American Journal of Engineering, Technology and Society
2015; 2(1): 1-8
Published online January 30, 2015 (http://www.opensciencesonline.com/journal/ajets)

Using artificial neural network for classification of high resolution remotely sensed images and assessment of its performance compared with statistical methods

Saeed Ojaghi[1, *], Hamid Ebadi[1], Farshid Farnood Ahmadi[2]

[1]Faculty of Geodesy and Geomatics Engineering, K. N. Toosi University of Technology, Tehran, Iran
[2]Department of Geomatics Engineering, Faculty of Civil Engineering, University of Tabriz, Tabriz, Iran

Email address

Saeed.ojaghi@chmail.ir (S. Ojaghi), ebadi@kntu.ac.ir (H. Ebadi), farnood@tabrizu.ac.ir (F. F. Ahmadi)

To cite this article

Saeed Ojaghi, Hamid Ebadi, Farshid Farnood Ahmadi. Using Artificial Neural Network for Classification of High Resolution Remotely Sensed Images and Assessment of Its Performance Compared with Statistical Methods. *American Journal of Engineering, Technology and Society*. Vol. 2, No. 1, 2015, pp. 1-8.

Abstract

Image classification is always one of the most important issues in remote sensing and obtained information from image classification is most widely used in this field and other applications like urban planning, natural resource management, agriculture and etc. The purpose of this study is to assess the performance of multi-layer perceptron neural networks to classify high resolution IKONOS image which covers mainly urban area of Shahriar city which is located in Tehran Province of Iran. The output of a neural network classifier has been compared with the results of support vector machine with the Gaussian kernel function and Maximum Likelihood Classification (MLC) algorithm which is most commonly used in statistical approach image classification. In the best situation, the classification outputs indicate that neural network algorithm, including 0.8775overall accuracy and 0.82 Kappa Coefficient is more accruable and reliable than both the support vector machine with 0.8557 and 0.8197 and maximum likelihood with 0.7836 and 0.7295 overall accuracy and Kappa Coefficient Respectively. Also results indicate that in these three methods, training data and model parameters play important roles in the classification accuracy.

Keywords

Artificial Neural Network, Support Vector Machine, Image Classification, Maximum Likelihood, Remote Sensing

1. Introduction

The objective of the classification process is to categorize all pixels in a digital image into one of the several land cover classes (e.g. Water, coniferous forest, deciduous forest, corn, wheat, etc.). The categorized data may later be used to produce thematic maps of the land cover contained in an image for application like urban planning, natural resource management and etc. [1],[2].

Image classification methods are categorized according to the type of learning approach (supervised or unsupervised classification), assumption on data distribution (parametric or nonparametric) and number of output class for each spatial unit (hard or soft classification)[2].

Traditional Statistical classification methods like maximum likelihood classification are parametric methods. The main deficiency of parametric methods is their dependence on the statistical distribution of the data and the effect of number of training samples on the estimation of this distribution. Also, they have low accuracy for image classification whereas Non-parametric classifiers like machine learning methods such as artificial neural network, (ANN), support vector machine (SVM) random forest (RF) do not rely on data belonging to any particular statistical distribution.

In the last decades, many studies have demonstrated that machine learning approaches such as ANN [3], SVM [4],[5] are more accurate than the traditional statistical classification approaches [6].

Some of them have investigated the use of neural-network (NN) classifier and have compared their performances with the ones of classical statistical methods[7],[8]and some have studied also comparison of machine learning algorithms for mapping land-cover modifications[9],[36].

The main purpose of this study is the classification of IKONOS image and preparation of land cover mapping using machine learning methods such as artificial neural network and support vector machines and traditional classification such as maximum likelihood and also comparing the performance and accuracy of these three methods using overall accuracy and Kappa coefficient parameters for land cover classification of the study area.

2. Neural Network

Neural network method is an algorithm in the area of machine learning and artificial intelligence, which is inspired by the human nervous system to analyze and model complex , nonlinear systems and parallel computations. Haykin [1] introduces an artificial neural network as a massively parallel learning machine, made up of simple processing units called neurons[2].

A neuron has a set of inputs $P_1, P_2, \ldots, P_R$ and each connection from the inputs to the processing element is affected by different connection strengths known as synaptic weights. A neuron is showed in mathematical and schematic terms by the following Eqs. (1)and Fig 1.

