IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 12, 2017, accepted May 9, 2017, date of publication May 23, 2017, date of current version June 28, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2706363

Remote Sensing Image Classification Based on Ensemble Extreme Learning Machine With Stacked Autoencoder

FEI LV[1], MIN HAN[1], (Senior Member, IEEE), AND TIE QIU[2], (Senior Member, IEEE)
[1]Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024, China
[2]School of Software, Dalian University of Technology, Dalian 116620, China

Corresponding author: Min Han (minhan@dlut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61374154 and in part by the National Basic Research Program of China under Grant 2013CB430403.

ABSTRACT Classification is one of the most popular topics in remote sensing. Consider the problems that the remote sensing data are complicated and few labeled training samples limit the performance and efficiency in the classification of remote sensing image. For these problems, a huge number of methods were proposed in the last two decades. However, most of them do not yield good performance. In this paper, a remote sensing image classification algorithm based on the ensemble of extreme learning machine (ELM) neural network, namely, stacked autoencoder (SAE)-ELM, is proposed. First, due to improve the ensemble classification accuracy, we adopt feature segmentation and SAE in the sample data to create high diversity among the base classifiers. Furthermore, ELM neural network is chosen as a base classifier to improve the learning speed of the algorithm. Finally, to determine the final ensemble-based classifier, Q-statistics is adopted. The experiment compares the proposed algorithm with Bagging, Adaboost, Random Forest et al., which results show that the proposed algorithm not only gets high classification accuracy on low resolution, medium resolution, high resolution and hyperspectral remote sensing images, but also has strong stability and generalization on UCI data.

INDEX TERMS Remote sensing classification, ensemble algorithm, extreme learning machine, Q-statistics, feature extraction.

I. INTRODUCTION

Classification of remote sensing images refers to attribute analysis and judgment based on information of different ground objects on remote sensing images, to fulfill the purpose of identifying the actual ground objects corresponding to the images, and extract information required on the ground object, which if one of the research focuses in the field of remote sensing [1]–[3]. Therefore, land classification and identification has become a new study hotspot through information acquired by remote sensing images. Remote sensing image classification is a way to distinguish class attributes and distribution of ground objects based on the feature of material electromagnetic radiation information in the remote sensing images. It's a hot topic in the field of remote sensing images.

Remote sensing is one of the most significant achievements of advanced earth observation technology. Traditionally, remote sensing image classification is implemented by a single classifier, for example, support vector machine(SVM),using original image data and other derived features as input signals. Those methods have proved their effectiveness for many applications, but there are still some problems.

First, each classifier has its own merits and limitations, and it is often difficult to achieve satisfied accuracy by a single classifier. Second, adjacent wavebands of remote sensing data are highly correlated, so the simultaneous use of all bands cant assure high accuracy. Owing to those limitations in classifiers and data, it is necessary to find some new ways to improve the classification performance. Based on a survey to remote sensing classification techniques and ensemble algorithms, some issues on remote sensing image classification based on ensemble algorithms are explored ensemble algorithms to solve the problem.

The main contributions of this paper are as follows:

- To solve the problem of ineffective classification of remote sensing images due to inadequate labeled

training samples, this paper proposes a remote sensing image classification algorithm, namely SAE-ELM.

- The proposed algorithm first makes in-depth division of the features of the initial training set to randomly divide the feature set of the original data into a plurality of small feature subsets, then makes SAE transformation on the feature subsets, and finally re-arrange the feature subsets based on the feature order of the initial training set. In other words, the training sets of each ELM base classifier are changed.After the above processing, variance among base classifiers will effectively improve.
- To solve the problem of classification ineffectiveness caused by the magnitude of remote sensing data, ELM with faster training speed can be used as base classifiers when selecting ensemble classifiers, while Q-Statistics [25] can be used for final selection of the ensemble base classifiers to obtain better classification effects.
- We use four remote sensing images data and UCI [29] data sets to do experiment to evaluate the performance of SAE-ELM, and also compare SAE-ELM with several existing neural network ensemble algorithms: Bagging [8], Adaboost [21], Random Forest [28] and Rotate Forest [20]. The proposed algorithm has higher classification accuracy and stronger generalization performance, and can adapt to the classification of different resolutions and hyperspectral remote sensing images.

The aforementioned facts motivated us to develop a novel remote sensing images classification method based on SAE-ELM. The rest of this paper is organized as follows. Section II briefly surveys related work.Section III presents a brief review of several related algorithms and gives the details of the proposed SAE-ELM algorithm. Section IV illustrates five examples, including remote sensing image data and UCI data classification to show the excellent performance of proposed SAE-ELM algorithm. Finally, discussions and conclusions are given in Seciton V.

II. RELATED WORK

As remote sensing images are characterized by massive data and complex data types, and have the prevalent drawback of inadequate labeled training samples, one single classifier can hardly achieve satisfactory results while processing such data.Therefore, some scholars put forward the concept of ensemble algorithm [4]–[7], i.e. to obtain the final classification result by combining results from a plurality of classifiers. Peter [8] and Ghorai *et al.* [9] suggested that the training set be done with n times of replacement sampling to construct different training sets for each base classifier so as to obtain better classification results through ensemble. Kim and Kang [10] suggested the use of Multi-layer Perceptrons (MLPs) as the base classifier of bagging so that the effectiveness of MLP gets improved. The ensemble algorithms are usually achieved by enhancing the variance among ensemble base classifiers to achieve better classification results [11], [12].

Merentitis and Debes [13] provided a discussion of the bias/variance tradeoff that is a key notion in machine learning and especially ensemble learning. Li *et al.* [14] proposed a classification paradigm to exploit rich texture information of hyperspectral imagery, the efficient extreme learning machine with a very simple structure is employed as the classifier. Chen *et al.* [15] proposed a regularized deep feature extraction method is presented fro hyperspectral image classification using a convolutional neural network, which with sparse constraints provide competitive results to state-of-the-art methods. Tan *et al.* [3] proposed an automatic method for change detection in high-resolution remote sensing images that uses a novel strategy for the selection of training samples and an ensemble of multiple classifiers, which gets a better performance. Damodaran *et al.* [16] proposed a new dynamic classifier selection/dynamic ensemble selection framework based on ELM regression and a new spectral-spatial classification model, which yields a significant increase in the accuracy when compared to the state-of-the-art approaches. In order to overcome the drawbacks of ELM caused by the randomness of input weights and bias, Samat *et al.* [17] proposed two new algorithms of ELM, Bagging-based and AdaBoost-based ELMs.

