



Hyperspectral image classification via active learning and broad learning system

Huifang Huang¹ · Zhi Liu¹ · C. L. Philip Chen² · Yun Zhang¹

Accepted: 28 August 2021 / Published online: 25 November 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Hyperspectral image (HSI) classification has continued to be a hot research topic in recent years, and the broad learning system (BLS) has been considered by scholars for the classification of HSIs due to its superior internal structure. Different from the traditional HSI classification mechanism, this paper proposes an active broad learning system approach for HSI classification. The spectral and spatial features of the image are extracted using principal component analysis and local binary patterns, respectively. Then, the vector fusion of the above two features is utilized as the input of the BLS and trained to obtain pre-labels of the samples. The next training samples are selected among the pre-labels by active learning. Unlike other classification algorithms, the method proposed in this paper utilizes active learning (AL) to select high-quality samples for training, thereby reducing the number of samples used and the cost of sample labeling. In addition, the use of incremental learning in broad learning significantly reduces the training time and improves the classification accuracy. The algorithm proposed in this paper is more effective compared to other state-of-the-art algorithms on three HSI datasets.

Keywords Hyperspectral image · Active learning · Broad learning system · Classification

1 Introduction

Different from traditional black & white images, RGB images and multispectral images [1], HSI provides higher-level spectral details for better classification and recognition of ground objects [2–4]. Currently, HSI with spectral and spatial information is utilized in various types of research

and exploration, such as atmospheric environment [5], land cover distribution [6], agriculture [7, 8], etc.

The previous HSI classification methods for extracting features mainly rely on the spectral information of each pixel, such as principal component analysis (PCA) [9, 10] and independent principal component analysis (ICA) [11]. However, these methods do not consider the spatial similarity among pixels, resulting in low classification accuracy. To alleviate this impact, the spatial information is introduced to improve the separability and classification accuracy of HSI. From this, many methods, such as super pixel [12–14], image segmentation [12, 15, 16], Markov random fields [17–19], sparse representation [20] and so on, are proposed. Experiments results of these methods confirm that employing spatial information can substantially improve the classification accuracy of the system.

In recent years, the deep learning [21–24] algorithms with spectral-spatial features are widely used to improve the accuracy in HSI classification. In [25], an attention network is introduced in the spectral-spatial to extract features. An end-to-end 3-D dense convolutional network with spectral-wise attention mechanism (MSDN-SA) approach was proposed in [26]. The method employs 3-D dilated convolution for feature extraction, and utilizes a spectral

✉ Zhi Liu
lz@gdut.edu.cn

Huifang Huang
m18438561038@163.com

C. L. Philip Chen
philip.chen@ieee.org

Yun Zhang
yz@gdut.edu.cn

¹ School of Automation, Guangdong University of Technology, Guangzhou 510006, China

² Faculty of Computer Science and Engineering, South China University of Technology, Guangzhou, Guangdong 510006, China

attention mechanism to suppress ineffective spectral features to increase separability. Although these classification methods have achieved significant improvements in accuracy, they also have a large degree of limitations.

Due to the disadvantages such as many parameters involved, complex models and long time consuming, traditional deep learning is replaced by a simple and efficient BLS [27–29]. The system, which was firstly adopted by Kong [30], now has been applied to many algorithms and achieves a superior performance in HSI classification [31–33]. Furthermore, Wang [34] proposed a domain adaptive broad learning (DABL) classification scheme. A discriminative locality preserving broad learning system (DPBLS) was proposed by Chu et al. [35]. But comparing with BLS models containing incremental learning, above algorithms require employing a significant amount of labeled samples and more training time. A point worth noting in classification is that the annotation of samples is time-consuming and expensive. And the quality of selected sample has not been a concern in many studies of existing BLS classification methods. As the sample selection is crucial to the classification performance, high-quality samples are selected for training in this paper.

For the sample selection problem, AL [36] is extensively utilized as an efficient sample selection approach. It mainly selects a batch of high quality samples by selection strategy, so as to achieve the purpose of decreasing the number of samples used and realizing better classification results. In our design methodology, AL is introduced to alleviate the problem of excess number with labeled samples in BLS for HSI classification and to ensure the accuracy of the classification. Many algorithms are available to classify HSI in combination with AL [37, 38] and achieve better classification accuracy. In [39], the CNN is fine-tuned in an iterative manner with newly labeled pixels, and MRF is applied to strengthen the category labeling smoothness. The experimental results demonstrated that the classification performance of HSI can be significantly improved using the AL approach. The BLS model with incremental learning has a simple structure and does not need to retrain the overall model when the network structure changes. Compared with other network models, it can greatly reduce time-consuming. In addition, the AL method can select higher quality samples to enhance the classification performance. Hence, a new active learning of broad learning system (AL-BLS) methodology is presented. To summarize, the principal contributions of the article as follows.

1. A novel AL-BLS classification method is proposed. The AL method aims to select a batch of samples with most information as new training samples from the candidate-pool, which are then trained as new inputs for BLS. The AL-BLS classification method

fully utilizes advantages of AL and BLS methods, and the experimental results sufficiently confirm its validity.

2. On three real public datasets, in this paper, we compare the classification performance of different active selection strategies to select the best one for subsequent experiments.

2 Related work

2.1 Broad learning system

Deep networks are extremely time consuming due to their huge number of parameters and complex structure. Therefore, Chen et al. [27] proposed a novel BLS network, which adds a direct connection from the input layer to the output layer based on a single-layer feedforward network and a nonlinear transformation from the input layer to the enhancement layer. The algorithm proposed in this paper mainly relies on the principle of BLS to realize HSI classification. The structural framework of BLS and its incremental learning are shown in Figs. 1 and 2. The following contents describe them in detail.

In Fig. 1, X , Y , n , m denote the input data, output results, the feature nodes and the enhancement nodes, respectively. Both ϕ_i and ξ_j are activation functions. W_{e_i} , W_{h_j} and β_{e_i} , β_{h_j} represent the randomly generated weights and biases, where the i th mapped feature is,

$$N_i = \phi_i(XW_{e_i} + \beta_{e_i}), i = 1, \dots, n. \quad (1)$$

All the mapped features are represented as $N^n = [N_1, \dots, N_n]$, where the j th enhancement nodes denotes as,

