

# Kernel Extreme Learning Machine joint Spatial-spectral Information for Hyperspectral Image Classification

Xumin Yu\*, Yan Feng\*, Zhongliang Wang† and Yanlong Gao\*

\*Northwestern Polytechnical University, Xi'an, China

E-mail: sycfy@nwpu.edu.cn

†Tongling University, Tongling 244061, Anhui, China

**Abstract**—Due to its fast learning speed and good generalization ability, extreme learning machine(ELM) has gained significant attention in machine learning and pattern recognition fields. However, when directly applied ELM to hyperspectral image classification (HSI), the accuracy is not high. In this paper, we propose a novel kernel ELM method, which joint spatial-spectral information together to investigate the performance of kernel ELM for HSI classification. In the proposed method, the spatial information are employed by extended morphological profiles. Experiments carried on two widely used hyperspectral datasets demonstrate that the proposed method outperform the SVM and kernel SVM methods. At the same time the cost of computation is much less than traditional methods.

## I. INTRODUCTION

Hyperspectral remote sensing images contain rich spatial and spectral information of objects, and the ability to recognize and classify the ground objects has been greatly improved. Classification is a hot research topic in the processing and application of hyperspectral remote sensing images, and its ultimate goal is to assign unique pixels to each pixel in the images. However, the supervised classification of hyperspectral images(HSI) is a challenging problem due to the high dimensionality, it usually suffers from the Hughes phenomenon[1]. To overcome the high-dimensional problem, many methods were recently introduced for HSI classification and had shown good performance, such as manifold learning, support vector machine(SVM)[2] and composite kernel-based methods [4-7]. In order to get more accurate result, spatial joint spectral information methods are widely used, especially, kernel-based methods. Y. Gu [8] has proposed use nonlinear multiple kernel learning. Pan [9] introduced an ensemble framework which combines spectral and spatial information in different scales. B. Pan [10] successfully use deep learning method on combining spatial-spectral information for HSI classification. Y. Yuan [11] proposed a method mainly focuses on multitask joint sparse representation (MJSR) and a stepwise Markov random field framework to tackle those problems. Compared to those kernel methods, extreme learning machine[12] has received much attention due to its advantages. ELM does not need to tune the hidden layer parameters if the network architecture is determined. The hidden layer parameters in ELM are randomly generated and independent of the training data and application environments.

By minimizing the training error and the norm of output weights simultaneously, ELM tends to have better generalization performance and has a unified analytic solution for binary, multiclass, and regression problems. Though, advantages above mentioned, when directly applied ELM on HSI datasets, the accurate is not high, since we only use spectral information. Some methods use spatial-spectral information based ELM for HSI classification methods are proposed. Zhou [13] has proposed two spatial-spectral composite kernel ELM for HSI classification. Chen [14] use Gabor filtering to extract spatial information, then joint spectral information as input of ELM. In [15], extended morphological profiles were employed for spatial information for ELM-based classification of HSI. M. Jiang [16] exploit a multiscale spatial weighted-mean-filtering-based approach to extract multiple spatial information.

Although those spatial-spectral with ELM-based methods have performed well, their performance can be further improved by using more efficient spatial information or more adopted kernels resembled.

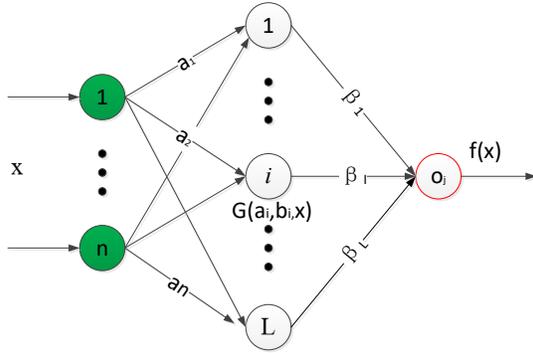
Considering the importance of spatial information, we use extended morphological profiles [17]. Two widely used morphological operators are opening and closing, which are based on the fundamental operations of erosion and dilation. From these basic operations, the so-called morphological profile (MP) can be constructed. An MP contains information of the structures of the image at different resolution sizes. In remote sensing, MPs are usually computed from hyperspectral data using principal component analysis (PCA) [18]. These attributes profiles extract information related to size, geometry, and homogeneity of regions.

