

# Time-series Processing of Large Scale Remote Sensing Data with Extreme Learning Machine over Cloud Infrastructure

XXX<sup>a</sup>, xxx<sup>a,b</sup> and xxx<sup>a</sup>

<sup>a</sup>*Zhejiang University*

<sup>b</sup>*School*

---

## Abstract

Nowadays, land-cover change detection plays a more and more important role in environment protection and many other fields. However, the current land-cover change detection methods encounter the problem of low efficiency and can't be expanded to parallel computing to deal with fast-growing remote sensing(RS) data. To solve the above problems, we propose a novel ELM-based land-cover change detection method, in which the supervised classification capability, fast training speed and high generalisation performance of ELM is utilized to efficiently train the land-cover classifier and apply the classifier to time-series satellite imageries' analysis. As one experiment, we apply our method to the analysis of rapid land use change in Changjiang River Delta over the past two decades due to accelerated urbanization. Moreover, we prove our method's scalability by another experiment of large scale RS imageries processing over the cloud infrastructure.

*Key words:* Extreme learning machine, Remote Sensing, Classification, Change Detection, Time-series, MapReduce, Hadoop.

---

## 1 Introduction

Nowdays, the available time-series RS images provide a new way for land-cover change detection which is widely required in various fields. However, except for the traditional challenge of high accuracy, large scale RS data processing also requires the land-cover change detection to be efficient and highly scalable.

---

\* Corresponding author

*Email address:* xxx@zju.edu.cn (XXX).

This paper proposes a novel elm-based land-cover change detection method with faster processing speed and higher scalability, which is practiced on cloud infrastructure.

### *1.1 Land-cover Change Detection with Time-series RS Images*

Change detection is the process of identifying differences in the state of an object or phenomenon by observing it at different times[1]. With the development of satellite technology, massive time-series high resolution multispectral RS images are available for applications. By comparing two sets of RS images, taken of the same area at different time, we can manually handle the change detection job by view. In fact, it is possible for the computer to process the digital RS images, and provide an automatical change detection technology, which is available for large scale processing of RS images. In summary, time-series RS image-based land-cover change detection method is to identify the interesting land-cover changes between "before" time images and "after" time images through RS image digital processing and comparison of RS images in an automatical way.

In the recent decades, timely and accurate change detection of Earth's surface features plays a more and more important role in better decision making. Automatic change detection can be used in such diverse applications as land usage analysis, disaster monitoring, snow-melt measurements, forest coverage monitoring and other environment changes. Especially, land-cover change detection is one of the major direction of change detection application. Various papers[2][3][4] have presented their work of applying change detection technology to the analysis of land-use and land-cover. Ross S. Lunetta[4] performed the change detection experiment in the biologically complex landscape of the Neuse River Basin, North Carolina using Landsat5 and 7 imagery collected in May of 1993 and 2000. Another typical application is performed by Qihao Weng[2], to analysis the rapid land use change taken place in Zhujiang Delta over the past decades with the help of change detection technology.

### *1.2 Challenges of Land-cover Change Detection*

In the applications of RS images, land-cover change detection has been a problem for some time. Among all the factors that cause land-cover change detectoin unavailable to real situations, change detection methods or algorithms used play a key role. Traditionally, the main challenge to land-cover change detection methods is the detection accuracy. However, with increase scale of RS images and these images' high resolution, the method's processing speed and scability rise to be another two major challenges.

Presently, large data processing has become a hot research topic of computing science. Similarly, in the field of RS data processing, increasing data size also brings some new challenges to processing methods. On one hand, efficient processing method is one of the best solutions to reduce processing time, on the other hand, parallel computing or distributed computing can greatly improve the processing efficiency of large scale data. In order to support parallel computing, the scalability of the processing method become another challenge.