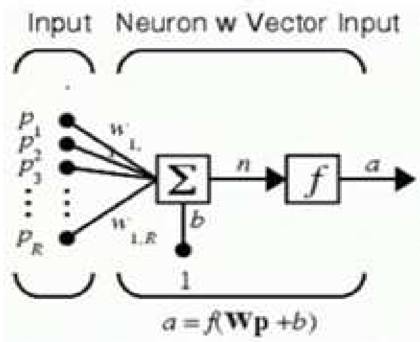

Fig. 1. *Structure of a neuron*

$$a = f(w_1p_1 + w_2p_2 + w_3p_3 + \cdots + w_Rp_R + b) \quad (1)$$

Where$P_1, P_2, \ldots, P_R$are the inputs $w_1, w_2, \ldots, w_R$and are the synaptic weights of the neuron L, also b is the bias, fis called activation function, and a is the output of the neuron.

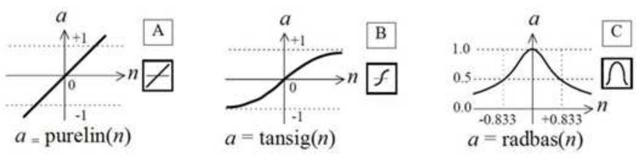

Fig. 2. *Different activation functions*

The output of the neuron is defined by activation function and terms of the linear combination of inputs have the role of argument in activation function. There are many different activation functions as shown in Fig 2.

The procedure in which neurons are arranged in a neural network is known as the network topology or architecture. There are many different types of neural network such as MLP, RBF ,and PNN.

One of the famous ANN architecture which is most widely used in remote sensing application is multilayer perceptron, or MLP that has been used in this study.

An MLP is a feed-forward neural network with one or more layers of neurons between the input and output layer called hidden layers[3].

2.1. Training Neural Network

The training process in an MLP network consists of determining the appropriate values for the weight vector *w* and bias vector *b*, using the training data and learning algorithms. There are several different training algorithms for feed forward networks. All of these algorithms use the gradient of the performance function to determine how to adjust the weight values to achieve minimal overall training error. The gradient is determined using a technique called back propagation [4], which involves performing computations backwards through the network.[5]

In this paper, Levenberg-Marquardt (LM) algorithm is used[32] to train network updating weight and bias values according to Levenberg-Marquardt optimization [6]and is often the fastest back propagation algorithm and is highly recommended as the first option for supervised algorithm, although requires more memory than the other algorithms.

3. Support Vector Machines

The Support vector machine method was introduced in 1995 by Vapnik [15]. Principles of support vector machine method are based on statistical learning theory and this method at first has been designed for binary-classification problems [16], [34]..

Consider the training set are linearly separable and consists of N vectors from the d dimensional feature space $x_i \in R^m$ (i=1,2,…,N) where x_i is the input data for the*i*th training sample and $y_i \in$ Error! Bookmark not defined. that $y_i = +1$ for classω_1 and y_i=-1 for class $\omega_2$toeach vector [7], the discriminating function in the form of a hyper plane as follows in Eq. (2):

$$g(x) = w^T x + b = 0 \quad (2)$$

Where w is an adjustable weight vector and b is a bias. Different hyper plans can be found to separate the two classes, but there is only one optimal hyper plane that well separates than other hyper planes. The optimal hyper plane should be as far away from the training data of both classes as possible. In other words, the optimal hyper plane maximizes the distance between the closest training data and separating hyper plane. It can be proved that this distance is as equal to 1/||w|| and the quantity 2/||w|| is called the margin (m). Thus, the optimal separating hyper plane is the one maximizing the margin [18].

Finding the optimal hyper plane by solving the following constrained optimization problem done by the method of Lagrange multipliers[18].(Eq. (3)).

$$\begin{aligned} &minimize: \ \tfrac{1}{2}\|w\|^2 \\ &Subjectto: y_i(w^T x + b) \geq 1 \quad i = 1,2,\ldots,N \end{aligned} \tag{3}$$