In this paper, based on ELM, extended morphological profiles are employed to extract spatial and spectral information. Then, we integrate spatial-spectral information for HSI classification. The rest of this paper is organized as follows. The ELM and kernel ELM are introduced in section II. The proposed spatial-spectral kernel ELM is detailed described in section III. Experimental results and analysis are shown in section IV. Conclusions of this paper are shown in section V.

## II. EXTREME LEARNING MACHINE

### A. Single Layer Feedback Neural network

Extreme learning machine is an algorithm for solving the hidden neural network proposed by G.B. Huang. The essence of ELM is a single hidden layer feedforward neural network (SLFN) that does not need to tune parameters. Figure .1 shows a single hidden feedback layer network.



N input neurons L Hidden neurons Output neuron

Fig. 1 Single layer feedback network

Given N training samples  $\{(x_i, y_i)\}_{i=1}^N$ , where  $x_i = [x_{i1}, \dots, x_{id}]^T \in R^n$  and  $y_i = [y_{i1}, \dots, y_{im}]^T \in R^m$ , the output function of a standard single layer feedback network with L hidden nodes can be presented as

$$f_L = \sum_{i=1}^L \beta_i G_i(x) = \sum_{i=1}^L \beta_i G(a_i, b_i, x) \quad (1)$$

Where  $a_i = [a_{i1}, \dots, a_{id}]^T$  is the weight vector connecting the input nodes to the i th hidden node,  $\beta_i = [\beta_1, \dots, \beta_{im}]^T$  is the weight vector connecting i th hidden node to the output node,  $b_i$  is the threshold of the i th hidden node,  $G_i(x) = G(a_i, b_i, x)$  is the hidden layer output activation of node i.

SLFN with L nodes can approximate N distinct samples, that is to say  $\sum_{j=1}^N \|f_L(x_j) - y_j\| = 0$ ,  $\exists \beta_i, a_i$  and  $b_i$  made the equation set up, such that

$$\sum_{i=1}^L \beta_i G(a_i, b_i, x) = y_j \quad (2)$$

The equation(2) can be Represented by matrix

$$H\beta = Y \quad (3)$$

$\beta_i = [\beta_1, \dots, \beta_L]^T \in R^{L \times m}$  and  $Y = [y_1, \dots, y_n]^T \in R^{N \times m}$

The hidden layer output matrix H

$$H = \begin{pmatrix} G(a_1, b_1, x_1) & \dots & G(a_L, b_L, x_L) \\ \vdots & \ddots & \vdots \\ G(a_1, b_1, x_N) & \dots & G(a_L, b_L, x_N) \end{pmatrix} \quad (4)$$

The matrix **H** is a function of hidden layer parameters  $a_i$  and  $b_i$  are unknown. :

$$\arg \min_{a_i, b_i, \beta} \|H(a_1, \dots, a_L; b_1, \dots, b_L) \beta - Y\|^2 \quad (5)$$

In traditional methods, equation (5) usually solved by gradient based iterative algorithm. However, in the process of iterating, parameters needed to be tuned, thus may cause gradient diffusion, local minima and overfitting.

### B. ELM and kernel ELM

Different from traditional SLFN, ELM is a generalized SLFN. The parameters  $a_i$  and  $b_i$  need not to be tuned in hidden layer, they are randomly generated in the beginning of the ELM algorithm learning process. Once the  $a_i$  and  $b_i$  are determined, the hidden layer output matrix H will be determined through  $a_i$  and  $b_i$ . Training a single layer feedback neuron network changes to solving a linear system, the output weight will be obtained by this linear analytic solutions.