### *1.3 Major Contributions of Elm-based method*

This paper’s major contribution is to propose a novel elm-based method for land-cover change detection. Using our method, the analysis of rapid land use change in Changjiang River Delta over the past two decades due to accelerated urbanization is done. Moreover, the land-cover change detection result is evaluated and compared to the current methods, which use the classifier of BP network and SVM. The novel approach supports high speed processing of single data set and can be extended to parallel computing on cloud infrastructure for the processing of large scale RS data. In our experiment, we expand the trained ELM network to hadoop based cloud infrastructure to prove the method’s scalability. Through the evaluation of the performance improvement of parallel computing, we prove the method’s capability to deal with increasing RS imageries.

ELM is a recently proposed and widely used machine learning method, with the capability of multiclass classification and universal approximation[5][6][7][8]. One of the major characteristic of ELM is its very fast training processing, which directly increases the processing efficiency of land-cover change detection method. The other advantage of ELM is the high generalisation performance. Generalisation is the key factor that will decide whether the trained network can be widely applied to large scale RS data. Moreover, our proposed change detection method also take advantage of ELM’s less human intervene for the process of network training. In conclusion, with ELM’s high speed of training, high generalisation and less manual intervene, we design our novel land-cover change detection method to overcome the challenges of large scale RS data processing.

## 2 Related Work

### 2.1 *Methods for Land-cover Change Detection*

Land usage analysis and land cover mapping has long been an area of research focus, and a wide range of methods including cross-correlation analysis, post-classification, image differencing, image ratioing, principal components analysis and so on have been explored. Both the traditional methods post-classification and cross-correlation determine the land-cover change through the "before" time and "after" time classification map, and produce comparable result. Although both methods are more automatic than the others like image differencing, the unsupervised classifier used result in low accuracy, which is one of the most important challenges that change detection methods face. The methods of image differencing, image ratioing and principal components analysis involve transformations of the original spectral bands so as to enhance the land cover changes. These methods and their improvement do increase the accuracy of land-cover change detection, but it is limited to small scale data processing and are lack of automation.

With the advances of machine learning, several alternative land-cover change detection methods using neural network are proposed and applied in the analysis of landscape. Presently, both supervised and unsupervised neural network algorithms have made progress. Compared to unsupervised land-cover transitions detection methods, the supervised ones provide more information about the kinds of transitions that occurred on the ground and are less affected by the difference atmospheric conditions, sensor calibration, and ground condition[9][10]. The paper[10] presents a technique for using predictive modeling, which predicates "before" and "after" pixel value, to identify unusual changes in imageries. Begum Demir[9] present a novel iterative active learning(AL) technique aimed at defining effective multitemporal training sets to be used for the supervised detection of land-cover transitions in a pair of time-series RS imageries. In comparison, unsupervised approaches require fewer manual interventions[11][12][13], and are quite suitable in the situations that the ground-truth is always unavailable. However, the low accuracy of unsupervised classification and predication severely limit the application of unsupervised approaches.

Although the classification and predication capability of neural network brings much improvement to the traditional land-cover change detection methods, the common and widely applied neural networks, like BP network and SVM, exist several inadequacies. The most two important aspects that exist room for improvement may be the low speed of training and low generalization.

## 2.2 ELM for Remote Sensing Application

Presently, little work of applying ELM to the field of RS imagery processing has been done. Mahesh Pal[14] did the land cover supervised classification experiment with ELM using remote sensing images. Wu Jun[15] revealed the equivalence between ELM and the positive and negative fuzzy rule system. Moreover, Wu Jun indicate the claim that ELM can be naturally used for training the positive and negative fuzzy rule system quickly for image classification. However, only simple comparison of classification cabability between ELM and other classification network was done in both efforts, and no time series images processing work are tried, not along large scale RS iamge processing.

## 2.3 ELM Based Methods for Large-scale Data Set Processing

Extreme learning machine is a new learning algorithm for single-hidden layer feedforward neural networkd(SLFNs) which randomly chooses the input weights and analytically determines the output weights of SLFNs[16]. In fact, ELM is to minimize the training error as well as the norm of the output weights[16][17], and thus to achieve better generalization performance. As the hidden nodes, including the input weights and hidden nodes' threshold are randomly generated,**H**, namely the hidden layer output matrix of the neural network[18][19], can be directly calcuated, thus greatly reducing the network training time.