Yu *et al.* [30] proposed a multi-feature representation method called diagonal structure descriptor, which is more suitable for intermediate feature extraction and conducive to multi-feature fusion. Tao *et al.* [18] investigated the application of deep neural networks to precipitation estimation from remotely sensed information, a stacked denoising autoencoder is used to automatically extract features from the infrared cloud images and estimate the amount of precipitation referred as PERSIANN-SDAE, it outperforms both the shallow neural network. Han *et al.* [19] proposed an efficient hierarchical convolutional sparse autoencoder algorithm considering all the features of the images integrally for scene classification, which adopts the unsupervised hierarchical idea based on the single-hierarchy convolutional sparse autoencoder.

Garcia-Pedrajas [21] weighted the base classifiers for each training to enhance variance by applying the concept of iterative weight, thus making ensemble possible. The algorithms proposed by Rodriguez *et al.* [20] rotate the training set feature axis of each base classifier by using the Principal Component Analysis(PCA)technology, resulting in base classifiers of greater variance, thereby improving the effect of final classification results. Lekamalage *et al.* [22] compared Autoencoder(AE) and PCA, it is found that AE has a better effect on feature extraction.

Kim *et al.* [23] used Q-statistic as a new evaluation measure that incorporates the stability of the selected feature subset in addition to the prediction accuracy, proposed the Booster of an feature selection algorithm that boosts the value of Q-Statistic of the algorithm applied. Wang and Yao [24] used Q-statistic to measure the diversity between base classifiers to improve overall accuracy.

In a word, due to the inadequacy of labeled training samples and the complexity of remote sensing data, how to improve ensemble variance and classification effects is an issue that demands immediate solution.

III. PROPOSED LEARNING ALGORITHM

A. AUTOENCODER

The autoencoder consists of an input layer, a hidden layer and an output layer. The objective of training and learning is to bring the output of the network as close to the input as possible. The autoencoder training process. The encoding process converts the input samples from linear mapping to non-linear mapping to obtain hidden-layer representation. Assume that the sample set is $X = \{x_i\}_{i=1}^{N}$, the hidden layer corresponding to the input sample x_i is represented as follows.

$$h_i = f(x_i) = sigmoid(W_1 x_i + b_1) \tag{1}$$

Where W_1 and b_1 represent the weight and bias between the input and the hidden layer, respectively, and $sigmoid()$ represents the excitation function of the hidden layer. The decoding process is to re-project the encoded representation into original signal space, and the decoded signal $\hat{x}_i$ can be represented as:

$$\hat{x}_i = g(x_i) = sigmoid(W_2 h_i + b_2) \tag{2}$$

Where W_2 and b_2 are the weight and bias of the input layer and the hidden layer, and $sigmoid()$ is the excitation function of the hidden layer. The objective of the self-encoder is to make the decoded output as close as possible to the input before encoding, and the parameters of the network can be optimized through minimization of the reconstruction error.The target function is as follows:

$$J(W_1, W_2, b_1, b_2) = \underset{W_1, W_2, b_1, b_2}{argmin} \sum_{i=1}^{N} \|x_i - \hat{x}_i\|_2^2 \tag{3}$$

After the training of one layer of autoencoder, the excitation value of the hidden layer is used as the input of the next layer of autoencoder so that the multi-layer autoencoder is stacked up to form a stacked set of autoencoder.

B. STACKED AUTOENCODER

The stacked autoencoder is a typical deep neural network, which is widely used in feature learning and representation. The network determines the parameters by greedy learning layer-by-layer, and then fine-tune the network-wide parameters through back propagation from the topmostlayer. If no label information is added to the topmost layer, then the learning process is an unsupervised learning process. If the label information of the sample is added to the topmost layer to reversely fine-tune the network-wide parameters, then the learning process becomes a supervised learning process. The stacked autoencoder as M layers, and the hidden layer of the topmost autoencoder is the network output, which is usually connected to the Softmax classifier behind the stacked autoencoder for classified identification. In the supervised learning network, the labels of the sample are also added into the Softmax classifier, and the network parameters are tuned as a whole by means of the back-propagation algorithm. The target function for fine-tuning the network parameters is as follows:

$$J(W_k|_{k-1}^{M}, b_k|_{k-1}^{M}) = \underset{W_k, b_k}{argmin} \sum_{i=1}^{N} \|y_i - g_M(f_M(\cdots f_1(x_i)))\|_2^2 \tag{4}$$

Where y_i represents the label corresponding to sample x_i, W_k and b_k represent the weight and bias of the network layers respectively. Optimizing the entire network by adding the label information into it can make feature representation of the network more suitable for the classifier i.e. different classes of samples will have greater feature variance.

C. ELM NETWORK STRUCTURE

The ELM is a new type of single-hidden-layer feedforward neural network that generates an uniquely optimal solution without having to fine-tune the network input weights and the hidden-layer bias value in the algorithm implementation process. All that is needed is to train the network's output weights.Thus, the training speed of the network can be greatly improved [26]. Suppose there are N different samples (x_i, l_i), where n is the number of the input-layer nodes of the network, $l_i = [l_{i1}, l_{i2}, \ldots, l_{im}]^T \in R^m$, and m is the number of the output-layer nodes of the network, then the mathematical expression is as shown in Equation(5):

$$l_i = \sum_{j=i}^{r} \beta_j g(w_j \cdot x_i + b_j), i = 1, 2, \ldots, N \tag{5}$$

In the equation, r represents the number of hidden-layer nodes of the network, $w_j = [w_{j1}, w_{j2}, \ldots, w_{jn}]$ represents the input weight vector connecting the input layer and the jth hidden-layer node, where $1 \le j \le r$, $\beta_j = [\beta_{j1}, \beta_{j2}, \ldots, \beta_{jm}]^T$ represents the output weight vector connecting the jth hidden-layer node and the output-layer node, where m is the number of output-layer nodes of the network. $l_i = [l_{i1}, l_{i2}, \ldots, l_{im}]^T$ represents the network output value, $g(\cdot)$ is the excitation function of the hidden-layer neuron, generally taken as the Sigmoid function, and b_j is the bias value of the hidden layer.