ELM ultimate goal is not only get the smallest training error but also get the smallest norm of output weight  $\beta$ .

$$\min \|H\beta - Y\|^2 \text{ and } \min \|\beta\|^2 \quad (6)$$

Based on optimization theory, equation (6) can be formulated as follow:

$$\min_{\beta} \frac{1}{2} \|\beta\|_2^2 + C \frac{1}{2} \xi_i^2 \quad (7)$$

$$s.t. h(x)\beta = y_i^T - \xi_i^T, i = 1, \dots, N$$

Where  $h(x) = [G(a_1, b_1, x), \dots, G(a_L, b_L, x)]$ ,  $\xi_i$  is the training error, C is regularization parameter.

According to Lagrange multiplier theory and KKT optimization conditions, the solution (7) can be expressed as

$$\beta = H^T \left( \frac{I}{C} + HH^T \right)^{-1} Y \quad (8)$$

After we get  $\beta$ , the output of ELM is obtained by equation(9).

$$f(x) = h(x)\beta = h(x)H^T \left( \frac{I}{C} + HH^T \right)^{-1} Y \quad (9)$$

To be similar as SVM kernel methods, kernel trick can be used in equation (9), kernel function  $h(x_i) \cdot h(x_j) = K(x_i, x_j)$  can replace the inner product  $h(x)H^T$  and  $HH^T$  in equation (9)

The kernel trick version of ELM is kernel ELM, which will be used in follow parts of this paper.

$$f(x) = K_x \left( \frac{I}{C} + K \right)^{-1} Y \quad (10)$$

### III. PROPOSED SPATIAL-SPATIAL KERNEL ELM BASED METHOD

In this section, we present a method joint spatial-spectral information together to investigate the performance of kernel ELM for HSI classification. Figure 2 shows the process of the method we proposed .

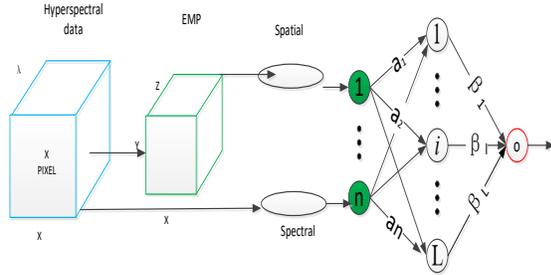


Fig. 2 Process of the spatial-spectral kernel ELM based HSI classification

Mathematical morphology is widely used to extract spatial information in images. Opening and closing are the most common morphological operators used to obtain morphological profiles. The precise details of the effect of the operator on the image are determined by a structuring element. Moreover, it is possible to define morphological operators that satisfy those property: if a spatial structure of the image contains the structuring element, then it is fully preserved; otherwise it is fully removed. In order to exploit all the elements with different sizes and shapes, it is necessary to use a range of structuring elements (SEs) with increasing sizes.

A morphological profile is composed of the opening profiles (OPs) and the closing profiles (CPs) [16]. The OP at the pixel  $x$  is defined as follows:

$$OP_i(x) = \gamma^{(i)}(x), i \in [0, n] \quad (11)$$

Where  $\gamma^{(i)}$  is the opening with an SE of size  $i$  and  $n$  is the total number of openings. Besides, the CP at the pixel  $x$  is defined just as follows:

$$CP_i(x) = \phi^{(i)}(x), i \in [0, n] \quad (12)$$

Where  $\phi^{(i)}$  is the closing with an SE of size  $i$ . So for a single pixel in an image, the MP is a vector with a length of  $2n+1$ , which can be defined as follows:

$$MP(x) = \{CP_n(x), \dots, I(x), \dots, OP_n(x)\} \quad (13)$$

Where  $I(x)$  is the original pixel. As indicated in (12), the definition of MP first came from a single grayscale image.

For hyperspectral images, if we build MPs for all bands on datasets, there will be a huge amount of data , and much of them are redundancy. To solve this problem, an effective approach is choosing some PCs before we build MPs. The EMP of a pixel  $x$  is a vector of  $m(2n+1)$  dimension ( $m$  is the number of PCs):

$$EMP(x) = \{MP_{PC_1}(x), \dots, MP_{PC_m}(x)\} \quad (14)$$

The EMPs of hyperspectral images contain both spatial and spectral information, and can be used as feature vectors for

classification. These attributes profiles extract information related to size, geometry, and homogeneity of regions.