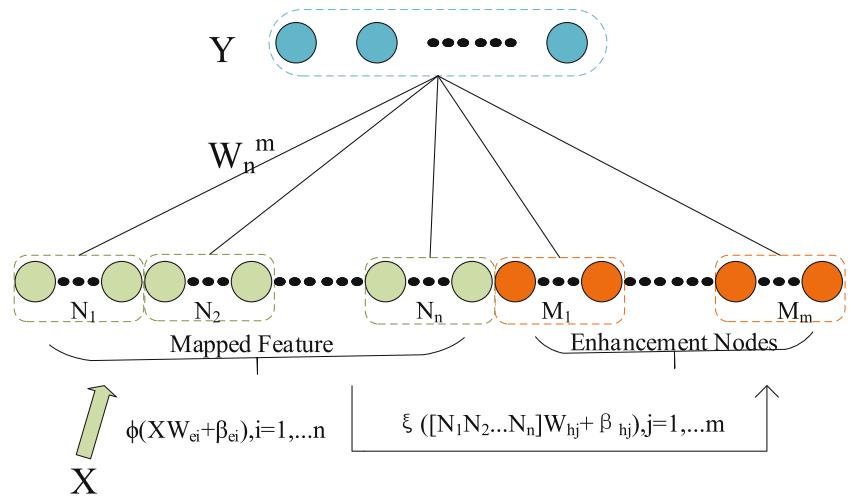
$$M_j = \xi_j(N^n W_{h_j} + \beta_{h_j}), j = 1, \dots, m. \quad (2)$$

All the enhancement nodes are represented as $M^m = [M_1, \dots, M_m]$. As a consequence, the BLS network structure can be formulated as the following function,

$$\begin{aligned} Y &= [N_1, N_2, \dots, N_n, M_1, M_2, \dots, M_m]W^m \\ &= [N^n, M^m]W^m \end{aligned} \quad (3)$$

The incremental learning algorithm of BLS has attracted the attention of many researchers by saving a large amount of training time. The whole framework of the BLS network with incremental learning is shown in Fig. 2. In this network, the new input samples are employed as training samples for the next iteration, which can save a huge amount of training time. A brief description of the dynamic update coefficients is given as follows.

A_n^m denotes the feature mapping nodes and the enhancement nodes within the initial network. The feature mapping

Fig. 1 Structure diagram of BLS

nodes and enhancement nodes generated from the new input samples are shown as:

$$N_{new}^n = [\phi(X_a W_{e1} + \beta_1), \dots, \phi(X_a W_{en} + \beta_n)] \quad (4)$$

$$A_{new} = [\phi(X_a W_{e1} + \beta_1), \dots, \phi(X_a W_{en} + \beta_n), \xi(N_{new}^n W_{h1} + \beta_{h1}), \dots, \xi(N_{new}^n W_{hm} + \beta_{hm})] \quad (5)$$

The weights of the output layer are shown as follows.

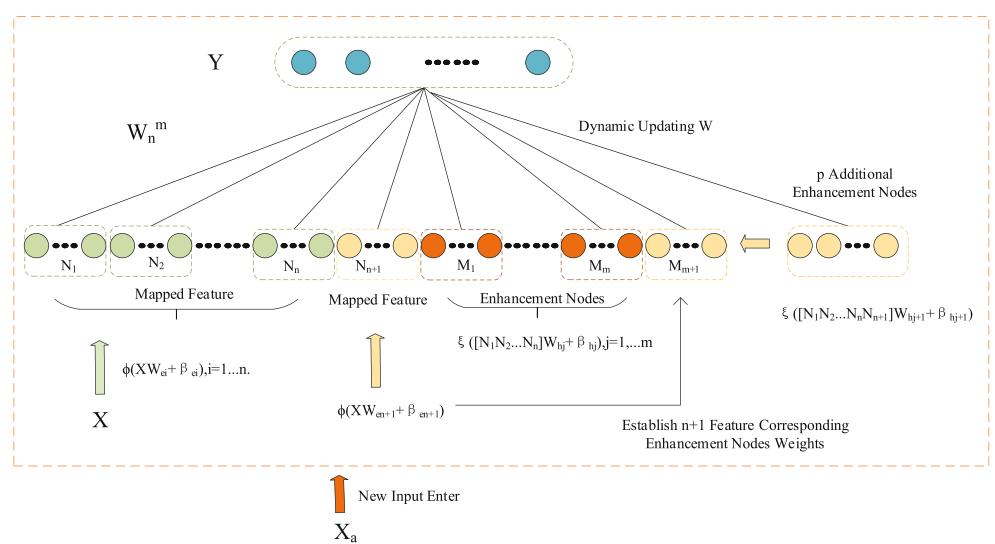
$${}^{new}W_n^m = W_n^m + (Y_A^T - A_{new}^T W_n^m)B \quad (6)$$

where

$$({}^{new}A_n^m)^+ = [(A_n^m)^+ - BD^T | B] \quad (7)$$

and

$$D^T = A_{new}^T A_n^m + \quad (8)$$

Fig. 2 BLS with increment of input data

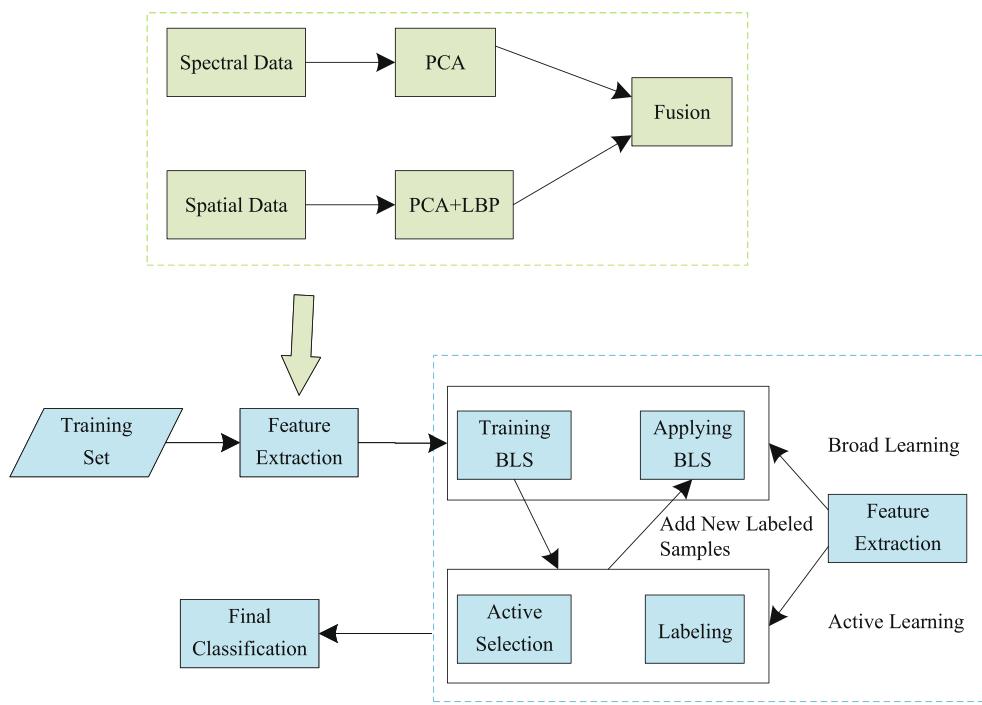
$$C = A_{new}^T - D^T A_n^m \quad (9)$$

$$B^T = \begin{cases} C^+ & \text{if } C \neq 0 \\ (1 + D^T D)^{-1}(A_n^m)^+ D & \text{if } C = 0 \end{cases} \quad (10)$$

2.2 Active learning

Labeling samples in HSI is more complicated and costly, which leads to the selection of information-rich samples as training samples is indispensable. Now, many algorithms introduce the AL method to improve the classification capability [40, 41]. Based on the above observations, the simple structured BLS method is adopted to alleviate the time consuming problem in the current classification algorithms [42, 43]. Although the algorithm for incremental learning of BLS is a bit more complex than BLS structure, the time consumed is still much shorter than that of deep learning [29, 44]. Finally, we propose an AL-BLS algorithm for HSI classification according to AL and incremental learning of BLS.

