

A COMPARATIVE STUDY OF SEVERAL SLFN-BASED CLASSIFICATION ALGORITHMS FOR URBAN AND RURAL LAND USE

Yi Lin¹, Guangshun Xie^{1, *}, Tinghui Zhang¹, Jie Yu¹, Hanchao Zhang², Jianqing Cai³

¹ College of Surveying and Geo-Informatics, Tongji University, Shanghai, 200092, China - {linyi, shunshun, zhang_th, 2011_jieyu}@tongji.edu.cn

² Key Laboratory of Surveying and Mapping Science and Geospatial Information Technology, Ministry of Natural Resources, Beijing, 100036, China - zhanghc@casm.ac.cn

³ Institute of Geodesy, University of Stuttgart, 70174 Stuttgart, Germany - cai@gis.uni-stuttgart.de

KEY WORDS: remote-sensing image classification, land use, extreme learning machine, kernel function, A-ELM, machine learning.

ABSTRACT:

In the study of urban sustainable development, accurate classification of land use has become an important basis for monitoring urban dynamic changes. Hence it is necessary to develop the appropriate recognition model for urban-rural land use. Although deep learning algorithms have become a research hotspot in image classification tasks in recent years, and many good results have been achieved. But other machine learning algorithms are not going away. Compared deep learning with machine learning, there are some advantages and disadvantages in data dependence, hardware dependence, feature processing, problem solving methods, execution time, and interpretability, etc. Especially in the classification for remote sensing images, the continuous research and development of traditional machine learning algorithms is still of great significance. In this paper, the performances of several SLFN-based classification algorithms were studied and compared, including ELM, RBF K-ELM, mixed K-ELM, A-ELM and SVM. Extreme Learning Machine (ELM) is a new algorithm for single-hidden-layer feedforward neural network (SLFN). It has simple structure, fast speed and is easy to train. In some applications, however, standard ELM is prone to be overfitting and its performance will be affected seriously when outliers exist. In order to explore the performance of ELM and its improved algorithm for urban-rural land use classification, comparative experiments between three improved ELM algorithms (RBF K-ELM, mixed K-ELM and A-ELM), ELM and SVM with image data from several study areas were performed, and the classification accuracy and efficiency were analysed. The results show that the three improved ELM algorithms perform better than the standard ELM and SVM both in overall accuracy and Kappa coefficient. However, it is worth noting that the computation efficiency of RBF K-ELM and mixed K-ELM decreases greatly with larger image, the time cost is much more than other algorithms. Compared with other algorithms, A-ELM has the advantages of higher Overall Accuracy and less classification time.

1. INTRODUCTION

With the progress of urbanization, farmland, forest, water and other agricultural land is occupied, instead, more buildings, roads, railways, viaducts, and other infrastructure were built, which caused a sharp increase in the complexity of urban-rural land use. How to quickly and accurately extract complex urban-rural land cover usage is of great significance to urban and rural planning, resource investigation, land reclamation, and rural illegal land investigation (Li et al., 2021).

In recent years, remote sensing technology has obvious advantages in large-scale land use dynamic monitoring due to its wide coverage and real-time data acquisition, which makes it an optimal approach to complex urban-rural land use investigation (Li et al., 2019). Image classification algorithm is an important part of remote sensing image processing and analysis. Commonly used image classification algorithms include maximum likelihood, SVM, random forest, etc. (Zhou et al., 2021). Maximum likelihood requires more training samples and has long classification time. Prior to the advent of deep learning, support vector machines (SVM) were considered the most successful and best performing algorithm in machine learning over the last decade or so. Classical SVM is only suitable for binary classification. For multi-class classification, one needs to combine multiple SVMs. The classification accuracy of random forest is prone to be overfitting with high

noise level, so it is not suitable for urban-rural complex land use classification.

At present, the deep learning model is developing rapidly, which shows strong learning ability and can mine information from massive data fully and intelligently (Chen et al., 2021). But how to choose the suitable network structure and hyperparameter is still a problem and it needs big datasets to achieve good result (Ma et al., 2019). In several remote sensing applications with limited sample data provided, the deep learning algorithm cannot estimate the law of data in the unbiased way. In some applications of image classification, especially for small study areas and small samples, compared with deep learning methods, traditional machine learning methods have the advantages of lower computational cost and higher efficiency, which are still worth studying (Yang et al., 2020, Yu et al., 2020, Xiao et al., 2019, Cai et al., 2020, Mou and Lie, 2019).