Train with N different samples(x_i, l_i). First, b_j is randomly generated and w_j stays fixed. It just needs to determine the value of β through training. The value can be found through one-step calculation using the pseudo-inverse algorithm, as shown in Equation(6).

$$\beta = H^{\dagger} L \tag{6}$$

$$H = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \cdots & g(w_r \cdot x_1 + b_r) \\ \vdots & \cdots & \vdots \\ g(w_1 \cdot x_N + b_1) & \cdots & g(w_r \cdot x_N + b_r) \end{bmatrix}_{N \times r} \tag{7}$$

Where H is the hidden-layer output matrix in the ELM, as shown in Equation(7), where $x_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^T \in R^n$, $1 \le i \le N$, $H^{\dagger}$ is the pseudo-inverse of H,

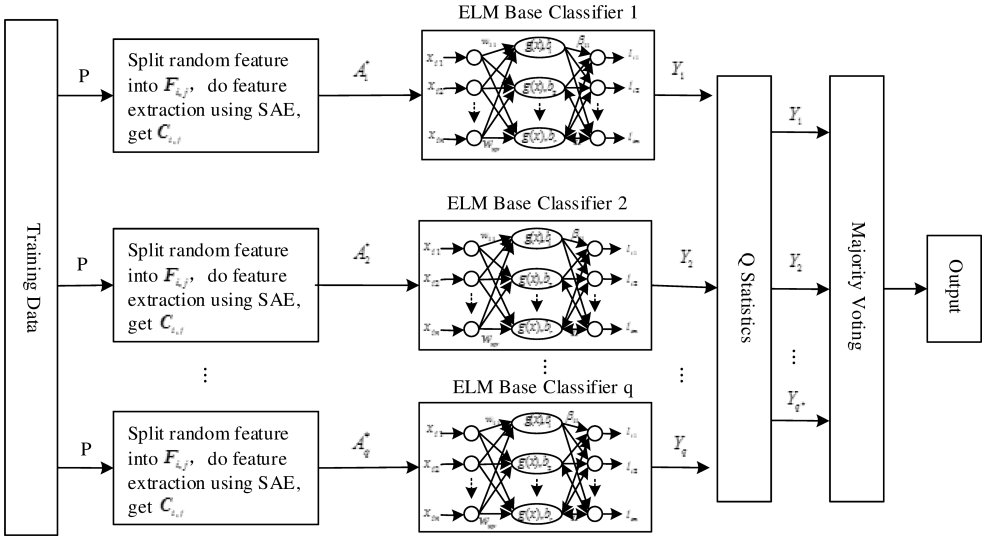

FIGURE 1. Flow diagram of SAE-ELM.

$\beta = [\beta_1, \beta_2, \ldots, \beta_r]^T, 1 \le j \le r; L = [l_1, l_2, \ldots, l_N]^T$ is the expected output vector, where $l_i = [l_{i1}, l_{i2}, \ldots, l_{im}]^T \in R^m$, m is the number of output-layer nodes in the network, $1 \le i \le N$. Train the ELM network and find the solution of β.

Traditional neural network algorithms (e.g. BP neural network) need to be manually equipped with a large number of network training parameters, and are prone to local optimal solution. The ELM neural network needs only to be configured with the number of hidden-layer nodes in the network.

During the process of algorithm implementation, there is no need to readjust the input weight of the network and the hidden-layer bias values or to generate a uniquely optimal solution. Therefore, the ELM has the advantages of fast learning and good generalization performance. Because the remote sensing image features massive data and complicated data type but not a quite large number of spatial dimensions, the ELM can be used in the classified processing of remote sensing images and effectively improve classification efficiency, without incurring problems such as memory overflow.

D. Q-STATISTICS DESCRIPTION

Given that there are N training samples and two classifiers (C_i, C_j), while N^{11}, N^{00} represent the number of samples that are all correctly classified and all incorrectly classified, respectively, by both C_i and C_j. N^{10} represents the number of samples that are correctly classified by C_i and incorrectly classified by C_j; and N^{01} represents the number of samples correctly classified by C_j but incorrectly classified by C_i. Definition of the Q-statistic value $Q_{i,j}$ of the classifiers C_i and C_j is as shown in Equation(8) [27].

$$Q_{i,j} = \frac{N^{11}N^{00} - N^{01}N^{10}}{N^{11}N^{00} + N^{01}N^{10}} \tag{8}$$

As seen from Equation(8), $Q_{i,j}$ is valued from the range-between -1 and 1. If the two classifiers are independent of each other, then the value of Q is 0. If both classifiers tend to assign a target correctly to the same class, then Q is positive, and if both classifiers tend to classify a target incorrectly into the same class, the Q is negative. Assuming there are k number of classifiers, then the Q-statistics refers to the value of all pairs of classifiers,Q_{av} is as shown in Equation(9):

$$Q_{av} = \frac{2}{k(k-1)} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} Q_{i,j} \tag{9}$$

The Q-statistics can be used to measure the difference between the base classifiers in the ensemble algorithm, and the computation is simple. Therefore, when the algorithm proposed in this paper selects the base classifiers with big difference for ensemble, Q-statistics is selected as the measurement index to obtain better classification results.

E. SAE-ELM ALGORITHM

Assume that $P = [p_1, p_2, \ldots, p_n]^T$ is an n-dimensional group of data sample points, P represents a set of training samples with $N*n$ dimensions, Y represents the class number corresponding to each sample, and is expressed as $Y = [y_1, y_2, \ldots, y_N]^T$, where y_i is the a class number in the class set $\{z_1, z_2, \ldots, z_j\}$, where j is the total number of classes.

The $C_1, \ldots, C_q$ are the ELM base classifiers in the ensemble process, where q is the number of base classifiers, and F is the total of all feature sets that these base classifiers are built upon. The SAE-ELM algorithm is used to process the training samples to get the new training set of each base classifiers, and the final classification results obtained are shown in Fig. 1. The algorithm consists of the following five specific steps.

- Step 1) F is randomly divided into K subsets, while M is the number of features contained in each feature

subset.Based on different means of segmentation, these subsets may be relatively independent, or overlapping and repetitive. Relatively independentİ means that each feature subset does not contain features of other feature subsets, and overlapping and repetitiveİ means that each feature subset may contain features of other feature subsets. In order to increase inter-feature diversity, relatively independent feature subsets are selected. To simplify the calculation, assume that M is an adjustable variable, and if M is fully divisible by n, then $K = n/M$, if the quotient is S and the remainder is T when n is divided by M, the $K = S + 1$, where S subsets contain M features and the last feature subset contains T features.