Fig. 3 The framework of AL-BLS



The primary purpose of AL is to achieve the target performance with as few labeled samples as possible. It majorly selects the most informative samples by a selection strategy [45–47], and then labels and adds them to the training set. Currently, sample selection strategies include Uncertainty Strategy (US) [39], Query-By-Committee (QBC) [48], and Expected Error Reduction, among others. In the experimental part of this paper, we compare the Entropy method and the Best vs Second Best (BvSB) method in US and the Kullback-Leibler divergence (KLD) method in QBC. These three strategies are detailed as follows.

The entropy strategy is capable of measuring the uncertainty of a system, and a higher entropy value indicates that the sample contains more information. This strategy selects the sample with the lower entropy value in the

sample as the training sample for the next iteration. The formula for the entropy strategy is shown in the following equation.

$$En(X_i) = - \sum_{d=1}^D P(X_{id}) \times \log(P(X_{id})) \quad (11)$$

where d , D and X correspond to the d th class of the sample, the class number of the sample and the i th sample in the test set, respectively. $P(x_{id})$ denotes the probability of the d th class of the i th sample.

KLD is a measure of the degree of difference between two asymmetric probability distributions. The greater the difference between the true and predicted distributions of the sample, the greater the KLD. The KDL is expressed as

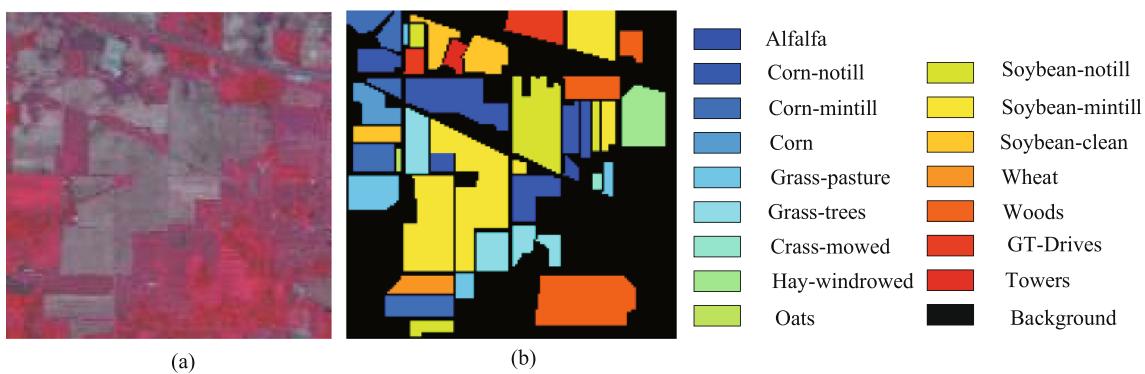


Fig. 4 Indian Pines dataset. (a) False color image; (b) The ground truth category diagram and colour codes

follows.

$$KLD(M||N) = \sum_{d=1}^D M(X_{id}) \times \log \frac{M(X_{id})}{P(X_{id})} \quad (12)$$

where M and N denote the true distribution and predicted probability distribution of the sample, respectively.

The BvSB is a strategy for multiclassification problems that obtains better classification results by considering only the classes with the first and second largest values of the sample prediction probability and ignoring the influence of other classes. The smaller the difference between the probabilities of these two classes means the richer the information contained in the sample. The BvSB formula is as follows.

$$\begin{aligned} BvSB &= P_B(X_i) - P_{SB}(X_i) \\ &= \arg \min_{X_i \in A} \left\{ \max_{u \in U} f(X_i, u) - \max_{u \in U \setminus u^+} f(X_i, u) \right\} \end{aligned} \quad (13)$$

where $P_B(X_i)$ and $P_{SB}(X_i)$ are the first and second largest values of the i th sample class probability, respectively. A , U , u and $f(X_i, u)$ denote the test set sample, a collection of classes, the class corresponding to the maximum predicted probability value and the u th class probability of sample X_i , respectively.

The three AL strategies mentioned above are used in an iterative process to select the most informative samples for labeling and allow them to train the model again. In the experimental part of this paper, the best performing strategy is selected by comparing the classification results.

3 Proposed AL-BLS model

This section gives a detailed description of the proposed active broad learning(AL-BLS). In the first place, spectral and spatial features are extracted separately for the normalized HSI data and fused as the input to the system. In the second place, the performance of different selection strategies is compared to choose the best one. In the end, AL is fused with BLS.

3.1 Spatial-spectral feature extraction

Spectral and spatial data describe HSI from two different perspectives, so it is important to extract them separately. Moreover, the classification accuracy of HSI can be dramatically improved by using spatial features. For spectral feature extraction, the singular value decomposition covariance matrix in the classical PCA method is utilized to diminish spectral data dimensionality. For spatial feature extraction, PCA is equally utilized to select the bands

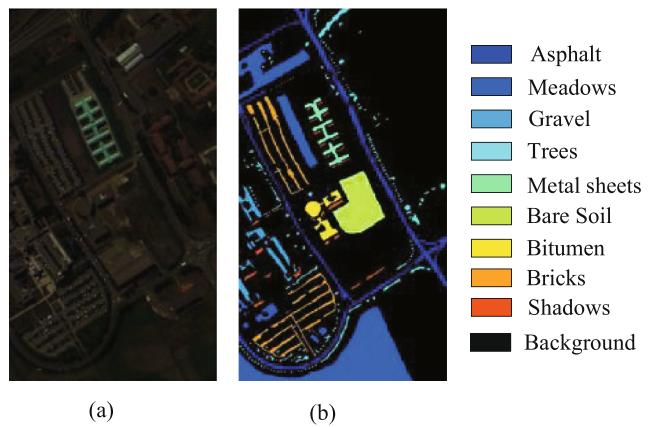


Fig. 5 Pavia University dataset. (a) False color image; (b) The ground truth category diagram and colour codes

with more information from the spectral information, and then utilizes the LBP method to improve the classification performance of the system.

LBP is a simple and efficient non-parametric local feature descriptor that converts neighborhood pixel values into local binary patterns by using the grayscale value of the central pixel as a threshold.

$$LBP_{w,r}(k_c) = \sum_{i=1}^{w-1} g(k_i - k_c) 2^i \quad (14)$$

where w denotes the quantity of neighboring pixels, and r and k_c denote the radius and center of the circle, respectively.

