Firefly-Algorithm-Inspired Framework With Band Selection and Extreme Learning Machine for Hyperspectral Image Classification

Hongjun Su, *Member, IEEE*, Yue Cai, and Qian Du, *Senior Member, IEEE*

***Abstract*—A firefly algorithm (FA) inspired band selection and optimized extreme learning machine (ELM) for hyperspectral image classification is proposed. In this framework, FA is to select a subset of original bands to reduce the complexity of the ELM network. It is also adapted to optimize the parameters in ELM (i.e., regularization coefficient *C*, Gaussian kernel σ, and hidden number of neurons *L*). Due to very low complexity of ELM, its classification accuracy can be used as the objective function of FA during band selection and parameter optimization. In the experiments, two hyperspectral image datasets acquired by HYDICE and HYMAP are used, and the experiment results indicate that the proposed method can offer better performance, compared with particle swarm optimization and other related band selection algorithms.**

***Index Terms*—Band selection, extreme learning machine (ELM), firefly algorithm (FA), hyperspectral image classification.**

I. Introduction

HYPERSPECTRAL remote sensing, with its high spectral resolution and wide range of spectral range, provides a rich spectral information for improving the accuracy of target detection, classification, and identification. The acquired hyperspectral imagery has been widely used in applications of vegetation, atmospheric environment, geological exploration, military reconnaissance, where better classification performance is the most effective factor for data mining [1]–[3]. However, curse of dimensionality, difficult operation for nonlinear sample data, limitations of small and unbalanced sample data, and redundancy among bands put forward a huge challenge for hyperspectral image classification [4], [5].

Dimensionality reduction is often used to reduce data redundancy and extract useful features. Commonly adopted transforms includes principal component analysis and independent component analysis, which lead the physical information loss of original bands [6]–[8]. Band selection is another way of dimensionality reduction by directly selecting a subset of original bands. Class separability is a widely used metric when a band subset is selected. Euclidean distance, Jeffreys–Matusita (JM), or spectral information divergence are also used as the criteria. Minimum endmember abundance covariance is another efficient measure using class signatures for band selection without training samples [9]. It is often prohibitive to use classification accuracy produced by a classifier [such as support vector machine (SVM)] as the metric [10]–[12] due to a time-consuming training process.

Recently, evolutionary methods, with global searching ability, provide alternative searching approaches for band selection. Particle swarm optimization (PSO) [13], [14] and its variants have been successfully used for band selection. The firefly algorithm (FA) is a new type of swarm intelligence optimization algorithm. It can adaptively adjust the radius of the induction and parallelly search multiple peaks, which has natural advantages in solving multimode optimization problem. The improved FA has been successfully used for band selection [15].

For hyperspectral image classification, many machine learning methods such as artificial neural networks (ANNs) [16], SVM [17] are popularly used, and semisupervised learning, clustering techniques, and sparse representation are also introduced for improving hyperspectral classification accuracy. However, it is still a challenge to provide both high efficiency and fast speed limited to high-dimensional and complex spectral data. Recently, extreme learning machine (ELM) has been proposed by Huang [18]–[21]. ELM is much more efficient than ANNs or SVM since it uses random weights in the hidden layer and the output layer is linear. A large number of existing experiments demonstrate its fast speed and capability of providing accurate classification. In [22]–[25], kernel-based ELM was proposed and achieved promising results. Similar to SVM, the parameters of ELM play a crucial role for the performance. The number of hidden neurons *L*, for example, will increase with more training samples according to empirical studies [26]. A large input layer (i.e., large spectral dimensionality) often requires more hidden neurons. Some swarm intelligence methods such as differential evolution, genetic algorithm, and PSO for parameters optimization of ELM have been studied [26], [27]. Here, we propose to use the FA due to its fast convergence [28], [29] and the fact that it is applicable for multiple parameter optimization.

Manuscript received March 18, 2016; revised June 04, 2016; accepted July 07, 2016. This work supported in part by the National Natural Science Foundation of China (41571325, 41201341), the Fundamental Research Funds for the Central Universities (2015B16814), the Open Research Fund of Key Laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences (2014LDE003), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD). *(Corresponding author: Hongjun Su.)*

H. Su and Y. Cai are with the School of Earth Sciences and Engineering, Hohai University, Nanjing, 211100, China (e-mail: hjsurs@163.com; cy_199107@sina.com).

Q. Du is with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS 39762 USA (e-mail: du@ece.msstate.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2016.2591004

In this paper, FA-optimized ELM framework is proposed. Due to very low complexity of ELM, its classification accuracy is adopted as the objective function of FA during band selection and parameter optimization. First, FA is applied for pre-selection of bands. Then, the kernel ELM is optimized by FA for parameters tuning. Finally, with convergence, the selected bands and optimized parameters are used to provide the final classification result.

Although there are some studies for hyperspectral band selection and ELM optimization for classification, most of them are conducted separately. Obviously, the performance of ELM is relevant to the number of input neurons, i.e., the number of bands; parameters are mutually interactive. Thus, in this paper, the number of bands to be selected, the number of hidden neurons, and the other two parameters of ELM are considered simultaneously to be optimized. Note that this is the first time for the ELM parameters being optimized simultaneously, which is done with an efficient evolutionary algorithm such as FA.

The remainder of this paper is organized as follows. Section II introduces FA and ELM. The FA-based framework is presented in Section III. Section IV shows the experimental results. Section V makes several concluding remarks.

II. Related Works

A. FA

FA is an evolutionary optimization algorithm proposed by Yang [28]; it is inspired by group's searching based on fireflies' biological characteristic of fluorescence. Fireflies have different flashing behaviors, which are used for communication and attracting the potential prey. There are two necessary elements, brightness I and attractiveness β. The I at a particular distance r obeys the inverse square law, which means that I decreases as r increases. In the FA, the brightness of a firefly is expressed in terms of its current position: if it is brighter, its position is preferred; this also means the value of the objective function is larger. The less bright ones will move toward the brighter ones. In the case that the brightness of fireflies has the same value, they will move randomly.

For simplicity, the following hypotheses are made in describing the FA: 1) all fireflies are unisex so that one firefly will be attracted to others regardless of their sex; 2) the attractiveness of a firefly is proportional to its brightness; and 3) the brightness is proportional to the objective function [11], [12]. During the process of fly movement, I and β are updated repeatedly, and randomly distributed points are moved gradually toward the extreme points. After a certain number of iterations, the less desired points are eliminated and the best positional points are finalized.

The brightness of a firefly varies with the value of an objective function, which can be defined as

$$I(r) = I_0 e^{-\gamma r_{ij}^2} \tag{1}$$

where I_0 is the maximum brightness when $r = 0$; it is related to the value of the objective function, and a larger value means brighter. Here, γ is light absorption coefficient, and r_{ij} is the distance between the ith and jth fireflies.

TABLE I
Optimal Parameters of Different Optimization Algorithms for ELM Classifier Using in the Experiment

Para.	ELM	FA	PSO
1	$C = [10^0, 10^3]$	Maximum iterations = 100	Maximum iterations = 100
2	$\sigma = (0,1]$	Population size = 10	Population size = 10
3	$L = [0, 10^3]$	Light absorbance = 1	Inertia weight = 1
4		Maximum attractiveness = 1	Accelerating factor C1 = 1.7
5		Step size = 0.2	Accelerating factor C2 = 1.5

TABLE II
A List of Methods for Comparison

Method	Description
3FA-ELM (OA)	**using 3FA optimized ELM for classification with bands selected by FA simultaneously (OA as objective function)**
FA-ELM(OA)	using FA optimized ELM for classification with bands selected by FA separately (OA as objective function)
PSO-ELM (OA)	using PSO optimized ELM for classification with bands selected by PSO separately (OA as objective function)
FA-ELM(JM)	using FA optimized ELM for classification with bands selected by FA separately (JM as objective function)
ELM	using ELM for classification with all bands
FA-ELM	using FA optimized ELM for classification with all bands

TABLE III
Ground Truth for HYDICE Data

Class	Name	Samples
1	Road	947
2	Grass	963
3	Shadow	589
4	Trail	624
5	Tree	679
6	Roof	1139
	Total	4941

