A Hyperspectral Image Classification Method Using Multifeature Vectors and Optimized KELM

Huayue Chen, *Member, IEEE*, Fang Miao, *Member, IEEE*, Yijia Chen ⬤ , *Member, IEEE*,
Yijun Xiong ⬤ , *Member, IEEE*, and Tao Chen ⬤ , *Member, IEEE*

Abstract—To improve the accuracy and generalization ability of hyperspectral image classification, a feature extraction method integrating principal component analysis (PCA) and local binary pattern (LBP) is developed for hyperspectral images in this article. The PCA is employed to reduce the dimension of the spectral features of hyperspectral images. The LBP with low computational complexity is used to extract the local spatial texture features of hyperspectral images to construct multifeature vectors. Then, the gray wolf optimization algorithm with global search capability is employed to optimize the parameters of kernel extreme learning machine (KELM) to construct an optimized KELM model, which is used to effectively realize a hyperspectral image classification (PLG-KELM) method. Finally, the Indian pines dataset, Houston dataset, and Pavia University dataset and an application of WHU-Hi-LongKou dataset are selected to verify the effectiveness of the PLG-KELM. The comparison experiment results show that the PLG-KELM can obtain higher classification accuracy, and takes on better generalization ability for small samples. It provides a new idea for processing hyperspectral images.

Index Terms—Gray wolf optimization (GWO), hyperspectral image classification, kernel extreme learning machine (KELM), local binary pattern (LBP), optimization, principal component analysis (PCA).

I. INTRODUCTION

THE hyperspectral sensor can obtain approximately continuous spectral curves of ground and objects in a large

Huayue Chen is with the Key Laboratory of Earth Exploration and Information Technology of Ministry of Education, Chengdu University of Technology, Chengdu 610059, China, with the School of Computer Science, China West Normal University, Nanchong 637002, China, and also with the Internet of Things Perception and Big Data Analysis Key Laboratory of Nanchong, China West Normal University, Nanchong 637002, China (e-mail: sunnyxiaoyue20@cwnu.edu.cn).

Fang Miao and Yijun Xiong are with the Key Laboratory of Earth Exploration and Information Technology of Ministry of Education, Chengdu University of Technology, Chengdu 610059, China (e-mail: mf@cdut.edu.cn; 15558187@qq.com).

Yijia Chen and Tao Chen are with the School of Computer Science, China West Normal University, Nanchong 637002, China, and also with the Internet of Things Perception and Big Data Analysis Key Laboratory of Nanchong, Nanchong 637002, China (e-mail: chenyijia943@stu.cwnu.edu.cn; chentao@cwnu.edu.cn).

number of electromagnetic bands, such as ultraviolet, visible, near-infrared, and mid-infrared by combining the spectral information of the reflection features of ground and objects. The 3-D data cube is composed of dozens of continuous band images, which has the characteristics of spectral integration. It can more effectively interpret ground features and improve the classification and monitoring abilities of ground and objects. Therefore, it has been widely applied in environment monitoring, disaster assessment, and other fields [1]–[3].

Hyperspectral image classification is an important part of hyperspectral applications, and it is one of the hot research field of hyperspectral image processing and interpretation. In recent years, a lot of classification methods are proposed for hyperspectral images. Li *et al.* [4] developed a new framework for the classification of hyperspectral scenes that pursues the combination of multiple features. Li *et al.* [5] proposed a discontinuity preserving relaxation strategy for postprocessing of class probability estimates. Liu *et al.* [6] proposed a new approach for accurate spatial spectral classification of hyperspectral images. Xu *et al.* [7] proposed a new technique based on multiple morphological component analysis. Liu *et al.* [8] proposed a new class-oriented spectral partitioning technique. Xue *et al.* [9] proposed a two novel sparse graph regularization methods, SGR and SGR with total variation. Liu *et al.* [10] proposed a new multiview active learning framework for hyperspectral image classification. Juan *et al.* [11] proposed several methods to deal with the hyperspectral image classification problem. Yu *et al.* [12] proposed a rather different approach from a viewpoint of mixed pixel classification to improve hyperspectral image classification for multiple classes. Chang *et al.* [13] introduced a new concept of band capacity of a hyperspectral image and further developed a theory for band capacity. Liu *et al.* [14] proposed a new semisupervised active learning approach. Chang *et al.* [15] proposed a new approach, which can implement automatic target generation process bandwise in a progressive manner. Zhong *et al.* [16] proposed a new spectral-spatial approach, called spectral-spatial feedback close network system. Yu *et al.* [17] proposed a novel approach called class signature-constrained background suppression approach to band selection. Yu *et al.* [18] proposed a convolutional neural network system embedded with an extracted hashing feature that utilizes the semantic information of the hyperspectral images. Song *et al.* [19] proposed a class information-based band selection approach. Yu *et al.* [20] proposed a novel hyperspectral image classification framework based on a simplified 2–3-D convolutional neural networks by

the cooperation between a 2-D convolutional neural network and a 3-D convolution layer. Yu et al. [21] proposed a simplified and equivalent model to the classic sparse representation-based classifier. Hong et al. [22] proposed a solution to address this issue by locally extracting invariant features from hyperspectral imagery in both spatial and frequency domains. Zhong et al. [23] proposed a hyperspectral image classification method with a generative adversarial network and conditional random field based framework. Luo et al. [24] proposed a hybrid-graph learning method to reveal the complex high-order relationships of the hyperspectral image. Liu et al. [25] proposed a content-guided convolutional neural network. Samat et al. [26] proposed an edge gradient-based active learning approach. Song et al. [27] proposed a new progressive band selection processing of hyperspectral image classification. Sun et al. [28] proposed a novel patch-based low rank component induced spatial-spectral kernel method. Chu et al. [29] proposed a new method based on the discriminative locality preserving broad learning system by exploiting the manifold structure between neighboring pixels of hyperspectral image. Song et al. [30] proposed a 3-D receiver operating characteristic analysis. Yang et al. [31] proposed a locality regularized robust-probabilistic collaborative representation classification based on the Euclidean distance between the training samples and the testing samples for hyperspectral images. Mou et al. [32] proposed a novel graph-based semisupervised network called nonlocal graph convolutional network. These proposed methods can better realize image classification and obtain better classification accuracy, but they still exist some shortcomings, such as poor generalization ability, higher complexity, lower accuracy for small samples, and so on.

In addition, the Hughes phenomenon caused by the high-dimensional characteristics and the limited training samples of hyperspectral images is always an important challenge in hyperspectral image classification. The feature extraction is usually performed on hyperspectral images before classification in order to obtain feature data to describe the hyperspectral image, and improve the accuracy of ground object classification. Jiang et al. [33] used principal component transformation to extract the image spectral features. Li et al. [34] proposed a local texture description operator to extract the local spatial texture features. In addition, support vector machine (SVM), extreme learning machine (ELM), deep belief network, and so on are used for constructing different classifiers [35]–[38]. Compared with artificial neural network, the ELM has simpler structure and faster training speed [39]–[41]. However, the number of hidden nodes is usually less than the number of training samples in practice, which could lead to the problem of multicollinearity. To solve these problems, Huang introduced kernel functions into ELM by comparing ELM and SVM to propose a KELM [42]. However, the kernel parameters and regular parameters of the KELM will affect the generalization ability and sensitivity of classifiers to a certain extent. Cao et al. [43] proposed a cascade classifier based on fusing ELM and adaptive sparse representation. Chen et al. [44] proposed a KELM classification algorithm with space spectrum information, which fused spectral information and spatial information in the construction of hyperspectral image feature vector to improve the classification

accuracy of KELM. In addition, the combination of different kernel parameters and regularization parameters has great influence on classification accuracy [45]–[47]. In recent years, some intelligent algorithms are widely applied in parameter optimization of KELM, such as particle swarm optimization, differential evolution, evolutionary algorithm, ant colony optimization, gray wolf optimization (GWO) and their improvements, and so on [48]–[56].

Due to the problems of same object with different spectra and different objects with same spectrum, it is difficult to fully describe the differences of each pixel using a feature space, so it is tough to obtain a better classification effect. Therefore, in order to improve the accuracy and efficiency of hyperspectral image classification, the principal component analysis (PCA), local binary pattern (LBP), GWO, and KELM are integrated to develop a new hyperspectral image classification method, namely PLG-KELM in this article. The PCA is employed to reduce the dimension of the spectral features. The LBP is used to extract the local spatial texture features to construct multifeature vectors. The GWO is employed to optimize the parameters of KELM to solve the unstable classification accuracy caused by the random selection of parameters of KELM. Finally, the comparative experiments are implemented to verify the effectiveness of PLG-KELM by three hyperspectral data.

The innovations and main contributions of this article are described as follows.

1) A feature extraction method integrating PCA and LBP is developed for hyperspectral images.
2) A parameter optimization method based on GWO with global search capability is proposed to optimize the parameters of KELM to construct an optimized KELM.
3) A hyperspectral image classification (PLG-KELM) method based on multifeature vectors and optimized KELM is proposed to realize the hyperspectral image classification.
4) Comprehensive experiments are designed and executed to comprehensively prove the effectiveness of PLG-KELM by different hyperspectral datasets and takes on better classification performance and generalization ability for small samples.

The rest of this article is organized as follows. Section II introduces the PCA, LBP, GWO, and KELM. Section III proposes a parameter optimization method for KELM. Section IV introduces the idea, model, and steps of the PLG-KELM in detail. Section V describes and analyzes the experimental results. An application in fine classification of crops is applied in Section VI. The discussion is given in Section VII. Finally, the conclusion and future work are discussed in Section VIII.

II. BASIC METHOD

A. PCA

The PCA is a common unsupervised learning method, which is the most basic dimensionality reduction method of hyperspectral data. In the PCA, the data are projected into a new coordinate space by using linear projection method, so that the new components are distributed according to the amount of

information. The measure of information quantity is the variance of data, which is described as follows:

$$\text{Var}(y_i) = a_i^T \sum a_i, i = 12, \ldots m \quad (1)$$

where a_i is the i-th transformation vector, Σ is the covariance matrix of the original data, y_i is the i-th principal component, $Var(y_i)$ is the variance of the i-th principal component.

The PCA uses orthogonal transformation to transform the observation data represented by linear correlation variables into a few data represented by linear independent variables. The linear independent variables are called principal components. The number of principal components is usually less than the number of original variables. Therefore, the PCA is a commonly used method to reduce data dimension.