- Step 2) Set $F_{i,j}$ as the jth feature subset used for training by the base classifier C_i,where $1 \leq i \leq q, 1 \leq j \leq K$. Using only M features in $F_{i,j}$ for SAE conversion, and using $C_{i,j} = [r_{i,j}^{(1)}, \ldots, r_{i,j}^{(M_j)}]$ to represent the resulting feature coefficients, the number of dimensions of each coefficient will be $M*1$, where M_j represents the number of features generated. For the resulting matrix, as the calculated feature values are likely to be zero, it may not be possible to obtain al M vectors, i.e. $M_j \leq M$ may occur. Therefore, the SAE feature is used only in the segmented feature subsets, but not in all of the data sets, so as to avoid generating similar coefficients for similar feature subsets in different base classifiers.
- Step 3) Construct a matrix A_i by synthesizing the vectors with feature coefficients.In order to construct a new training set of the base classifier C_i, the columns of the matrix A_i are first rearranged so that the arranged form is consistent with the arrangement of the original feature set, and the rearranged matrix is represented as A_i^* with $N * n$ dimensions. The training dataset of the base classifier C_i is (A_i^*, Y), where Y is the class corresponding to each sample in the dataset.
- Step 4) Given that a new training set is (A_i^*, Y) for an ELM base classifier C_i, and the training of ELM network can be summarized into three points according to the description in Section C:a.Random selection of the input weight w_i and the hidden-layer bias value b_i. b.Calculate the output matrix H of the hidden layer based on Equation(6).C. Obtain the output weight β based on Equation(5).
- Step 5) Granted that the number of ensemble base classifiers is q, select q^* base classifiers from q base classifiers, and we will get the Q-statistics values of the q^* base classifiers through Equation(9). Select the ensemble classifier set closest to 0 as the final ensemble classifier set, and use the voting method to get the classification results.

As seen from the above description of the proposed SAE-ELM algorithm, the training set is completed with relatively independent feature segmentation and SAE transformation in the initial part of the SAE-ELM algorithm. Therefore, randomness is introduced into each ELM base classifier, which also enhances the difference between the base classifiers. From the basic theory of ensemble learning, we can see that when the classification accuracy of each base classifier remains constant in the ensemble algorithm, increase of variance among basic classifiers will effectively improve the classification result [6]. Therefore, the proposed SAE-ELM algorithm can generate better classification result.

IV. SIMULATION RESULTS

In order to verify the effectiveness of the proposed algorithm, the first step is to apply it to the classification processing of true remote sensing image. The selected remote sensing data are Landsat ETM+ remote sensing images of Zhalong Wetland, the SPOT5 remote sensing images of a certain development zone and the ALOS remote sensing images of Haicheng River, which resolutions are 60m, 10m and 2.5m respectively, and Indian Pines hyperspectral remote sensing images. To verify the generalization of the proposed algorithm, experimental simulation has been performed on the UCI standard dataset.

A. SIMULATION OF ZHALONG WETLAND IMAGE DATA

The remote sensing images of Zhalong wetland are as shown in Fig. 2 (a). The image samples total 262,114 and the sizes are 512*512 pixels and 1-8 bands of the images are selected for the experiments. Visual interpretation and the status of land use show that the remote sensing images of Zhalong Wetland mainly include five land features: grassland, marsh, fire area, saline-alkali land and water body. According to the survey results, 3,000 image elements are selected from 262,114 samples as labeled sample points, of which 2,400 are used as training samples and 600 as test samples. The three most important parameters in the SAE-ELM algorithm proposed in this paper are the number of sub-features selected for constructing the transformation matrix, the number of hidden-layer nodes in the ELM neural network and the number of base classifiers selected by Q-statistics. As far as the current study is concerned, there is no specific theory to explain how to select these three parameters, and the best parameters are usually obtained on the basis of experience or experiments. As known through previous research and experimental simulation, the optimal number of sub-features selected during construction of the transformation matrix is between two and five.

The number of hidden-layer nodes in the ELM neural network is related to the corresponding experimental data and is indefinite; while the number of ensemble base classifiers determined by Q-statistics can be kept at one third of the original base classifiers. Therefore, in the experiment part of this paper, the number of sub-features selected for constructing transformation matrix is 3, the number of hidden-layer nodes selected by ELM is 100, and the number of base classifiers selected by Q-statistics is 8.

The comparing algorithms used are Bagging [8], Adaboost [21], Random Forest [28] and ELM [26] algorithm, the parameters are used with the best parameters in the corresponding literature, i.e. The number of iterations in

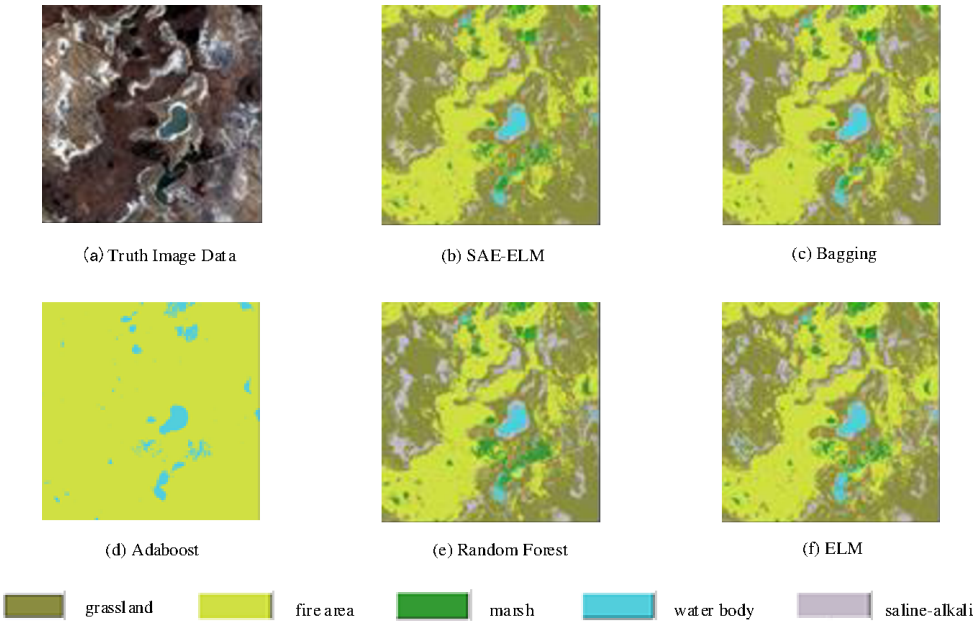

FIGURE 2. Classification results by different algorithms on zhalong wetlands remote sensing image data.

TABLE 1. Overall accuracy and kappa coefficient of zhalong wetlands remote sensing image.

	Overall Accuracy	Kappa Coefficient
SAE-ELM	0.953	0.9462
Bagging[8]	0.925	0.9027
Adaboost[21]	0.485	0.3108
Random Forest[28]	0.925	0.9027
ELM[26]	0.926	0.9031

Bagging, Adaboost and Random Forest is 10, all the base classifiers adopt the decision tree, and the number of hidden-layer nodes of ELM neural network is 100. The experiment adopts averaging 50 times to estimate the precision of the calculation. The number of base classifiers is 20 in the proposed algorithm SAE-ELM and the other three comparing ensemble algorithms.