The attractiveness of a firefly is proportional to its light brightness observed by adjacent fireflies; it can be expressed as

$$\beta(r) = \beta_0 e^{-\gamma r_{ij}^2} \tag{2}$$

where β_0 is the attractiveness when the distance between two fireflies is zero. In this research, we simply set $I(r) = \beta(r)$. The equation that updates the jth firefly's location based on the ith firefly's attraction can be described as

$$m_j^{(i)} = m_j + \beta_0 e^{-\gamma r_{ij}^2} (m_j - m_i) + w \left(\mathbf{rand} - \frac{1}{2} \right) \tag{3}$$

where m_i and m_j are the initial position of the ith and jth firefly, respectively, w is a constant within [0, 1], *rand* is a random number within [0, 1] and r_{ij} is the distance between the ith and jth fireflies. In this research, a firefly represents selected band indices, which is a vector. Then, (3) becomes

$$\mathbf{M}_j^{(i)} = \mathbf{M}_j + \beta_0 e^{-\gamma r_{ij}^2} (\mathbf{M}_j - \mathbf{M}_i) + w \left(\mathbf{rand} - \frac{1}{2} \cdot \mathbf{1} \right) \tag{4}$$

where the distance r_{ij} is the Euclidean distance between the indices of two sets of selected bands, i.e., $\mathbf{M}_i$ and $\mathbf{M}_j$. For n fireflies, the new location of $\mathbf{M}_j$ can be determined after

TABLE IV
CLASSIFICATION PERFORMANCE OF DIFFERENT METHODS FOR HYDICE DATA WITH DIFFERENT BANDS AND TRAINING SAMPLES

(A) 2% AS TRAINING SAMPLES

Samples = 2%	Overall classification accuracy(OA)		
Bands = 5	Minimum	Maximum	Average
3FA-ELM(OA)	**0.9711**	**0.9756**	**0.9737**
FA-ELM(OA)	0.9602	0.9661	0.9637
PSO-ELM(OA)	0.9372	0.9399	0.9387
FA-ELM(JM)	0.9022	0.9060	0.9037
Bands = 10	Minimum	Maximum	Average
3FA-ELM(OA)	**0.9713**	**0.9736**	**0.9724**
FA-ELM(OA)	0.9540	0.9550	0.9548
PSO-ELM(OA)	0.9353	0.9374	0.9362
FA-ELM(JM)	0.8922	0.9207	0.9063
Bands = 15	Minimum	Maximum	Average
3FA-ELM(OA)	**0.9733**	**0.9755**	**0.9742**
FA-ELM(OA)	0.9522	0.9578	0.9554
PSO-ELM(OA)	0.9389	0.9434	0.9416
FA-ELM(JM)	0.9237	0.9288	0.9262
Bands = 20	Minimum	Maximum	Average
3FA-ELM(OA)	**0.9693**	**0.9731**	**0.9710**
FA-ELM(OA)	0.9584	0.9608	0.9595
PSO-ELM(OA)	0.9519	0.9571	0.9541
FA-ELM(JM)	0.9203	0.9250	0.9229
Bands = Allbands	Minimum	Maximum	Average
ELM	0.8585	0.9110	0.8941
FA-ELM	0.8803	0.9312	0.9208

(B) 5% AS TRAINING SAMPLES

Samples = 5%	Overall classification accuracy(OA)		
Bands = 5	Minimum	Maximum	Average
3FA-ELM(OA)	**0.9736**	**0.9777**	**0.9753**
FA-ELM(OA)	0.9715	0.9736	0.9726
PSO-ELM(OA)	0.9422	0.9489	0.9458
FA-ELM(JM)	0.9298	0.9330	0.9311
Bands = 10	Minimum	Maximum	Average
3FA-ELM(OA)	**0.9724**	**0.9798**	**0.9769**
FA-ELM(OA)	0.9732	0.9806	0.9761
PSO-ELM(OA)	0.9304	0.9387	0.9342
FA-ELM(JM)	0.8965	0.9102	0.8997
Bands = 15	Minimum	Maximum	Average
3FA-ELM(OA)	**0.9724**	**0.9796**	**0.9783**
FA-ELM(OA)	0.9762	0.9802	0.9787
PSO-ELM(OA)	0.9432	0.9493	0.9479
FA-ELM(JM)	0.9312	0.9385	0.9342
Bands = 20	Minimum	Maximum	Average
3FA-ELM(OA)	**0.9768**	**0.9865**	**0.9800**
FA-ELM(OA)	0.9704	0.9796	0.9770
PSO-ELM(OA)	0.9601	0.9677	0.9627
FA-ELM(JM)	0.9504	0.9597	0.9517
Bands = Allbands	Minimum	Maximum	Average
ELM	0.9117	0.9310	0.9276
FA-ELM	0.9290	0.9366	0.9331

(C) 8% AS TRAINING SAMPLES

Samples = 8%	Overall classification accuracy(OA)		
Bands = 5	Minimum	Maximum	Average
3FA-ELM(OA)	**0.9724**	**0.9861**	**0.9807**
FA-ELM(OA)	0.9725	0.9794	0.9734
PSO-ELM(OA)	0.9102	0.9215	0.9128
FA-ELM(JM)	0.8975	0.8997	0.8986
Bands = 10	Minimum	Maximum	Average
3FA-ELM(OA)	**0.9832**	**0.9896**	**0.9879**
FA-ELM(OA)	0.9743	0.9791	0.9787
PSO-ELM(OA)	0.9453	0.9485	0.9462
FA-ELM(JM)	0.9233	0.9279	0.9258
Bands = 15	Minimum	Maximum	Average
3FA-ELM(OA)	**0.9804**	**0.9884**	**0.9867**
FA-ELM(OA)	0.9793	0.9864	0.9813
PSO-ELM(OA)	0.9498	0.9547	0.9513
FA-ELM(JM)	0.9211	0.9279	0.9237
Bands = 20	Minimum	Maximum	Average
3FA-ELM(OA)	**0.9877**	**0.9912**	**0.9899**
FA-ELM(OA)	0.9745	0.9815	0.9794
PSO-ELM(OA)	0.9524	0.9607	0.9589
FA-ELM(JM)	0.9203	0.9287	0.9238
Bands = Allbands	Minimum	Maximum	Average
ELM	0.9198	0.9414	0.9327
FA-ELM	0.9392	0.9525	0.9468

(D) 10% AS TRAINING SAMPLES

Samples = 10%	Overall classification accuracy(OA)		
Bands = 5	Minimum	Maximum	Average
3FA-ELM(OA)	**0.9812**	**0.9859**	**0.9844**
FA-ELM(OA)	0.9732	0.9785	0.9778
PSO-ELM(OA)	0.9552	0.9597	0.9573
FA-ELM(JM)	0.9201	0.9254	0.9233
Bands = 10	Minimum	Maximum	Average
3FA-ELM(OA)	**0.9837**	**0.9897**	**0.9873**
FA-ELM(OA)	0.9802	0.9877	0.9830
PSO-ELM(OA)	0.9533	0.9594	0.9564
FA-ELM(JM)	0.9425	0.9491	0.9478
Bands = 15	Minimum	Maximum	Average
3FA-ELM(OA)	**0.9865**	**0.9889**	**0.9879**
FA-ELM(OA)	0.9802	0.9874	0.9834
PSO-ELM(OA)	0.9513	0.9577	0.9541
FA-ELM(JM)	0.9624	0.9715	0.9691
Bands = 20	Minimum	Maximum	Average
3FA-ELM(OA)	**0.9896**	**0.9913**	**0.9903**
FA-ELM(OA)	0.9822	0.9896	0.9863
PSO-ELM(OA)	0.9412	0.9479	0.9446
FA-ELM(JM)	0.9623	0.9701	0.9668
Bands = Allbands	Minimum	Maximum	Average
ELM	0.9398	0.9435	0.9413
FA-ELM	0.9511	0.9644	0.9570

considering all other fireflies:

$$\mathbf{M}_j^{\text{new}} = \sum_{i=1, i\neq j}^{m} \mathbf{M}_j^{(i)} / (n-1). \tag{5}$$

B. ELM

ELM is a single-hidden-layer feedforward neural network. Given a set of training data $(\boldsymbol{x}_s, \boldsymbol{t}_s)$, where $x_s = [x_{s1}, x_{s2}, \ldots, x_{sn}]^T \in R^n$, $t_s \in R^m, s = 1, \ldots, N$, the structure of ELM is composed of *N*-dimensional input layers and hidden layer of *L* nodes. The output of ELM is

$$\sum_{s=1}^{N} \mathbf{h}_s(\mathbf{x})\beta_s = \mathbf{t}_s, s = 1, \ldots, N. \tag{6}$$

It can be written as

$$\mathbf{H}\beta = \mathbf{T} \tag{7}$$

where $b = [b_1, b_2, \ldots, b_L]^T$ is the vector of output weights to connect the hidden layer of *L* nodes and the output node. **H** is the hidden-layer output matrix

$$\mathrm{H} = \begin{bmatrix} \mathbf{h}(\mathbf{x_1}) \\ \vdots \\ \mathbf{h}(\mathbf{x_n}) \end{bmatrix} = \begin{bmatrix} h_1(x_1) & \ldots & h_L(x_1) \\ \vdots & \ddots & \vdots \\ h_1(x_n) & \ldots & h_L(x_n) \end{bmatrix}. \tag{8}$$

The *s*th column in **H** represents the output vector, which is the *s*th hidden node according to the input vector of $(x_1, x_2, \ldots, x_N)$. In fact, **h(x)** is a kind of feature mapping

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

that reflects the samples from N-dimensional input space to the L-dimensional hidden-layer feature space.