B. LBP

The LBP is an operator, which is used to describe the local texture features of image and analyze image texture features. It uses the structure analysis method to analyze the fixed window features, and then statistics is used to extract the overall features. The mathematical expression of LBP is described as follows:

$$\text{LBP}(x_c, y_c) = \sum_{p=0}^{p-1} S(i_p - i_c) * 2^p \quad (2)$$

where (x_c, y_c) is the coordinate of the center pixel, P is the p-th pixel of the neighborhood, i_p is the gray value of the neighborhood pixel, i_c is the gray value of the center pixel. $S(x)$ is a symbolic function, which is defined as follows:

$$S(x) = \begin{cases} 1 \ x \geq 0 \\ 0 \ x < 0 \end{cases}. \quad (3)$$

C. GWO

The GWO is inspired by the predatory activities of gray wolves, it simulates the predatory behavior of gray wolves and achieves the optimization goal based on the cooperation mechanism of wolves [49]. Due to the existence of adaptive convergence factor and information feedback mechanism, the balance between local optimization and global search can be achieved. The GWO has the characteristics of simple structure and less adjusted parameters, and is easy to realize. In the GWO, in order to simulate the leadership mechanism of gray wolves, the gray wolves are divided into four levels of α, β, δ, and ω, which are corresponded to the first optimal solution, second optimal solution, common solution, and chosen solution. The optimization process of GWO includes the social hierarchy, tracking, encircling, and attacking prey. The GWO calculates the fitness of each individual in the population, and marks the three gray wolves with the best fitness values as α, β, and δ in turn, the rest gray wolves are marked as ω.

D. Kernel Extreme Learning Machine

The ELM has a network structure, which is used to solve the problem at one time. Given the number of hidden layer nodes as L, the output function of hidden layer

is H(x) $= [h_1(x), \ldots, h_L(x)]$, the output weight of hidden layer is $\beta = [\beta_1, \ldots, \beta_L]^T$, the input training sample is $T = \{(x_i, w_i, b_i) x_i \in R^d, w_i, b_i \in R^m, i = 1, \ldots, N\}$. That is, N sample points are given, $w_i \ and \ b_i$ are the node parameters of hidden layer. $g(x_i, w_i, b_i)$ is the activation function. The mathematical model of ELM is described as follows:

$$\begin{cases} h_i(x) = g(x_i, w_i, b_i) = g(w_i x_i + b_i) \\ f(x) = \sum_{i=1}^{L} \beta_i h_i(x) = \beta H(x) \end{cases}. \quad (4)$$

The goal of ELM is to minimize the norm of training error and hidden layer output weight, which is defined as follows:

$$\min \|\beta \cdot h(x_i) - t_i\|^2 \quad (5)$$

In order to further enhance the generalization ability and stability of ELM, Huang proposed a kernel extreme learning machine (KELM) by comparing the principle of ELM and SVM. The kernel matrix is defined as follows:

$$\begin{cases} \Omega_{\text{ELM}} = HH^T \\ \Omega_{ij} = h(x_i) \cdot h(x_j) = K(x_i, x_j) \end{cases}. \quad (6)$$

The kernel matrix Ω replaces the random matrix HH^T in ELM, and uses kernel function to map all input samples from N-dimensional input space to high-dimensional hidden layer feature space. After the kernel parameters are determined, the mapping value of Ω is fixed. $h_i(x)$ is the output function of hidden layer nodes. The kernel function $K(x_i, x_j)$ includes radial basis function (RBF), linear kernel function, polynomial kernel function, and so on. The kernel function is usually set as RBF, which is described as follows:

$$K(x_j, x_j) = exp\left(-\frac{\|x_i - x_j\|^2}{\sigma}\right). \quad (7)$$

The parameter I/C is added to the main diagonal of the unit diagonal matrix HH^T so that its characteristic root is not zero. Then, the weight vector β^* is obtained, which makes KELM to be more stable and better generalization. At this time, the output weight of KELM is changed as follows:

$$\begin{aligned} f(x) &= h(x) H^T \left(I/C + HH^T\right)^{-1} T \\ &= \begin{bmatrix} K(x, x_1) \\ \cdots \\ K(x, x_N) \end{bmatrix}^T \left(I/C + \Omega_{\text{ELM}}\right)^{-1} T. \end{aligned} \quad (8)$$

III. PARAMETER OPTIMIZATION OF THE KELM

A. Idea of Parameter Optimization

In this article, RBF is used as the activation function of KELM, the kernel parameter K of RBF is set, and the output weight is simply calculated and analyzed to replace the original tedious iterative process. According to the principle of ridge regression, when the output matrix H of hidden layer is calculated, the regular term operator C is added in order to keep more stable solution of generalized inverse of H. Therefore, the combination of K and C in KELM affects the generalization performance and the sensitivity of the classifier. Traditional parameter selection
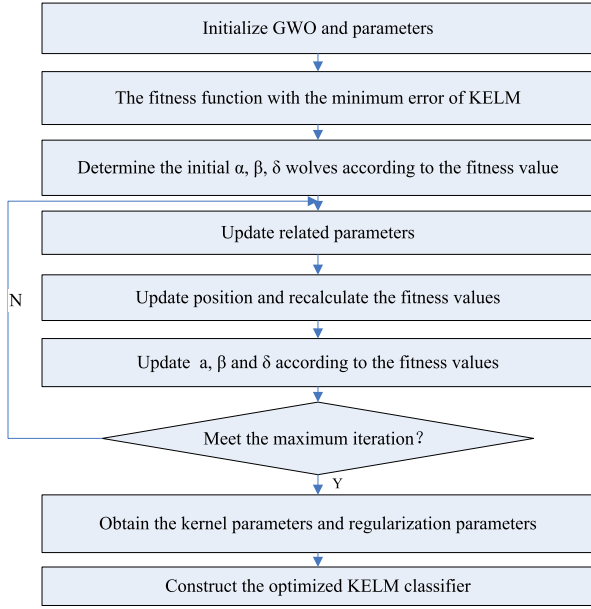
Fig. 1. Parameter optimization model.

methods use samples to repeat experiments and analysis to manually select the best parameters. However, it lacks theoretical guidance and takes a long time. Some numerical simulation methods (such as Newton method, gradient descent algorithm, and so on) are sensitive to the selection of initial value. The GWO has the simple structure, few adjusted parameters, and is easy to realize. There exists a self-adaptive convergence factor and information feedback mechanism, which can achieve a balance between local optimization and global search. Therefore, the GWO is employed to optimize the K and C of KELM. The parameter optimization model is shown in Fig. 1.

B. Steps of Parameter Optimization

The GWO is used to iteratively find the location of the maximum value of the wolf, that is, the optimal value of classification. The steps are described as follows.

Step 1: Determine the fitness function $f(x) = [\begin{smallmatrix} K(x,x_1) \\ \cdots \\ K(x,x_N) \end{smallmatrix}]^T (I/C + \Omega_{ELM})^{-1} T$ where K and C are the optimized objects. In the search space, the initial population of gray wolf is composed of random initialization individuals. The values of coefficient vector $\vec{A}$, $\vec{C}$ and convergence factor $\vec{a}$ are initialized.

Step 2: The root mean square error function is used to calculate the fitness value of each individual, which are ranked. The positions of the individual corresponding to the three minimum fitness values in the population error are taken as $\vec{X}_\alpha$, $\vec{X}_\beta$, and $\vec{X}_\delta$.

Step 3: Calculate the nonlinear variation parameter $\vec{r}_1$ and $\vec{r}_2$. Then, the values of $\vec{A}$ and $\vec{C}$ are updated according to $\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a}$ and $\vec{C} = 2 \vec{r}_2$. $\vec{a}$ is a convergence factor, which is linearly decreased from 2 to 0 with the increasing of the number of iterations. $\vec{r}_1$ and $\vec{r}_2$ are the random number in [01].

Step 4: The position of each individual is updated

$$\begin{cases} \tilde{X}_1 = \tilde{X}_\alpha - \tilde{A}_1 \cdot \tilde{D}_\alpha \\ \tilde{X}_2 = \tilde{X}_\beta - \tilde{A}_2 \cdot \tilde{D}_\beta \\ \tilde{X}_3 = \tilde{X}_\delta - \tilde{A}_3 \cdot \tilde{D}_\delta \end{cases} \tag{9}$$

$$\begin{cases} \tilde{D}_\alpha = \left| \tilde{C}_1 \cdot \tilde{X}_\alpha - \tilde{X} \right| \\ \tilde{D}_\beta = \left| \tilde{C}_2 \cdot \tilde{X}_\beta - \tilde{X} \right| \\ \tilde{D}_\delta = \left| \tilde{C}_3 \cdot \tilde{X}_\delta - \tilde{X} \right| \end{cases} . \tag{10}$$

Among them, $\vec{D}_\alpha$, $\vec{D}_\beta$, $\vec{D}_\delta$ represent the distance from α, β, and δ to other individuals, respectively. $\vec{X}_\alpha$, $\vec{X}_\beta$, and $\vec{X}_\delta$ represent the current position of α, β, and δ, respectively. $\vec{C}_1$, $\vec{C}_2$, and $\vec{C}_3$ are the random vector, and $\vec{X}$ is the current position of gray wolf.

Step 5: The fitness value of each individual is recalculated to update the individual extremum and population extremum, which further update $\vec{X}_\alpha$, $\vec{X}_\beta$, and $\vec{X}_\delta$.

Step 6: Determine whether the number of iterations is equal to the maximum value. If it is not, execute $t = t+1$ and return to Step 3. Otherwise, the optimal solution α is output. The parameters of KELM are determined in order to establish the optimal classifer.