For single-hidden-layer feedforward neural network (SLFN), gradient-based optimization methods are commonly used to optimize network parameters, but more iterative steps are needed to obtain better generalization ability. In view of the above shortcomings, Huang et al. (2006) proposed a fast construction and learning algorithm which was called Extreme Learning Machine (ELM) because of its fast training speed. However, ELM model is simple and the generalization ability is

* Corresponding author

poor. Researchers applied the kernel function to ELM and obtained kernel ELM (K-ELM), which improved the generalization performance significantly (Yao et al., 2014, Huang et al., 2012, Lin et al., 2018). In order to estimate the optimal regularization parameters of ELM, Qian (2018) studied the A-optimal design regularization ELM (A-ELM). The regression analysis and image classification experiments show that the A-ELM has excellent performance. Some scholars applied ELM to remote sensing image classification and got good results. Lin et al. (2017) used the bionic fish swarm algorithm (AF) to optimize the internal parameters of ELM, which achieved good results in classification experiments. Wang et al. (2018) applied multi-kernel extreme learning machine (MK-ELM) to forest land information extraction with high classification accuracy.

This paper explores the performance of ELM and its improved algorithms in urban-rural complex land use classification. Three improved ELM algorithms (RBF K-ELM, mixed K-ELM and A-ELM) are compared with ELM and SVM by using four groups of multispectral remote sensing data in different research areas. The efficiency and accuracy of these algorithms are studied, and the applicability and superiority of the algorithm are summarized.

2. ELM AND ITS IMPROVEMENTS

2.1 ELM

ELM is based on SLFN, whose structure is shown in Figure 1.

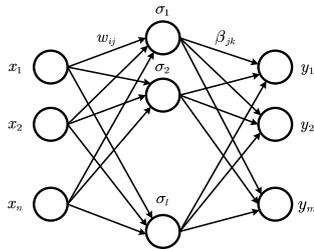


Figure 1. Structure of SLFN.

Huang et al. (2006) proved that a standard SLFN with L hidden layer neuron and an infinitely differentiable activation function $g(x)$ can approximate N training samples without error. For N random training samples, the mathematical model of SLFN with L hidden nodes and activation functions $g(x)$ is as follows:

$$\sum_{i=1}^L \beta_i g_i(x_j) = \sum_{i=1}^L \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = y_j, \quad j = 1, 2, \dots, N \quad (1)$$

where $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in \mathbf{R}^n$, $\mathbf{y}_i = [y_{i1}, y_{i2}, \dots, y_{im}]^T \in \mathbf{R}^m$, $\mathbf{w}_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$ is the weight between the i th hidden node and the input layer, $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ is the weight between the i th hidden node and the output layer, b_i is the bias of the i th hidden node, the above N equations can be written in the form of matrix:

$$\mathbf{H} \boldsymbol{\beta} = \mathbf{Y} \quad (2)$$

$\begin{matrix} \mathbf{H} & \boldsymbol{\beta} & = & \mathbf{Y} \\ \begin{matrix} N \times L & L \times m & & N \times m \end{matrix} \end{matrix}$

where

$$\mathbf{H} = \begin{bmatrix} g(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1) & g(\mathbf{w}_L \cdot \mathbf{x}_1 + b_L) \\ \vdots & \vdots \\ g(\mathbf{w}_1 \cdot \mathbf{x}_N + b_1) & g(\mathbf{w}_L \cdot \mathbf{x}_N + b_L) \end{bmatrix}_{N \times L} \quad (3)$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m} \quad \text{and} \quad \mathbf{Y} = \begin{bmatrix} y_1^T \\ \vdots \\ y_N^T \end{bmatrix}_{N \times m}$$

ELM's strategy is to generate \mathbf{w}_i and b_i randomly and remain constant. Regularization method such as ridge estimation is used to solve equation (2), then we can get:

$$\hat{\boldsymbol{\beta}} = (\mathbf{H}^T \mathbf{H} + \frac{\mathbf{I}}{C})^{-1} \mathbf{H}^T \mathbf{Y} \quad (4)$$

where \mathbf{I} is the identity matrix, and C is the regularization parameter.