The training criterion is

$$\text{Minimize} : \|\mathbf{H}\beta - \mathbf{T}\| . \tag{9}$$

We can find the properties that ELM tends to reach not only the smallest training error but also the smallest norm of output weights. According to Bartlett's theory, to achieve a smaller training error for feedforward neural networks, the networks should have smaller norms of weights. The only minimal norm least squares solution is

$$\tilde{\beta} = \mathbf{H}^{+}\mathbf{T} \tag{10}$$

where $\mathbf{H}^{+}$ is the Moore–Penrose generalized inverse of matrix $\mathbf{H}$.

With a multioutput node, the constraint optimization based ELM can be described as follows:

$$\begin{aligned} &\text{Minimize} : L_{P_{\mathrm{ELM}}} = \tfrac{1}{2}\|\beta\|^2 + C\tfrac{1}{2}\sum_{i=1}^{N}\|\xi_i\|^2 \\ &\text{subject to} : \mathbf{h}(\mathbf{x}_s)\beta = \mathbf{t}_s^T - \xi_s^T, s = 1, \ldots, N. \end{aligned} \tag{11}$$

According to the KKT theorem, to train ELM is equivalent to solve the following dual optimization problem:

$$\begin{aligned} L_{D_{ELM}} = &\frac{1}{2}\|\beta\|^2 + C\frac{1}{2}\sum_{s=1}^{N}\|\xi_s\|^2 \\ &- \sum_{s=1}^{N}\sum_{s=1}^{N}\alpha_{s,k}(\mathbf{h}(\mathbf{x}_s))\beta_k - t_{s,k} + \xi_{s,k}). \end{aligned} \tag{12}$$

For the partial derivatives of L

$$\begin{cases} \frac{\partial L_{D_{ELM}}}{\partial \beta_k} = 0 \rightarrow \beta_k = \sum_{s=1}^{N}\alpha_{s,k}\mathbf{h}(\mathbf{x}_s)^T \rightarrow \beta = \mathbf{H}^T\alpha \\ \frac{\partial L_{D_{ELM}}}{\partial \beta_k} = 0 \rightarrow \alpha_s = C\xi_s, \qquad s = 1, \ldots, N \\ \frac{\partial L_{D_{ELM}}}{\partial \beta_k} = 0 \rightarrow \mathbf{h}(\mathbf{x}_s)\beta - \mathbf{t}_s^T + \xi_s^T = 0, \quad s = 1, \ldots, N. \end{cases} \tag{13}$$

Transforming the above formula, the problem equals to

$$\left(\frac{\mathbf{I}}{C} + \mathbf{H}\mathbf{H}^T\right)\alpha = \mathbf{T} \tag{14}$$

where

$$\mathbf{T} = \begin{bmatrix} \mathbf{t}_1^{\mathbf{T}} \\ \vdots \\ \mathbf{t}_N^{\mathbf{T}} \end{bmatrix} = \begin{bmatrix} t_{11} & \ldots & t_{1L} \\ \vdots & \ddots & \vdots \\ t_{N1} & \ldots & t_{NL} \end{bmatrix}. \tag{15}$$

From (13), we can find

$$\beta = \mathbf{H}^T\left(\frac{\mathbf{I}}{C} + \mathbf{H}\mathbf{H}^T\right)^{-1}\mathbf{T} \tag{16}$$

Or

$$\beta = \left(\frac{\mathbf{I}}{C} + \mathbf{H}\mathbf{H}^T\right)^{-1}\mathbf{H}^T\mathbf{T}. \tag{17}$$

The output function of ELM classifier is

$$f(x) = \mathbf{h}(\mathbf{x})\beta = \mathbf{h}(\mathbf{x})\mathbf{H}^{\mathbf{T}}\left(\frac{\mathbf{I}}{C} + \mathbf{H}\mathbf{H}^{\mathbf{T}}\right)^{-1}\mathbf{T} \tag{18}$$

Or

$$f(x) = \mathbf{h}(\mathbf{x})\beta = \mathbf{h}(\mathbf{x})\left(\frac{\mathbf{I}}{C} + \mathbf{H}^{\mathbf{T}}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathbf{T}}\mathbf{T}. \tag{19}$$

If a feature mapping $\mathbf{h}(\mathbf{x})$ is unknown to users, Mercer's conditions can be used on ELM, We can define a kernel matrix for ELM as follows:

$$\Omega_{\mathrm{ELM}} = \mathbf{H}\mathbf{H}^T : \Omega_{\mathrm{ELM}_{s,k}} = h(x_s) \cdot h(x_k). \tag{20}$$

Finally, the output function of ELM classifier can be written as

$$\begin{aligned} f(x) = &\ \mathbf{h}(\mathbf{x})\mathbf{H}^{\mathbf{T}}\left(\frac{\mathbf{I}}{C} + \mathbf{H}\mathbf{H}^{\mathbf{T}}\right)^{-1}\mathbf{T} \\ = &\begin{bmatrix} K(x, x_1) \\ \vdots \\ K(x, x_N) \end{bmatrix}^T \left(\frac{\mathbf{I}}{C} + \mathbf{\Omega_{ELM}}\right)^{-1}\mathbf{T}. \end{aligned} \tag{21}$$

C. Band Selection Objective Function

The objective of band selection is to select an n-dimension vector from a p-dimension vector by some criterion functions ($n < p$). In this paper, in the view of extreme fast classification characteristics and excellent performance of ELM, the overall accuracy (OA) derived from ELM can be used for objective function directly. OA is a basic evaluation index based on a confusion matrix. For each random sample, OA means that the probability of the classification results is consistent with the test data types. The formula can be expressed as follows:

$$P_{OA} = \sum_{k=1}^{d} P_{kk} \Big/ S_n \tag{22}$$

where d means the number of classes, k represents the kth class, S_{n} is the total number of samples, P_{kk} is the sum of the diagonal, which is the sum of pixels classified correctly.

III. Proposed FA-Optimized Band Selection and Classification

A. FA-Optimized ELM

In the ELM classifier, regularization coefficient C, kernel parameter, and the number of hidden layer nodes L are three important parameters. A positive value **1/C** can be added to the diagonal of $\mathbf{H}^{\mathrm{T}}\mathbf{H}$ or $\mathbf{H}\mathbf{H}^{\mathrm{T}}$ of the Moore-Penrose to ensure the invertibility, so the solution is more stable and tends to have better generalization performance. When the RBF kernel is chosen, the Gaussian kernel σ will decide the new feature space distribution after data mapping. The number of hidden layer nodes will directly decide the size of the hidden layer of matrix and solution of the beta. Therefore, the parameter optimization plays a significant role for classification performance.

The overall steps can be described as follows.