If the data are not linearly separable, a penalty value C that trades off between the number of misclassifications in the training data and the maximization the margin and positive slack variables($\xi_1, \ldots \xi_N$) Which measure the degree of misclassification the training data are introduced. The new optimal hyper plane formulation with incorporating slack variables and a penalty term as given by Eqs. (4):

$$\begin{cases} Minimize: \left\{\tfrac{1}{2}||W||^2 + C\sum_{i=1}^{N}\xi_i\right\} \\ y_i(w^T x + b) \geq 1 + \xi_i \quad , i = 1,2,..N \end{cases} \tag{4}$$

To generalize the aforesaid method to non-linear discernment functions, the input vector x is mapped into a higher dimension feature space using a proper nonlinear transformation function and then constructs the optimal separating hyper plane in that space. Suppose, function $\emptyset : X \rightarrow$ F maps the data into a higher dimension feature space, the decision function after using Lagrange multiplier methods is as follows Eq. (5).

$$g(x) = sign(\textstyle\sum_{xi\epsilon SV} \alpha i \, yi\langle\emptyset(xi), \emptyset(x)\rangle + b) \tag{5}$$

Where αi are the Lagrange multipliers and x_i are the support vectors that have $\alpha i \neq 0$.

The mapping function $\emptyset$ is not given explicitly in most cases. Instead, a kernel function

Gives the inner product value of x and x_i in the feature space [19] in the form of Eqs. (6).

$$\mathrm{K\ (xi,\ x)} = \emptyset\ (\mathrm{xi}).\ \emptyset\ (\mathrm{x}) \tag{6}$$

There are different kernel functions while generally used kernel functions are:

$$polynomial \ \ K(x_i, x_j) = \left(1 + x_i^T x_j\right)^P \tag{7}$$

$$sigmoidK(x_i, x_j) \ = \tanh(\beta_0 x_i^T x_j \ + \beta_1) \tag{8}$$

$$RadialbasisfunctionK = \exp\,(-\frac{\|x_i - x_j\|^2}{2\sigma^2}) \tag{9}$$

3.1. SVM Multi-Class Classification

SVM is basically binary classifier. To apply SVM for multi-class classification, the problem can be divided into sub-problems which are binary classification. The first approach is called “one against all” regarding N is the number of classes; N binary classifiers are trained to separate each class from all others. The class to be chosen is the one corresponding to the SVM with the highest discriminant function value [34].

The second approach is called “one against one, that builds $N(N-1)/2$ classifiers where each one classifies applied to each pair of classes. Each classifier gives a vote to the winning class.The class that has the maximum votes will be considered as label class of data [20].

4. Maximum Likelihood

One of the simplest but widely used statistical approaches and supervised classification methods for image classification is Maximum Likelihood Classification (MLC), derived from the Bayes theorem, assumes the image data for each class in each band is normally distributed. During the MLC procedure, a given pixel has a probability that belongs to a specific class. Hence, the probability of each pixel using of a discriminant function calculated and each pixel is assigned to the class that has the highest probability (that is, the maximum likelihood). The probability that a pixel with feature vector x belongs to classω_i, calculated by Bayes formula[8] in the form of Eq. (10).

$$P(\omega_i|x) = \frac{p(x|\omega_i)P(x)}{p(\omega_i)} \tag{10}$$

Where $p(x|\omega_i)$ is class-conditional probability density function called the likelihood function which describes the probability of finding a pixel from classω_i, at the position x, that estimated from training data. $p(\omega_i)$ is the a priori information, i.e., the probability that class ω_i occurs in the image and P(x) is the probability of finding a pixel from any class at location x[9] which can be written as Eq. (11).

$$P(x) = \textstyle\sum_{i=1}^{M} P(x|\omega_i)P(\omega_i) \tag{11}$$

Where M is the total number of classes. Regarding that P (x) is a common factor in estimating probability and is not important in the following, so it can be removed from. ML often assumes the distribution of the data within a given class ω_i (Probability distribution function) which obeys a multivariate Gaussian distribution; therefore density function is using the following Eq. (12).

$$g_i\ (\mathrm{x}) = \ln\ (P(\omega_i|x)) = \ln\ \mathrm{p}\ (\mathrm{x}|\omega_i\ ) + \ln\ \mathrm{p}\ (\omega_i\ ) \tag{12}$$

Where ln is the natural logarithm. The final form of the discriminant function for maximum likelihood classification, based upon the assumption of normal statistics, is given by Eq. (13)

$$g_i(\mathrm{x}) = \ln\ \mathrm{p}\ (\omega_i) - \tfrac{1}{2}\ln|\Sigma_i| - \tfrac{1}{2}(\mathrm{x}\ -\ m_i)^t \Sigma_i^{-1}(x - m_i) \tag{13}$$

Where m_i and Σ_i are the mean vector and Covariance matrix of each class that are obtained from the training dataset at the learning step.