2.2 K-ELM

In K-ELM, a kernel trick is applied. The kernel function is defined as follows, which contains a mapping from low dimension to high dimension (MARIÉTHOZ and BENGIO, 2007).

$$K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) \quad (5)$$

Similar to equation (1), we replace the activation function $g(x)$ with kernel function $K(x_1, x_2)$, then a kernel function model can be obtained (Sugiyama, 2015):

$$\sum_{i=1}^L \beta_i K(\mathbf{x}, \mathbf{x}_j) = y_j, \quad j = 1, 2, \dots, N \quad (6)$$

and the kernel matrix \mathbf{K} is:

$$\mathbf{K} = \begin{bmatrix} K(\mathbf{x}_1, \mathbf{x}_1) & K(\mathbf{x}_1, \mathbf{x}_N) \\ \vdots & \vdots \\ K(\mathbf{x}_N, \mathbf{x}_1) & K(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix}_{N \times N} \quad (7)$$

Then equation (8) can be derived by replacing the matrix \mathbf{H} with the kernel matrix \mathbf{K} in equation (4):

$$\hat{\boldsymbol{\beta}} = (\mathbf{K}^T \mathbf{K} + \frac{\mathbf{I}}{C})^{-1} \mathbf{K}^T \mathbf{Y} \quad (8)$$

In K-ELM, the activation function $g(x)$ and the number of hidden-layer neurons in the L need not be given, as long as a kernel function is given, and the random mapping is replaced by kernel mapping, which makes K-ELM perform more stable than ELM.

2.3 A-ELM

The regularization parameter C is of great importance for ELM model. ELM tends to be overfitting when regularization parameter is too small and is prone to be underfitting in case of large regularization parameter. This study also carefully explores several commonly used methods for determining regularization parameters. It mainly includes L-curve method, GCV (Generalized Cross-Validation) method and Ridge Trace method, which are numerical exploration and approximation.

But by the L-curve method, it is sometimes difficult to obtain the corresponding inflection point. By the GCV method, the change is too gentle and even difficult to converge, and the obtained regularization parameter is only a local optimum parameter, not a global optimum. By the ridge trace method is too subjective. Cai et al. and Cai (2004, 2004) derived the optimal method of regularization parameters in geodesy inversion problem under generalized conditions by mean square error matrix trace minimization (A-optimal design). Qian (2018) studied an A-optimal regularization method to obtain the optimal regularization parameter of ELM model based on the above research:

$$\lambda = \frac{\text{trace}(\mathbf{H}^T \mathbf{H} (\mathbf{H}^T \mathbf{H} + \lambda \mathbf{I})^{-3})}{\hat{\boldsymbol{\beta}}^T (\mathbf{H}^T \mathbf{H} + \lambda \mathbf{I})^{-2} \mathbf{H}^T \mathbf{H} (\mathbf{H}^T \mathbf{H} + \lambda \mathbf{I})^{-1} \hat{\boldsymbol{\beta}}} \quad (9)$$

where $\lambda = 1/C$. When dealing with the binary classification problem, the optimal regularization parameter can be directly obtained by using the above equation. For multi-class classification problem, we can rewrite equation (4) as:

$$[\hat{\boldsymbol{\beta}}_{C1}, \hat{\boldsymbol{\beta}}_{C2}, \dots, \hat{\boldsymbol{\beta}}_{Cm}] = (\mathbf{H}^T \mathbf{H} + \lambda \mathbf{I})^{-1} \mathbf{H}^T [\mathbf{y}_{C1}, \mathbf{y}_{C2}, \dots, \mathbf{y}_{Cm}] \quad (10)$$

where

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{C1} &= (\mathbf{H}^T \mathbf{H} + \lambda_1 \mathbf{I})^{-1} \mathbf{H}^T \mathbf{y}_{C1} \\ \hat{\boldsymbol{\beta}}_{C2} &= (\mathbf{H}^T \mathbf{H} + \lambda_2 \mathbf{I})^{-1} \mathbf{H}^T \mathbf{y}_{C2} \\ \hat{\boldsymbol{\beta}}_{Cm} &= (\mathbf{H}^T \mathbf{H} + \lambda_m \mathbf{I})^{-1} \mathbf{H}^T \mathbf{y}_{Cm} \end{aligned} \quad (11)$$

where $\lambda_1, \lambda_2, \dots, \lambda_m$ can be calculated by equation (9). Thus, the training result of A-ELM for multi-class classification is

$$\hat{\boldsymbol{\beta}} = [\hat{\boldsymbol{\beta}}_{C1}, \hat{\boldsymbol{\beta}}_{C2}, \dots, \hat{\boldsymbol{\beta}}_{Cm}] \quad (12)$$

Therefore, this study also tried the A-optimal regularization method to solve the regularization parameters in the classification algorithm.