1) *Band selection initialization*: FA-based initial band selection with the JM distance as the objective function.
2) *ELM parameters initialization*: the main parameters includes regularization coefficient C, kernel parameter σ,

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE V
CONFUSION MATRIX OF DIFFERENT METHODS FOR HYDICE EXPERIMENT

(A) 3FA-ELM(OA)

3FA-ELM(OA)		Ground Truth Road	Grass	Trail	Tree	Shadow	Roof	No. of classified pixels	Users accuracy(%)
classified	Road	911	0	8	1	0	8	928	98.17%
	Grass	0	925	0	0	18	0	943	98.09%
	Trail	1	0	577	0	0	0	578	99.83%
	Tree	0	2	0	608	0	1	611	99.51%
	Shadow	0	10	0	0	655	0	665	98.50%
	Roof	58	0	2	19	0	1037	1116	92.92%
No. of ground truth pixels		970	937	587	628	673	1046	OA = 97.36%	Kappa = 0.9679
Producers accuracy(%)		93.92%	98.72%	98.30%	96.82%	97.33%	99.14%		

(B) FA-ELM(OA)

FA-ELM(OA)		Ground Truth Road	Grass	Trail	Tree	Shadow	Roof	No. of classified pixels	Users accuracy(%)
classified	Road	908	0	6	0	0	14	928	97.84%
	Grass	0	928	0	1	14	1	944	98.31%
	Trail	25	0	553	0	0	0	578	95.67%
	Tree	0	0	0	554	1	57	612	90.52%
	Shadow	0	24	0	0	641	0	665	96.39%
	Roof	41	0	0	22	0	1053	1116	94.35%
No. of ground truth pixels		974	952	559	577	656	1125	OA = 95.50%	Kappa = 0.9483
Producers accuracy(%)		93.22%	97.48%	98.93%	96.01%	97.71%	93.60%		

(C) PSO-ELM(OA)

PSO-ELM(OA)		Ground Truth Road	Grass	Trail	Tree	Shadow	Roof	No. of classified pixels	Users accuracy(%)
classified	Road	908	0	16	0	0	4	928	97.84%
	Grass	0	916	0	1	25	2	944	97.03%
	Trail	8	0	569	0	0	0	577	98.61%
	Tree	0	2	0	541	1	67	611	88.54%
	Shadow	0	11	0	0	654	0	665	98.35%
	Roof	49	4	4	109	0	950	1116	85.13%
No. of ground truth pixels		965	933	589	651	680	1023	OA = 93.74%	Kappa = 0.9241
Producers accuracy(%)		94.09%	98.18%	96.60%	83.10%	96.18%	92.86%		

(D) FA-ELM(JM)

FA-ELM(JM)		Ground Truth Road	Grass	Trail	Tree	Shadow	Roof	No. of classified pixels	Users accuracy(%)
classified	Road	904	0	7	0	0	17	928	97.41%
	Grass	0	918	0	3	10	13	944	97.25%
	Trail	16	0	561	0	0	0	577	97.23%
	Tree	0	14	0	526	0	71	611	86.09%
	Shadow	1	45	4	0	615	0	665	92.48%
	Roof	22	0	1	160	0	933	1116	83.60%
No. of ground truth pixels		943	977	573	689	625	1034	OA = 92.07%	Kappa = 0.9038
Producers accuracy(%)		95.86%	93.96%	97.91%	76.34%	98.40%	90.23%		

(E) ELM USING ALLBANDS

ELM		Ground Truth Road	Grass	Trail	Tree	Shadow	Roof	No. of classified pixels	Users accuracy(%)
classified	Road	914	0	9	0	0	5	928	98.49%
	Grass	0	751	0	2	191	0	944	79.56%
	Trail	10	0	567	0	0	0	577	98.27%
	Tree	3	5	0	601	1	1	611	98.36%
	Shadow	0	14	1	0	651	0	666	97.75%
	Roof	95	0	1	175	0	845	1116	75.72%
No. of ground truth pixels		1022	770	578	778	843	851	OA = 89.41%	Kappa = 0.8723
Producers accuracy(%)		89.43%	97.53%	98.10%	77.25%	77.22%	99.29		

(F) FA-ELM USING ALLBANDS

FA-ELM		Ground Truth							
		Road	Grass	Trail	Tree	Shadow	Roof	No. of classified pixels	Users accuracy(%)
classified	Road	880	0	12	0	0	8	900	97.78%
	Grass	0	848	0	1	67	0	916	92.58%
	Trail	22	0	538	0	0	0	560	96.07%
	Tree	3	0	0	586	1	3	593	98.82%
	Shadow	0	15	0	0	630	0	645	97.67%
	Roof	53	0	0	187	0	842	1082	77.82%
No. of ground truth pixels		958	863	550	774	698	853	OA = 92.08%	Kappa = 0.9043
Producers accuracy(%)		91.86%	98.26%	97.82%	75.71%	90.26%	98.71%		

and the number of hidden neurons L; empirical values are used as initials [26].

3) *FA-based band selection*: the overall classification accuracy produced by ELM is used as the objective function; and the selected bands are updated with the FA algorithm.
4) *FA-based ELM parameter optimization*: the ELM parameter combination is updated with the selected bands in Step 3.
5) FA-based hidden neurons optimization: the number of hidden neurons is updated with the selected bands in Step 3 and the optimized parameters in Step 4.
6) Repeat Steps 3–5 until reaching the maximum number of iterations.

B. Basic FA Optimization Process

The proposed system uses the FA to optimize three different parameters. The band selection optimization can be described as follows:

1) *Initialization*: Maximum iterations $t_s = 100$, step size $\alpha = 0.5$, light absorbance $\gamma = 1$, the numbers of fireflies m, the number of selected bands p, overall accuracy as I_0.
2) Compute the brightness (i.e., attractiveness) with (1). The objective function I_0 is evaluated by using the b bands whose indices are included in a firefly; in total, p fireflies (i.e., m sets of selected bands) are evaluated.
3) Estimate the movement state using (3) and (4).
4) Update the objective function according to the classification accuracy with updated fireflies (i.e., updated selected band indices).
5) Repeat Steps 2–4 until reaching the maximum number of iterations. The final selected bands are the p bands whose indices are included in a firefly that generates the largest I_0.

C. Proposed FA-Inspired Classification Framework

In this paper, a 3FA optimization system is designed to optimize all the parameters for classification. In this framework, FA is used three times for band selection, parameters optimization, and hidden neurons optimization in each iteration. According to the basic principle of FA, OA is chosen as the brightness values, i.e., objective function in the 3FA system. The values to be optimized can be regarded as the location of the fireflies, and the values yielding the maximum OA are the optimal position.

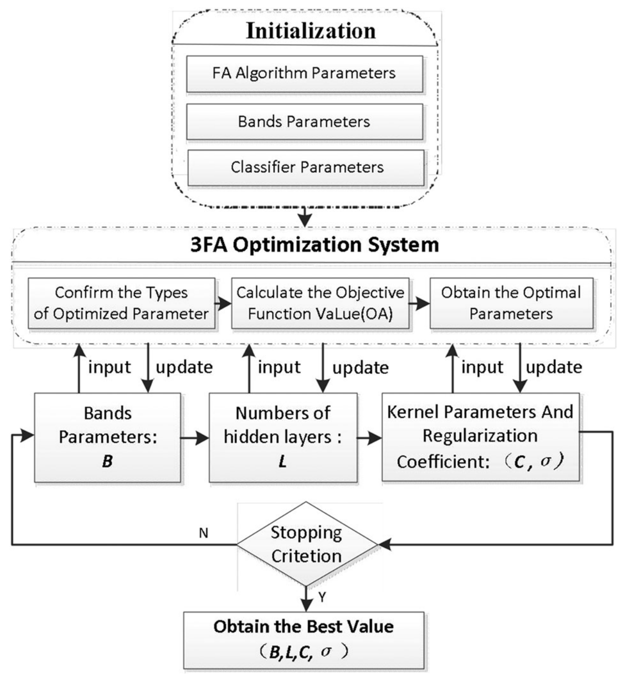


Fig. 1. Proposed FA-inspired classification framework.

After multiple iterations, the algorithm will approach the brightest fireflies location, which are the best values. The proposed FA-inspired classification framework is shown in Fig. 1. The proposed 3FA framework is described as follows.

1) Initial step:
   a) Use the FA algorithm for band pre-selection [15], where the JM distance is the objective due to its reliable performance [15], [30].
   b) Based on the selected bands in 1a), optimize the number of hidden neurons, the parameters in ELM are set to be the empirical values in [26].
2) 3FA system:
   a) *First FA*: select bands with the ELM accuracy being the objective and parameters from 1b).
   b) *Second FA*: optimize the number of hidden neurons in ELM based on the selected bands in 2a).
   c) *Third FA*: optimize the parameters in ELM based on the selected bands in 2a).
   d) Go back to 2a).

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

D. Algorithms for Comparison

1) Objective Function Based on JM Metric: The JM distance between two classes ω_i and ω_j is defined as

$$J_{i,j} = \int \left\{ \sqrt{p(\boldsymbol{r}|\omega_i)} - \sqrt{p(\boldsymbol{r}|\omega_j)} \right\}^2 d\mathrm{r} \tag{23}$$

where $\boldsymbol{r}$ is the feature vector of dimension k (a k-band subset of the spectrum), and $p(\boldsymbol{r}|\omega_i)$ and $p(\boldsymbol{r}|\omega_j)$ are two class-conditional probability distributions of r. When $p(\boldsymbol{r}|\omega_i)$ and $p(\boldsymbol{r}|\omega_j)$ are Gaussian distributions, the JM distance can be simplified as

$$J_{i,j} = 2(1 - e^{-B_{i,j}}) \tag{24}$$

where

$$B_{i,j} = \frac{1}{8}(\mu_i - \mu_j)^T \left(\frac{\Sigma_i + \Sigma_j}{2} \right)^{-1} (\mu_i - \mu_j) + \frac{1}{2} \ln \left(\frac{|(\Sigma_i + \Sigma_j)/2|}{|\Sigma_i|^{\frac{1}{2}} |\Sigma_j|^{\frac{1}{2}}} \right) \tag{25}$$

is the Bhattacharyya distance between ω_i and ω_j. Here, μ_i and μ_j are class means, and Σ_i and Σ_j are class covariance matrices. Class samples are required such that class means and covariance matrices can be reliably estimated.