Hence, for fully utilizing the information in HSI and ensuring the classification accuracy of the proposed approach, the extracted spectral features are superimposed onto the end of the spatial features to form the spatial-spectral feature.

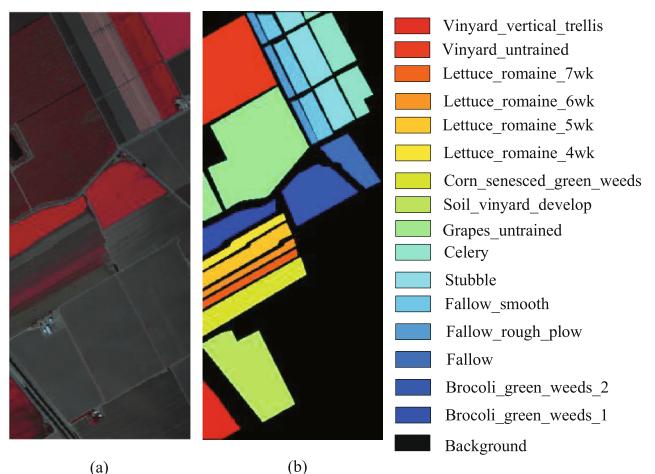


Fig. 6 Salinas dataset. (a) False color image; (b) The ground truth category diagram and colour codes

3.2 Select strategy

In the selection strategy of AL, different strategies select samples according to different methods, which have different degrees of impact on the classification performance of the system. In this paper, we employ three strategies, Entropy, KLD and BvSB, which can calculate the uncertainty of the sample according to the predicted probability value of the training sample, and then get the amount of information contained. Hence, these methods can select the maximum informative samples to achieve the purpose of reducing the sample usage and improving the performance of the system.

Input: sample labels Y ; initial number of training samples nts ; number of active learning cycles num

Output: prediction labels Y_k

- 1: Normalize the HSI data
 - 2: Extraction of spectral and spatial features to fuse them
 - 3: Select the training sample X
 - 4: Calculate N_i and set feature map grouping N by (1)
 - 5: Calculate M_j and set feature map grouping M by (2)
 - 6: Get the initial prediction labels Y_1 in (3)
 - 7: **while** the active learning runs are smaller than num **do**
 - 8: Get new training sample data X_a by AL in (11),(12) or (13)
 - 9: Calculate N_{new} and A_{new} with (4)(5)
 - 10: Get the prediction labels Y_k
 - 11: **end**
-

Algorithm 1 AL-BLS.

3.3 AL-BLS algorithm

The whole framework in this model is illustrated in Fig. 3. Algorithm 1 describes the process of AL-BLS methodology in details.

To begin with, a normalization process is performed to reduce the complexity of the HSI data and improve the performance of the system. In the next step, a certain number of feature samples with spectral and spatial information are chosen stochastically for the training samples of the initial system. After that, the selection strategy in AL is employed to select the most informative sample X_a from the test set by using the prediction probability of the system. Subsequently, X_a is removed from the candidate pool and utilized for the next training of this model. At last, the above sample selection and training steps are repeated until the loop exit requirement is satisfied.

The whole process described above has many advantages, the most important of which include the following two. The first one is to take advantage of the incremental learning of BLS to replace the overall training sample with a newly selected labeled sample X_a for the next training sample. This can avoid repeated training for a large number of samples and reduce a large amount of training time. The second one is that the AL strategy has the capability to determine the uncertainty of the samples based on the predicted probability values, and the samples with the most effective information can be selected as the training collection to improve the accuracy of HSI classification.

4 Experiments

4.1 Data description

In this section, there are a series of experiments conducted mainly on three real public HSI datasets, Indian Pines, Pavia University and Salinas. A detailed description of these datasets is given as follows.

The Indian pines is the first dataset used for spectral classification. It has a wavelength scale of 0.4–2.5 m, a spatial resolution of about 20 m, and contains 220 bands for continuous imaging features, 20 of which cannot be reflected by water, so the remaining 200 bands are mainly studied. The data has a size of 145x145 and 21025 pixels with 10249 feature pixels and 10776 background pixels,

Table 1 Class accuracy (%), overall accuracy (%), average accuracy (%), kappa (%) and time (s) by three selection strategies for 25 (%) samples of Indian Pines dataset and 10 (%) samples of Pavia University, Salinas dataset, respectively

Dataset	Indian Pines			Pavia University			Salinas		
	Strategy	Entropy	KLD	BvSB	Entropy	KLD	BvSB	Entropy	KLD
OA	88.19	97.55	99.97	93.59	98.38	99.98	88.37	98.49	99.99
AA	91.70	92.45	99.98	94.57	95.60	99.99	92.12	96.55	99.99
Kappa	90.36	97.14	99.97	92.72	93.93	99.98	91.23	96.14	99.98
Time	5.60	6.13	5.63	41.77	40.64	40.46	51.77	52.34	52.67

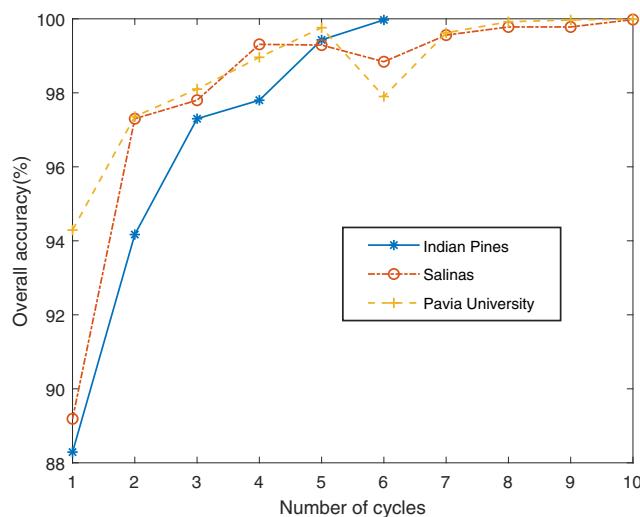


Fig. 7 The number of cycles of active learning corresponds to the change in classification performance of datasets

which contains 16 categories. The basic information of the dataset is given in Fig. 4.

The Pavia University has a 0.43–0.86 m wavelength scale and a spatial resolution with 1.3 m. It contains 115 bands of continuous imaging, 12 of which are influenced of noise, so the other 103 bands will be studied. The data has a size of 610 × 340 pixels and includes a large number of background

pixels in 2207400 pixels, which contains 9 categories. The basic information of the dataset is given in Fig. 5.

The Salinas has a spatial resolution with 3.7 m, and it contains 224 bands, 20 of which are equally not reflected by water, and the other 204 bands are used for the study. It has a size of 512 × 217 and contains 111,104 pixels, including 56,975 background pixels and 54,129 feature pixels, which are classified into 16 categories. The basic information of the dataset is given in Fig. 6.

