection for hyperspectral imagery target detection," *Acta Electronica Sinica*, vol. 45, no. 10, pp. 2368–2374, 2017, doi: 10.3969/j.issn.0372-2112.2017.10.009.

[7] Y. H. Zhang, G. Yang, J. H. Huang, and Y. B. Yang, "Data dimension reduction method for hyperspectral images based on segmented column-and-line 2D-PCA," *Comput. Eng.*, vol. 43, no. 9, pp. 256–262, 2017, doi: 10.3969/j.issn.1000-3428.2017.09.045.

[8] J. Zabalza, J. Ren, J. Zheng, H. Zhao, C. Qing, Z. Yang, P. Du, and S. Marshall, "Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging," *Neurocomputing*, vol. 185, pp. 1–10, Apr. 2016, doi: 10.1016/j.neucom.2015.11.044.

[9] T. Qiao, J. Ren, Z. Wang, J. Zabalza, M. Sun, H. Zhao, S. Li, J. A. Benediktsson, Q. Dai, and S. Marshall, "Effective denoising and classification of hyperspectral images using curvelet transform and singular spectrum analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 119–133, Jan. 2017, doi: 10.1109/TGRS.2016.2598065.

[10] L. Windrim, R. Ramakrishnan, A. Melkumyan, and R. J. Murphy, "A physics-based deep learning approach to shadow invariant representations of hyperspectral images," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 665–677, Feb. 2018, doi: 10.1109/TIP.2017.2761542.

[11] Q. Ran, H. Y. Yu, L. R. Gao, W. Li, and B. Ghang, "Superpixel and subspace projection-based support vector machines for hyperspectral image classification," *J. Image Graph.*, vol. 23, no. 1, pp. 0095–0105, 2018, doi: 10.11834/jig.170201.

[12] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "A spatial–spectral kernel-based approach for the classification of remote-sensing images," *Pattern Recognit.*, vol. 45, no. 1, pp. 381–392, Jan. 2012, doi: 10.1016/j.patcog.2011.03.035.

[13] D. D. Mou and L. Liu, "Comparative study of ELM and SVM in hyperspectral image supervision classification," *Remote Sens. Technol. Appl.*, vol. 34, no. 1, pp. 115–124, 2019.

[14] P. Li, L. Zhi, C. Xu, Y. Q. Fang, and F. Zhang, "Research on space object's materials multi-color photometry identification based on the extreme learning machine algorithm," *Spectrosc. Spectra Anal.*, vol. 39, no. 2, pp. 363–369, 2019.
[15] M. Jiang, F. Cao, and Y. Lu, "Extreme learning machine with enhanced composite feature for spectral-spatial hyperspectral image classification," *IEEE Access*, vol. 6, pp. 22645–22654, 2018, doi: 10.1109/ACCESS.2018.2825978.
[16] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "A new deep convolutional neural network for fast hyperspectral image classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 120–147, Nov. 2018.
[17] Y. Y. Liu *et al.*, "3D object recognition via convolutional-recursive neural network and kernel extreme learning machine," *Pattern Recognit. Artif. Intell.*, vol. 30, no. 12, pp. 1091–1099, 2017, doi: 10.16451/j.cnki.issn1003-6059.201712004.
[18] C. Cervellera and D. Maccio, "An extreme learning machine approach to density estimation problems," *IEEE Trans. Cybern.*, vol. 47, no. 10, pp. 3254–3265, Oct. 2017, doi: 10.1109/TCYB.2017.2648261.
[19] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst., Man, Cybern. B. Cybern.*, vol. 42, no. 2, pp. 513–529, Apr. 2012.
[20] J. Li, Q. Du, W. Li, and Y. Li, "Optimizing extreme learning machine for hyperspectral image classification," *J. Appl. Remote Sens.*, vol. 9, no. 1, Mar. 2015, Art. no. 097296, doi: 10.1117/1.JRS.9.097296.
[21] Y. X. Liu, J. J. Fang, and X. J. S. J. Zhang, "Application of extreme learning machine in the nonlinear error compensation of magnetic compass," *Chin. J. Sci. Instrum.*, vol. 36, no. 9, pp. 1921–1927, 2015, doi: 10.3969/j.issn.0254-3087.2015.09.001.
[22] Y. Zhou, J. Peng, and C. L. P. Chen, "Extreme learning machine with composite kernels for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2351–2360, Jun. 2015, doi: 10.1109/JSTARS.2014.2359965.
[23] F. Cao, Z. Yang, J. Ren, M. Jiang, and W.-K. Ling, "Linear vs. Nonlinear extreme learning machine for spectral-spatial classification of hyperspectral images," *Sensors*, vol. 17, no. 11, p. 2603, Nov. 2017, doi: 10.3390/s17112603.
[24] S. Alim, D. Peijun, L. Sicong, and J. Li, "E2LMs: Ensemble extreme learning machines for hyperspertral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1060–1069, 2014.
[25] Y. Gu, Y. Xu, and B. F. Guo, "Hyperspectral image classification by combination of spatial-spectral features and ensemble extreme learning machines," *Acta Geodaetica et Cartographica Sinica*, vol. 47, no. 9, pp. 1238–1249, 2018, doi: 10.11947/j.AGCS.2018.20170476.
[26] Y. Mei and C. B. Lu, "Adaptive weighted online extreme learning machine for imbalance data steam," *Pattern Recognit. Artif. Intell.*, vol. 32, no. 2, pp. 144–150, 2019, doi: 10.16451/j.cnki.issn1003-6059.201902006.
[27] W. Guo, T. Xu, K. M. Tang, and J. J. Yu, "Online sequential extreme learning machine with generalized regularization and forgetting mechanism," *Control Decis.*, vol. 32, no. 2, pp. 247–254, 2017.
[28] Y. J. Chen, C. Y. Ma, L. Sun, and T. M. Zhan, "Edge-modified superpixel based spectral-patial kernel method for hyperspectral image classification," *Acta Electro-Nica Sinica*, vol. 47, no. 1, pp. 73–81, doi: 10.3969/j.issn.0372-2112.2019.01.010.
[29] C. Shi and C.-M. Pun, "Multi-scale hierarchical recurrent neural networks for hyperspectral image classification," *Neurocomputing*, vol. 294, pp. 82–93, Jun. 2018.
[30] B. Schlkopf and A. Smola, *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2002, doi: 10.7551/mitpress/4175.001.0001.
[31] F. Lv and M. Han, "Hyperspectral remote sensing image classification based on deep extreme learning machine," *J. Dalian Univ. Technol.*, vol. 58, no. 2, pp. 166–173, 2018, doi: 10.7511/dllgxb201802009.

YUPING YIN received the M.S. degree in power electronics and the Ph.D. degree in information visualization technology from Liaoning Technical University, China, in 2007 and 2020, respectively. She has published more than ten articles. Her current research interests include deep learning and image processing.

LIN WEI received the Ph.D. degree in safety administration from Liaoning Technical University, China, in 2018. He is currently an Associate Professor and a Master Supervisor with Liaoning Technical University. He has published more than 16 articles. His current research interests include machine learning and image processing.