3. EXPERIMENT

3.1 Data

In order to fully analyse and compare various improved ELM models, four Landsat images with different geographical features in different regions are selected for comparison experiments. Details are as follows.

Hamburg: Hamburg(53°33' N, 9°59' E) has a temperate maritime climate and is mild and humid all year round. Hundreds of small canals form a dense network of rivers throughout the city, and in the northeast, there are many ports and waterways with an extremely dense distribution of containers. Land use is quite complex for Hamburg.

Karlsruhe: Karlsruhe is located at 49°0' 50" N, 8°24' 15" E, with an average annual temperature of 10.7°C. It is one of the warmest cities in Germany. It mainly includes rivers, forests, cultivated land, buildings and other ground objects.

Hefei: Hefei(30°57' N-32°32' N, 116°41' E-117°58' E) has complex topography, abundant vegetation resources and

developed water system. It is an important city in the economic development of the Yangtze River Delta. So urban expansion is obvious there.

Chongming Dongtan: Dongtan(31°29' 39.34" N, 121°57' 31.99" E) is located at the entrance of the Yangtze River, at the eastern end of Chongming island. The siltation of the Yangtze River has created a large area of freshwater to brackish marshland, tidal ditches and intertidal mudflats. There are numerous farmlands, fish ponds, crab ponds and reed ponds in the area, with lush marsh vegetation and abundant benthic fauna, making it an excellent stopover and staging area for migratory birds in spring and autumn in the Asia-Pacific region, as well as an important wintering ground for migratory birds. However, the introduced *S. alterniflora* has threatened the survival of local vegetation.

Images of the study area are shown in Figure 2, which were obtained by Landsat 8 satellite and were pre-processed by ENVI software, including radiometric correction, atmospheric correction and image clipping. Detail of images are shown in Table 1.



Figure 2. Images of the study area.

Research region	Date	Size(pixel)	Resolution(m)
Hamburg	2018.08.29	725×705	30
Karlsruhe	2018.08.06	641×661	30
Hefei	2018.10.30	591×554	30
Chongming Dongtan	2018.05.28	289×395	30

Table 1. Image details.

According to field survey, for the first three images, the study area is classified into six categories: bare land, cultivated land, forest, building, concrete land and water. In the experiment, nearly 400 samples were manually selected from images using ENVI ROI tool. But for Chongming Dongtan image, classes are background, bare land, water, *S. alterniflora*, *S. mariqueter* and reed and about 150 samples were generated in view of the image size. For all images, 75 % of the samples were randomly chosen as training samples, and 25 % of the samples were used to validate results.

3.2 Feature Space

With the development of remote sensing technology, the spectral resolution of images is becoming higher and higher. But not all bands are useful for image classification. Therefore, it is a key step for classification to analyse the original image data to construct the optimal feature space. In this paper, for each image, three image bands with maximum OIF values were firstly calculated by the optimal index method (Dwivedi, 1992). Then, according to the features of the ground object category in the research area, K-T transformation was carried out on the

image. The first three components of k-T transformation (brightness, greenness and humidity) were taken. After that, NDVI and MNDWI (Xu, 2005) were calculated, which are helpful to distinguish water and plants. The above features were combined into eight-dimensional feature space for the experiment in this paper.

3.3 Comparison of Classification Algorithms

SVM basically dominated the classification models until algorithms like integrated learning and neural networks showed superior performance. In the current era of big data and large samples, SVM has lost some of its popularity due to its huge computational power for large samples, but it is still a common machine learning algorithm. In view of the good classification performance of SVM, SVM was added into the comparison experiment with improved ELM.