4.2 Parameter settings

In order to estimate the validity of the presented AL-BLS method, we performed the series of experiments for three datasets separately. The classification performance is compared with some advanced HSI classification algorithms, such as SVM, LORSAL [49], BLS, CCJSR [50], SMLR-SpTV [49], ELM-ECF [51], and FDSI [52]. We select class accuracy (CA), overall accuracy (OA), average accuracy (AA), kappa and time as significant measures of HSI classification performance. Each algorithm was executed on a Matlab R2019b with 2.00 GHz and 8 GB computer.

For the proposed approach, PCA is employed to down-dimensioning the data, and 15 and 3 of the bands are selected in the spectral and spatial data, respectively. The

Table 2 Class accuracy (%), overall accuracy (%), average accuracy (%), kappa (%) and time (s) by the several approaches on a 25 (%) sample of the Indian Pines dataset

Class	SVM	LORSAL	BLS	CCJSR	SMLR-SpTV	ELM-ECF	FDSI	AL-BLS
1	100.00	52.94	100.00	100.00	94.12	100.00	100.00	100.00
2	78.11	81.61	95.77	96.48	99.25	98.97	99.81	99.90
3	75.23	79.74	94.55	95.63	99.04	99.84	99.36	100.00
4	73.23	65.17	100.00	93.99	100.00	98.87	100.00	100.00
5	95.21	99.37	100.00	94.07	96.13	100.00	99.72	100.00
6	90.97	96.89	97.79	94.77	96.63	100.00	99.45	100.00
7	100.00	61.91	100.00	89.47	76.19	100.00	100.00	100.00
8	92.51	98.88	100.00	100.00	100.00	100.00	100.00	100.00
9	100.00	66.67	100.00	100.00	86.67	86.67	100.00	100.00
10	75.18	79.15	96.34	97.24	98.49	97.26	99.31	99.84
11	78.10	87.40	94.55	97.60	99.67	99.89	99.14	100.00
12	80.04	82.70	95.74	92.08	99.10	99.77	98.23	100.00
13	99.34	98.05	98.31	100.00	100.00	100.00	100.00	100.00
14	92.83	94.52	99.08	100.00	100.00	100.00	99.89	100.00
15	79.45	70.24	98.43	91.23	100.00	98.27	100.00	100.00
16	100.00	77.14	97.14	86.21	98.57	97.10	98.46	100.00
OA	82.78	85.99	96.63	96.55	99.21	99.40	99.47	99.97
AA	88.14	80.40	97.98	95.55	96.68	98.54	99.59	99.98
Kappa	80.27	84.00	96.15	96.07	99.09	99.32	99.39	99.97
Time	2.94	36.08	3.63	950.64	54.15	80.30	202.23	5.63

Table 3 Class accuracy (%), overall accuracy (%), average accuracy (%), kappa (%) and time (s) by the several approaches on a 10 (%) sample of the Pavia University dataset

Class	SVM	LORSAL	BLS	CCJSR	SMLR-SpTV	ELM-ECF	FDSI	AL-BLS
1	66.62	90.57	97.35	96.12	99.63	99.30	99.50	99.93
2	87.83	95.22	94.98	98.38	100.00	100.00	99.67	99.98
3	91.65	72.84	94.29	84.52	89.20	94.97	100.00	100.00
4	95.60	93.55	97.61	98.42	94.67	98.80	99.92	100.00
5	99.67	99.83	99.75	96.87	100.00	99.75	99.92	100.00
6	97.49	78.37	95.11	93.86	99.76	99.73	99.98	100.00
7	81.82	81.04	99.20	89.19	94.90	99.83	100.00	100.00
8	83.86	80.18	93.75	80.89	97.71	99.03	99.46	100.00
9	100.00	57.28	100.00	86.28	81.46	100.00	99.07	100.00
OA	84.57	88.87	95.78	94.46	98.24	99.43	99.70	99.98
AA	89.39	83.21	96.89	91.61	95.26	99.05	99.73	99.99
Kappa	78.89	85.17	94.35	92.65	97.66	99.25	99.60	99.98
Time	7.61	117.53	15.05	7653.57	171.32	218.00	388.53	40.46

size of patch is set to 19 in the process of LBP feature extraction. In the specific operation, we select 10 samples for each category of data in each dataset as initial samples, and then a different number of the samples for the training are selected actively respectively. In the Indian Pines training set, 250, 250, 400, 400, 550, and 550 samples are actively selected in per cycle. The training sets of Salinas

and Pavia University are actively selected 10 times, with 420 and 520 samples each time, respectively. So these three datasets have about 25%, 10% and 10% of the overall amount of training samples, respectively.

With other parameters held constant on each dataset, AL-BLS is chosen to perform classification tests in three selection strategies, Entropy, KLD and BvSB. The

Table 4 Class accuracy (%), overall accuracy (%), average accuracy (%), kappa (%) and time (s) by the several approaches on a 10 (%) sample of the Salinas dataset

Class	SVM	LORSAL	BLS	CCJSR	SMLR-SpTV	ELM-ECF	FDSI	AL-BLS
1	61.18	98.84	100.00	99.94	100.00	100.00	100.00	100.00
2	99.94	99.85	99.26	100.00	100.00	99.97	100.00	100.00
3	93.53	97.47	99.21	99.05	99.78	100.00	100.00	100.00
4	98.26	98.17	99.13	97.20	99.04	99.68	97.13	100.00
5	98.69	99.00	99.63	98.21	100.00	99.54	99.96	100.00
6	100.00	99.75	99.89	99.89	100.00	100.00	100.00	99.97
7	100.00	99.66	98.28	99.97	100.00	99.78	100.00	100.00
8	76.50	84.38	92.91	99.16	99.93	99.28	99.92	100.00
9	99.50	99.53	98.32	99.87	100.00	100.00	100.00	100.00
10	96.78	97.12	97.01	98.92	99.46	99.59	99.73	99.83
11	97.95	97.71	98.64	99.58	100.00	100.00	99.79	100.00
12	97.25	99.88	97.85	98.80	100.00	100.00	99.48	100.00
13	93.35	98.91	98.91	99.63	100.00	100.00	100.00	100.00
14	98.74	95.02	97.08	97.92	99.79	98.96	100.00	100.00
15	84.11	72.70	89.34	99.34	99.27	98.32	100.00	99.98
16	99.75	98.09	98.21	99.57	99.63	99.57	100.00	100.00
OA	89.76	92.34	96.24	99.34	99.80	99.52	99.87	99.99
AA	93.47	96.00	97.73	99.19	99.81	99.67	99.75	99.99
Kappa	88.59	91.47	95.80	99.26	99.78	99.46	99.85	99.98
Time	13.23	228.28	17.96	14398.23	342.05	524.23	491.75	52.67