IV. HYPERSPECTRAL IMAGE CLASSIFICATION METHOD

A. Idea of the PLG-KELM

Hyperspectral images have the characteristics of large number of bands, large data, redundant information, high dimension, and so on. Although the increase of spectral bands brings more abundant information of categories of ground and objects in hyperspectral images, the high redundancy and high correlation between bands make the low classification accuracy. The PCA can map N-dimensional features to the k-dimensional features to reduce the dimension of data features and the impact of spectral information redundancy. The LBP has the advantages of rotation invariance and gray level invariance. It can effectively extract the local spatial texture features of hyperspectral images. Compared with the traditional feed forward neural networks, the KELM has the fast training speed and good generalization performance for high-dimensional data. The GWO has the characteristics of less optimization parameters, simple implementation, strong convergence ability, fast optimization speed, and strong global optimization ability. To improve the classification accuracy and generalization ability, a hyperspectral image classification (PLG-KELM) method based on integrating merits of PCA, LBP, GWO, and KELM is proposed in this article. First, the spectral information is combined with the local spatial context information, and different feature extraction methods are adopted in the spectral dimension and spatial dimension of the hyperspectral images. The PCA is used to select the principal component bands of spectral information to reduce the impact of spectral information redundancy. The LBP operator is used to describe the local spatial texture information of the hyperspectral images. The complementarity of different features are combined, and the feature level fusion and decision level fusion are analyzed to construct multifeature vector of hyperspectral images, which
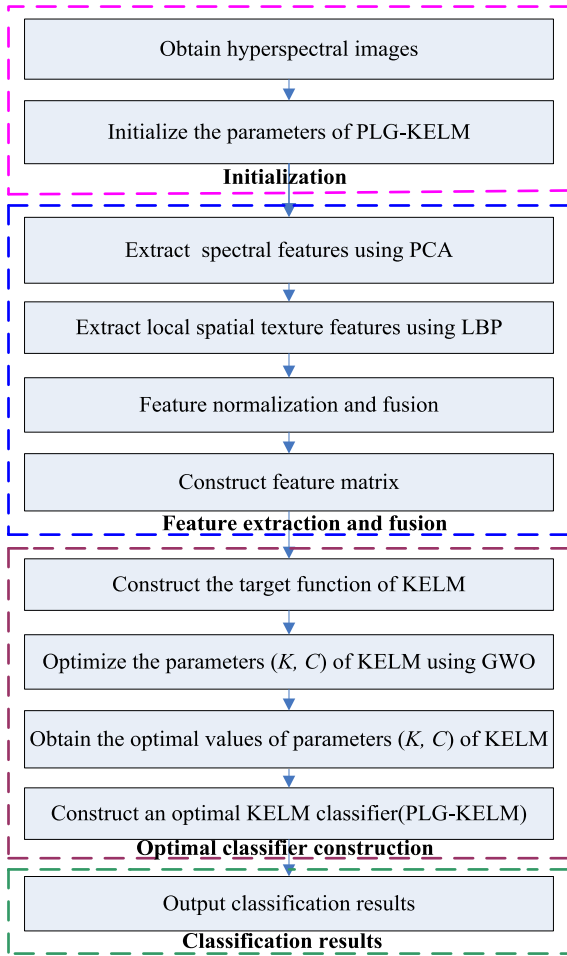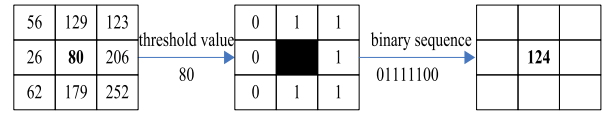
Fig. 2. Model of the PLG-KELM.

Fig. 3. Calculation example of LBP operator.

4) The obtained eigenvectors are arranged into a matrix according to the corresponding eigenvalues, and the values of the first k lines are taken to form matrix Z.
5) The hyperspectral data G = ZX with spectral dimension of K is obtained.

Step 2: Extract local spatial texture features

The LBP operator is defined on a central pixel and its surrounding rectangular neighborhood system with a size of 3×3. According to the formula (2) and formula (3), each neighborhood pixel value and gray value of the center pixel are binarized. In this way, a local binary mode is formed. Start from the upper left corner, a series of binary numbers are obtained in clockwise direction, and the decimal number corresponding to the binary number is used to uniquely identify the central pixel. Each pixel can be calculated to obtain a local binary mode. An example of the calculation is given in Fig. 3.

Step 3: Feature fusion

In feature fusion, the feature normalization is realized in order to unify the scale of eigenvalues. The minimum–maximum technique is used to map feature data to the range [01]. Then, the extracted texture and spectral features of dimensionality reduction are directly combined. The multifeature fusion method is adopted to stack multiple features into a combined feature, which is concatenated into a vector to obtain the feature fusion $X = [x_1, x_2 \cdots x_p]$.

Step 4: Construct an optimal KELM classifier

The multifeature vectors are considered as training set, and the classification accuracy of KELM is used as fitness function of the GWO. The kernel parameter and regularization parameter (K, C) are combined as the optimal parameters of KELM, and an optimal KELM classifier is constructed.

Step 5: Obtain the classification results

V. EXPERIMENTAL RESULTS AND ANALYSIS

To verify the effectiveness of the PLG-KELM, the Indian pine, Houston 2013, and Pavia University are selected in here. The classification models of SVM, broad learning system (BLS), contractive auto-encoder and convolutional neural network (CAE-CNN), PCA-CNN, and KELM are selected to compare with the PLG-KELM [57].

A. Experimental Data Sources

Indian pines data are one of the most commonly used datasets for hyperspectral images. The image is cut the size of 145 × 145 for labeling. Houston 2013 dataset was acquired by

is used as the training set of classifier. Then, the GWO with strong convergence and optimization ability is used to obtain the optimal parameter combination values of the KELM in order to construct an optimized KELM classifier to realize a PLG-KELM classification method.

B. Model of the PLG-KELM

In order to improve the classification accuracy, the hyperspectral image features based on spatial spectrum information is extracted and a PLG-KELM method based on PCA, LBP, GWO, and KELM is developed for hyperspectral image classification. The model of the PLG-KELM is shown in Fig. 2.

The specific realization process of the PLG-KELM is described as follows.

Step 1: Extract spectral features

1) The spectral dimension is quantized to obtain the matrix $X = [X_1, X_2, \cdots, X_p]$, where p is the dimension.
2) Calculate the covariance matrix C of X, $C = \frac{1}{p} X X^T$.
3) The eigenvalues and eigenvectors of covariance matrix C are solved.

TABLE I
BASIC INFORMATION OF THREE DATASETS

Data	Indian Pines	Houston	Pavia University
Collection location	Indiana, USA	Houston, USA	Northern Italy
Acquisition equipment	AVIRIS	CASI-1500	ROSIS
Spectral coverage(μm)	0.4~2.5	0.38~1.05	0.43~0.86
Data size(pixel)	145×145	349×1905	610x340
Spatial resolution(m)	20	2.5	1.3
Number of bands	220	144	115
Number of bands after denoising	200	144	103
Sample size	10249	15029	42776
Number of categories	16	15	9

TABLE II
GROUND AND OBJECTS OF 16 TYPES IN INDIAN PINES

Category	Class name	Number of samples
1	Alfalfa	46
2	Corn-notill	1428
3	Corn-min	830
4	Corn	237
5	Grass-pasture	483
6	Grass-trees	730
7	Grass-pastue-mowed	28
8	Hay-windrowed	478
9	Oats	20
10	Soybean-notill	972
11	Soybean-min	2455
12	Soybean-clean	593
13	Wheat	205
14	Woods	1265
15	Bldg-Grass-Tree-Drives	386
16	Stone-Steel-Towers	93

TABLE III
GROUND AND OBJECTS OF 15 TYPES IN HOUSTON

Category	Class name	Number of samples
1	Healthy grass	1251
2	Stressed grass	1254
3	Synthetic grass	697
4	Trees	1244
5	Soil	1242
6	Water	325
7	Residential	1268
8	Commercial	1244
9	Road	1252
10	Highway	1227
11	Railway	1235
12	Parking Lot 1	1233
13	Parking Lot 2	469
14	Tennis Court	428
15	Running Track	660

TABLE IV
GROUND AND OBJECTS OF 9 TYPES IN PAVIA UNIVERSITY

Category	Class name	Number of samples
1	Asphalt	6631
2	Meadows	18649
3	Gravel	2099
4	Trees	3064
5	Painted metal sheets	1345
6	Bare Soil	5029
7	Bitumen	1330
8	Self-Blocking Bricks	3682
9	Shadows	947

B. Experimental Environment and Parameter Settings

The experiment is on Intel (R) i7-9750h CPU, RAM 16G, Win10. The programming language is MATLAB R2018a.

The LBP operator uses the original LBP operator, which selects 3×3 neighborhood size. Each neighborhood pixel value takes the gray value of the central pixel as the threshold value for binary quantization. The range of regularization parameters of SVM is $[10^{-1} \sim 10^7]$, the input nodes and output nodes are the classification numbers of the corresponding hyperspectral data. The lower bound is 0.01, and the upper bound is 100, and the convergence factor $\vec{a} \in [20]$, $\vec{C}_1 = \vec{C}_2 = \vec{C}_3 = 2 * rand()$, the number of iterations is 400.

C. Experimental Results and Analysis

To verify the classification performance of the PLG-KELM, the experiments are executed on three datasets. For each dataset, random 10% and 5% of all labeled pixels are chosen from each class of ground and objects as the training set, and the remaining labeled pixels are the test set. The overall accuracy (OA), average accuracy (AA), and Kappa coefficient of classification results are used as evaluating indicators under different training set and test set.

1) Indian Pines Dataset: It contains 16 different feature categories. The experimental results of the all methods are shown in Table V, where the obtained best results are bold highlight. The comparison results of the OA, AA, and Kappa coefficient of different methods for 10% training samples are shown in Fig. 4. The classification effects of the all methods are shown in Fig. 5.

In Fig. 5, (a)–(g) are the original image, BLS classification effect, SVM classification effect, PCA-CNN classification effect, CEA-CNN classification effect, KELM classification effect, and PLG-KELM classification effect. From Table V and Fig. 4, it can be seen that the OA, AA, and Kappa coefficient of the PLG-KELM are increased by 3.35%, 5.56%, and 3.82% for 10% training set, and 3.76%, 5.81%, and 4.32% for 5% training set by comparing with the BLS, SVM, PCA-CNN, CAE-CNN, and KELM, respectively. The experiment results show that the PLG-KELM has higher classification accuracy than the compared methods. The OA, AA, and Kappa coefficient of KELM reach 99%, 98.65%, and 98.86% for 10% training set, respectively, which are obviously better than the SVM, BLS, PCA-CNN, and CAE-CNN. However, the classification performance of the PLG-KELM is further improved by using GWO algorithm,

hyperspectral image analysis team and NCALM on the campus of the University of Houston and the nearby urban area. The size of the image is 349 × 1905. Pavia University dataset is a hyperspectral image dataset from Pavia University area in northern Italy. The size is 610 × 340. The basic information of three datasets are shown in Table I–Table IV.