By exploiting the information in spatial domain and spectral domain, the kernel method is usually used to perform the spatial-spectral classification .In the kernel method, the original spectral features are used to compute spatial and spectral kernels, which are combined to form kernels.

Given a pixel  $x_i$ , a pixel is a sample consisting of the spectral characteristics across a continuous range of spectral bands, we denote its spectral and spatial features as  $x_i^w$  and  $x_i^s$  respectively. The spectral feature vector  $x_i^w$  is the original  $x_i$ , which consists of spectral reflection values across all bands. The spatial feature vector  $x_i^s$  is extracted by EMP of pixel  $x_i$ .

Once the spatial and spectral features  $x_i^s$  and  $x_i^w$  are constructed, we can compute the spatial kernel  $K_s$  and spectral kernel as follows:

$$k_s(x_i, x_j) = \exp\left(-\frac{\|x_i^s - x_j^s\|^2}{2\sigma_s^2}\right) \quad (15)$$

$$k_w(x_i, x_j) = \exp\left(-\frac{\|x_i^w - x_j^w\|^2}{2\sigma_w^2}\right) \quad (16)$$

The RBF kernel is used,  $\sigma_s$  and  $\sigma_w$  are the width of the RBF kernels. Kernel ELM is represented as:

$$K = \mu K_s + (1 - \mu) K_w \quad (17)$$

When the spatial-spectral kernel in equation (14) is computed, the kernel ELM model resolves equation (18).

$$\alpha = \left(\frac{I}{C} + K\right)^{-1} Y \quad (18)$$

And we get outputs:

$$f(x) = K_x \alpha \quad (19)$$

Each test sample  $x_i$  will be assigned to the highest value in  $f_i(x_i) = [f_1(x_i), \dots, f_m(x_i)]$  according to the index during the prediction phase.

#### IV. EXPERIMENTAL RESULT

##### A. Hyperspectral Image Datasets

In this section, we evaluate the generalization of the proposed approach by two widely used datasets, Indian pines and University of Pavia. The two datasets are public available hyperspectral datasets.

1) *Indian Pines*: The dataset was acquired by the AVIRIS sensor in 1992. The image scene contains  $145 \times 145$  pixels and 220 spectral bands, where 20 channels were discarded because of the atmospheric affection. The spatial resolution of the data is 20 m per pixel. There are 16 classes and totally 10 249 labeled samples in the dataset. The ground-truth map are

shown in Fig. 3.

2) *University of Pavia*: This dataset was acquired in 2001 by the ROSIS instrument over the city of Pavia, Italy. This image scene corresponds to the University of Pavia and has the size of  $610 \times 340$  pixels and 115 spectral bands. The spatial resolution is 1.3 m per pixel. After removing noisy and water absorption bands, 103 bands are retained. The data contain 9 ground-truth classes and 42776 labeled samples in total. The ground-truth map are shown in Fig. 3.

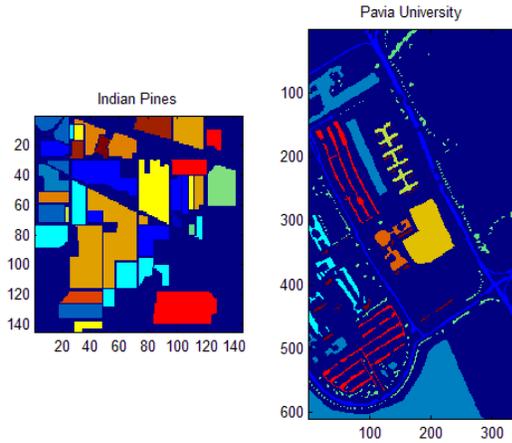


Fig.3 The ground truth of Indian Pines and Pavia University

**B. Parameter setting**

The proposed kernel ELM method is compared with the classical classifiers, SVM with CK (SVM-CK), and general ELM method. The classification performance of different algorithms is assessed on the testing set by the overall accuracy (OA) which is the number of correctly classified testing samples divided by the number of total testing samples, by the average accuracy (AA), which represents the average of the classification accuracies for the individual classes, and by the kappa ( $\kappa$ ) coefficient, which measures the accuracy of classification agreement. The experiments are conducted using MATLAB R2017b, 2.8 GHz dual core and 16GB RAM.