For K-ELM, Different kernel functions have different advantages (Zhou, 2016). To explore the influence of different kernel functions on K-ELM, in this paper, radial basis kernel function and mixed kernel function are selected to construct RBF K-ELM and mixed K-ELM. The mixed kernel function is the linear combination of the local kernel function, namely the radial basis kernel function, and a global kernel function, namely the polynomial kernel function. The construction equation is as follows:

$$\begin{aligned}
 K_{poly}(\mathbf{x}_i, \mathbf{x}_j) &= (\mathbf{x}_i^T \mathbf{x}_j)^d \\
 K_{RBF}(\mathbf{x}_i, \mathbf{x}_j) &= \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right) \\
 K_{mix}(\mathbf{x}_i, \mathbf{x}_j) &= \lambda K_{poly}(\mathbf{x}_i, \mathbf{x}_j) + (1-\lambda)K_{RBF}(\mathbf{x}_i, \mathbf{x}_j)
 \end{aligned}
 \quad (13)$$

where $K_{poly}(\mathbf{x}_i, \mathbf{x}_j)$ is polynomial kernel function, $K_{RBF}(\mathbf{x}_i, \mathbf{x}_j)$ is radial basis kernel function and λ weight coefficient.

In the experiment, the classification algorithm of ELM, RBF K-ELM, mixed K-ELM, A-ELM and SVM were compared and analysed. Because hyper-parameters have an important impact on the performance of the model. In this paper, the heuristic particle swarm optimization (PSO) algorithm (Zhang et al., 2019) is used to optimize the parameters of the above five models.

The experiments in this paper were performed step by step according to the flow chart shown in Figure 3.

3.4 Performance Evaluation

The performance of the model should be evaluated from two aspects, i.e. accuracy and efficiency. In the field of remote sensing, confusion matrix is commonly used to evaluate the classification model from which the Kappa coefficient and the overall accuracy can be calculated. Kappa coefficient is a multivariate discrete method to evaluate the classification accuracy and error matrix of remote sensing images. In addition, time consumption is an important indicator to evaluate the efficiency of classifier because users may prefer a model with fast speed and good generalization performance. Therefore, this paper selects training time, overall accuracy, Kappa coefficient and classification time as the evaluation indexes.

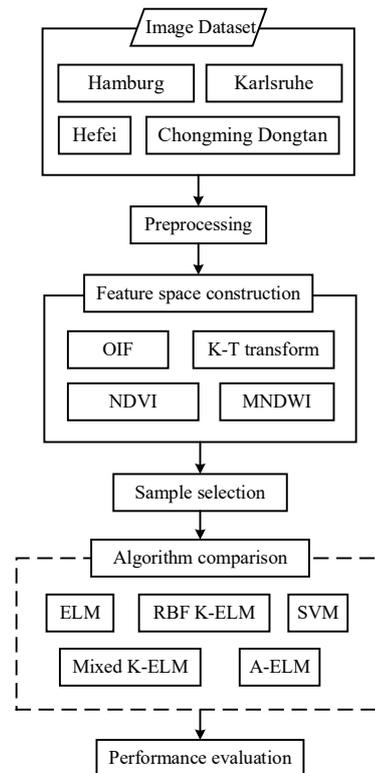


Figure 3. Flowchart of comparison experiment

3.5 Results

The classification results of images are shown in Figure 4 – Figure 7. The performance evaluation result of each model is shown in Table 2.

Based on the classification results and performance evaluation results, the following conclusions can be drawn:

- 1) In terms of accuracy, the overall accuracy and kappa coefficient of three improved ELM algorithms (RBF K-ELM, mixed K-ELM and A-ELM) are significantly higher than those of ELM and SVM. Among the three improved algorithms, RBF K-ELM has the highest overall accuracy and Kappa coefficient, followed by mixed K-ELM and A-ELM.
- 2) In terms of typical ground object recognition, there are some obvious errors of the ELM. For example, the port is identified as water and in Figure 4 and the concrete land is identified as bare land in Figure 5. Compared to ELM, the improved ELM algorithms can differentiate land use correctly, and distinguish urban land from agricultural land better. It shows that the three improved ELM algorithms have significantly improved the recognition effect of urban-rural complex land use.
- 3) In term of efficiency, the performance of the three improved ELM algorithms is different. The training time of RBF K-ELM and mixed K-ELM was slightly higher than that of ELM and SVM. A-ELM needs to calculate the optimal regularization parameters iteratively, so the training time is longer than other models. But the extra time is within acceptable range considering the good classification results achieved by A-ELM. Although the overall accuracy of RBF K-ELM and mixed K-ELM was very high, the classification time was much longer than that of other models, and the difference was more pronounced in larger size images (Hamburg, Karlsruhe and Hefei). Besides, the classification time of mixed K-ELM was more than three times that of RBF K-ELM, indicating that the performance of different kernel functions is quite different. In