TABLE V
CLASSIFICATION RESULTS ON INDIAN PINES DATASET (%)

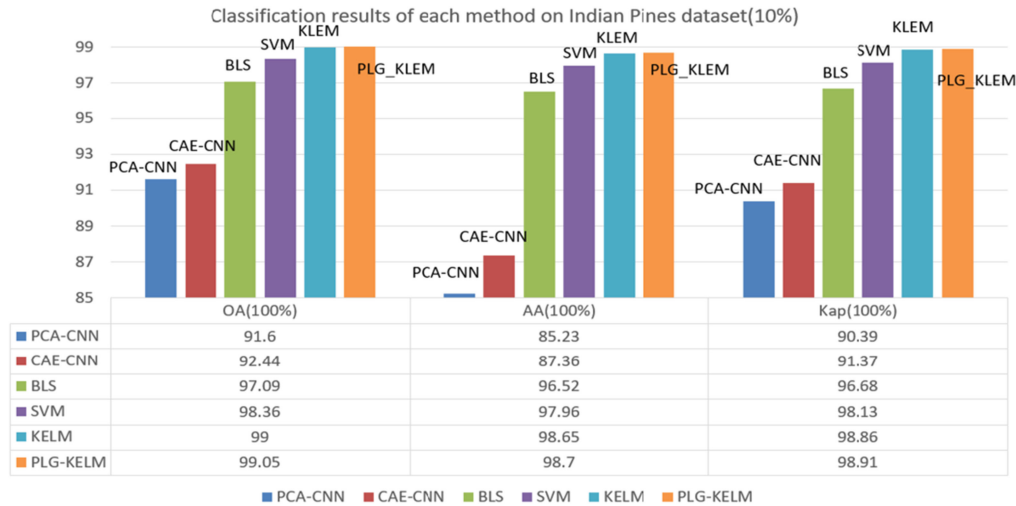| Class	Training set (5%)						Training set (10%)					
	SVM	BLS	PCA-CNN	CAE-CNN	KELM	PLG-KELM	SVM	BLS	PCA-CNN	CAE-CNN	KELM	PLG-KELM
1	**100.0**	**100.0**	62.75	72.55	96.08	**100.0**	100.0	97.92	81.25	54.17	**100.0**	**100.0**
2	96.77	94.57	84.65	85.24	96.77	**96.84**	98.84	96.82	86.36	89.92	99.69	**99.77**
3	98.74	95.96	92.55	87.88	98.86	**99.24**	98.93	98.67	87.87	88.13	**99.47**	99.33
4	**95.05**	93.24	83.78	81.08	**95.05**	**95.05**	**96.19**	94.76	84.29	91.43	95.71	95.71
5	96.61	94.70	74.36	89.62	**97.67**	97.46	97.99	95.97	94.63	93.51	98.43	**99.11**
6	96.47	93.94	**99.58**	98.73	95.91	96.05	**99.85**	97.62	98.66	98.51	**99.85**	**99.85**
7	**100.0**	**100.0**	91.67	**100.0**	**100.0**	**100.0**	100.0	91.30	60.87	91.30	**100.0**	**100.0**
8	**100.0**	**100.0**	99.78	96.34	**100.0**	**100.0**	100.0	**100.0**	99.32	99.55	**100.0**	**100.0**
9	78.95	**90.47**	0.00	0.00	84.21	84.21	94.44	**100.0**	33.33	38.89	**100.0**	**100.0**
10	88.36	**93.80**	80.96	71.38	90.75	90.86	96.21	93.69	83.70	82.32	98.16	**98.28**
11	**97.99**	95.69	90.02	88.31	97.95	97.87	98.56	98.42	96.04	94.06	99.55	**99.59**
12	**79.76**	76.84	67.58	76.67	77.36	77.87	95.29	91.30	75.91	88.59	**95.65**	**95.65**
13	95.02	95.52	99.00	**99.50**	96.02	95.52	98.42	96.84	**100.0**	98.95	99.47	99.47
14	**99.19**	98.70	97.64	98.37	98.86	98.94	99.19	98.88	**99.66**	98.71	99.23	99.14
15	89.81	86.23	81.54	69.97	95.87	**96.97**	99.42	99.71	86.55	93.27	99.12	99.12
16	86.67	82.22	**93.33**	**93.33**	87.78	87.78	94.12	85.88	95.29	**96.47**	94.12	94.12
OA(%)	95.12	94.24	88.10	87.24	95.71	**95.84**	98.36	97.09	91.60	92.44	99.00	**99.05**
AA(%)	93.71	93.24	81.20	81.81	94.32	**94.67**	97.96	96.52	85.23	87.36	98.65	**98.70**
Kapp(%)	94.38	93.35	86.43	85.44	95.12	**95.26**	98.13	96.68	90.39	91.37	98.86	**98.91**
Tra-time(s)	114.91	56.94	250.45	301.91	57.49	175.59	156.08	57.57	377.26	372.01	57.86	224.86

Fig. 4. Classification comparison results on Indian pines dataset.

and its OA, AA, and kappa coefficient are averagely increased by 0.05%. The OA, AA, and Kappa coefficient of KELM are 95.71%, 94.32%, and 95.12%, respectively, when the training set data are reduced by 5% in small samples. The PLG-KELM improves OA, AA, and Kappa coefficient by 0.13%, 0.35%, and 0.14%, respectively. It can be known that the optimized KELM has some advantages of the KELM in classification of high-dimensional data. It can also be known that the OA and kappa coefficient of the PLG-KELM are, respectively, improved by 0.08%, 0.3%, and 0.09%, which indicates that the PLG-KELM has better generalization ability.

2) Houston 2013 Dataset: It contains 15 different feature categories. The experimental results of the all methods are shown

in Table VI. The comparison results of the OA, AA, and Kappa coefficient for 10% training samples are shown Fig. 6. The classification effects of the all methods are shown in Fig. 7.

In Fig. 7, (a)–(g) are the original image, BLS classification effect, SVM classification effect, PCA-CNN classification effect, CEA-CNN classification effect, KELM classification effect, and PLG-KELM classification effect. From Table VI and Fig. 6, it can be seen that the OA, AA, and Kappa coefficient of the PLG-KELM are increased by 3.27%, 2.77%, and 4.06% for 10% training set, and 2.14%, 2.17%, and 2.32% for 5% training set by comparing with the BLS, SVM, PCA-CNN, CAE-CNN, and KELM, respectively. The PLG-KELM shows better classification results than the compared methods. For
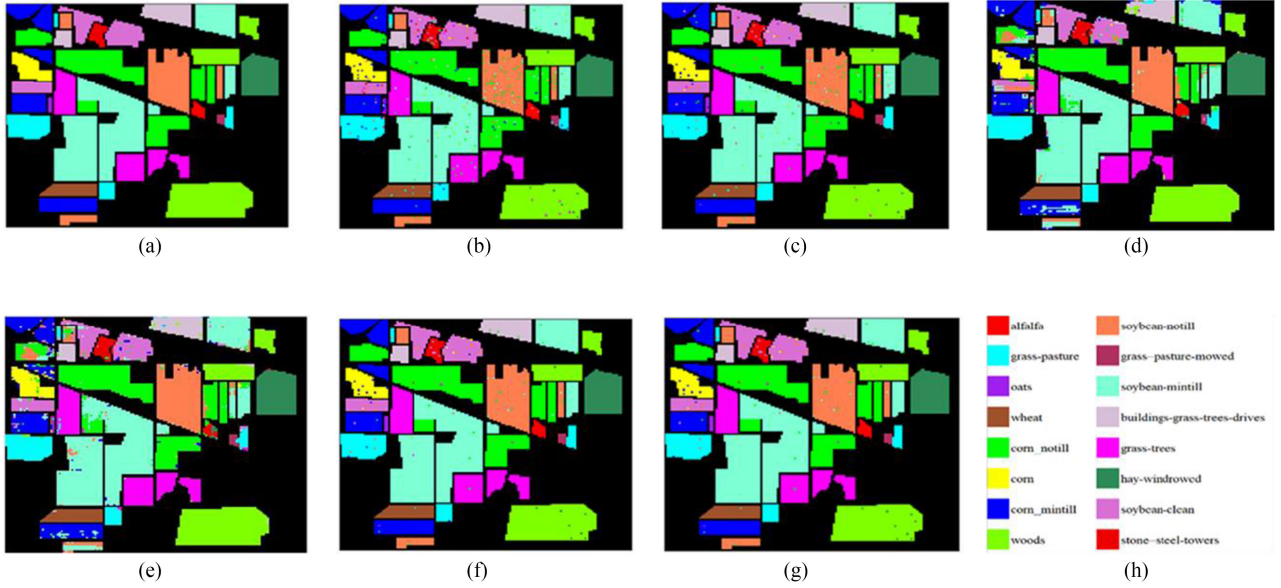
Fig. 5. Classification effect of the all methods on Indian Pines dataset.

TABLE VI
CLASSIFICATION RESULTS ON HOUSTON DATASET (%)

Class	Training set (5%)						Training set (10%)					
	SVM	BLS	PCA-CNN	CAE-CNN	KELM	PLG-KELM	SVM	BLS	PCA-CNN	CAE-CNN	KELM	PLG-KELM
1	**95.03**	88.03	94.86	85.16	93.51	92.92	94.83	84.83	**97.59**	98.84	93.49	94.29
2	**95.71**	88.23	86.88	85.87	92.26	93.94	98.93	91.81	94.93	87.19	98.84	98.58
3	**99.55**	99.70	99.85	96.39	**99.55**	98.95	98.89	99.53	**100.0**	100.0	98.42	98.89
4	84.99	83.97	93.89	**98.81**	83.38	83.72	92.01	82.14	**99.37**	94.25	88.24	90.13
5	98.22	90.99	**100.0**	**100.0**	98.13	98.22	**100.0**	96.58	99.91	99.28	**100.0**	**100.0**
6	96.47	82.69	93.59	88.14	97.76	**98.40**	98.33	95.65	95.32	96.66	99.00	**99.33**
7	89.94	79.80	87.45	91.44	89.03	**92.02**	90.95	88.75	91.65	90.25	91.65	**92.62**
8	90.33	82.78	74.55	79.39	90.92	**91.86**	96.77	90.13	91.65	80.97	96.23	96.41
9	85.85	85.76	86.77	88.12	85.17	**85.93**	98.31	83.78	94.92	89.13	97.50	98.13
10	99.74	100.00	83.99	88.12	99.91	**100.0**	99.73	99.91	87.24	63.45	**100.0**	**100.0**
11	99.74	98.97	96.50	89.15	**100.0**	**100.0**	**100.0**	99.46	99.00	97.83	**100.0**	**100.0**
12	95.38	93.84	93.84	**97.17**	95.89	95.72	**98.37**	93.20	97.19	96.19	98.73	98.01
13	80.70	71.49	**93.20**	92.54	86.18	84.65	90.97	71.11	92.78	**95.71**	95.94	95.26
14	**100.0**	91.81	99.76	99.76	**100.0**	**100.0**	**100.0**	99.75	99.75	99.50	**100.0**	**100.0**
15	97.13	95.86	99.84	**100.0**	97.13	99.52	**100.0**	100.0	100.0	100.0	99.83	**100.0**
OA(%)	93.76	89.34	91.24	91.32	93.40	**93.95**	97.12	88.27	95.81	91.33	96.83	**97.14**
AA(%)	93.92	88.93	92.33	92.00	93.92	**94.39**	97.21	90.25	96.09	92.62	97.19	**97.44**
Kapp(%)	93.25	88.47	90.54	90.62	92.87	**93.47**	96.88	84.67	95.48	90.63	96.58	**96.91**
Tra-time(s)	66.21	7.96	258.96	263.78	0.90	19.00	74.36	8.27	391.80	403.02	1.24	31.00

10% training set, the OA, AA, and Kappa coefficient of the KELM reach 96.83%, 97.19%, and 96.58%, respectively. The experiment results show that the LBP is more effective in extracting local spatial texture features in high-resolution dataset. On this basis, the classification accuracy of the PLG-KELM is further improved, and the OA, AA, and Kappa coefficient are increased by 0.31%, 0.25%, and 0.33%, respectively. The OA, AA, and Kappa coefficient of KELM are 93.40%, 93.92%, and 92.87%, when the training set is reduced by 5%. On this basis, the OA, AA, and Kappa coefficient of the PLG-KELM

are increased by 0.55%, 0.47%, and 0.6%, respectively, which still has better classification accuracy in small samples. The experiment results show that the PLG-KELM can effectively extract spatial spectrum features, and has better classification accuracy for hyperspectral images.