For the Kernel methods, SVM-CK, kernel ELM the combination coefficient  $\mu$  is set to be 0.8. For all kernel-based algorithms, the RBF kernel is used. The parameter  $\sigma$  varies in the range  $\{2^{-4}, 2^{-3}, \dots, 2^4\}$ ,  $C$  ranges from  $10^0$  to  $10^5$ , three fold cross-validation with a grid search method employed to select the optimal parameters. The experiment performs the following cross-validation operations:

- 1) Original training set is randomly divided into three equally sized subsets;
- 2) Two subsets are used to train the model and the remaining subset is used as the validation data for testing the model and outputting OA;
- 3) Step 2) is repeated three times (folds) such that each of

the three subsets are used as the validation data once;

- 4) Three results from the folds are averaged to produce a single OA.

Finally, the parameter pair with the highest OA obtained. By the cross-validation process is set as the optimal parameter pair. The optimal parameter pair corresponds to the highest empirically cross-validation OA and is used for training and testing.

In the general ELM method, the sigmoid function is used, the hidden layer parameters  $(a_i, b_i)_{i=1}^L$  are randomly generated based on uniform distribution from the range  $[-1, 1]$ , and the number of hidden nodes  $L$  is 1000.

**C. Experiment Results and Analysis**

The total number of pixels of Indian Pines available in the reference data is 10,366, but some classes just contain very small training samples. To evaluate the performance of different algorithms in this challenging case, we try to get nearly 30 labeled training samples (small sample class less than 30) per class randomly, since some classes total sample are very small, we chose 10 samples in order keep the training samples balanced. The remaining labeled samples are used for testing.

The classification accuracy of KELM, CK-SVM and EMP kernel ELM measures are provided in Table I.

Table I  
OA, AA and Kappa Obtained by Different Approaches on Indian Pines

Class No.	Train	KELM	CK-SVM	EMP-KELM
1	10	88.89	87.80	<b>91.12</b>
2	30	66.03	<b>91.74</b>	88.58
3	30	59.70	94.32	<b>96.09</b>
4	30	89.71	92.12	<b>94.23</b>
5	30	89.51	94.28	<b>96.94</b>
6	30	91.07	95.60	<b>99.50</b>
7	10	100.00	99.22	<b>96.96</b>
8	30	96.08	99.35	<b>99.93</b>
9	10	<b>100.00</b>	<b>100.00</b>	99.80
10	30	69.30	<b>94.56</b>	92.30
11	30	53.65	94.46	<b>98.14</b>
12	30	73.97	88.35	<b>95.51</b>
13	10	99.45	98.01	<b>99.38</b>
14	30	91.14	<b>99.64</b>	99.80
15	10	74.29	90.20	<b>96.35</b>
16	10	86.23	<b>92.22</b>	87.29
OA		83.25	93.17	<b>94.52</b>
AA		83.63	92.54	<b>93.40</b>
$\kappa$		0.6905	0.9311	<b>0.9426</b>
Time(s)		<b>3.2445</b>	39.87	42.16

From Table I, we can see that, EMP kernel ELM shows slightly better results than SVM, and ELM provides the worst results especially for the classes with limited training samples. This demonstrates the kernel used in kernel ELM or SVM is more powerful than the randomly generated. When additional spatial information is available, the performance of the

spectral-based classifiers is dramatically improved.

The results for the classification of the Indian Pine are shown in figure.4.