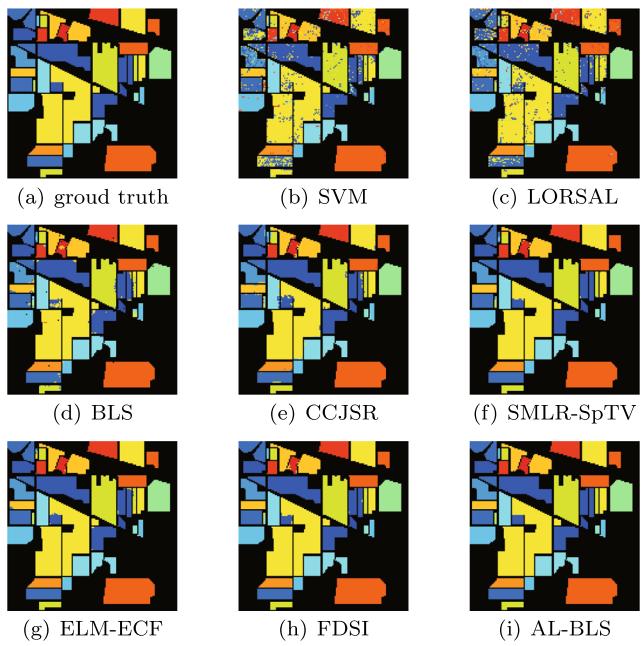


Fig. 8 The results on the different classification approaches in the Indian Pines dataset

performance results of the three strategies are shown in Table 1, from which it is observed that BvSB has better classification performance than the other two. Thus, in the following experimental results of AL-BLS are chosen to perform on the BvSB strategy in AL. The samples are selected on these three datasets by using the BvSB strategy with better performance, whose OA results corresponding to each training are shown in Fig. 7.

To check whether there are significant performance differences among AL-BLS and SVM, LORSAL, BLS, CCJSR, SMLR-SpTV, ELM-ECF, and FDSI, we validated the algorithms on three datasets based on their classification results by using the Friedman method. The equation of the test value is as follows.

$$\chi^2 = \frac{12L}{R(R+1)} \left(\sum_{i=1}^R K_i^2 - \frac{R(R+1)^2}{4} \right) \quad (15)$$

where R is the number of all algorithms compared on L datasets. K_i denotes the average ordinal value of the i th algorithm.

4.3 Results analysis

The results of AL-BLS method and other advanced HSI classification methods are shown in Tables 2, 3 and 4. The classification diagram is shown in Figs. 8, 9 and 10, in which each color represents a class of ground objects except for the background color, and the larger the difference with the true ground distribution, the lower the classification accuracy. The AL-BLS method has the highest accuracy,

so the difference with the true ground distribution is small. Table 2 and Fig. 8 represent the classification results of each algorithm for selecting 25% samples in the Indian Pines dataset. Table 3, Fig. 9 and Table 4, Fig. 10 represent the classification results of each algorithm for selecting 10% samples on the other two datasets, respectively. The following conclusions are obtained by comparing the experimental results.

The proportion of training samples selected in Indian Pine dataset is about 15% more than that of Pavia University and Salinas datasets to achieve roughly the same classification performance. The main reason is that the total number of samples in the Indian pine data set is small and the higher spatial resolution results in more mixed image elements, which leads to classification difficulties and a larger proportion of samples.

It is first assumed that there is no significant difference between AL-BLS and the comparison algorithm. The test is performed at the significance level of $\alpha = 0.05$ to check the table and get $\chi^2_{0.05(7)} = 14.067$. According to the (15), it is calculated that $\chi^2 = 20.556 > 14.067$, so the original hypothesis of the test is rejected and considered that these seven algorithms differed significantly in their classification results. The significant difference between the algorithm test results indicates that the AL-BLS algorithm is effective in terms of classification performance.

In this paper, the SVM and LORSAL algorithms only utilize spectral information, while the BLS, CCJSR, SMLR-SpTV, ELM-ECF, FDSI and AL-BLS algorithms utilize

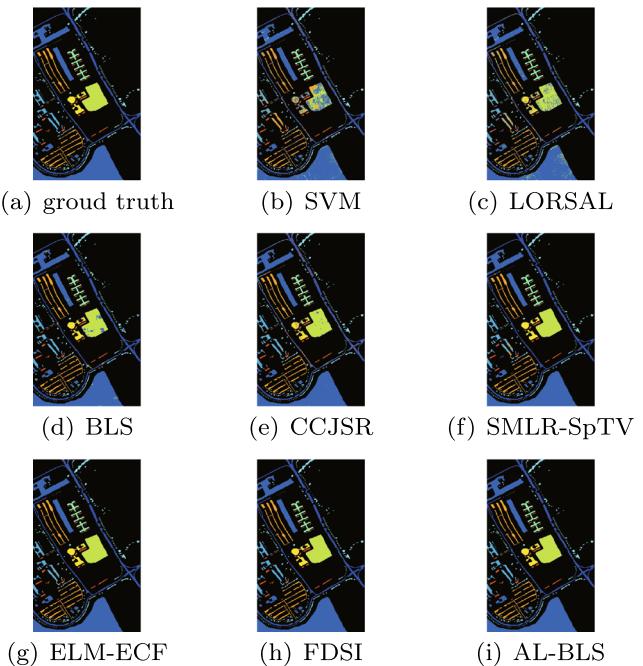


Fig. 9 The results on the different classification approaches in the Pavia University dataset

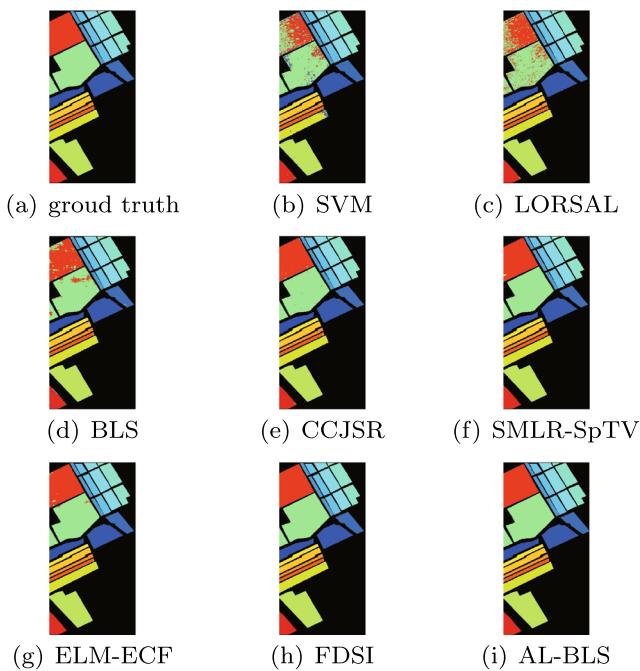


Fig. 10 The results on the different classification approaches in the Salinas dataset

both spatial and spectral information, and the experimental results show that the OA, AA, and Kappa of the former two algorithms are inferior. Thus, considering spatial information in classification can improve the classification accuracy to a significant degree.