Table 1 shows the comparison between the overall classification accuracy and the KAPPA coefficient obtained from the test samples. As can be seen from Table 1, compared with other ensemble algorithms, the SAE-ELM algorithm has some improvements over the overall accuracy and KAPPA coefficient, 0.953 and 0.9462, respectively, which are better than 0.925 and 0.9027 of Bagging, 0.485 and 0.3108 of Adaboost, 0.925 and 0.9027 of Random Forest and 0.926 and 0.9031 of ELM, respectively. The overall classification accuracy shows that the algorithm is effective in classifying remote sensing images, and the high Kappa coefficient indicates the stability of the algorithm. As can be seen from the comparison of the above two indexes, the proposed algorithm of SAE-ELM is more suitable for classifying the low-resolution remote sensing images than the other four algorithms.

In order to show intuitively the classification effect of the proposed algorithm in remote sensing image, the entire remote sensing image is classified with the 3,000 labeled sample points selected as training samples, and the classification effect diagram for Zhalong Wetland is as shown in Fig. 2, of which(b) is the classification result diagram obtained in the algorithm proposed in this text, and (c), (d), (e) and (f) are the classification results of the Bagging algorithm, Adaboost algorithm, Random Forest algorithm and the ELM algorithm. We can see that there is no problem of unclear classification in Adaboost algorithm, and the farmland, marsh and saline-alkali land are not distinctively classified; and the Random Forest algorithm has too many misclassifications. The algorithm proposed in this paper can better classify the remote sensing images for Zhalong Wetland and enjoys certain advantages over other comparison algorithms.

B. SIMULATION OF A CERTAIN DEVELOPMENT ZONE IMAGE DATA

The remote sensing image of a certain development zone is shown in Fig. 3(a) in the same size of 512 * 512 pixels. The green band, red band and near-infrared band of the images are selected for the experiments. As found through field research and visual interpretation, the remote sensing images of the development zone are classified into five types: sea, development land, residential area, green belt and bare land. The experiment also selected 3,000 image elements from 262,114 samples as the labeled sample points, of which 2,400 were used as training samples and 600 as test samples. The methods for selection of the proposed algorithm parameters are the same as that for the experiment in Section A, and the parameters are also compared with other four algorithms e.g. Bagging.

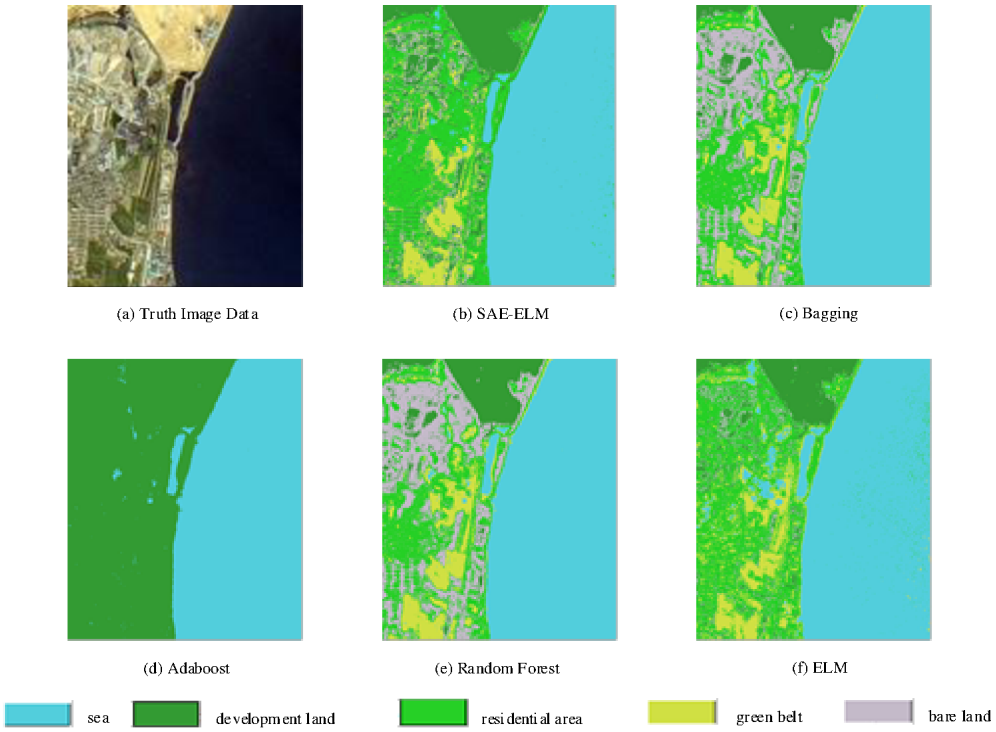

FIGURE 3. Classification results by different algorithms on a certain development zone remote sensing image data.

TABLE 2. Overall accuracy and kappa coefficient of a certain development zone remote sensing image data.

	Overall Accuracy	Kappa Coefficient
SAE-ELM	0.932	0.9100
Bagging[8]	0.918	0.8926
Adaboost[21]	0.518	0.3698
Random Forest[28]	0.901	0.8732
ELM[26]	0.895	0.8619

Table 2 shows the comparison between the overall classification accuracy and the KAPPA coefficient obtained from the test samples. As can be seen from Table 2, compared with other comparison algorithms, the SAE-ELM algorithm has advantages in overall accuracy and KAPPA coefficient, reaching 0.932 and 0.9100, respectively, better than other algorithms such as Bagging, Adaboost, Random Forest and ELM. The classification accuracy and KAPPA coefficients reflect the classification effect and stability of the algorithm on remote sensing images. Comparison of the said two indicators shows that the proposed algorithm of SAE-ELM is better than the other four contrast algorithms in the classification of remote sensing images with medium resolution.

Just as in Experiment A, the whole image is classified using 3,000 labeled sample points as training samples to obtain the classification effect, as shown in Fig. 3 of which (b) is the classification result diagram obtained from the algorithm proposed in this paper, and (c), (d), (e) and (f) are the classification results obtained by Bagging, Adaboost, Random Forest and ELM, respectively. As can be seen from the figure, the algorithm proposed in this paper can classify each type of objects, has no unclear classification like the algorithm of Adaboost, and has rendered good classification effect.

C. SIMULATION OF HAICHENG RIVER IMAGE DATA

The remote sensing images of Haicheng River are as shown in Fig. 4(a), in sizes of 512 * 512 pixels. The blue band, green band, red band and near-infrared band of the images are selected for the experiments. Combined with visual interpretation and field research, the remote sensing images of the Haicheng River are classified into six types: farmland, dry sandy land, residential area, river, wetland and grassland. 3,000 image elements are also selected from 262,114 samples for the experiment as the labeled sample points, of which 2,400 are used as training samples and 600 as test samples. The parameters of the proposed algorithms are the same as those in Experiment A and they have been compared with the algorithms of Bagging, Adaboost, Random Forest and ELM as well.