2) PSO for Band Selection: PSO is a kind of intelligent optimization algorithms, which is based on the behavior of birds feeding. Through iterative search, the PSO uses a fitness value to evaluate the quality of an optimal solution. With its easy implementation and fast convergence, PSO algorithm is also widely used in solving practical problems.

IV. Experiments

A. Parameter Tuning

For optimization, the parameters of each algorithm are set as: FA (absorbance $\gamma = 1$, maximum attractiveness $\beta_0 = 1$, step size $\alpha = 0.2$); The range of the regularization coefficient in ELM is $[0, 10^3]$, Gaussian kernel ($\gamma = {}^{-1}\!/_{2\sigma^2}$) is $(0, 1]$. Meanwhile, the fixed parameters of FA for the classifier parameters optimization are set as previously mentioned. We choose 5, 10, 15, 20 as the number of selected bands and 2%, 5%, 8%, 10% as the number of training samples. The specific parameters of each algorithm are shown in Table I.

In the experiment, six methods are used for comparison. 3FA-ELM (OA) is our proposed method to jointly select bands and optimize ELM parameters. FA-ELM (OA) and PSO-ELM (OA) use FA or PSO for parameter optimization only once, and bands are selected with the OA criterion separately. FA-ELM(JM) means the bands are selected by the JM distance and then the parameters of ELM are optimized by FA. The ELM method uses all bands for ELM classification, and the FA-ELM uses the FA for parameters optimization with all bands for classification. The details can be found in Table II.

Fig. 2. HYDICE DC MALL.

B. HYDICE DC MALL Experiment

The HYDICE subimage scene in Fig. 2 with 304 $\times$ 301 pixels over the Washington DC Mall area was used in the first experiment. After bad band removal, 191 bands were left in the experiment. There are six classes: roof, tree, grass, water, road, and trail. These six class centers were used for band selection. The available training and test samples are shown in Table III.

The classification results are shown in Table IV. The comparison results from four methods with 5, 10, 15, 20 bands and 2%, 5%, 8%, 10% training samples are listed. We can see that the proposed 3FA-ELM (OA) provided the best results, which was better than three other methods using the same number of bands and the same training samples for classification. Take 2% training data and ten bands as example, the classification accuracy of 3FA-ELM (OA) is much higher than that of FA-ELM (OA), which means after a constant number of iterations, classifier can find better parameters to improve the classification performance. Classification accuracy increased from 95.5% to 97.36%. On the other hand, the classification accuracy of FA-ELM(OA) are higher than that of PSO-ELM(OA) in most cases, which means FA has more excellent search performance. Compared to the traditional way using JM distance as objective function, OA can get better classification accuracy results. In addition, the performance of using all bands with ELM classification is very poor.

For each algorithm, the selected bands numbers were from 5 to 20 with ten individual runs. With 2% samples, ten bands, for example, the best overall accuracy were chosen as the final results and the corresponding bands were illustrated in Fig. 3. Table V tabulates the confusion matrices for the best overall accuracy of each method, and the corresponding land cover classification maps are shown in Fig. 4.

To further evaluate the performance, the running time of each algorithm with ten bands are listed in Table VI. The number of hidden layer nodes L is increased when the number of training samples becomes larger, which expands the size of the hidden layer matrix and affects the running time. The running time of

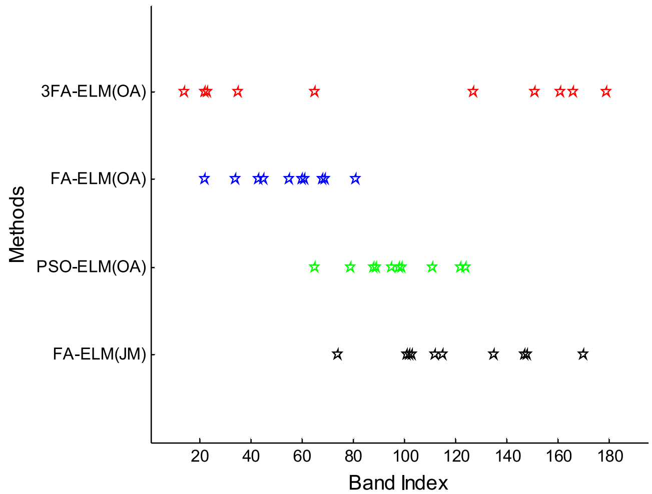


Fig. 3. Band selected by different methods for HYDICE data (2% training samples, ten bands).

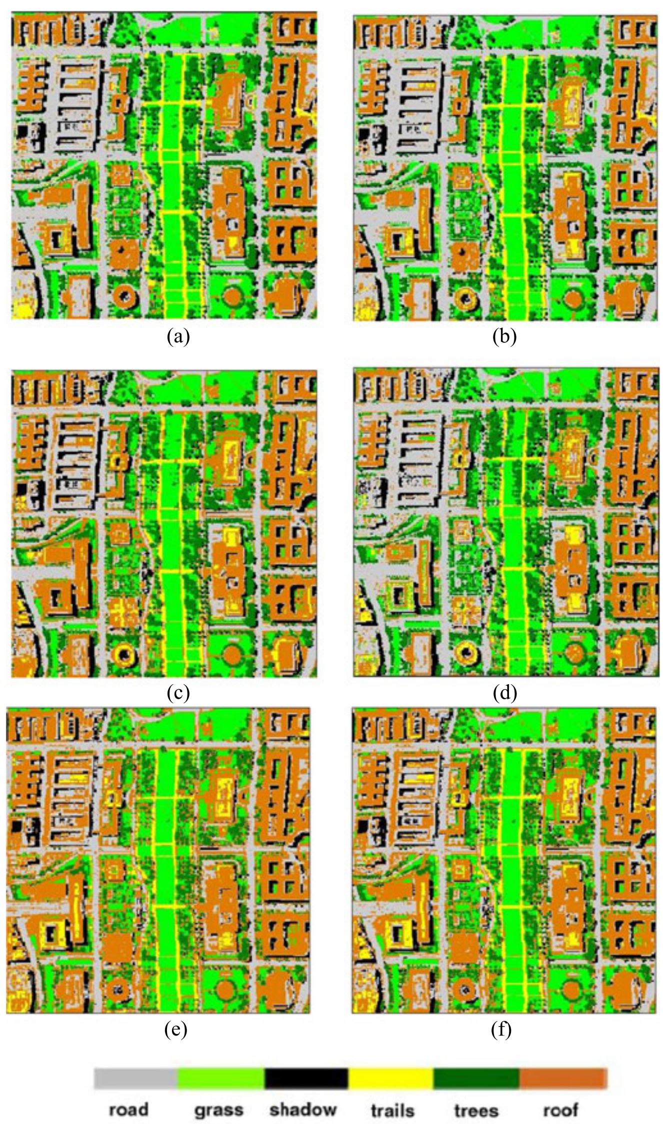


Fig. 4. Classification map of different algorithms HYDICE data with 2% training samples and ten bands. (a) 3FA-ELM(OA) (97.36%), (b) FA-ELM(OA)(95.5%), (c) PSO-ELM(OA)(93.74%), (d) FA-ELM(JM) (92.07%), (e) ELM (89.41%), and (f) FA-ELM (92.08%).

Fig. 5. HyMap Purdue Campus.

TABLE VI
RUNNING TIME OF DIFFERENT METHODS FOR HYDICE DATA WITH TEN BANDS (SECONDS)

Samples	3FA-ELM(OA)	FA-ELM(OA)	PSO-ELM(OA)	FA-ELM(JM)
2%	554.36	10.40	19.82	1.51
5%	653.04	11.67	24.32	1.52
8%	722.13	12.76	27.85	1.51
10%	767.15	15.15	28.86	1.50

TABLE VII
GROUND TRUTH FOR HYMAP DATA

Class	Name	samples
1	Road	1287
2	Grass	1114
3	Shadow	219
4	Soil	379
5	Tree	1351
6	Roof	1285
	Total	5635

3FA-ELM (OA) under different proportion of samples after 30 times iterations is improved with longer computing time. The running time of FA-ELM (OA) is less than that of PSO-ELM (OA), which means the FA algorithm has faster search ability. Because FA-ELM (JM) does not involve a classification process, its running time with different proportional samples are similar.