As large-scale data set processing has been the trend of present applications, some researchers have paid their attentions to the scapability of ELM methods. Through improving the scapability, they extend the ELM variant methods to parallel computing infrastructures to deal with large-scale data set training. Qing He[20] proposed to implement an efficient paralalled ESVM based on the current and powerful parallel programming framework MapReduce. Moreover, an incremental learning algorithm for ESVM, which can meet the requirement of online learning to update the existing network, as well as its parallel version, is proposed in [20]. Mark van Heeswijk[21] compared three distinct ways of accelerating the evaluation of this ensemble of ELMs, and concluded that the GPU-accelerated and parallelized ELM ensemble achieved attractive speedups over using single CPU. Yongjiao Sun[21] propose an OS-ELM based ensemble classification framework for distributed classification in a hierarchical P2P network. In spite of the above contributions to the scability improvement, their proposed parallel computing methods are limited to the model building process. Meanwhile, no real applications are applied in and their efficiency on remote sensing imagery processing are known.

### 3 ELM for Change Detection and Approximation

Fundamentally, our way of change detection is based on the land-cover classification of the pair of time-series RS images. Huang et al[5][6][7] have proved that ELM has the capability of multiclass classification.

#### 3.1 Land Cover Classification

According the standard of XXX, the land cover can be classified into several types according to the specifical area. In order to assure the precision, statistical property and representative of the sample, training sample of each category should be selected through visual interpretation and field survey[22]. According to method of paper[23], we use image texture to interpret each ground object, and the are corresponding with Table 1.

Ground object category	Image features	Ground object description
Road	Light grey, stripy	Urban road
Water	Black or deep blue, stripy	River or pond
Vegetation	Crimson or pink, block	Flower bed or green belt
Building	Bright gray, block	Urban building
Vacant land	Gray, block	Vacant land

Table 1

Image Interpretation of Each Ground Object Category

Here is the process of land cover classifier training and testing:

#### First, Training and Testing Data Preparation

The RS images of the eara xxx are downloaded from the website of XXX. To represent the categories of the land cover, every type of the result is assigned to an integer, ranging from 1 to 5. As we use multi-spectral RS images, every pixel is multi-dimensional, and thus every sample can be described:

$$\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in \mathbf{R}^n$$

$$\mathbf{t}_i = f(x) = \begin{cases} [1, 0, 0, 0, 0], & \text{if type is Road} \\ [0, 1, 0, 0, 0], & \text{if type is Water} \\ [0, 0, 1, 0, 0], & \text{if type is Vegetation} \\ [0, 0, 0, 1, 0], & \text{if type is Building} \\ [0, 0, 0, 0, 1], & \text{if type is Vacant land} \end{cases}$$

where  $\mathbf{x}_i$  is the input,  $\mathbf{t}_i$  is the output,  $n$  is the number of spectralis of the RS images,  $i$  ranges from 1 to  $N$ , and  $N$  is the number of samples. In order to obtain enough training data, we annotate five  $30 \times 30$  squares in the chosen RS image by visual, each of which represents one type of land cover. In result,

$30 \times 30 \times 5$  number of samples are provided for the ELM network training, namely,  $N = 4500$ . Enough samples for network testing can also be obtained in a similar way, but from different image individuals in the same image type, in order to assure that the network has enough generalization performance to fulfill large-scale change detection.