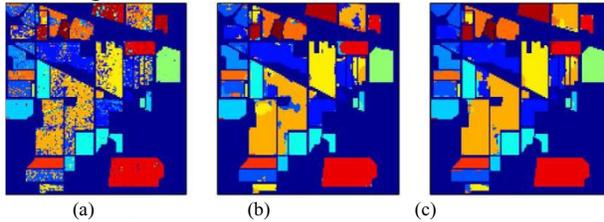


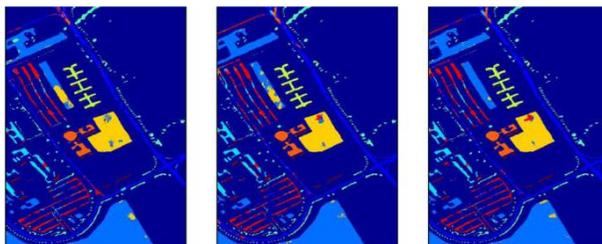
Fig.4 The classification map of Indian Pines(a) KELM (b)CK-SVM(c) EMP-kernel ELM

The results for the classification of the Indian Pine are shown in figure.4. It can be clearly seen that the classification maps of EMP kernel ELM is more coherent in the homogeneous regions than other methods. In addition, the spatial-spectral methods provide better results than spectral methods in terms of consistent classification results with little noise. In particular, the improvement is typically arisen for classes with similar spectral signatures.

The classification results for the Pavia Centre image are shown in figure 5 and the accuracy measures in Table II. The total number of pixels available in the reference data is 414815. Accordingly, a training set of 30 samples per class. Regarding Table II, the accuracy measures of the ELM-based proposed technique can provide equally competitive and even better classification results when compared to the traditional approach.

TABLE II  
OA, AA AND KAPPA BY DIFFERENT APPROACHES ON PAVIA UNIVERSITY FOR DIFFERENT CLASSES

Class	Train	KELM	CK-SVM	EMP-KELM
Asphalt	30	95.12	90.23	<b>100</b>
Meadow	30	98.63	91.19	<b>95.82</b>
Gravel	30	76.80	88.21	<b>93.75</b>
Trees	30	93.15	<b>95.02</b>	94.63.
Metal sheets	30	99.56	<b>99.73</b>	99.50
Bare soil	30	84.10	93.99	<b>96.30</b>
Bitumen	30	78.93	97.35	<b>99.52</b>
Bricks	30	89.56	86.90	<b>95.26</b>
Shadows	30	99.95	<b>99.98</b>	99.87
OA		79.16	91.45	<b>95.28</b>
AA		85.28	92.03	<b>96.27</b>
$\kappa$		0.7239	0.8990	<b>0.9780</b>
Time(s)		<b>2.9349</b>	38.26	20.36



(a) (b) (c)

Fig.5 The classification map of University of Pavia (a) KELM (b) kernel-SVM (c) EMP-kernel ELM

From Table II , we can see that when spatial information is used ,the accurate of classification is dramatically increased . Such as bare soil , from 84.10% to 96.30%,bitumen ,from 78.93 %to 99.52%.The reason is that spatial information helps to discriminant the samples with similar spectral curves.

In all of the experiments, the spatial-spectral methods provide more accurate results than the spectral methods. It indicates that spatial information is necessary to complement the spectral features for identifying the subtle differences of similar objects.

V. CONCLUSIONS

In this paper, we have proposed a new EMP based spatial-spectral information joint kernel ELM framework for HSI classification. In particular, the spatial-spectral information are employed by using a weighted summation. Experimental results have shown that the proposed EMP kernel ELM is more accurate and faster than the SVM-CK for the classification of HSI.

VI. ACKNOWLEDGEMENT

The author would like to thank the Editor-in-chief, the associate editor, the anonymous reviewers for their suggestions and insightful comments for this paper.