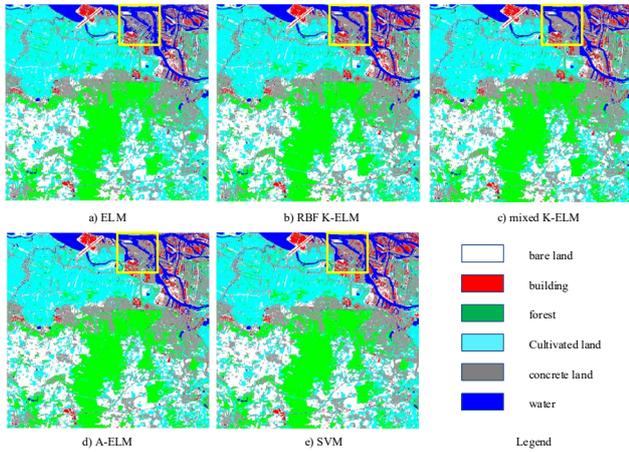


Figure 4. Classification result of Hamburg dataset.

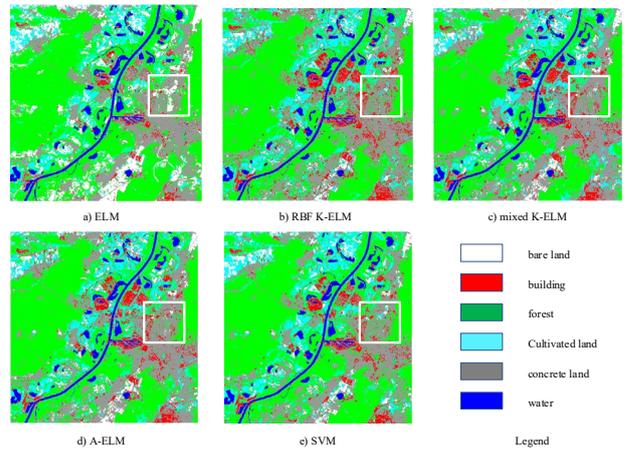


Figure 5. Classification result of Karlsruhe dataset.

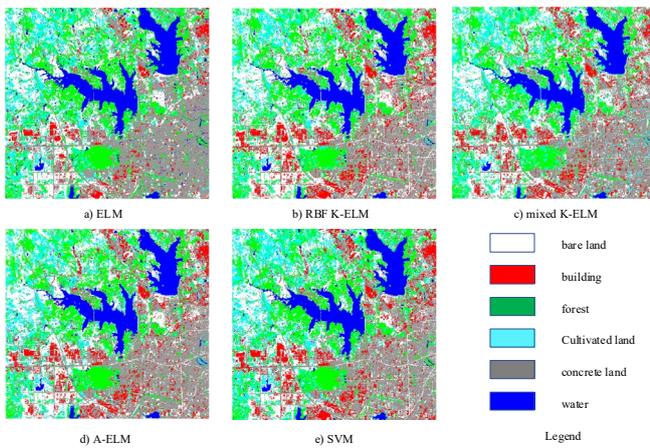


Figure 6. Classification result of Hefei dataset.

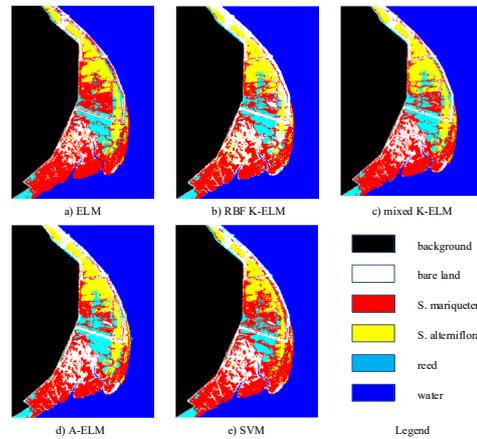


Figure 7. Classification result of Chongming Dongtan dataset.