C. Hyperspectral Mapper (HyMap) Purdue Campus

The second dataset is a flightline over the Purdue University West Lafayette campus. The hyperspectral data were collected on September 30, 1999 with the airborne HyMap system, providing image data in 128 spectral bands in the visible and infrared regions (0.4–2.4 μm). In this experiment, except the atmospheric water absorption bands, the 126 bands are used. The system was flown at an altitude such that the pixel size is about 3.5 m. An image of the scene is shown in Fig. 5. The information of training and testing samples are listed in Table VII.

In this experiment, the classification results are shown in Table VIII. The results list the highest accuracy, the lowest accuracy and the average accuracy with ten runs for different

TABLE VIII
CLASSIFICATION PERFORMANCE OF DIFFERENT METHODS FOR HYMAP DATA WITH DIFFERENT BANDS AND TRAINING SAMPLES

(A) 2% AS TRAINING SAMPLES			
Samples = 2%	Overall classification accuracy(OA)		
Bands = 5	Minimum	Maximum	Average
3FA-ELM(OA)	**0.9441**	**0.9463**	**0.9455**
FA-ELM(OA)	0.9407	0.9445	0.9421
PSO-ELM(OA)	0.8798	0.8912	0.8843
FA-ELM(JM)	0.6224	0.7315	0.6729
Bands = 10	Minimum	Maximum	Average
3FA-ELM(OA)	**0.9377**	**0.9558**	**0.9503**
FA-ELM(OA)	0.9133	0.9477	0.9326
PSO-ELM(OA)	0.8809	0.8877	0.8851
FA-ELM(JM)	0.7544	0.8187	0.7833
Bands = 15	Minimum	Maximum	Average
3FA-ELM(OA)	**0.9464**	**0.9502**	**0.9475**
FA-ELM(OA)	0.9379	0.9480	0.9420
PSO-ELM(OA)	0.8319	0.8823	0.8515
FA-ELM(JM)	0.8098	0.8402	0.8216
Bands = 20	Minimum	Maximum	Average
3FA-ELM(OA)	0.9433	0.9548	0.9498
FA-ELM(OA)	0.9177	0.9455	0.9301
PSO-ELM(OA)	0.8386	0.8803	0.8675
FA-ELM(JM)	0.7698	0.8266	0.8024
Bands = Allbands	Minimum	Maximum	Average
ELM	0.7487	0.8002	0.7865
FA-ELM	0.7905	0.8413	0.8233
(B) 5% AS TRAINING SAMPLES			
Samples = 5%	Overall classification accuracy(OA)		
Bands = 5	Minimum	Maximum	Average
3FA-ELM(OA)	**0.9689**	**0.9733**	**0.9715**
FA-ELM(OA)	0.9410	0.9456	0.9432
PSO-ELM(OA)	0.9003	0.9081	0.9040
FA-ELM(JM)	0.7506	0.8227	0.7927
Bands = 10	Minimum	Maximum	Average
3FA-ELM(OA)	**0.9734**	**0.9793**	**0.9768**
FA-ELM(OA)	0.9591	0.9665	0.9607
PSO-ELM(OA)	0.9462	0.9492	0.9478
FA-ELM(JM)	0.7885	0.8406	0.8138
Bands = 15	Minimum	Maximum	Average
3FA-ELM(OA)	**0.9768**	**0.9814**	**0.9790**
FA-ELM(OA)	0.9670	0.9698	0.9688
PSO-ELM(OA)	0.9106	0.9371	0.9203
FA-ELM(JM)	0.8197	0.8609	0.8484
Bands = 20	Minimum	Maximum	Average
3FA-ELM(OA)	0.9776	0.9793	0.9787
FA-ELM(OA)	0.9695	0.9752	0.9714
PSO-ELM(OA)	0.9256	0.9399	0.9329
FA-ELM(JM)	0.8229	0.8803	0.8576
Bands = Allbands	Minimum	Maximum	Average
ELM	0.7678	0.8280	0.8008
FA-ELM	0.7923	0.8598	0.8254
(C) 8% AS TRAINING SAMPLES			
Samples = 8%	Overall classification accuracy(OA)		
Bands = 5	Minimum	Maximum	Average
3FA-ELM(OA)	**0.9700**	**0.9741**	**0.9727**
FA-ELM(OA)	0.9558	0.9662	0.9608
PSO-ELM(OA)	0.9154	0.9369	0.9203
FA-ELM(JM)	0.8088	0.8475	0.8235
Bands = 10	Minimum	Maximum	Average
3FA-ELM(OA)	**0.9792**	**0.9809**	**0.9801**
FA-ELM(OA)	0.9531	0.9651	0.9604
PSO-ELM(OA)	0.9418	0.9490	0.9450
FA-ELM(JM)	0.7993	0.8305	0.8194
Bands = 15	Minimum	Maximum	Average
3FA-ELM(OA)	**0.9811**	**0.9836**	**0.9824**
FA-ELM(OA)	0.9708	0.9784	0.9752
PSO-ELM(OA)	0.9359	0.9406	0.9385
Bands = 20	Minimum	Maximum	Average
3FA-ELM(OA)	**0.9827**	**0.9853**	**0.9840**
FA-ELM(OA)	0.9735	0.9777	0.9758
PSO-ELM(OA)	0.9524	0.9556	0.9540
FA-ELM(JM)	0.9131	0.9204	0.9186
Bands = Allbands	Minimum	Maximum	Average
ELM	0.8011	0.8814	0.8385
FA-ELM	0.8102	0.8683	0.8379
(D) 10% AS TRAINING SAMPLES			
Samples = 10%	Overall classification accuracy(OA)		
Bands = 5	Minimum	Maximum	Average
3FA-ELM(OA)	**0.9721**	**0.9766**	**0.9748**
FA-ELM(OA)	0.9655	0.9691	0.9672
PSO-ELM(OA)	0.9541	0.9650	0.9590
FA-ELM(JM)	0.7517	0.8012	0.7812
Bands = 10	Minimum	Maximum	Average
3FA-ELM(OA)	**0.9741**	**0.9787**	**0.9767**
FA-ELM(OA)	0.9722	0.9780	0.9744
PSO-ELM(OA)	0.9406	0.9542	0.9489
FA-ELM(JM)	0.8703	0.8923	0.8827
Bands = 15	Minimum	Maximum	Average
3FA-ELM(OA)	**0.9777**	**0.9801**	**0.9792**
FA-ELM(OA)	0.9754	0.9792	0.9776
PSO-ELM(OA)	0.9479	0.9588	0.9522
FA-ELM(JM)	0.9102	0.9203	0.9152
Bands = 20	Minimum	Maximum	Average
3FA-ELM(OA)	**0.9789**	**0.9825**	**0.9811**
FA-ELM(OA)	0.9741	0.9898	0.9776
PSO-ELM(OA)	0.9639	0.9690	0.9669
FA-ELM(JM)	0.9094	0.9214	0.9173
Bands = Allbands	Minimum	Maximum	Average
ELM	0.8230	0.8907	0.8622
FA-ELM	0.8449	0.9106	0.8743
FA-ELM(JM)	0.8773	0.9019	0.8994

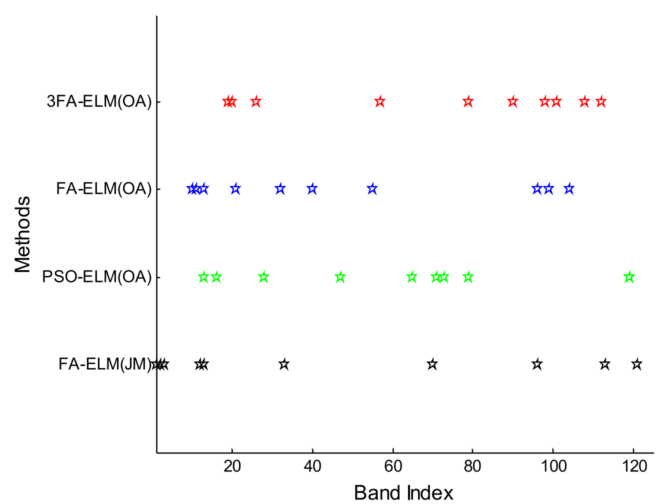


Fig. 6. Band selected by different methods for HyMap (2% training samples, ten bands).