3) Pavia University Dataset: It contains 9 different feature categories. The experimental results of the all methods are shown in Table VII. The comparison results of OA, AA, and Kappa coefficient for 10% training samples are shown in Fig. 8. The classification effect of the all methods are shown in Fig. 9.

Classification results of each method on Houston dataset(10%)

	OA(100%)	AA(100%)	Kap(100%)
■ CAE-CNN	91.33	92.62	90.63
■ BLS	88.27	90.25	84.67
■ PCA-CNN	95.81	96.09	95.48
■ KELM	96.83	97.19	96.58
■ SVM	97.12	97.21	96.88
■ PLG-KELM	97.14	97.44	96.91

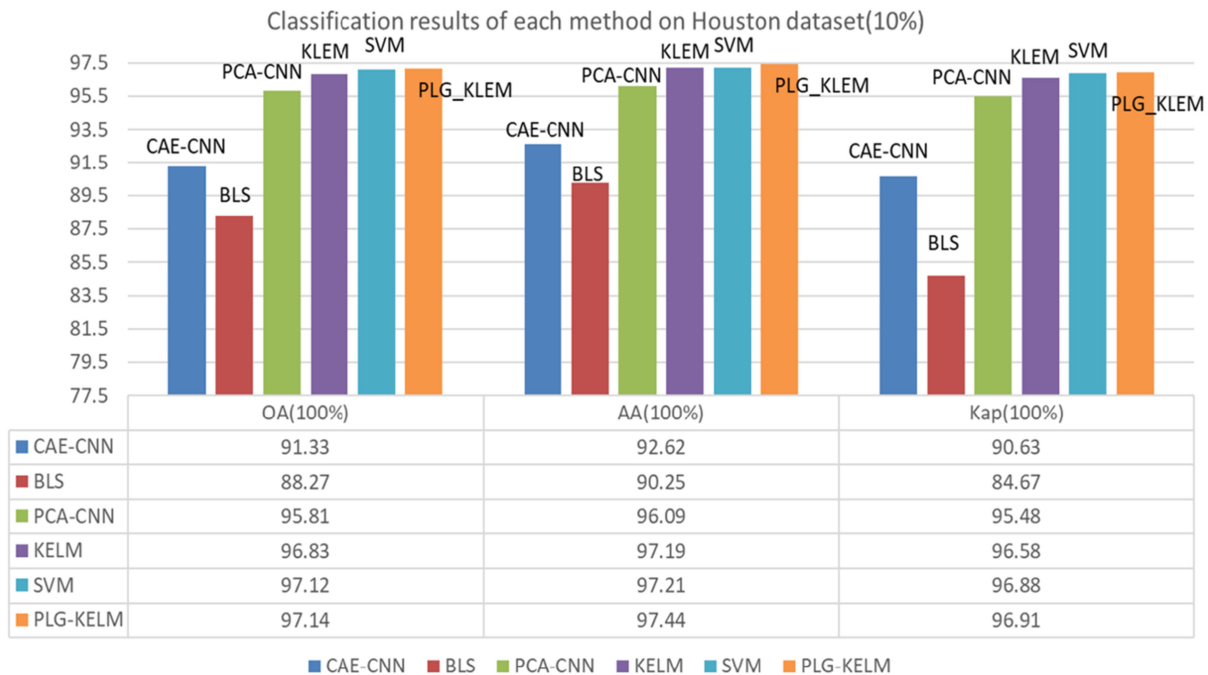■ CAE-CNN ■ BLS ■ PCA-CNN ■ KELM ■ SVM ■ PLG-KELM

Fig. 6.   Classification comparison results on Houston dataset.
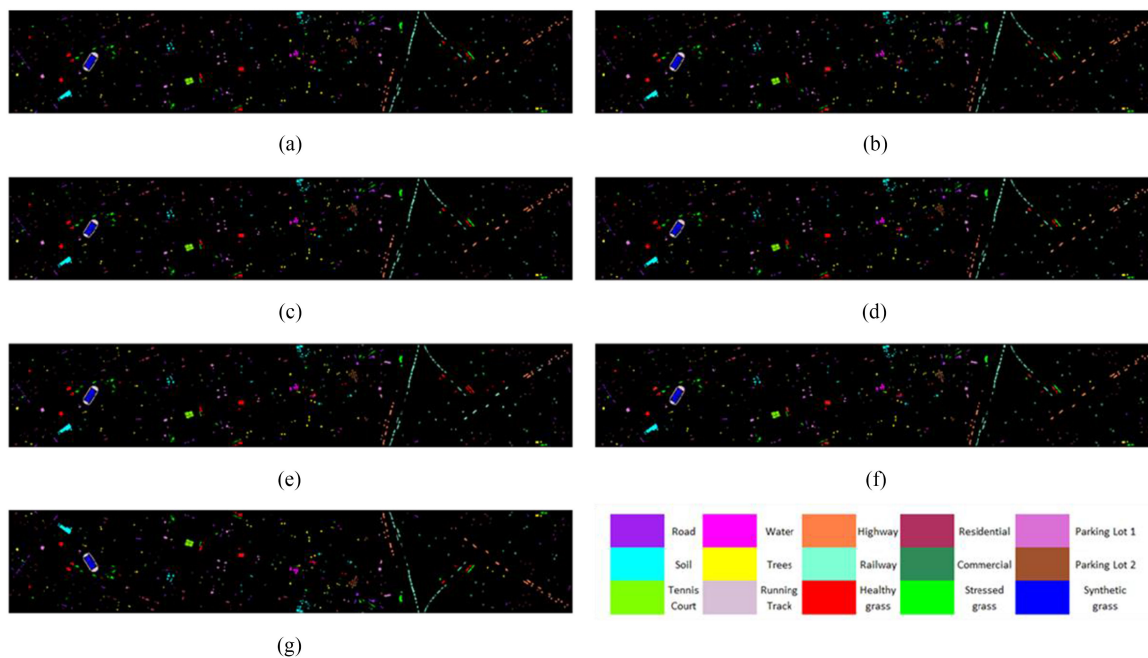

Fig. 7.   Classification effect of the all methods on Houston dataset.

In Fig. 9, (a)–(g) are the original image, BLS classification effect, SVM classification effect, PCA-CNN classification effect, CEA-CNN classification effect, KELM classification effect, and PLG-KELM classification effect. There are only 9 kinds of features in Pavia University dataset, which is 7 less than Indian pines dataset. The number of available effective wavebands is the least among three datasets. From Table VII and Fig. 8, it can be known that the OA, AA, and Kappa coefficient of the PLG-KELM are increased by 3.32%, 2.31%, and 4.10% for 10% training set and 3.80%, 2.23%, and 4.53% for 5% training set by comparing with the BLS, SVM, PAC-CNN, CAE-CNN, and KELM, respectively. The experiment results show that the PLG-KELM has better classification effect than the other compared methods. For 10% training set, compared with the classification results of the SVM, the OA, AA, and Kappa coefficient of the KELM are decreased by 0.71%, 1.19%, and 0.98%, and the OA, AA, and Kappa coefficient of the PLG-KELM are increased by 0.18%, 0.04%, and 0.24%, respectively. The experiment results

TABLE VII
CLASSIFICATION RESULTS ON PAVIA UNIVERSITY DATASET (%)