5. Study Area and Data

The study area is located in the Shahriar city in West of Tehran in Iran with 51’4’30’ East Longitude and ’35 ‘39’3’

North latitude. The study area is composed of variable land cover types, mainly including agricultural, building, trees roads and bare soil. This area is covered by the georefrenced and radio metrically calibrated IKONOS image (RGB+NIR) with 4 meter spatial resolution.The Training data set is selected on image by visually identifying and manually digitizing multiple polygons for each class. Training areas are selected on the image for five different land use classes named agricultural, building, trees, roads and bare soil. (Fig 3)

Fig. 3. *The satellite image used in the study*

6. Results

6.1. ANN Results

In this research, neural network with a hidden layer failed to obtain the minimum required accuracy, so a neural network with two hidden layers is used. Also back-propagation learning and gradient descent algorithm are used to train artificial neural networks. Inputs are spectral band values for each pixel of training data, and outputs are corresponding to land cover classes contained in the image, so the network has four input neurons and five output neurons, which are respectively equal to the number of spectral bands and the number of different land covers contained in the image as shown in Fig4.

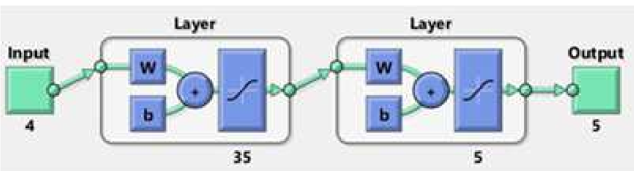

Fig. 4. *Number of input and output neurons and hidden layer*

The suitable number of neurons in the hidden layer is very important to achieve reasonable accuracy, since fewer neurons in the hidden layer can increase the training error and prevent convergence (something called under-fitting). On the other hand, extra neurons in the hidden layer may cause over-fitting which occurs when the neural network has so much information processing capacity and the limited piece of information contained in the training set is not enough to train all the neurons in the hidden layers[10].

As number of neurons in the hidden layer of the neural network depend on the number of inputs, outputs, the nature of the problem, network training method, there is no certain rule to exactly find the number of hidden layers and their neurons.

Nevertheless, cross-validation, early-stopping [11] , genetic algorithm [12], purning [26]are the common methods of determining whether the number of hidden neurons is optimal or not. There is a review of methods to choose the optimal number of hidden neurons in neural networks in [27], [33].

This paper presents an analytical assessment of optimal hidden layers and hidden neurons in neural network for image classification .To perform the analysis, at first many cases with various hidden neurons are examined to estimate overall accuracy which is used as the factor to measure accuracy of image classifications [29].So, the effect of the change in the number of neurons of the hidden layers on the classification accuracy in 100 epoch is evaluated. The results are shown in Fig 5.