algorithms. We can see that 3FA-ELM (OA) provided the best results among all methods, which can reach 98.53% with 8% training samples and 20 selected bands. With the increasing of band number and the sample proportion, classification accuracy is improved especially when the sample proportion from 2% to 5%, the overall accuracy increased from 95.03% to 97.68%. On the performance of the search algorithm, FA method has more obvious advantages in the classification accuracy, which

TABLE IX
CONFUSION MATRIX OF DIFFERENT METHODS FOR HYMAP EXPERIMENT

(A) 3FA-ELM(OA)

		Ground Truth							
3FA-ELM(OA)		Road	Grass	Trail	Tree	Shadow	Roof	No. of classified pixels	Users accuracy(%)
Classified	Road	1197	3	0	1	0	61	1262	94.85%
	Grass	0	1044	0	41	7	0	1092	95.60%
	Trail	0	1	201	0	5	7	214	93.93%
	Tree	0	6	0	365	0	0	371	98.38%
	Shadow	0	10	0	0	1314	0	1324	99.24%
	Roof	67	25	3	4	3	1157	1259	91.90%
No. of ground truth pixels		1264	1089	204	411	1329	1225	OA = 95.58%	Kappa = 0.9443
Producers accuracy(%)	94.70%	95.87%	98.53%	88.81%	98.87%	94.45%			

(B) FA-ELM(OA)

		Ground Truth							
FA-ELM(OA)		Road	Grass	Trail	Tree	Shadow	Roof	No. of classified pixels	Users accuracy(%)
Classified	Road	1166	3	0	3	0	89	1261	92.47%
	Grass	1	1035	0	25	20	11	1092	94.78%
	Trail	0	0	202	0	5	7	214	94.39%
	Tree	1	19	0	346	0	6	372	93.01%
	Shadow	3	8	0	1	1310	2	1324	98.94%
	Roof	62	15	3	5	0	1174	1259	93.25%
No. of ground truth pixels		1233	1080	205	380	1335	1289	OA = 94.77%	Kappa = 0.9340
Producers accuracy(%)	94.57%	95.83%	98.54%	91.05%	98.13%	91.08%			

(C) PSO-ELM(OA)

		Ground Truth							
PSO-ELM(OA)		Road	Grass	Trail	Tree	Shadow	Roof	No. of classified pixels	Users accuracy(%)
Classified	Road	1206	19	0	4	0	33	1262	95.56%
	Grass	0	1081	0	0	11	0	1092	98.99%
	Trail	0	2	204	0	5	4	215	94.88%
	Tree	30	142	0	199	0	0	371	53.64%
	Shadow	0	4	0	0	1320	0	1324	99.70%
	Roof	325	33	4	0	4	893	1259	70.93%
No. of ground truth pixels		1561	1281	208	203	1340	930	OA = 88.77%	Kappa = 0.8577
Producers accuracy(%)	77.26%	84.39%	98.08%	98.03%	98.51%	96.02%			

(D) FA-ELM(JM)

		Ground Truth							
FA-ELM(JM)		Road	Grass	Trail	Tree	Shadow	Roof	No. of classified pixels	Users accuracy(%)
Classified	Road	1069	61	0	0	0	131	1261	84.77%
	Grass	0	949	0	0	137	6	1092	86.90%
	Trail	0	1	152	0	11	50	214	71.03%
	Tree	0	329	0	38	0	4	371	10.24%
	Shadow	0	1	0	0	1323	0	1324	99.92%
	Roof	243	22	0	0	5	990	1260	78.57%
No. of ground truth pixels		1312	1363	152	38	1476	1181	OA = 81.87%	Kappa = 0.7683
Producers accuracy (%)	81.48%	69.63%	100.00%	100.00%	89.63%	83.83%			

(E) ELM USING ALLBANDS

		Ground Truth							
ELM		Road	Grass	Trail	Tree	Shadow	Roof	No. of classified pixels	Users accuracy(%)
Classified	Road	914	0	9	0	0	5	928	98.49%
	Grass	0	751	0	2	191	0	944	79.56%
	Trail	10	0	567	0	0	0	577	98.27%
	Tree	3	5	0	601	1	1	611	98.36%
	Shadow	0	14	1	0	651	0	666	97.75%
	Roof	95	0	1	175	0	845	1116	75.72%
No. of ground truth pixels		1022	770	578	778	843	851	OA = 78.65%	Kappa = 0.7341
Producers accuracy(%)	89.43%	97.53%	98.10%	77.25%	77.22%	99.29%			

(F) FA-ELM USING ALLBANDS

FA-ELM		Ground Truth Road	Grass	Trail	Tree	Shadow	Roof	No. of classified pixels	Users accuracy(%)
Classified	Road	872	4	0	22	0	332	1230	70.89%
	Grass	0	1017	0	30	19	6	1072	94.87%
	Trail	0	0	207	0	3	3	213	97.18%
	Tree	2	32	0	337	0	0	371	90.84%
	Shadow	0	14	48	0	1246	13	1321	94.32%
	Roof	309	14	111	0	0	802	1236	64.89%
No. of ground truth pixels		1183	1081	366	389	1268	1156	OA = 82.33%	Kappa = 0.7789
Producers accuracy(%)	73.71%	94.08%	56.56%	86.63%	98.26%	69.38%			

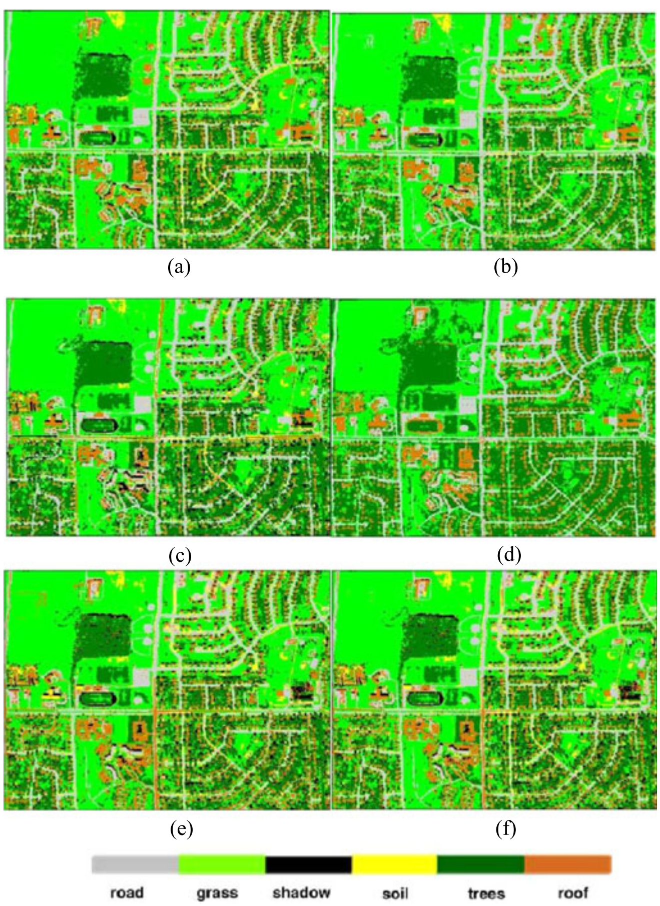


Fig. 7. Classification maps of HyMap data with 2% training samples and ten bands. (a) 3FA-ELM(OA) (95.58%), (b) FA-ELM(OA) (94.77%), (c) PSO-ELM(OA) (88.77%), (d) FA-ELM(JM) (81.87%), (e) ELM (78.65%), and (f) FA-ELM (82.33%).

reflects its better adaptability and generalization performance. Another comparison was between as different band selection measures, OA is also better than JM band selection criterion. In addition, using all bands with ELM classification provides the worst performance.

The selected band numbers were also varies from 5 to 20 with ten runs for FA and PSO method as in the first experiment. With 2% samples, ten bands, for example, the corresponding selected numbers of bands are illustrated in Fig. 6. Confusion matrices for each method and classification maps are shown in Table IX and Fig. 7.

TABLE X
RUNNING TIME OF DIFFERENT METHODS FOR HYMAP DATA WITH TEN BANDS (SECONDS)

Samples	3FA-ELM(OA)	FA-ELM(OA)	PSO-ELM(OA)	FA-ELM(JM)
2%	730.86	12.86	29.34	1.27
5%	739.18	13.24	31.33	1.30
8%	854.01	14.32	32.03	1.27
10%	1104.21	16.60	33.73	1.24

Once again, the running time of different methods for HyMap Purdue Campus data with ten bands are listed in Table X. From the table, we can see that due to the three FAs used in the proposed 3FA-ELM (OA) method, it takes much time, but it provides the highest classification accuracy. For FA-ELM (OA), the running time is much less than those of PSO-ELM (OA), indicating the search speed of FA is much faster than that of PSO. It is noted that, FA-ELM (JM), as one of the filter methods, is much faster, although its accuracy is less than FA-ELM (OA).