### Second, Hidden Nodes Generation

According to the theorem of Huang, et al [17], for  $N$  arbitrary distinct samples  $(\mathbf{x}_i, \mathbf{t}_i)$  described above, standard SLFNs with  $M$  hidden nodes and activation function  $g(x)$  are mathematically modeled as

$$\sum_{i=1}^M \beta_i g_i(\mathbf{x}_j) = \sum_{i=1}^M \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = \mathbf{o}_j, j = 1, \dots, N,$$

where  $\mathbf{w}_i = [w_{i1}, w_{i2}, \dots, w_{im}]^T$  is the weight vector connecting the  $i$ th hidden node and the input nodes,  $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ ,  $m = 5$  is the weight vector connecting the  $i$ th hidden node and the output nodes, and  $b_i$  is the threshold of the  $i$ th hidden node.  $\mathbf{w}_i \cdot \mathbf{x}_j$  denotes the inner product of  $\mathbf{w}_i$  and  $\mathbf{x}_j$ .

In fact, ELM is to minimize the training error as well as the norm of the output weights [16][17], and thus to achieve better generalization performance.

$$\text{Minimize : } \left\| \mathbf{H}\beta - \mathbf{T} \right\|^2 \text{ and } \left\| \beta \right\|$$

That means

$$\sum_{i=1}^M \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = \mathbf{t}_j, j = 1, \dots, N, N = 4500$$

namely,

$$\mathbf{H}\beta = \mathbf{T},$$

where

$$\mathbf{H}(\mathbf{w}_1, \dots, \mathbf{w}_M, b_1, \dots, b_M, \mathbf{x}_1, \dots, \mathbf{x}_N) = \begin{bmatrix} g(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1) & \cdots & g(\mathbf{w}_M \cdot \mathbf{x}_1 + b_M) \\ \vdots & \ddots & \vdots \\ g(\mathbf{w}_N \cdot \mathbf{x}_1 + b_1) & \cdots & g(\mathbf{w}_M \cdot \mathbf{x}_N + b_N) \end{bmatrix}_{N \times M},$$

$$\beta = \begin{bmatrix} \beta_i^T \\ \vdots \\ \beta_M^T \end{bmatrix}_{M \times m}, \mathbf{T} = \begin{bmatrix} \mathbf{t}_1^T \\ \vdots \\ \mathbf{t}_N^T \end{bmatrix}_{N \times m} \quad \text{and } m = 5, N = 4500$$

As the  $M$  hidden node are randomly generated, the weight  $\mathbf{w}_i$  between the  $i$ th hidden node and the input nodes, as well as the  $i$ th hidden node's threshold

$b_i$ , can be randomly generated. In common case, the values of  $\mathbf{w}_i$  and  $b_i$  range from 0 to 1. Thus,  $\mathbf{H}$ , namely the hidden layer output matrix of the neural network[18][19], can be specified when  $\mathbf{w}_i$  and  $b_i$  are randomly generated.

### Third, Output Weights Calculation

In fact, we can calculate the output weights  $\beta$  through the smallest norm least-squares solution of the above linear system:  $\mathbf{H}\beta = \mathbf{T}$ .

$$\hat{\beta} = \mathbf{H}^\dagger \mathbf{T},$$

where  $H^\dagger$  is the Moore-Penrose generalized inverse of matrix  $\mathbf{H}$ . The methods to calculate the Moore-Penrose generalized inverse of  $\mathbf{H}$  may includes but are not limited to orthogonal projection, orthogonalization method, iterative method, and singular value decomposition(SVD). However, in ELM algorithm, the SVD method is proposed by the paper[17].

### Forth, Test

In order to assure highest generalization performance of the trained ELM network, testing samples from different RS image individuals of the same type are needed to calculate the testing generalization. ELM network with too few hidden nodes has no capability to achieve enough training classification accuracy, while too many hidden nodes will cause over-fitting, causing low testing classification accuracy. In a word, an algorithm with network testing is required to adjust the number of hidden nodes and activation function, finally to obtain an ELM network with highest generalization performance.

Here is the incremental algorithm to adjust the number of hidden nodes to achieve the highest testing generalization performance.