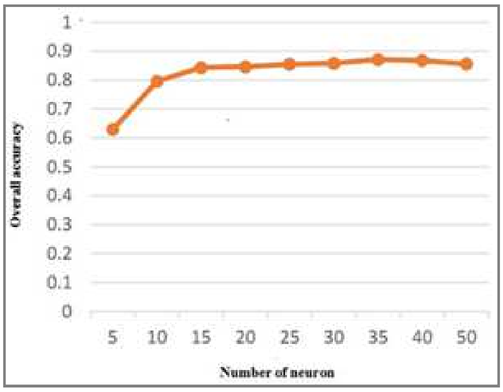

Fig. 5. *Effect of number of hidden neurons in overall accuracy*

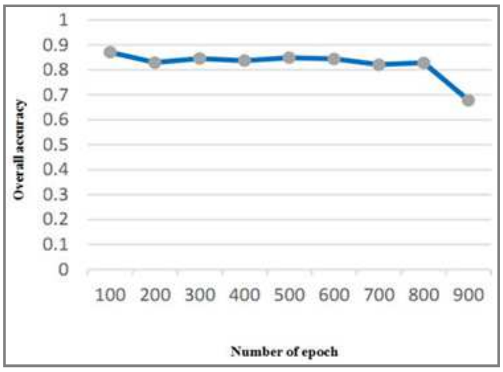

Fig. 6. *Effect of the number of epoch in overall accuracy*

Increasing the number of neurons to 35, results in increase in overall accuracy and after choosing more than 35, the classification accuracy decreases slightly. Considering that the highest classification accuracy is obtained from the 35 neurons case, at the second part of the accuracy assessment, the number of neurons is fixed at 35 and the number of epoch changes. According to the Fig6, the increase in the number of epoch to a certain one increases classification overall accuracy and for an interval after that no significant change in classification accuracy is observed even with regard to a slight fluctuation in this interval. It is not guaranteed that the increase in the number of epoch to more than that interval leads in improvement of overall accuracy. Even in 1000 epoch a considerable reduction in overall accuracy occurs.

The confusion matrix and classified image for the classifier neural network with 35 neurons according to the table 1 and Fig7were obtained.

Table 1. *The confusion matrix using 35 neurons*

Class accuracy	Road	Building	Tree	Agricultural	Barren soil	Class
0.82	1	33	0	1	166	Barren soil
0.78	0	0	54	190	0	agricultural
0.95	0	0	157	8	0	Tree
0.86	0	158	0	0	26	Building
0.98	202	7	1	0	1	Road

overall accuracy=0.8775
kappa coefficient =0.8469

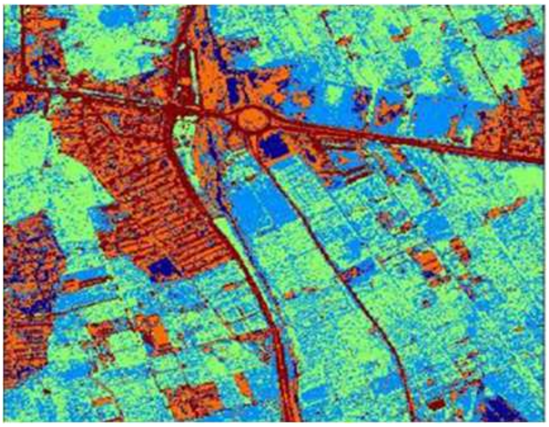

Fig. 7. *The image classified using ANN*

Also MSE (Mean Square Error) is used as an error measurement function for the classifier neural network with 35 neurons which is to be minimized. It can be observed in Fig. 8 rapid convergence in the first few steps and only slight improvements and oscillatory behavior for the following steps.

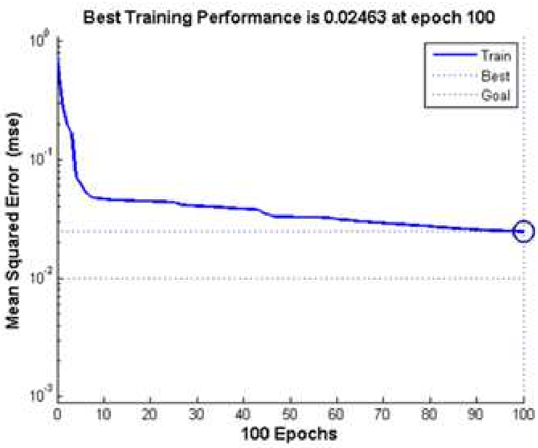

Fig. 8. *Neural network Training's performance using MSE*

6.2. SVM Result

Selecting a kernel function and the kernel-specified parameters and choosing the parameter values such as penalty term C, which controls the trade-off between maximizing the margin and minimizing the training error, are important in SVM application and depends on the type of the problem. Grid search and Genetic algorithm (GA) are typical procedures to improve image classification results in SVMs optimization.[30]In[13], a GA based feature selection and parameter optimization to improve classification accuracy of SVM is studied. Similar to neural network approach, there is no established heuristics and theoretical method for determining a kernel function and its parameters. Furthermore, there is no much guidance in the literature for selecting these SVM parameters and we are to some extent made to follow a trial-and-error approach [14],[35].

The most common kernel functions used in remote sensing image classification are Radial Basis and polynomial kernelFunction (RBF). Also has been reported that the "one-against one" strategy for SVM multi-class classification can be has upper classification accuracies [15]. In this research, polynomial function had low overall accuracy and performance; therefore we used a Gaussian radial basis kernel function including sigma parameter. And also one-against one" strategy for SVM multi-class classification is implemented. The proposed methodology is analytic parameter selection based on trial-and-error approach. The overall accuracy for different values of σ and C is obtained and the results are shown in table 2 and Fig9.