As can be seen in Table 3, the classification accuracy and KAPPA coefficient from the proposed SAE-ELM algorithm have reached 0.923 and 0.9142, respectively, better than the comparing algorithms of Bagging, Adaboost, Random Forest and ELM. Similarly, the high overall classification accuracy indicates that the algorithm has good effects in classifying remote sensing images, and the high Kappa coefficient indicates good stability of the algorithm. Therefore, the proposed SAE-ELM algorithm also performs well in the classification processing of high-resolution remote sensing images.

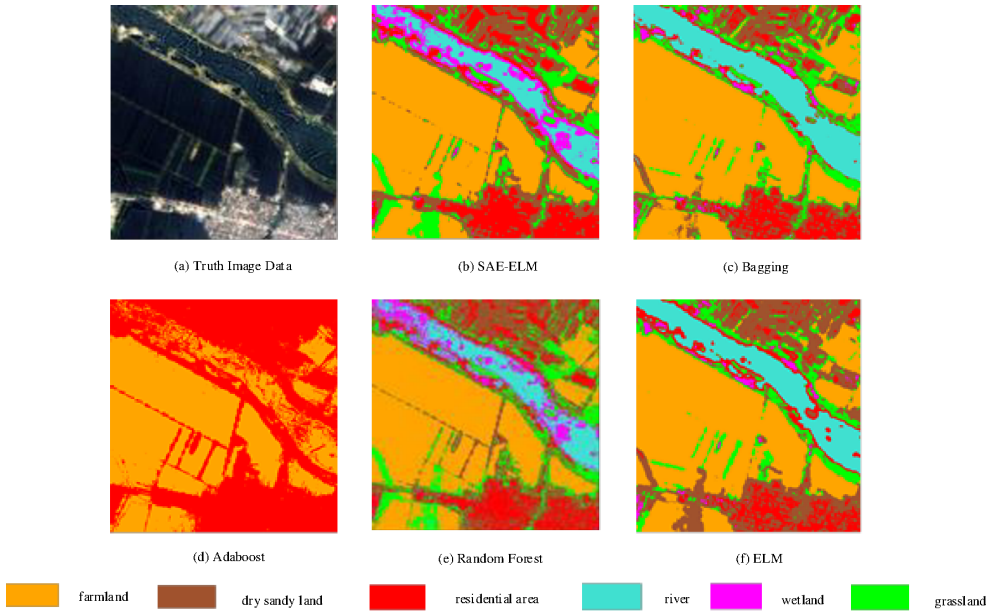

FIGURE 4. Classification results by different algorithms on remote sensing image data of haicheng river.

TABLE 3. Overall accuracy and kappa coefficient of haicheng rivers remote sensing image data.

	Overall Accuracy	Kappa Coefficient
SAE-ELM	0.923	0.9142
Bagging[8]	0.852	0.8329
Adaboost[21]	0.610	0.4419
Random Forest[28]	0.886	0.8518
ELM[26]	0.833	0.8114

Classification of the entire remote sensing image of Haicheng River is done in the same manner as for Experiments A and B. The obtained classification chart is shown in Fig. 4, of which (b) is the classification result chart obtained by the algorithm of this paper, and (c), (d) and (e) and (f) are the classification results obtained using the algorithms of Bagging, Adaboost, Random Forest and ELM respectively. It can be seen from the figure that the proposed algorithm can classify each kind of terrain clearly, and the integrity is also rather good. Moreover, there are no excessive salt and pepper noises.

As generally shown in Tables 1 to 3 and Fig. 2, Fig. 3 and Fig. 4, the algorithm proposed in this paper has good results in classifying remote sensing images with fewer labeled training samples. Whats worth mentioning is that the resolutions of the three remote sensing images vary widely, among which the ground resolution of the remote sensing image of Zhalong Wetland is 30m, that of a certain development zone is10m and that of Haicheng River is 2.5m. The results show that the proposed algorithm has rather good results on three images, which means the generalization of the algorithm is better than other algorithms. Achievement of good classification results through algorithms depends on the improvement of ensemble variance.

TABLE 4. Overall accuracy and kappa coefficient of indian pines remote sensing image data.

	Overall Accuracy	Kappa Coefficient
SAE-ELM	0.9467	0.9389
Bagging[8]	0.8787	0.8420
Random Forest[28]	0.8576	0.8366
Rotate Forest[20]	0.7569	0.7239
ELM[26]	0.9138	0.9013
SVM	0.8794	0.8628

D. SIMULATION OF INDIAN PINES IMAGE DATA

In order to further assess the performance of SAE-ELM algorithm, we conduct the experiments with hyperspectral remote sensing images obtained from NASAs Airborne Visible Infra-Red Imaging Spectrometer (AVIRIS). AVIRIS dataset is captured over a vegetation area of Indian Pines, Indiana, USA. The image contains pixels, with 200 spectral bands after removing 20 water absorption bands (104-108, 150-163, and 220). The spatial resolution is 20m/pixel. The classification data of AVIRIS is shown in Table.4. 10000 image elements are also selected from all samples for the experiment as the labeled sample points, of which 8000 are used as training samples and 2000 as test samples. The parameters of the proposed algorithms are the same as those in Experiment A and they have been compared with the algorithms of Bagging, Random Forest, Rotate Forest, SVM and ELM as well.

As can be seen in Table 5, the classification overall accuracy and KAPPA coefficient from the proposed SAE-ELM algorithm have reached 94.67 and 93.89, respectively, better than the comparing algorithms of Bagging, Random Forest, Rotate Forest, SVM and ELM. Similarly, the high overall classification accuracy indicates that the algorithm has good effects on classifying hyperspectral remote sensing images, Therefore, the proposed SAE-ELM algorithm also