## 3.2 Land Cover Change Detection

With the trained and hidden nodes number adjusted ELM network, we can apply the network to the application of land cover change detection. Input every pixel of the RS image to be processed, namely  $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in \mathbf{R}^n$ , to the above ELM network, and get the calculated output:  $\mathbf{t}_i$ , an vector that represents the type of land cover.

With this method, we can calculate both "before" time and "after" time RS images' type of land cover. Through comparison of the statistics of land cover type between "befor" time and "after" time RS images, we can calculate the change of five types of land cover, such as the growth or decline rate of vegetation. Moreover, with the land cover type of every pixel calculated, the land cover maps of the time-series images can be drawn, which give direct trend of land cover in visual.



---

**Algorithm 1** Incremental Algorithm for SLFNs With Random Additive or RBF Nodes

---

**Require:** Given a training samples set  $\chi = \{(\mathbf{x}_i, \mathbf{t}_i) | \mathbf{x}_i \in \mathbf{R}^n, \mathbf{t}_i \in R^m, i = \{1, \dots, N\}\}$ , a testing samples set  $\psi = \{(\mathbf{x}_i, \mathbf{t}_i) | \mathbf{x}_i \in \mathbf{R}^n, \mathbf{t}_i \in R^m, i = \{1, \dots, \widehat{N}\}\}$ , activation function  $g(x)$ , maximum hidden node number  $M_{max}$ , minimum hidden node number  $M_{min}$ , and expected testing accuracy  $\epsilon$ :

Step 1) **Initialization:** Let  $M = M_{min}$ , testing accuracy  $E = 0$ , number of hidden nodes with highest testing accuracy  $M_{highest} = 0$ , highest testing accuracy  $E_{highest} = 0$

Step 2) **Learning Step:**

**while**  $M < M_{highest}$  and  $E_{highest} < \epsilon$  **do**

a) increase by one the number of hidden nodes  $M : M = M + 1$

b) training the new network with the training samples set  $\chi$

c) calculate the testing accuracy  $E$  of the trained network with the testing samples set  $\psi$

**if**  $E > E_{highest}$  **then**

d)  $E_{highest} \leftarrow E$

e)  $M_{highest} \leftarrow M$

**end if**

**end while**

---

## 4 Large Scale RS Image Processing

With the progress of earth observation satellite, large scale RS images are available for applications. Moreover, multi-spectral RS images, which increase the data size of every pixel by increasing the dimension of input vector and require more computing capability to process, tend to be the mainstream in the field of RS images processing. In order to process an increasing amount of RS data, new computing technologies, such as cloud computing, need to be applied in the field.

In the above described land cover change detection application, change detection of a specific land with a pair of time-series RS images is possible. With the higher resolution RS images, more accurate change detection result can be provided. Because of the size limit of a single RS image, accurate change detection of a large area only through a pair of time-series RS images is unreliable.

However, a series of high resolution time-series RS images, which cover a large area when combined together, are available now. Basically, we can divide the large set of time-series RS images into pairs of RS images, and do the change detection process of every pair of RS images in a distributed way.

Every pair of time-series RS images can be represented as  $d_i = (I_{before}, I_{after}), i =$

$1, 2, \dots, l$ , where  $l$  is the number of pairs of time-series RS images,  $I_{before}$  and  $I_{after}$  represent the "before" time RS image and "after" time RS image individually. Thus we can represent the total RS images set as:

$$D = \{d_1, d_2, \dots, d_l\}$$

With the trained and tested ELM network, we can detect the change of every sub-area that is covered by a pair of RS image  $d_i$  separately. Similarly, we represent the detected change statistics of every sub-area as

$$\mathbf{r}_i = [s_{i1}, \dots, s_{im}], i = 1, 2, \dots, l$$

where  $m$  is the number of land cover types, and  $s_{ij}, j = 1, 2, \dots, m$  is the land cover statistic of  $j$ th land cover type defined. Moreover, the detected change statistics of the whole area, combined by the sub-area, is

$$\mathbf{R} = [s_1, \dots, s_m] = \mathbf{r}_1 + \mathbf{r}_2 + \dots + \mathbf{r}_l$$

Here is the algorithm for the processing of large scale RS image set  $D$  with MapReduce.