REFERENCES

- [1] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE Signal Proc. Mag.*, vol. 31, no. 1, pp. 45-54, Jan. 2014.
- [2] J. A. Gualtieri and R. F. Crompt, "Support vector machines for hyperspectral remote sensing classification," in *Proc. SPIE Workshop Adv. Comput. Assisted Recog.*, 1998, vol. 3584, pp. 221-232.
- [3] G. Camps-Valls, L. Gomez-Chova, J. Muñoz Maré, J. Vila-Francés, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93-97, Jan. 2006.
- [4] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Evaluation of kernels for multiclass classification of hyperspectral remote sensing data," in *Proc. IEEE ICASSP*, May 2006, pp. II-813-II-816.
- [5] M. Fauvel, B. Arbelot, J. A. Benediktsson, D. Sheeren, and J. Chanussot, "Detection of hedgerows in rural landscape using a local orientation feature: From linear opening to path opening," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 1, pp. 15-26, Feb. 2013.
- [6] S. Valero, J. Chanussot, J. A. Benediktsson, H. Talbot, and B. Waske, "Advanced directional mathematical morphology of the detection of the road network in very high resolution remote sensing images," *Pattern Recognit. Lett.*, vol. 31, no. 10, pp. 1120-1127, Jul. 2010.
- [7] L. Fang, S. Li, W. Duan, J. Ren, and J. A. Benediktsson, "Classification of hyperspectral images by exploiting spectral-spatial information of super pixel via multiple kernels,"

- IEEETrans.Geosci.RemoteSens.*,vol.53, no. 12, pp. 6663–6674, Dec. 2015.
- [8] Y. Gu, T. Liu, X. Jia, J. A. Benediktsson and J. Chanussot, "Nonlinear Multiple Kernel Learning With Multiple-Structure-Element Extended Morphological Profiles for Hyperspectral Image Classification," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 6, pp. 3235-3247, June 2016.
- [9] Pan, Bin; Shi, Zhenwei; Xu, Xia. Nicole, "Hierarchical Guidance Filtering-Based Ensemble Classification for Hyperspectral Images ," *IEEETrans.Geosci.RemoteSens.*,vol.55, no. 7, pp. 4177–4189, Jul. 201
- [10] B. Pan, Z. Shi and X. Xu, "R-VCANet: A New Deep-Learning-Based Hyperspectral Image Classification Method," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 5, pp. 1975-1986, May 2017..
- [11] Y. Yuan, J. Lin and Q. Wang, "Hyperspectral Image Classification via Multitask Joint Sparse Representation and Stepwise MRF Optimization," in *IEEE Transactions on Cybernetics*, vol. 46, no. 12, pp. 2966-2977, Dec. 2016.
- [12] G. B. Huang, H. Zhou, X. Ding and R. Zhang, "Extreme Learning Machine for Regression and Multiclass Classification," in *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 2, pp. 513-529, April 2012.
- [13] Y. Zhou, J. Peng and C. L. P. Chen, "Extreme Learning Machine With Composite Kernels for Hyperspectral Image Classification," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2351-2360, June 2015.
- [14] Chen, C.; Li,W.; Su, H.; Liu, K. "Spectral-spatial classification of hyperspectral image based on kernel extreme learning machine." *Remote Sens. vol. 6,no 6, pp.5795–5814,Jun 2014*
- [15] Arguello, Francisco; Heras, Dora B. "ELM-based spectral-spatial classification of hyperspectral images using extended morphological profiles and composite feature mappings." *Remote Sens. vol. 36, no. 2, pp. 645-664, Jan 2015.*
- [16] M. Jiang, F. Cao and Y. Lu, "Extreme Learning Machine With Enhanced Composite Feature for Spectral-Spatial Hyperspectral Image Classification," in *IEEE Access*, vol. 6, pp. 22645-22654, 2018.
- [17] P. R. Marpu, M. Pedernana, M. Dalla Mura, J. A. Benediktsson and L. Bruzzone, "Automatic Generation of Standard Deviation Attribute Profiles for Spectral–Spatial Classification of Remote Sensing Data," in *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 2, pp. 293-297, March 2013
- [18] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Phil. Mag.*, vol. 2, no. 6, pp. 559–572, 1901.