With the same method of extracting features, the OA of the AL-BLS algorithm was improved by 3.34%, 4.20%, and 3.75% in the three datasets compared to BLS. Selecting information samples using the AL method contributes significantly to the classification effectiveness of the system.

In the classification results of these three datasets, firstly, the SVM algorithm is the least time-consuming of all classification methods, but the overall performance is inferior. the concise and effective structure of BLS dictates a small amount of time consumption in the experiments. The AL-BLS algorithm, which adds the AL method to the incremental learning of BLS, has a more complex structure leading to increased time consumption, but still holds a great advantage in terms of time consumption compared with other algorithms.

From the evaluation criteria of classification accuracy, OA, AA, Kappa and Time in the table, and from the differences between the classification maps of experimental results and the real classification maps in Figs. 8, 9 and 10, it is obvious that the classification performance of the AL-BLS method proposed in this paper is significantly better than other classification methods. The major factors are as follows. In the first place, the data is preprocessed and the samples of spatial spectral features are fed to the system. In

the second place, the strategy in active learning that helps the most to improve the performance is selected. Finally, the incremental learning system in BLS is utilized, which can reduce the labeled samples and also improve the accuracy to reduce the time consumption.

5 Conclusion

A novel AL-BLS method is proposed in the article for HSI classification. Firstly, the proposed method obtains the spatial-spectral features through PCA and LBP, and trains the extracted initial training samples by BLS. Subsequently, the proposed method chooses information-rich samples in former training results, which are trained again by BvSB decision method. The experimental results that are shown in three datasets validate that AL-BLS can significantly improvement the classification performance of HSI.

Acknowledgements This work was supported in part by Guangdong Province Universities and Colleges Pearl River Scholar Funded Scheme, and in part by the National Key Research and Development Program of China under Project 2020AAA0108303.

References

- Wicaksono P, Aryaguna PA (2020) Analyses of inter-class spectral separability and classification accuracy of benthic habitat mapping using multispectral image. *Remote Sensing Applications: Society and Environment* 19:100335
- Huang B, Ge L, Chen G, Radenkovic M, Wang X, Duan J, Pan Z (2021) Nonlocal graph theory based transductive learning for hyperspectral image classification. *Pattern Recogn* 116:107967
- Chu Y, Lin H, Yang L, Sun S, Diao Y, Min C, Fan X, Shen C (2021) Hyperspectral image classification with discriminative manifold broad learning system. *Neurocomputing* 442:236–248
- Fang J, Cao X (2020) Multidimensional relation learning for hyperspectral image classification. *Neurocomputing* 410:211–219
- Brunet D, Sills D (2016) A generalized distance transform: Theory and applications to weather analysis and forecasting. *IEEE Trans Geosci Remote Sens* 55(3):1752–1764
- Hong D, Yokoya N, Chanussot J, Zhu XX (2019) Cospace: Common subspace learning from hyperspectral-multispectral correspondences. *IEEE Trans Geosci Remote Sens* 57(7):4349–4359
- Meng S, Wang X, Hu X, Luo C, Zhong Y (2021) Deep learning-based crop mapping in the cloudy season using one-shot hyperspectral satellite imagery. *Comput Electron Agric* 186:106188
- Hong G, El-Hamid HTA (2020) Hyperspectral imaging using multivariate analysis for simulation and prediction of agricultural crops in ningxia, china. *Comput Electron Agric* 172:105355
- Kang X, Xiang X, Li S, Benediktsson JA (2017) Pca-based edge-preserving features for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 55(12):7140–7151
- Dadon A, Mandelmilch M, Ben-Dor E, Sheffer E (2019) Sequential pca-based classification of mediterranean forest plants