---

**Algorithm 2** Distributed Processing Algorithm for Large Scale RS Images with MapReduce

---

**Require:** number of hadoop job workers:  $v$

- a) **Map:** divide  $D$  into subsets:  $s_k = d_{(k-1) \times \frac{l}{v}}, \dots, d_{k \times \frac{l}{v}}$  and  $k = 1, \dots, v$
  - b) **Process:** Calculate  $r_i$  of every pair of time-series RS images on the slaver of hadoop cluster
  - c) **Merge:** Merge all the result,  $r_i$  of  $j$ th worker node into result set  $rs_j$
  - c) **Reduce:** Reduce the merged result  $rs_i, i = 1, \dots, v$  into the final result set,  $R$
- 

## 5 Time-series RS Image Prediction

## 6 Experiments

### 6.1 Classification and change detection of a pair of remote sensing images

Annotate the samples with five kinds of land-cover types: water, Road, Vegetation, Building, and Vacant land Draw the land-cover map and statistics table using the trained network. Compare the result of two time-series images, and finally get the change result

### *6.2 Using hadoop to accelerate the change detection processing on a large scale of images*

Building the hadoop platform Calculate the change detection result of every pair of time-series images in a distributed way using the network trained in the first experiment. Merge the change detection results of every sub-area and finally get the result of the whole large area

### *6.3 Predicate the land-coverly using ELM*

Get a pair of time-series samples from a pair of remote sensing images Training the approximation network using the pair of samples Calculate the land survey of the after-time image through the before-time image using the above trained network. Do another experiment using SVM and BP, compare the traing speed and generalisation performance

## **7 Performance Verification**

### *7.1 1*

Nearly no time-series processing of remote sensing application has been done by ELM. We can compare it with the traditional way of change detection, such as the approach using image diffence.

### *7.2 2*

Using hadoop to accelerate the change detection processing has also not been tried

### *7.3 3*

Compare the approximation performance of ELM to SVM and BP network

## 8 Conclusions

This paper studies ELM for classification with the standard optimization method.

## References

- [1] A. SINGH, “Review article digital change detection techniques using remotely-sensed data,” *International Journal of Remote Sensing*, vol. 10, no. 6, pp. 989–1003, 1989.
- [2] Q. Weng, “Land use change analysis in the zhujiang delta of china using satellite remote sensing, gis and stochastic modelling,” *Journal of Environmental Management*, vol. 64, no. 3, pp. 273 – 284, 2002.
- [3] J. M. Read and N. S.-N. Lam, “Spatial methods for characterising land cover and detecting land-cover changes for the tropics,” *International Journal of Remote Sensing*, vol. 23, no. 12, pp. 2457–2474, 2002.
- [4] R. S. Lunetta, J. Ediriwickrema, D. M. Johnson, J. G. Lyon, and A. McKerrow, “Impacts of vegetation dynamics on the identification of land-cover change in a biologically complex community in north carolina, usa,” *Remote Sensing of Environment*, vol. 82, no. 23, pp. 258 – 270, 2002.
- [5] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, “Extreme learning machine for regression and multiclass classification,” *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 42, pp. 513 –529, april 2012.
- [6] G.-B. Huang, Y.-Q. Chen, and H. A. Babri, “Classification ability of single hidden layer feedforward neural networks,” *IEEE Transactions on Neural Networks*, vol. 11, pp. 799–801, 2000.
- [7] R. Zhang, G.-B. Huang, N. Sundararajan, and P. Saratchandran, “Multicategory classification using an extreme learning machine for microarray gene expression cancer diagnosis,” *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, vol. 4, pp. 485–495, July 2007.
- [8] G.-B. Huang, “Universal approximation using incremental constructive feedforward networks with random hidden nodes,” 2006.
- [9] B. Demir, F. Bovolo, and L. Bruzzone, “Detection of land-cover transitions in multitemporal remote sensing images with active-learning-based compound classification,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. PP, no. 99, pp. 1 –12, 2011.
- [10] C. Clifton, “Change detection in overhead imagery using neural networks,” *Applied Intelligence*, vol. 18, pp. 215–234, March 2003.