	Training set (5%)						Training set (10%)					
Class	SVM	BLS	PCA-CNN	CAE-CNN	KELM	PLG-KELM	SVM	BLS	PCA-CNN	CAE-CNN	KELM	PLG-KELM
1	95.58	78.13	96.28	**97.21**	90.81	96.25	99.01	74.55	98.64	**99.35**	96.85	98.86
2	96.10	86.48	91.31	88.10	**97.44**	97.29	98.28	80.16	98.09	**99.70**	97.98	98.71
3	98.68	95.51	95.56	99.00	98.20	**98.78**	99.51	96.31	98.52	**99.51**	99.01	99.45
4	94.17	83.69	99.21	**99.39**	90.95	93.17	99.40	95.37	**100.0**	99.95	97.79	99.35
5	**100.0**	99.72	**100.0**	**100.0**	**100.0**	**100.0**	99.81	98.97	**100.0**	100.00	98.88	99.91
6	**100.0**	98.84	**100.0**	99.79	99.96	**100.0**	**100.0**	98.34	99.92	94.45	99.94	**100.0**
7	99.63	98.08	99.54	99.45	**99.73**	**99.73**	99.72	96.11	**100.0**	99.91	99.53	99.81
8	99.12	98.01	97.51	98.33	98.89	**99.24**	99.50	97.10	97.32	98.38	99.22	**99.53**
9	97.79	73.78	**100.0**	**100.0**	81.15	97.79	99.79	73.31	99.79	100.00	95.13	99.79
OA(%)	96.95	88.27	93.35	93.89	96.26	**97.54**	98.92	85.26	98.05	98.48	98.21	99.10
AA(%)	97.90	90.25	97.71	97.92	95.24	**98.03**	99.45	90.02	99.14	99.03	98.26	99.49
Kapp(%)	95.91	84.67	93.35	91.94	94.97	**96.70**	98.52	80.72	98.05	98.48	97.54	98.76
Tra-time(s)	330.59	18.90	517.25	312.53	6.44	89.53	533.08	19.27	814.15	829.72	9.80	150.27
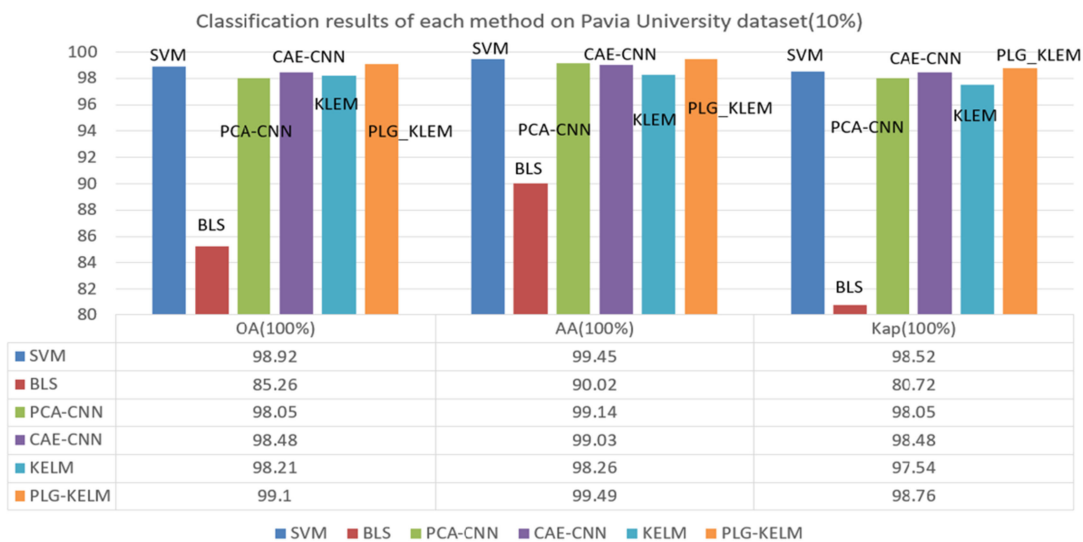
Fig. 8. Classification comparison results of the all methods on Pavia University dataset.
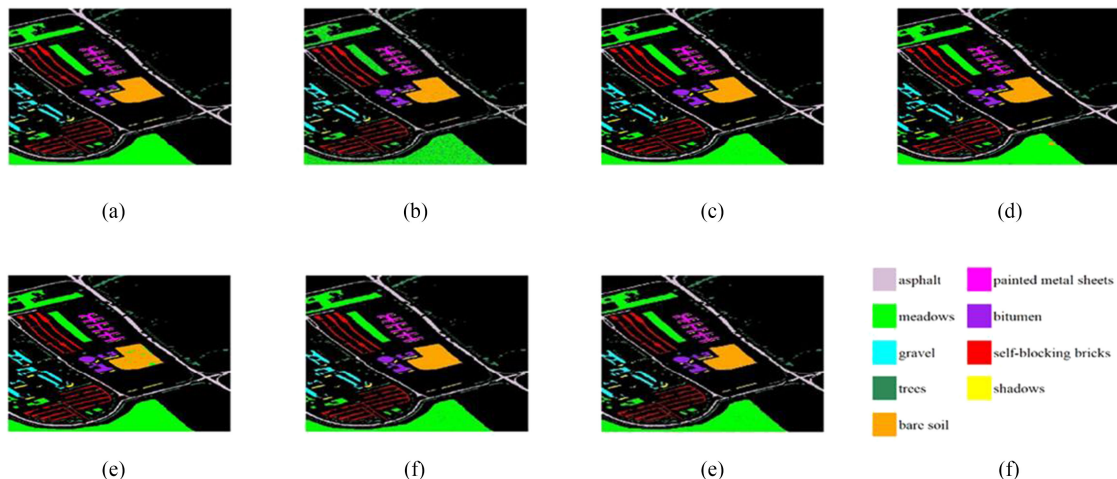

(a) (b) (c) (d)

(e) (f) (e) (f)

Fig. 9. Classification effect of the all methods on Pavia University dataset.

TABLE VIII
COMPARISON RESULTS OF KELM AND PLG-KELM FOR DIFFERENT DATA (%)

Indicators	Models	Indian pines		Houston		Pavia University	
		5%	10%	5%	10%	5%	10%
OA(%)	KELM	95.71	99.00	93.40	96.83	96.26	98.21
	PLG-KELM	95.84	99.05	93.95	97.14	97.54	99.10
AA(%)	KELM	94.32	98.65	93.92	97.19	95.24	98.26
	PLG-KELM	94.67	98.70	94.39	97.44	98.03	99.49
Kapp(%)	KELM	95.12	98.86	92.87	96.58	94.97	97.54
	PLG-KELM	95.26	98.91	93.47	96.91	96.70	98.76

TABLE IX
BASIC INFORMATION AND FEATURES STATISTICS

Data	WHU-Hi-LongKou
Collection location	Hubei province, China
Acquisition equipment	Headwall Nano-Hyperspec
Spectral coverage(μm)	0.4~1
Data size(pixel)	550×400
Spatial resolution(m)	0.463
Number of bands	270
Sample size	204542
Number of categories	9

TABLE X
GROUND AND OBJECTS OF 9 TYPES IN WHU-HI-LONGKOU

Category	Class name	Number of samples
1	Corn	34511
2	Cotton	8374
3	Sesame	3031
4	Broad-leaf soybean	63212
5	Narrow-leaf soybean	4151
6	Rice	11854
7	Water	67056
8	Roads and houses	7124
9	Mixed weed	5229

show that the parameter optimization of the KELM can further improve the classification performance. For 5% training set, compared with the classification results of the KELM, the OA, AA, and Kappa coefficient of the PLG-KELM are increased by 1.28%, 2.79%, and 1.73%, respectively. Compared with the 10% training set, the improvement of PLG-KELM is more significant, and the PLG-KELM can obtain better classification result for high-dimensional data. For small-scale training set, for the PLG-KELM, the OA, AA, and Kappa coefficient for 5% training set are 0.39%, 1.56%, and 0.51% higher than those for 10% training set, which indicates that the PLG-KELM has better generalization ability for hyperspectral image classification.

D. Comparison of Parameter Optimization Effect for KLEM

In order to verify the optimization performance of GWO for parameter optimization of the KLEM, the comparison results of OA, AA, and Kappa coefficient of the KELM and PLG-KELM for Indian pines dataset, Houston dataset, and Pavia University dataset are shown in Table VIII.

It can be seen from the Table VIII, for all indicators of OA, AA, and Kappa coefficient, the classification accuracies of the PLG-KELM are better than those of the KELM for three datasets of Indian pines, Houston 2013, and Pavia University for 5% training set and 10% training set. The experiment results show that the PLG-KELM has better classification ability. The reason is that the GWO algorithm with global optimization capability can effectively optimize the parameters of the KELM and obtain the ideal values of the parameters for KELM. Therefore, the GWO algorithm is an intelligent optimization algorithm based on population, it is easy to implement and has good global performance, which can effectively optimize the parameters of classification models.

VI. APPLICATION IN FINE CLASSIFICATION OF CROPS

In order to verify the classification performance and generalization ability of the PLG-KELM for fine classification of crops, an actual application of WHU-Hi-LongKou hyperspectral images is selected in this article. The dataset of WHU-Hi-LongKou was obtained by UAV hyperspectral remote sensing observation in LongKou, Hubei Province, China, by RSIDEA research team of State Key Laboratory of surveying, mapping and Remote Sensing Information Engineering [58], [59]. This dataset started from the application demand of fine classification of crops. It is an open and shared hyperspectral hyperspace remote sensing

images of fine classification dataset. The image size is 550×400 pixels. The basic information and features statistics of the dataset are shown in Table IX. The ground and objects of 9 types in WHU-Hi-Longkou is shown in Table X.

In the actual application, the classification models of SVM, BLS, PCA-CNN, CAE-CNN, and KELM are still selected to compare with the PLG-KELM. In addition, in order to verify the classification performance and generalization ability of the PLG-KELM for small samples, random 0.5% and 1% of all labeled pixels in WHU-Hi-LongKou dataset are chosen from each class of ground and objects as the training samples, and the remaining labeled pixels are the test samples. The OA, AA, and Kappa coefficient of classification results are still used as evaluating indicators. The experimental results of the all methods are shown in Table XI. The comparison results of the OA, AA, and Kappa coefficient for 1% training set are shown Fig. 10. The classification effect of the all methods are shown in Fig. 11.

In Fig. 11, (a)–(g) are the original image, BLS classification effect, SVM classification effect, PCA-CNN classification effect, CEA-CNN classification effect, KELM classification effect, and PLG-KELM classification effect. There are only 9 kinds of features in WHU-Hi-LongKou dataset. As can be seen from Table XI and Fig. 10, compared with SVM, BLS, PCA-CNN, CAE-CNN, and KELM, the OA, AA, and Kappa coefficient of the PLG-KELM are increased by 1.65%, 6.17%, and 2.17% for 1% training set, and 5.01%, 6.88%, and 6.15% for 0.5% training set, which show that the PLG-KELM has higher classification accuracy than those of the other compared

TABLE XI
CLASSIFICATION RESULTS ON WHU-HI-LONGKOU DATASET (%)

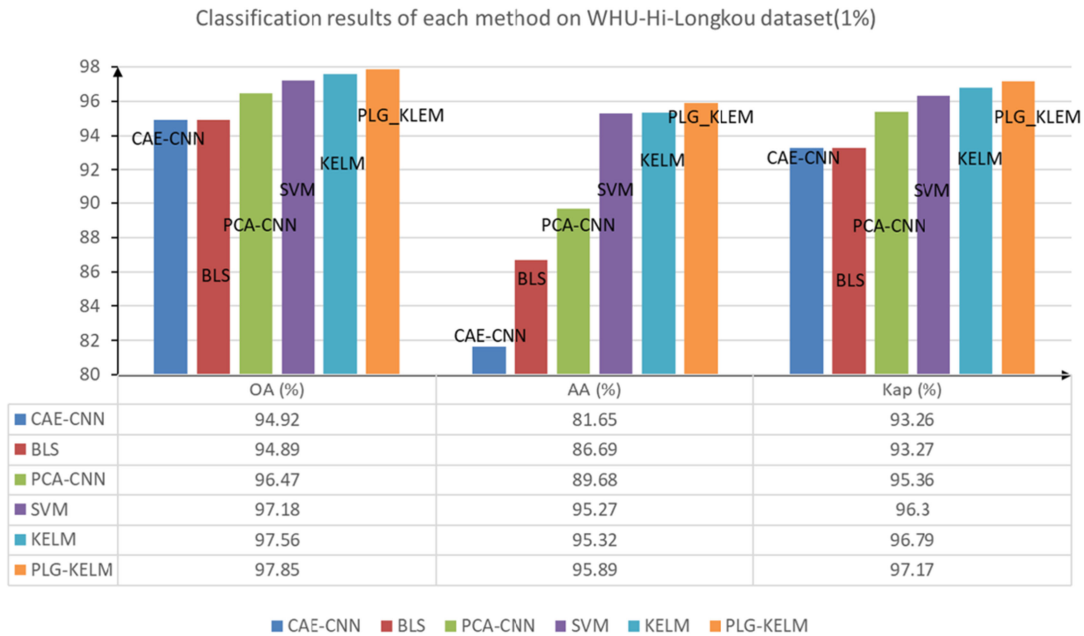| Class	Training set (0.5%)						Training set (1%)					
	SVM	BLS	PCA-CNN	CAE-CNN	KELM	PLG-KELM	SVM	BLS	PCA-CNN	CAE-CNN	KELM	PLG-KELM
1	96.76	96.88	**98.99**	96.76	97.44	97.68	98.06	97.20	**99.92**	99.86	98.58	99.08
2	97.98	97.35	19.55	73.26	98.18	**98.26**	98.89	96.79	82.61	47.88	98.64	98.71
3	95.51	77.70	96.15	0.23	96.89	**98.06**	99.18	69.19	80.55	49.00	99.49	**99.97**
4	96.15	96.18	36.95	**99.33**	97.05	97.18	96.57	96.36	96.33	**98.40**	97.56	97.83
5	88.72	67.07	**95.93**	87.94	88.41	92.28	97.09	71.83	74.64	90.54	97.69	**98.55**
6	89.66	87.41	95.10	**98.36**	87.72	88.57	94.92	86.65	96.10	**98.33**	95.16	95.72
7	98.59	97.36	99.85	**99.83**	98.86	98.95	99.29	98.53	99.95	**99.98**	99.34	99.34
8	84.30	77.16	66.21	**90.87**	82.84	83.87	88.05	80.67	**94.02**	84.38	87.71	87.68
9	82.41	76.18	94.73	**87.54**	78.18	80.32	85.37	82.95	83.05	66.47	83.69	**86.15**
OA(%)	95.83	94.20	75.26	95.66	96.06	**96.41**	97.18	94.89	96.47	94.92	97.56	**97.85**
AA(%)	92.23	85.92	78.16	81.57	91.73	**92.80**	95.27	86.69	89.68	81.65	95.32	**95.89**
Kapp(%)	94.53	92.36	69.74	94.22	94.83	**95.29**	96.30	93.27	95.36	93.26	96.79	**97.17**
Tra-time(s)	52.51	21.31	363.66	372.08	5.13	5.58	82.78	21.93	397.59	395.63	15.37	17.37