- using airborne hyperspectral remote sensing. *Remote Sens* 11(23):2800
11. Xia J, Bombrun L, Adali T, Berthoumieu Y, Germain C (2016) Spectral-spatial classification of hyperspectral images using ica and edge-preserving filter via an ensemble strategy. *IEEE Trans Geosci Remote Sens* 54(8):4971–4982
 12. Cao X, Lu H, Ren M, Jiao L (2019) Non-overlapping classification of hyperspectral imagery with superpixel segmentation. *Appl Soft Comput* 83:105630
 13. Yu H, Gao L, Liao W, Zhang B, Pižurica A, Philips W (2017) Multiscale superpixel-level subspace-based support vector machines for hyperspectral image classification. *IEEE Geosci Remote Sens Lett* 14(11):2142–2146
 14. Shi C, Pun C-M (2018) Superpixel-based 3d deep neural networks for hyperspectral image classification. *Pattern Recogn* 74:600–616
 15. Nalepa J, Myller M, Kawulok M (2019) Training-and test-time data augmentation for hyperspectral image segmentation. *IEEE Geosci Remote Sens Lett* 17(2):292–296
 16. Moghaddam SHA, Mokhtarzade M, Beirami BA (2020) A feature extraction method based on spectral segmentation and integration of hyperspectral images. *Int J Appl Earth Obs Geoinf* 89: 102097
 17. Cao X, Xu L, Meng D, Zhao Q, Xu Z (2017) Integration of 3-dimensional discrete wavelet transform and markov random field for hyperspectral image classification. *Neurocomputing* 226:90–100
 18. Wang L, Huang X, Zheng C, Zhang Y (2017) A markov random field integrating spectral dissimilarity and class co-occurrence dependency for remote sensing image classification optimization. *ISPRS J Photogramm Remote Sens* 128:223–239
 19. Pan C, Gao X, Wang Y, Li J (2018) Markov random fields integrating adaptive interclass-pair penalty and spectral similarity for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 57(5):2520–2534
 20. Zhang S, Li S, Fu W, Fang L (2017) Multiscale superpixel-based sparse representation for hyperspectral image classification. *Remote Sens* 9(2):139
 21. Zhang M, Li W, Du Q (2018) Diverse region-based cnn for hyperspectral image classification. *IEEE Trans Image Process* 27(6):2623–2634
 22. Zhong Z, Li J, Luo Z, Chapman M (2017) Spectral-spatial residual network for hyperspectral image classification: A 3-d deep learning framework. *IEEE Trans Geosci Remote Sens* 56(2):847–858
 23. Wang Q, Yuan Z, Du Q, Li X (2018) Getnet: A general end-to-end 2-d cnn framework for hyperspectral image change detection. *IEEE Trans Geosci Remote Sens* 57(1):3–13
 24. Audebert N, Le Saux B, Lefèvre S (2019) Deep learning for classification of hyperspectral data: A comparative review. *IEEE Geosci Remote Sens Mag* 7(2):159–173
 25. Sun H, Zheng X, Lu X, Wu S (2019) Spectral-spatial attention network for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 58(5):3232–3245
 26. Fang B, Li Y, Zhang H, Chan JC-W (2019) Hyperspectral images classification based on dense convolutional networks with spectral-wise attention mechanism. *Remote Sens* 11(2):159
 27. Chen PCL, Liu Z (2017) Broad learning system: An effective and efficient incremental learning system without the need for deep architecture. *IEEE Trans Neural Netw Learn Syst* 29(1):10–24
 28. Zhao H, Zheng J, Deng W, Song Y (2020) Semi-supervised broad learning system based on manifold regularization and broad network. *IEEE Trans Circ Syst I: Regular Papers* 67(3): 983–994
 29. Chen PCL, Liu Z, Feng S (2018) Universal approximation capability of broad learning system and its structural variations. *IEEE Trans Neural Netw Learn Syst* 30(4):1191–1204
 30. Kong Y, Wang X, Cheng Y, Chen CL (2018) Hyperspectral imagery classification based on semi-supervised broad learning system. *Remote Sens* 10(5):685
 31. Feng S, Chen CLP (2018) Fuzzy broad learning system A novel neuro-fuzzy model for regression and classification. *IEEE Trans Cybern* 50(2):414–424
 32. Jin J, Liu Z, Chen CLP (2018) Discriminative graph regularized broad learning system for image recognition. *Sci China Inf Sci* 61(11):1–14
 33. Kong Y, Cheng Y, Chen CLP, Wang X (2019) Hyperspectral image clustering based on unsupervised broad learning. *IEEE Geosci Remote Sens Lett* 16(11):1741–1745
 34. Wang H, Wang X, Chen CLP, Cheng Y (2020) Hyperspectral image classification based on domain adaptation broad learning. *IEEE J Sel Top Appl Earth Obs Remote Sens* 13:3006–3018
 35. Chu Y, Lin H, Yang L, Zhang D, Diao Y, Fan X, Shen C (2020) Hyperspectral image classification based on discriminative locality preserving broad learning system. *Knowl Based Syst* 206:106319
 36. Cohn DA, Ghahramani Z, Jordan MI (1996) Active learning with statistical models. *J Artif Intell Res* 4:129–145
 37. Li Y, Ting L, Li S (2020) Subpixel-pixel-superpixel-based multiview active learning for hyperspectral images classification. *IEEE Trans Geosci Remote Sens* 58(7):4976–4988
 38. Haut JM, Paoletti ME, Plaza J, Li J, Plaza A (2018) Active learning with convolutional neural networks for hyperspectral image classification using a new bayesian approach. *IEEE Trans Geosci Remote Sens* 56(11):6440–6461
 39. Cao X, Yao J, Xu Z, Meng D (2020) Hyperspectral image classification with convolutional neural network and active learning. *IEEE Trans Geosci Remote Sens* 58(7):4604–4616
 40. Munoz-Mari J, Tuia D, Camps-Valls G (2012) Semisupervised classification of remote sensing images with active queries. *IEEE Trans Geosci Remote Sens* 50(10):3751–3763
 41. Li J, Huang X, Chang X (2020) A label-noise robust active learning sample collection method for multi-temporal urban land-cover classification and change analysis. *ISPRS J Photogramm Remote Sens* 163:1–17
 42. Ma Y, Liu Z, Chen CLP (2021) Multiscale random convolution broad learning system for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*
 43. Zhao G, Wang X, Kong Y, Cheng Y (2021) Spectral-spatial joint classification of hyperspectral image based on broad learning system. *Remote Sens* 13(4):583
 44. Fan J, Wang X, Wang X, Zhao J, Liu X (2019) Incremental wishart broad learning system for fast polsar image classification. *IEEE Geosci Remote Sens Lett* 16(12):1854–1858
 45. Zhou X, Prasad S (2017) Active and semisupervised learning with morphological component analysis for hyperspectral image classification. *IEEE Geosci Remote Sens Lett* 14(8):1348–1352
 46. Liu P, Zhang H, Eom KB (2016) Active deep learning for classification of hyperspectral images. *IEEE J Sel Top Appl Earth Obs Remote Sens* 10(2):712–724
 47. Zhang Y, Cao G, Li X, Wang B, Fu P (2019) Active semi-supervised random forest for hyperspectral image classification. *Remote Sens* 11(24):2974
 48. Freund Y, Seung HS, Shamir E, Tishby N (1997) Selective sampling using the query by committee algorithm. *Mach Learn* 28(2):133–168
 49. Sun L, Wu Z, Liu J, Xiao L, Wei Z (2014) Supervised spectral-spatial hyperspectral image classification with weighted markov random fields. *IEEE Trans Geosci Remote Sens* 53(3):1490–1503
 50. Tu B, Zhang X, Kang X, Zhang G, Wang J (2018) Hyperspectral image classification via fusing correlation coefficient and joint sparse representation. *IEEE Geosci Remote Sens Lett* 15(3):340–344

51. Jiang M, Cao F, Lu Y (2018) Extreme learning machine with enhanced composite feature for spectral-spatial hyperspectral image classification. *IEEE Access* 6:22645–22654
52. Duan P, Ghamisi P, Kang X, Rasti B, Li S, Gloaguen R (2020) Fusion of dual spatial information for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



Huifang Huang received the B.S. degree from Luoyang Normal University, Luoyang, China, in 2019, the M.S. degree in automation from Guangdong University of Technology, Guangzhou, China, in 2022.



Zhi Liu received the B.S. degree from the Huazhong University of Science and Technology, Wuhan, China, in 1997, the M.S. degree from Hunan University, Changsha, China, in 2000, and the Ph.D. degree from Tsinghua University, Beijing, China, in 2004, all in electrical engineering. He is currently a Professor with the Department of Automation, Guangdong University of Technology, Guangzhou, China. His research interests include fuzzy logic systems, neural networks, robotics, and robust control.



C.L. Philip Chen (Fellow, IEEE) received the M.S. degree in electrical engineering from the University of Michigan, Ann Arbor, MI, USA, in 1985, and the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1988. He is currently the Head of the School of Computer Science and Engineering, South China University of Technology, Guangzhou, Guangdong, China. His current research interests include

systems, cybernetics, and computational intelligence. Dr. Chen is a fellow of the American Association for the Advancement of Science (AAAS), the International Association for Pattern Recognition (IAPR), the Chinese Association of Automation (CAA), and the Hong Kong Institution of Engineers (HKIE). He was the Chair of TC 9.1 Economic and Business Systems of the International Federation of Automatic Control from 2015 to 2017 and a Program Evaluator of the Accreditation Board of Engineering and Technology Education (ABET), USA, for computer engineering, electrical engineering, and software engineering programs. He has been the Editor-in-Chief of the *IEEE TRANSACTION ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS* since 2014 and an associate editor of several IEEE TRANSACTIONS.



Yun Zhang received the B.S. and M.S. degrees in automatic engineering from Hunan University, Changsha, China, in 1982 and 1986, respectively, and the Ph.D. degree in automatic engineering from the South China University of Science and Technology, Guangzhou, China, in 1998. He is currently a Professor with the Department of Automation, Guangdong University of Technology, Guangzhou. His research interests include intelligent control systems, network systems, and signal processing.