V. CONCLUSION

In this paper, a 3FA-inspired framework with selected bands and optimized ELM for hyperspectral image classification is proposed. The FA is introduced for band selection with overall classification accuracy as objective function to extract useful information, and the 3FA-optimized ELM is used for hyperspectral images classification. Two different dataset experiments show the significantly improved classification accuracy with the proposed method. It is demonstrated that the 3FA system is more concise than other methods for band selection and ELM parameter optimization for classification. Compared with other optimization algorithms, FA has better search ability and adaptability. The proposed method offers accurate classification for hyperspectral remote sensing images. In the future, we plan to implement it in parallel to further reduce execution time.

REFERENCES

[1] J. Bioucas-Dias *et al.*, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.
[2] C.-I. Chang, *Hypersepctral Imaging: Techniques for Spectral Detection and Classification*. New York, NY, USA: Kluwer, 2003.
[3] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification: Earth monitoring withstatistical learning methods," *IEEE Signal Proc. Mag.*, vol. 31, no. 1, pp. 45–54, Jan. 2014.

[4] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. S110–S122, 2009.

[5] X. Briottet, Y. Boucher, A. Dimmeler, A. Malaplate, A. Cini, M. Diani, H. Bekman, P. Schwering, T. Skauli, I. Kasen, I. Renhorn, L. Klase ´n, M. Gilmore, and D. Oxford, B, "Military applications of hyperspectral imagery," *Proc. SPIE-Int. Soc. Opt. Eng.*, vol. 6239, Jun. Art. no. 62390B, 2006.

[6] A. Villa, J. A. Benediktsson, J. Chanussot, and C. Jutten, "Hyperspectral image classification with independent component discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 4865–4876, Dec. 2011.

[7] J. S. Tyo, A. Konsolakis, D. I. Diersen, and R. C. Olsen, "Principal components- based display strategy for spectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 3, pp. 708–718, Mar. 2003.

[8] Q. Du and H. Yang, "Similarity-based unsupervised band selection for hyperspectral image analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 564–568, Oct. 2008.

[9] H. Yang, Q. Du, H. Su, and Y. Sheng, "An efficient method for supervised hyperspectral band selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. l, pp. 138–142, Jan. 2011.

[10] X. Bai, Z. Guo, Y. Wang, Z. Zhang, and J. Zhou, "Semisupervised hyperspectral band selection via spectral—spatial hypergraph model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2774–2783, Jun. 2015.

[11] X. Bai, H. Zhang, and J. Zhou, "VHR object detection based on structural feature extraction and query expansion," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6508–6520, Oct. 2014.

[12] Y. Qian, M. Ye, and J. Zhou, "Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2276–2291, Apr. 2013.

[13] S. T. Monteiro and Y. Kosugi, "A particle swarm optimization-based approach for hyperspectral band selection," in *Proc. IEEE Congr. Evol. Comput.*, 2007, pp. 3335–3340.

[14] H. Yang, Q. Du, and G. Chen, "Particle swarm optimization-based hyperspectral dimensionality reduction for urban land cover classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 544–554, Apr. 2012.

[15] H. Su, B. Yong, and Q. Du, "Hyperspectral band selection using improved firefly algorithm," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 68–72, Jan. 2016.

[16] F. Ratle, G. Camps-Valls, and J. Weston. "Semisupervised neural networks for efficient hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, 2271–2282, May 2010.

[17] C. Cortes and V. Vapnik, "Support vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[18] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst., Man Cybern. B*, vol. 42, no. 2, pp. 513–529, Oct. 2011.

[19] G.-B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, pp. 489–501, 2006.

[20] G.-B. Huang, D. Wang, and Y. Lan, "Extreme learning machines: A survey," *Int. J. Mach. Lean. Cybern.*, vol. 2, no. 2, pp. 107–122, 2011.

[21] G.-B. Huang, X. Ding, and H. Zhou, "Optimization method based extreme learning machine for classification," *Neurocomputing*, vol. 74, pp. 155–163, 2010.

[22] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.

[23] M. Pal, A. E. Maxwell, and T. A. Warner, "Kernel-based extreme learning machine for remote-sensing image classification," *Remote Sens. Lett.*, vol. 4, no. 9, pp. 853–862, 2013.

[24] 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.

[25] C. Chen, W. Li, H. Sun, and K. Liu, "Spectral-spatial classification of hyperspectral image based on kernel extreme learning machine," *Remote Sens.*, vol. 6, no. 6, pp. 5795–5814, 2014.

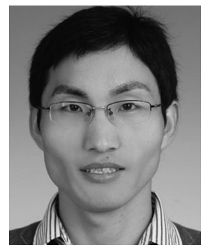
[26] J. Li, Q. Du, and W. Li, "Optimizing extreme learning machine for hyperspectral image classification," *J. Appl. Remote Sens.*, vol. 9, p. 097296, 1–13, 2015.

[27] Y. Bazi, N. Alajlan, F. Melgani, H. Al Hichri, S. Malek, and R. R. Yager, "Differential evolution extreme learning machine for the classification of hyperspectral images," *IEEE Geosc. Remote Sens. Lett.*, vol. 11, no. 6, pp. 1066–1070, Jun. 2014.

[28] X. S. Yang, "Firefly algorithms for multimodal optimization," *Stochastic Algorithms: Foundations and Applications.* New York, NY, USA: Springer, 2009, pp. 169–178.

[29] J. Senthilnath, S. N. Omkar, and V. Mani, "Clustering using firefly algorithm: Performance study," *Swarm Evol. Comput.*, vol. 1, no. 3, pp. 164–171, 2011.

[30] H. Su, Q. Du, G. Chen, and P. Du, "Optimized hyperspectral band selection using particle swarm optimization," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2659–2670, Jun. 2014.



Hongjun Su (S'09–M'11) received the B.S. degree in geography information system from the China University of Mining and Technology, Xuzhou, China, and the Ph.D. degree in cartography and geography information system from the Key Laboratory of Virtual Geographic Environment (Ministry of Education), Nanjing Normal University, Nanjing, China, in 2006 and 2011, respectively. He was a Visiting Student in the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA, from 2009 to 2010. He was also a Postdoctoral Fellow in the Department of Geographical Information Science, Nanjing University, Nanjing, from 2012 to 2014.

He is currently an Associate Professor at the School of Earth Sciences and Engineering, Hohai University, Nanjing. His main research interests include hyperspectral remote sensing dimensionality reduction, classification, and spectral unmxing.

Dr. Su is an Active Reviewer for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, the *International Journal of Remote Sensing*, and the *Optics Express*.

Yue Cai received the B.S. degree in surveying and mapping engineering from the Heilongjiang University of Science and Technology, Harbin, China, and the M.S. degrees in photogrammetry and remote sensing from the School of Earth Sciences and Engineering, Hohai University, Nanjing, China, in 2013 and 2016, respectively.

He is currently at Anhui Survey and Design Institute of Water Conservancy and Hydropower, Bengbu, Anhui, China. His main research interests include hyperspectral remote sensing dimensionality reduction, and classification. He has published four papers about hyperspectral remote sensing classifications.

Qian Du (S'98–M'00–SM'05) received the Ph.D. degree in electrical engineering from the University of Maryland Baltimore County, Baltimore, MD, USA, in 2000.

She is the Bobby Shackouls Professor in the Department of Electrical and Computer Engineering at Mississippi State University, Starkville, MS, USA. Her research interests include hyperspectral remote sensing image analysis, pattern classification, data compression, and neural networks.

Dr. Du serves as a Co-Chair for the Data Fusion Technical Committee of IEEE Geoscience and Remote Sensing Society (GRSS) (2009–2013), and the Chair for Remote Sensing and Mapping Technical Committee of International Association for Pattern Recognition. She also serves as an Editor-in-Chief for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING since 2016, and an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS. She is the Guest Editor for several journal special issues on remote sensing data processing and analysis.

Dr. Du received the 2010 Best Reviewer Award from IEEE GRSS for her service to IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. She is the General Chair of the 4th GRSS Workshop on Hyperspectal Image and Signal Processing: Evolution in Remote Sensing.