Fig. 10.   Classification results on WHU-Hi-LongKou dataset.

methods. For 1% training set, the OA, AA, and Kappa coefficient of the KELM are 97.85%, 95.89%, and 97.17%, respectively, which show that the KELM is superior to SVM, BLS, PCA-CNN, CAE-CNN, and KELM. The OA, AA, and Kappa coefficient of the PLG-KELM are further improved by 0.29%, 0.57%, and 0.38% than those of the KELM. For 0.5% training set, the OA, AA, and Kappa coefficient of the KELM are 96.06%, 91.73%, and 94.83%, respectively. The OA, AA, and Kappa coefficient of the PLG-KELM are further improved by 0.35%, 1.07%, and 0.46% than those of the KELM. Compared with the 1% training set, the improvement of the PLG-KELM is more significant. It can be concluded that the PLG-KELM can better play the classification advantage of KELM for high-dimensional data. For the PLG-KELM, it can be seen that the OA, AA, and Kappa coefficient of the 0.5% training set are improved by 0.06%, 0.5%, and 0.08% than those of 1% training set in small samples, which indicate that the PLG-KELM has better classification accuracy and generalization ability.

VII. DISCUSSION

It can be known from the results that the PLG-KELM is compared with the SVM, BLS, PCA-CNN, CAE-CNN, and KELM methods. For Indian pines dataset, Houston dataset, and Pavia University dataset, the OA, AA, and Kappa coefficient of the PLG-KELM are increased in various degrees, respectively. For actual WHU-Hi-LongKou dataset, the OA, AA, and Kappa coefficient of the PLG-KELM are increased by 1.65%, 6.17%, and 2.17% for 1% training set, and 5.01%, 6.88%, and 6.15% for 0.5% training set, respectively. The experiment results show
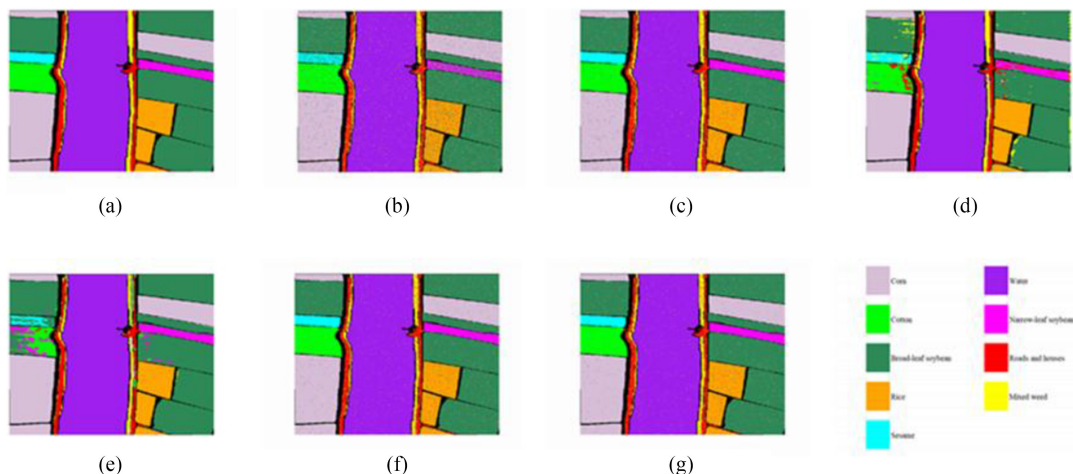
Fig. 11.   Classification effect on WHU-Hi-LongKou dataset.

that the PCA and LBP can better extract the complex features from the hyperspectral images, and the PLG-KELM can effectively remove the phenomenon of salt and pepper. At the same time, the PLG-KELM can also effectively improve the classification accuracy, avoid the sample misclassification at the class boundary, and maintain the smoothness and consistency of the classification results. And it takes on better classification performance and generalization ability for small samples. But, the operation efficiency is slightly lower than three compared methods.

VIII. CONCLUSION AND FUTURE WORK

For the hyperspectral images with the serious same object and different spectrum and foreign matter with the same spectrum phenomenon, the complex distribution of ground and objects, the large spatial scale difference, the few labeled samples, the diverse noise types, and the traditional image classification methods are difficult to realize better classification effect. Therefore, a hyperspectral image classification method based on integrating PCA, LBP, GWO, and KELM, namely PLG-KELM is proposed in this article. Compared with the BLS, SVM, PCA-CNN, CAE-CNN, and KELM, the OA, AA, and Kappa coefficient of the PLG-KELM are, respectively, increased for 10% training set and 5% training set for Indian pines dataset, Houston, and Pavia University dataset, and for 1% and 0.5% training set for WHU-Hi-LongKou dataset. The experiment results show that the PLG-KELM has a significant advantage of classification effect, which can realize a balance between local optimization and global search. And it achieves a better classification effect, and effectively avoid the "pepper and salt misclassification" problem. In addition, the PLG-KELM takes on better classification performance and generalization ability for small samples.

Due to the lower operation efficiency of the PLG-KELM, we think about how to improve the operation efficiency of the PLG-KELM in the future work.

REFERENCES

[1] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Proc. Mag.*, vol. 19, no. 1, pp. 17–28, Aug. 2002.

[2] F. Liu and Z. Zhou, "A new data classification method based on chaotic particle swarm optimization and least square-support vector machine," *Chemometr. Intell. Lab.*, vol. 147, no. 1, pp. 147–156, Oct. 2015.

[3] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state-of-the-art," *IEEE Trans. Geosci. Remote Sens*, vol. 4, no. 2, pp. 22–40, Feb. 2016.

[4] J. Li, X. Huang, and P. Gamba, "Multiple feature learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1592–1606, Mar. 2015.

[5] J. Li, M. Khodadadzadeh, A. Plaza, X. Jia, and J. M. Bioucas-Dias, "A discontinuity preserving relaxation scheme for spectral-spatial hyperspectral image classification," *IEEE J. Sel. Topic Appl. Earth Obs. Remote Sens.*, vol. 9, no. 2, pp. 625–639, Feb. 2016.

[6] J. J. Liu, Z. B. Wu, J. Li, A. Plaza, and Y. Yuan, "Probabilistic-kernel collaborative representation for spatial-spectral hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 2371–2384, Apr. 2016.

[7] X. Xu, J. Li, X. Huang, M. Dalla Mura, and A. Plaza, "Multiple morphological component analysis based decomposition for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 3083–3102, May 2016.

[8] Y. Liu, J. Li, P. J. Du, A. Plaza, X. Jia, and X. Zhang, "Class-oriented spectral partitioning for remotely sensed hyperspectral image classification," *IEEE J. Sel. Topic Appl. Earth Obs. Remote Sens.,* vol. 10, no. 2, pp. 695–711, Feb. 2017.

[9] Z. H. Xue, P. J. Du, J. Li, and H. Su, "Sparse graph regularization for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 2351–2366, Apr. 2017.

[10] X. Xu, J. Li, and S. T. Li, "Multiview intensity-based active learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 669–680, Feb. 2018.

[11] H. M. Juan, M. E. Paoletti, J. Plaza, J. Li, and A. Plaza, "Active learning with convolutional neural networks for hyperspectral image classification using a new Bayesian approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6440–6461, Nov. 2018.

[12] C. Y. Yu, B. Xue, M. P. Song, Y. Wang, S. Li, and C. Chang, "Iterative target-constrained interference-minimized classifier for hyperspectral classification," *IEEE J. Sel. Topic Appl. Earth Obs. Remote Sens.,* vol. 11, no. 4, pp. 1095–1117, Apr. 2018.

[13] C. I. Chang, "Spectral inter-band discrimination capacity of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1749–1766, Mar. 2018.

[14] C. Y. Liu, J. Li, and L. He, "Superpixel-based semisupervised active learning for hyperspectral image classification," *IEEE J. Sel. Topic Appl. Earth Obs. Remote Sens.,* vol. 12, no. 1, pp. 357–370, Jan. 2019.

[15] C. I. Chang and Y. Li, "Recursive band processing of automatic target generation process for finding unsupervised targets in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 9, pp. 5081–5094, Sep. 2019.

[16] S. W. Zhong, Y. Zhang, and C. I. Chang, "A spectral-spatial feedback close network system for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 10056–10069, Dec. 2019.

[17] C. Y. Yu, Y. L. Wang, M. P. Song, and C. Chang, "Class signature-constrained background-suppressed approach to band selection for classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 14–31, Jan. 2019.

[18] C. Y. Yu *et al.*, "Hyperspectral image classification method based on CNN architecture embedding with hashing semantic feature," *IEEE J. Sel. Topic Appl. Earth Obs. Remote Sens.*, vol. 12, no. 6, pp. 1866–1881, Jun. 2019.