- [11] S. Patra, S. Ghosh, and A. Ghosh, “Unsupervised change detection in remote-sensing images using one-dimensional self-organizing feature map neural network,” in *Information Technology, 2006. ICIT '06. 9th International Conference on*, pp. 141–142, dec. 2006.
- [12] S. Patra, S. Ghosh, and A. Ghosh, “Unsupervised change detection in remote-sensing images using modified self-organizing feature map neural network,” in *Computing: Theory and Applications, 2007. ICCTA '07. International Conference on*, pp. 716–720, march 2007.
- [13] K. Chen, C. Huo, Z. Zhou, and H. Lu, “Unsupervised change detection in high spatial resolution optical imagery based on modified hopfield neural network,” in *Natural Computation, 2008. ICNC '08. Fourth International Conference on*, vol. 4, pp. 281–285, oct. 2008.
- [14] M. Pal, “Extreme learning machine for land cover classification,” *CoRR*, vol. abs/0802.1412, 2008.
- [15] W. Jun, W. Shitong, and F.-l. Chung, “Positive and negative fuzzy rule system, extreme learning machine and image classification,” *International Journal of Machine Learning and Cybernetics*, vol. 2, pp. 261–271, 2011. 10.1007/s13042-011-0024-1.
- [16] G.-B. Huang, “Extreme learning machine: a new learning scheme of feedforward neural networks,” 2004.
- [17] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, “Extreme learning machine: Theory and applications,” *Neurocomputing*, vol. 70, no. 13, pp. 489 – 501, 2006. jce:title;Neural Networks;ce:title; jce:subtitle;Selected Papers from the 7th Brazilian Symposium on Neural Networks (SBRN '04);ce:subtitle; jxocs:full-name;7th Brazilian Symposium on Neural Networks;jxocs:full-name;.
- [18] G.-B. Huang, “Learning capability and storage capacity of two-hidden-layer feedforward networks,” 2003.
- [19] G.-B. Huang, “Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions,” 1998.
- [20] Q. He, C. Du, Q. Wang, F. Zhuang, and Z. Shi, “A parallel incremental extreme svm classifier,” *Neurocomputing*, vol. 74, no. 16, pp. 2532 – 2540, 2011. jce:title;Advances in Extreme Learning Machine: Theory and Applications;ce:title; jce:title;Biological Inspired Systems. Computational and Ambient Intelligence;ce:title; jce:subtitle;Selected papers of the 10th International Work-Conference on Artificial Neural Networks (IWANN2009);ce:subtitle;.
- [21] M. van Heeswijk, Y. Miche, E. Oja, and A. Lendasse, “Gpu-accelerated and parallelized elm ensembles for large-scale regression,” *Neurocomputing*, vol. 74, no. 16, pp. 2430 – 2437, 2011. jce:title;Advances in Extreme Learning Machine: Theory and Applications;ce:title; jce:title;Biological Inspired Systems. Computational and Ambient Intelligence;ce:title; jce:subtitle;Selected papers

of the 10th International Work-Conference on Artificial Neural Networks (IWANN2009);i/ce:subtitlej.

- [22] Z. J.-L. J. X.-G. Yu Xiu-Lan, Qian Guo-Hui, "Learning sample selection in multispectral remote sensing image classification using the bp neural network," 1999.
- [23] G. Y. D. Z.-L. Z. Jiefeng Jiang, Jing Zhang, "Application of back propagation neural network in the classification of high resolution remote sensing image: Take remote sensing image of beijing for instance," 2011.