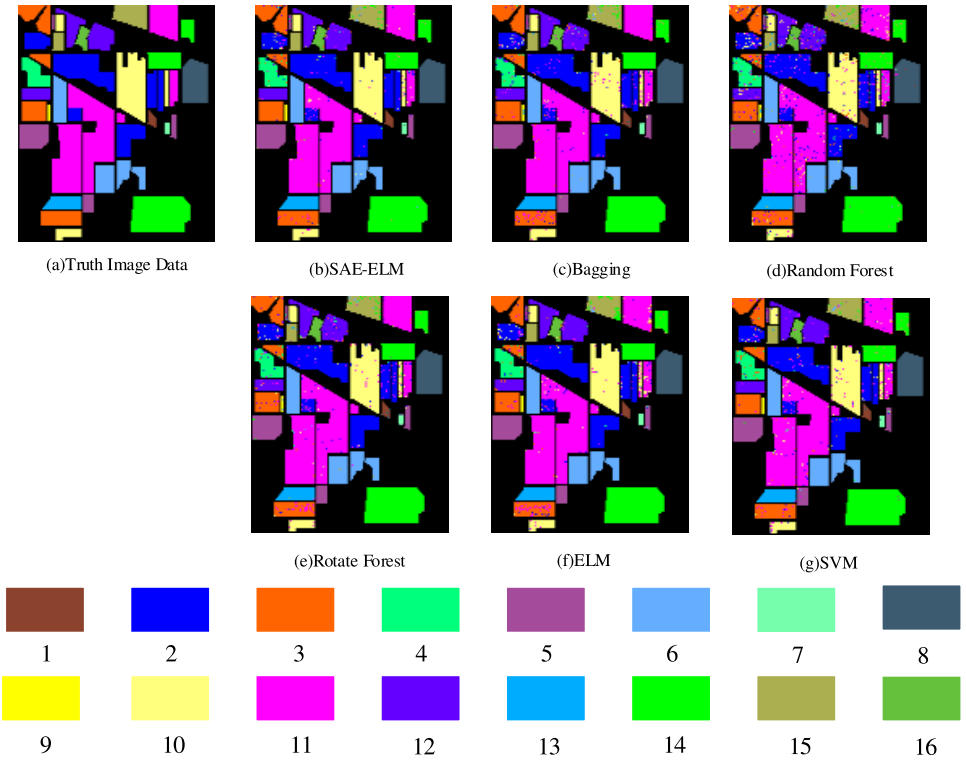

FIGURE 5. Classification results by different algorithms on indian pines remote sensing image data.

TABLE 5. Attributes of experiment data.

	Samples	Features	Classes	Type
Balance scale	625	4	3	Categorical
Car_E	1728	6	4	Categorical
ILPD	583	10	2	Integer, Real
Breast-w	699	10	2	Integer
Tic-tac-toe	958	9	2	Categorical

TABLE 6. Comparison of accuracy for different ensemble algorithms in this paper.

	SAE-ELM	Bagging[8]	Adaboost[21]
Balance scale	0.9225	0.7640	0.7240
Car_E	0.9327	0.8459	0.7062
ILPD	0.7345	0.6983	0.6724
Breast-w	0.9593	0.9499	0.9302
Tic-tac-toe	0.8634	0.7206	0.7507

performs better robustness on hyperspectral remote sensing images.

Classification of the entire remote sensing image of Indian Pines is done in the same manner as for Experiments A,B and C. The obtained classification chart is shown in Fig. 5, of which (b) is the classification result chart obtained by the algorithm of this paper, and (c), (d), (e), (f) and (g) are the classification results obtained using the algorithms of Bagging, Random Forest, Rotate Forest, ELM and SVM respectively. It can be seen from the figure that the proposed algorithm can get better classification result.

E. SIMULATION OF UCI DATA

To further verify the validity of the proposed algorithm, the algorithms are also compared in the UCI database [29], and a total of five sets of data are selected, and the data attributes in each group are as shown in Table 4. In the experiment, 80% of the data are used as the training data, and 20%as test data, and are compared with the classical algorithms of Bagging and Adaboost respectively. The parameters are selected in the same manner as for the parameters of Experiment A, and the results obtained are shown in Table 7. As can be seen in Table 7, the proposed algorithm achieves the highest accuracy in the five groups of UCI data. Table 6 shows that these five sets of data include Integer data, Real data and Categorical data, and the number of classes are not always the same, which means the generalization performance of the proposed algorithm. The above-mentioned experiment results show that the proposed SAE-ELM algorithm has strong generalization performance, which has good results not only for classifying remote sensing data, but also rather favorable results for classifying data with more dimensions and more classes.

V. CONCLUSION

In this paper, a remote sensing image classification algorithm based on the ensemble of ELM is proposed to solve the problem of poor data classification and low classification efficiency due to the complex data type and the small number

of training samples in classifying remote sensing images. The algorithm firstly segments the features in the original training set, and then transforms the segmented sub-features for classification to improve variance among the base classifiers. Then we use the transformed training set to train each ELM base classifier, and finally use Q-statistics to measure the difference between base classifiers, and the classifier with greater difference is selected to synthesize the results and to obtain the final classification results.

According to the classification experiment results on the remote sensing images for Zhalong Wetland, a certain development zone, Haicheng River and Indian Pines, compared with other ensemble algorithms, the proposed algorithm has higher classification accuracy and stronger generalization performance, and can adapt to the classification of different resolutions and Hyperspectral remote sensing images.

We will further focus on the selection of the number of base classifiers for Q-statistics. Now it is by manual selection only, and the theoretical basis is far from many, which will call for solution in the next step of work.