[19] M. P. Song, X. D. Shang, Y. L. Wang, C. Yu, and C. Chang, "Class information-based band selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens*, vol. 57, no. 11, pp. 8394–8416, Nov. 2019.

[20] C. Y. Yu, R. Han, M. P. Song, C. Liu, and C. -I. Chang, "A simplified 2D-3D CNN architecture for hyperspectral image classification based on spatial-spectral fusion," *IEEE J. Sel. Topic Appl. Earth Obs. Remote Sens.*, vol. 13, no. 4, pp. 2485–2501, Apr. 2020.

[21] H. Y. Yu, X. Zhang, C. Y. Yu, L. Gao, and B. Zhang, "Neighborhood activity-driven representation for hyperspectral imagery classification," *IEEE J. Sel. Topic Appl. Earth Obs. Remote Sens.*, vol. 13, no. 4, pp. 4506–4517, Aug. 2020.

[22] D. F. Hong, X. Wu, P. Ghamisi, J. Chanussot, N. Yokoya, and X. Zhu, "Invariant attribute profiles: A spatial-frequency joint feature extractor for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 3791–3808, Jun. 2020.

[23] Z. L. Zhong, J. Li, D. A. Clausi, and A. Wong, "Generative adversarial networks and conditional random fields for hyperspectral image classification," *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 3318–3329, Jul. 2020.

[24] F. L. Luo, L. P. Zhang, B. Du, and L. Zhang, "Dimensionality reduction with enhanced hybrid-graph discriminant learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5336–5353, Aug. 2020.

[25] Q. C. Liu, L. Xiao, J. X. Yang, and J. C. -W. Chan, "Content-guided convolutional neural network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 9, pp. 6124–6137, Sep. 2020.

[26] A. Samat, J. Li, C. Lin, S. Liu, and E. Li, "Edge gradient-based active learning for hyperspectral image classification," *IEEE Geosci. Remote. Sens.*, vol. 17, no. 9, pp. 1588–1592, Sep. 2020.

[27] M. P. Song, C. Y. Yu, H. Y. Xie, and C. -I. Chang, "Progressive band selection processing of hyperspectral image classification," *IEEE Geosci. Remote. Sens.*, vol. 17, no. 10, pp. 1762–1766, Oct. 2020.

[28] L. Sun *et al.*, "Low rank component induced spatial-spectral kernel method for hyperspectral image classification," *IEEE Trans. Circ. Syst. Video Tech.*, vol. 30, no. 10, pp. 3829–3842, Oct. 2020.

[29] Y. H. Chu *et al.*, "Hyperspectral image classification based on discriminative locality broad," *Knowl. Based Syst.*, vol. 206, Oct. 2020, Art. no. 106319.

[30] M. P. Song, X. D. Shang, and C. I. Chang, "3-D receiver operating characteristic analysis for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 80937–88115, Nov. 2020.

[31] Z. J. Yang, F. X. Cao, Y. Q. Cheng, W. -K. Ling, and R. Hu, "Locality regularized robust-PCRC: A novel simultaneous feature extraction and classification framework for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8567–8582, Dec. 2020.

[32] L. C. Mou, X. Q. Lu, X. L. Li, and X. X. Zhu, "Nonlocal graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8246–8257, Dec. 2020.

[33] J. Jiang, J. Ma, C. Chen, Z. Wang, Z. Cai, and L. Wang, "Super PCA: A superpixelwise PCA approach for unsupervised feature extraction of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 4581–4593, Aug. 2020.

[34] W. Li, C. Chen, H. Su, and Q. Du, "Local binary patterns and extreme learning machine for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3681–3693, Jul. 2015.

[35] K. Satoru, Z. Abdallah, and M. Farid, "Spatial and structured SVM for multilabel image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5948–5960, Oct. 2018.

[36] W. Deng, H. L. Liu, J. J. Xu, H. Zhao, and Y. Song, "An improved quantum-inspired differential evolution algorithm for deep belief network," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 10, pp. 7319–7327, Oct. 2020.

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

[38] H. M. Zhao, H. D. Liu, J. J. Xu, and W. Deng, "Performance prediction using high-order differential mathematical morphology gradient spectrum entropy and extreme learning machine," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 7, pp. 4165–4172, Jul. 2020.

[39] G.B. Huang, Q. Y. Zhu, and CK. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing,* vol. 70, no. 1/3, pp. 489–501, Mar. 2006.

[40] Y. J. Song *et al.*, "MPPCEDE: Multi-population parallel co-evolutionary differential evolution for parameter optimization," *Energy Convers. Manage.,* vol. 228, Jan. 2021, Art. no. 1136611.

[41] A. Samat, P. Du, S. Liu, J. Li, and L. Cheng, "(ELMs)-L-2: Ensemble extreme learning machines for hyperspectral image classification," *IEEE J. Sel. Topic Appl. Earth Obs. Remote Sens.*, vol. 7, no. 4, pp. 1060–1069, Apr. 2014.

[42] G. B. Huang, "An insight into extreme learning machines: Random neurons, random features and kernels," *Cogn. Comput.*, vol. 6, no. 3, pp. 376–390, Mar. 2014.

[43] J. Cao *et al.*, "Extreme learning machine and adaptive sparse representation for image classification," *Neural Netw.*, vol. 81, pp. 91–102, 2016.

[44] C. Chen *et al.*, "Spectral-spatial classification of hyperspectral image based on kernel extreme learning machine," *Remote Sens.*, vol. 6, no. 6, pp. 5795–5814, Jun. 2014.

[45] X. R. Zhang, Y. J. Sun, and K. Jiang, "Spatial sequential recurrent neural network for hyperspectral image classification," *IEEE J. Sel. Topic Earth Obs. Remote Sens.*, vol. 11, no. 11, pp. 4141–4155, Nov. 2018.

[46] W. Deng *et al.*, "Differential evolution algorithm with wavelet basis function and optimal mutation strategy for complex optimization problem," *Appl. Soft Comput.*, vol. 100, Jan. 2021, Art. no. 106724.

[47] Y. Zhou, J. Peng, and C. Chen, "Extreme learning machine with composite kernels for hyperspectral image classification," *IEEE J. Sel. Topic Appl. Earth Obs. Remote Sens.*, vol. 8, no. 6, pp. 2351–2360, Jun. 2017.

[48] T. Jin *et al.*, "Reliability index and option pricing formulas of the first hitting time model based on the uncertain fractional-order differential equation with Caputo type," *Fractals*, vol. 29, no. 1, Jul. 2020, Art. no. 2150012, doi:10.1142/S0218348X21500122.

[49] S. M. Mirjalili and A. Lewis, "Grey wolf optimizer," *Adv. Eng. Softw.*, vol. 69, pp. 46–61, 2014.

[50] W. Deng, J. J. Xu, H. M. Zhao, and Y. Song, "A novel gate resource allocation method using improved PSO-based QEA," *IEEE Tran. Intell. Transp. Syst.*, to be published, doi: 10.1109/TITS.2020.3025796.

[51] T. Jin *et al.*, "Valuation of interest rate ceiling and floor based on the uncertain fractional differential equation in Caputo sense," *J. Intell. Fuzzy Syst.*, pp. 1–10, Dec. 2020, doi: 10.3233/jifs-201930.

[52] W. Deng, J. J. Xu, X. Z. Gao, and H. Zhao, "An enhanced MSIQDE algorithm with novel multiple strategies for global optimization problems," *IEEE Trans. Syst. Man. Cybern. Syst.*, to be published, doi: 10.1109/TSMC.2020.3030792.

[53] T. Jin *et al.*, "Reliability index and Asian barrier option pricing formulas of the uncertain fractional first-hitting time model with Caputo type," *Chaos Soliton Fractals*, vol. 142, Jan. 2021, Art. no. 110409, doi: 10.1016/j.chaos.2020.110409.

[54] Y. J. Song *et al.*, "Enhanced success history adaptive DE for parameter optimization of photovoltaic models," *Complexity,* vol. 2021, Jan. 2021, Art. no. 6660115.

[55] W. Deng, S. F. Shang, and X. Cai, "An improved differential evolution algorithm and its application in optimization problem," *Soft Comput.*, Dec. 2020, doi: 10.1007/s00500-020-05527-x.

[56] T. Jin, H. X. Xia, and H. Chen, "Optimal control problem of the uncertain second-order circuit based on first hitting criteria," *Math. Method Appl. Sci.*, vol. 44, no. 1, pp. 882–900, Jun. 2021.

[57] B. Rasti *et al.*, "Feature extraction for hyperspectral imagery: The evolution from shallow to deep," Jun. 19, 2020. [Online]. Available: https://arxiv.org/abs/2003.02822v1

[58] Y. Zhong *et al.*, "WHU-Hi: UAV-borne hyperspectral with high spatial resolution (H2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with CRF," *Remote Sens. Environ.*, vol. 250, Dec. 2020, Art. no. 112012.

[59] Y. Zhong *et al.*, "Mini-UAV-borne hyperspectral remote sensing: From observation and processing to applications," *IEEE Geosci. Remote Sens. Mag.*, vol. 6, no. 4, pp. 46–62, Dec. 2018.

Huayue Chen (Member, IEEE) received the M.S. degree in computer software and theory from Chongqin University, Chongqin, China, in 2005. Since 2015, she has been working toward the Ph.D. degree in geodetection and information technology, from Chengdu Technology University, Chengdu, China.

Since 2012, she has been an Associate Professor with China West Normal University, Nanchong, China. Her research interest includes artificial intelligence, optimization method, and hyperspectral image processing.

Yijun Xiong (Member, IEEE) received the B.S. degree from the Chengdu University of Information Technology, Chengdu, China, in 2007, and the M.S. degree in engineering from the Chengdu University of Technology, in 2010.

Since 2010, he has been a Teacher with Chengdu University of Technology. His research interests include combinatorics and statistical analysis.

Fang Miao (Member, IEEE) received the Ph.D. degree in radioactive geology and exploration from the Chengdu Institute of Geology, Chengdu, China, in 1990.

Since 1995, he has been a Professor with Chengdu Technology University, Chengdu, China. His research interest includes artificial intelligence, optimization method, and hyperspectral image processing.

Tao Chen (Member, IEEE) received the Ph.D. degree in geodetection and information technology from the Chengdu University of Technology, Chengdu, China, in 2013.

Since 2013, he has been a Professor with China West Normal University, Nanchong, China. His research interest includes 3S technology, ecological economy, and environmental remote sensing.

Yijia Chen (Member, IEEE) is an undergraduate student at the School of Computer Science, China West Normal University, Nanchong, China since 2018.

Her research interest includes computer vision and natural language processing.