Research region	Algorithms	Training time(s)	Overall accuracy(%)	Kappa	Classification time(s)
Hamburg	ELM	0.0045	95.5823	0.9470	0.3273
	RBF K-ELM	0.0849	97.8581	0.9743	13.9685
	Mixed K-ELM	0.2440	98.1258	0.9775	49.5322
	A-ELM	4.2463	97.0549	0.9647	0.3802
	SVM	0.0136	96.5194	0.9582	3.8370
Karlsruhe	ELM	0.0045	91.0067	0.8921	0.2653
	RBF K-ELM	0.0880	97.047	0.9646	11.9063
	Mixed K-ELM	0.2393	97.047	0.9646	60.0326
	A-ELM	4.5540	95.9732	0.9517	0.3262
	SVM	0.0186	94.7651	0.9372	3.4137
Hefei	ELM	0.0036	86.3388	0.8360	0.2126
	RBF K-ELM	0.0583	91.8033	0.9016	8.0916
	Mixed K-ELM	0.1551	90.4372	0.8852	33.3805
	A-ELM	2.4661	88.7978	0.8656	0.2386
	SVM	0.0132	86.8852	0.8426	1.5113
Chongming Dongtan	ELM	0.0013	97.4265	0.9691	0.0752
	RBF K-ELM	0.0111	100.0000	1.0000	1.2042
	Mixed K-ELM	0.0325	99.6324	0.9956	3.8370
	A-ELM	1.1085	99.6324	0.9956	0.0847
	SVM	0.0018	97.7941	0.9735	0.1729

Table 2. Performance evaluation of different algorithms.

contrast, A-ELM has much shorter classification time than RBF K-ELM and mixed K-ELM and has higher overall accuracy than ELM and SVM, which is a superiority of A-ELM model. Considering all performance metrics, A-ELM is what we call a fast model with high accuracy and good generalization performance.

4. CONCLUSIONS

In this paper, four groups of image data with different characteristics are used to explore the performance of three improved ELM algorithms (RBF K-ELM, mixed K-ELM and A-ELM) in extracting complex urban-rural land use, and they are compared with ELM and SVM in accuracy and efficiency.

The experimental results show that the performance of the three improved ELM models is better than that of the ELM and SVM. Among the three improved ELM methods, A-ELM has high classification accuracy and less classification time. It also has higher accuracy and good generalization performance.

For K-ELM, when the image size is larger, the classification time of RBF K-ELM and mixed K-ELM is much more than ELM, A-ELM and SVM. Moreover, the classification time of mixed K-ELM is more than three times that of RBF K-ELM, indicating that the classification performance of K-ELM models constructed by different kernel functions is quite different. Subsequent researches are planned to address the kernel function problem by analysing the characteristics and laws of different kernel functions and how to choose different combinations of kernel functions to improve the performance of the model.

ACKNOWLEDGEMENTS

This study was funded by the National Natural Science Foundation (NSFC) Project (No. 41771449), the DAAD Thematic Network Project (No. 57421148), the Key Laboratory of Surveying and Mapping Science and Geospatial Information Technology, Ministry of Natural Resources, Beijing, China (No. 2020-2-3), the Open Fund Project of State Key Laboratory of Geo-Information Engineering (No. SKLGIE2018-K-3-1) and Key Project of Shanghai Science and technology Innovation Action (No. 20dz1201202).

REFERENCES

LI Mingjie, WANG Mingchang, WANG Fengyan, et al. Urban land use classification of multi-features random forest[J]. *Science of Surveying and Mapping*, 2021: 1–8.

LI Xiaobin, JIANG Bitao, WANG Shengjin. A Review and Comparison of Optical Remote Sensing Scene Classification[J]. *Radio Engineering*, 2019, 49(4):265 – 271.

Zhou Ke, Yang Yongqing, Zhang Yanna, et al. Review of land use classification methods based on optical remote sensing images[J]. *Science Technology and Engineering*, 2021, 21(32): 13603-13613.

Chen Ni, Ying Feng, Wang Jing, et al. Research on Land Use Information Extraction based on U-Net[J]. *Remote Sensing Technology and Application*, 2021, 36(2): 285-292.

MA L, LIU Y, ZHANG X, et al. Deep Learning in Remote Sensing Applications: A Meta-Analysis and Review[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2019, 152: 166–177. DOI:10.1016/j.isprsjprs.2019.04.015.

YANG Mingli, FAN Yugang, LI Baoyun. Research on dimensionality reduction and classification of hyperspectral images based on LDA and ELM[J]. *Journal of Electronic Measurement and Instrumentation*, 2020, 34(05): 190–196.

YU Fenghua, FENG Shuai, ZHAO Yiran, et al. Inversion model of chlorophyll content in japonica rice canopy based on PSO-ELM and hyper-spectral remote sensing [J]. *Journal of South China Agricultural University*, 2020, 41(06): 59–66.