REFERENCES

[1] G. T. Kaya, O. K. Ersoy, and M. E. Kamasak, "Support vector selection and adaptation for remote sensing classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2071–2079, Jun. 2011.
[2] Z.-X. Panm, J. Yu, S.-X. Hu, and W.-D. Sun, "Single image super resolution based on multi-scale structural self-similarity," *Acta Autom. Sinica*, vol. 40, no. 4, pp. 549–603, 2014.
[3] K. Tan, X. Jin, A. Plaza, X. Wang, L. Xiao, and P. Du, "Automatic change detection in high-resolution remote sensing images by using a multiple classifier system and spectral–spatial features," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 8, pp. 3439–3451, Aug. 2016.
[4] E. Othman, Y. Bazi, H. AlHichri, and N. Alajlan, "A deep learning approach for unsupervised domain adaptation in multitemporal remote sensing images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2015, pp. 2401–2404.
[5] M. A. Bencherif, Y. Bazi, A. Guessoum, N. Alajlan, F. Melgani, and H. AlHichri, "Fusion of extreme learning machine and graph-based optimization methods for active classification of remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 527–531, Mar. 2015.
[6] X. Liu, C. Deng, S. Wang, G.-B. Huang, B. Zhao, and P. Lauren, "Fast and accurate spatiotemporal fusion based upon extreme learning machine," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 2039–2043, Dec. 2016.
[7] P. Du and A. Samat, "Multiple instance ensemble learning method for high-resolution remote sensing image classification," *J. Remote Sens.*, vol. 17, no. 1, pp. 77–97, 2013.
[8] B. Peter, *Handbook of Computational Statistics*. Berlin, Germany: Springer, 2012, pp. 985–1022.
[9] S. Ghorai, A. Mukherjee, S. Sengupta, and P. K. Dutta, "Cancer classification from gene expression data by NPPC ensemble," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 8, no. 3, pp. 659–671, May/Jun. 2011.
[10] M.-J. Kim and D.-K. Kang, "Ensemble with neural networks for bankruptcy prediction," *Expert Syst. Appl.*, vol. 37, no. 4, pp. 3373–3379, Apr. 2010.
[11] S. Khakabimamaghani, F. Barzinpour, and M. R. Gholamian, "A high diversity hybrid ensemble of classifiers," in *Proc. 2nd Int. Conf. Softw. Eng. Data Mining*, Jun. 2010, pp. 461–466.
[12] M. Han and B. Liu, "A remote sensing image classification method based on extreme learning machine ensemble," in *Advances in Neural Networks*. Berlin, Germany: Springer, 2013, pp. 447–454.
[13] A. Merentitis and C. Debes, "Many hands make light work—On ensemble learning techniques for data fusion in remote sensing," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 86–99, Sep. 2015.
[14] 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.
[15] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016.
[16] B. B. Damodaran, R. R. Nidamanuri, and Y. Tarabalka, "Dynamic ensemble selection approach for hyperspectral image classification with joint spectral and spatial information," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2405–2417, Jun. 2015.
[17] A. Samat, P. Du, S. Liu, J. Li, and L. Cheng, "E^2LMs: Ensemble extreme learning machines for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1060–1069, Apr. 2014.
[18] Y. Tao, X. Gao, A. Ihler, K. Hsu, and S. Sorooshian, "Deep neural networks for precipitation estimation from remotely sensed information," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jul. 2016, pp. 1349–1355.
[19] X. Han, Y. Zhong, B. Zhao, and L. Zhang, "Unsupervised hierarchical convolutional sparse auto-encoder for high spatial resolution imagery scene classification," in *Proc. 11th Int. Conf. Natural Comput. (ICNC)*, Aug. 2015, pp. 42–46.
[20] J. J. Rodriguez, L. I. Kuncheva, and C. J. Alonso, "Rotation forest: A new classifier ensemble method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1619–1630, Oct. 2006.
[21] N. García-Pedrajas, "Supervised projection approach for boosting classifiers," *Pattern Recognit.*, vol. 42, no. 9, pp. 1742–1760, Sep. 2009.
[22] L. L. C. Kasun, Y. Yang, G.-B. Huang, and Z. Zhang, "Dimension reduction with extreme learning machine," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3906–3918, Aug. 2016.
[23] H. J. Kim, B. S. Choi, and M. Y. Huh, "Booster in high dimensional data classification," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 1, pp. 29–40, Jan. 2016.
[24] S. Wang and X. Yao, "Relationships between diversity of classification ensembles and single-class performance measures," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 1, pp. 206–219, Jan. 2013.
[25] C. Yang, X. Yin, and H. Hao, "Classifier ensemble with diversity: Effectiveness analysis and ensemble optimization," *Acta Autom. Sinica*, vol. 40, no. 4, pp. 660–674, 2014.
[26] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, Dec. 2006.
[27] L. I. Kuncheva and C. J. Whitaker, "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy," *Mach. Learn.*, vol. 51, no. 2, pp. 181–207, 2003.
[28] M. Bacauskiene, A. Verikas, A. Gelzinis, and A. Vegiene, "Random forests based monitoring of human larynx using questionnaire data," *Expert Syst. Appl.*, vol. 39, no. 5, pp. 5506–5512, Apr. 2012.
[29] C. L. Blake and C. J. Merz. *UCI Repository of Machine Learning Databases*. [Online]. Available: http://www.ics.uci.Edu/ mlearn/MLR-
[30] L. Yu, L. Feng, C. Chen, T. Qiu, L. Li, and J. Wu, "A novel multi-feature representation of images for heterogeneous IoTs," *IEEE Access*, vol. 4, pp. 6204–6215, 2016.

FEI LV received the M.S. degree from the School of Electronic and Information Engineering, Dalian University of Technology, Dalian, China, in 2012, where he is currently pursuing the Ph.D. degree with the Faculty of Electronic Information and Electrical Engineering. His current research interests include neural networks and ensemble classification.

MIN HAN (M'95–A'03–SM'06) received the B.S. and M.S. degrees from the Department of Electrical Engineering, Dalian University of Technology, Dalian, China, in 1982 and 1993, respectively, and the M.S. and Ph.D. degrees from Kyushu University, Fukuoka, Japan, in 1996 and 1999, respectively. Since 2003, she has been a Professor with the Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology. She has visited to Washington University, St. Louis, USA, in 2009. She serves as a Deputy Director of the Chinese Society of Instrumentation Youth Work Committee, the Director of the Institute of Liaoning Province System Simulation, a Committee Member of the Chinese Society of Artificial Intelligence, a Consultant of Jiangsu Province Department of Science and Technology, a Deputy Director of the Institute of Fuzzy Information Processing and Machine Intelligence, Dalian University of Technology, and an Organizing Chair of ISNN2013, ICICIP 2014, and ICIST2016. She has authored five books and over 300 articles in international journals and conference proceedings. She has made a deep research on the multivariate chaotic time series prediction, remote sensing image interpretation, and optimization of industrial process. For example, she proposes a support vector echo state machine to solve the problem of multivariate chaotic time series modeling of complex systems. She has applied the migration learning to the multi period remote sensing image classification, which realizes the reuse of historical data and reduces the demand of the current sample. Her current research interests include complex system modeling and forecasting method, time series analysis and forecasting, artificial intelligence method, neural networks, and chaos and their applications to control and identification.

TIE QIU (M'12–SM'16) received the M.Sc. and Ph.D. degrees from the Dalian University of Technology (DUT), China, in 2005 and 2012, respectively. He was a Visiting Professor of Electrical and Computer Engineering with Iowa State University, USA, from 2014 to 2015. He is currently an Associate Professor with the School of Software, DUT. He has authored/co-authored eight books and over 60 scientific papers in international journals and conference proceedings. He has contributed to the development of four copyrighted software systems and invented 15 patents. He is a Senior Member of China Computer Federation and the ACM. He serves as a TPC Member of Industrial IoT15, IoT16, AIA13, EEC14, EEC15, EEC16, and ICSN16, and the Workshop Chair of CISIS13 and ICCMSE15. He serves as the Program Chair of iThings2016 and iThings2017. He serves as an Associate Editor of the IEEE Access Journal, *Computers & Electrical Engineering* (Elsevier Journal), and *Human-Centric Computing and Information Sciences* (Springer Journal), an Editorial Board Member of *Ad Hoc Networks* (Elsevier Journal), the *Journal of Advanced Computer Science & Technology*, and the *International Journal on AdHoc Networking Systems*, and a Gest Editor of *Computers and Electrical Engineering* (Elsevier Journal) and *Ad Hoc Networks* (Elsevier Journal).