XIAO Dongsheng, LU Enming, LIU Fuzhen. Classification Method of Remote Sensing Imagery in Rotation Forest and Limit Learning[J]. *Remote Sensing Information*, 2019, 34(03): 93–98.

CAI Heng, CHU Heng, SHAN Deming. ELM-based urban road extraction from remote sensing images[J]. *Computer Engineering & Science*, 2020, 42(01): 125–130.

MOU Duoduo, LIU Lei. Comparative Study of ELM and SVM in Hyperspectral Image Supervision Classification[J]. *Remote Sensing Technology and Application*, 2019, 34(01): 115–124.

HUANG G-B, ZHU Q-Y, SIEW C-K. Extreme Learning Machine: Theory and Applications[J]. *Neurocomputing*, 2006, 70(1): 489–501. DOI: 10.1016/j.neucom.2005.12.126.

YAO W, ZENG Z, LIAN C, et al. A Kernel ELM Classifier for High-Resolution Remotely Sensed Imagery Based on Multiple Features[C]//*Advances in Neural Networks – ISNN 2014*. Cham: Springer International Publishing, 2014: 270–277. DOI:10.1007/978-3-319-12436-0_30.

HUANG G-B, ZHOU H, DING X, et al. Extreme Learning Machine for Regression and Multiclass Classification[J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 2012, 42(2): 513–529. DOI:10.1109/TSMCB.2011.2168604.

LIN Y, YU J, CAI J, et al. Spatio-Temporal Analysis of Wetland Changes Using a Kernel Extreme Learning Machine Approach[J]. *Remote Sensing*, 2018, 10(7): 1129. DOI:10.3390/rs10071129.

QIAN K. The Optimal Regularization and its Application in Extreme Learning Machine for Regression Analysis and Multiclass Classification[D]. University of Stuttgart, 2018.

LIN Yi, JI Haowei, SNEEUW N, et al. Optimization of ELM Classification Model for Remote Sensing Image Based on Artificial Fish-swarm Algorithm[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2017, 48(10): 156–164.

WANG Chuanli, ZHANG Xiaofang, TANG Nai, et al. Hyperspectral remote sensing images classification based on multi-kernel extreme learning machine[J]. *Journal of Central South University of Forestry & Technology*, 2018, 38(09): 20–25.

MARIÉTHOZ J, BENGIO S. A kernel trick for sequences applied to text-independent speaker verification systems[J]. *Pattern Recognition*, 2007, 40(8): 2315–2324.

SUGIYAMA M. *Introduction to Statistical Machine Learning*[J]. Morgan Kaufmann Publishers Inc., 2015.

Cai J., Grafarend E. and Schaffrin B. (2004): The A-optimal regularization parameter in uniform Tykhonov-Phillips regularization - Alpha-weighted BLE, IAG Symposia 127 “V Hotine-Marussi Symposium on Mathematical Geodesy”, Matera, Italy, 17-21 June 2002, edited by F. Sanso, Springer, 309-324

Cai J. (2004): *Statistical inference of the eigenspace components of a symmetric random deformation tensor*, Dissertation, Deutsche Geodätische Kommission (DGK) Reihe C, Heft Nr. 577, 138 Seiten, München, 2004

DWIVEDI R S, RAO B R M. The Selection of the Best Possible Landsat TM Band Combination for Delineating Salt-Affected Soils[J]. *International Journal of Remote Sensing*, 1992, 13(11): 2051–2058. DOI:10.1080/01431169208904252.

XU Hanqiu. A Study on Information Extraction of Water Body with the Modified Normalized Difference Water Index (MNDWI)[J]. *Journal of Remote Sensing*, 2005(05): 589–595.

LIN Y, SHEN M, LIU B, et al. Remote sensing classification method of wetland based on an improved SVM[J]. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2013, XL-7/W1: 179–183. DOI:10.5194/isprsarchives-XL-7-W1-179-2013.

YE Qiaolin, XU Dengping, ZHANG Dong. Remote sensing image classification based on deep learning features and support vector machine[J]. *Journal of Forestry Engineering*, 2019, 4(02): 119–125.

ZHOU Zhihua. *Machine Learning*[M]. Tsinghua University Press, 2016.

ZHANG Tinghui, YU Jie, YE Zhanglin, et al. Classification Model Research of Mixed Kernel Extreme Learning Machine Based on Particle Swarm Optimization[J]. *Journal of Geomatics Science and Technology*, 2019, 36(01): 56–61