Table 2. *Overall accuracy for different value of C and σ*

C \ σ	0.001	10	25	50	100	250
10	40/61	72/92	76/18	80/13	82/70	84/28
100	49/70	77/76	82/70	85/17	84/68	81/84
1000	61/85	81/62	85/27	84/88	84/78	78/85
10000	70/35	85/57	84/38	81/48	79/64	81/12
100000	70/15	84/58	83/20	79/24	81/91	80/33

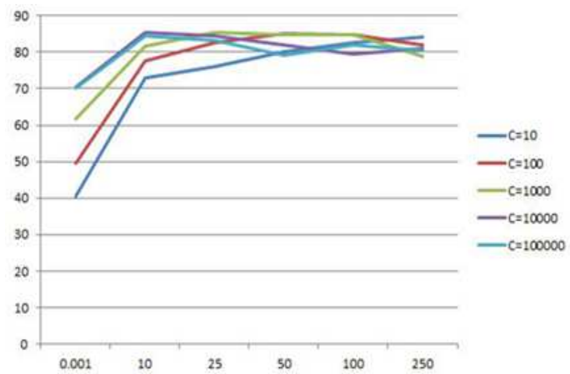

Fig. 9. *Overall accuracy for different value of C and σ*

Increasing the value of Gama for all values of penalty term C, the overall accuracy increases to a certain extent and then the slope of overall accuracy changes is approximately zero.

As a result, a radial basis kernel function with parameters

C=10000 and $\sigma = 10$ and 'one against one' strategy gives the highest overall classification accuracy equal to 0.85.57. Also confusion matrix and classified image are in Table 3 and Fig 10

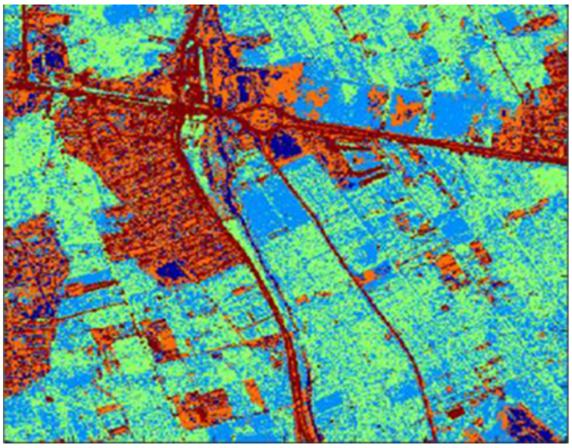

Fig. 10. *The classified image using SVM*

Table 3. *The Confusion matrix for SVM classification with C=10000 and*$\sigma = 10$

Class accuracy	Road	Building	Tree	Agricultural	Barren soil	Class
0.75	1	55	0	0	177	Barrensoil
0.79	0	0	49	191	0	Agricultural
0.95	0	0	162	8	0	Tree
0.84	3	137	0	0	23	Building
0.96	199	6	1	0	0	Road

Overall accuracy=0. 8557
Kappa coefficient =0. 8197

6.3. MLC Results

In this study, to implement the maximum likelihood classification, discrimination function in Eq. (13) is used[8]. With regard to the fact that there is no useful information about the P(ω_i), an equal prior probabilities is assumed for all land cover classes contained in the image, so P(ω_i) can be removed from Eq. (13) , since it is the same for all classes. Classification is performed as Eq. (14).

$$x \in \omega_i \text{ if } g_i(x) > g_j(x) \text{ for all } j \neq i \tag{14}$$

i.e., the pixel at x belongs to class ω_i if g_i (x) is the largest [7]. Confusion matrix and image classification for maximum likelihood classification are reported in Table 4 and Fig 11.

Table 4. *Confusion matrix for MLC*

Class accuracy	Road	Building	Tree	Agricultural	Barren soil	Class
0.62	2	109	0	0	179	Barren soil
0.78	0	1	53	182	0	Agricultural
0.90	1	0	159	15	0	Tree
0.70	14	87	0	2	20	Building
0.98	186	1	0	0	1	Road

Overall accuracy =0.7836
Kappa coefficient = 0.7295

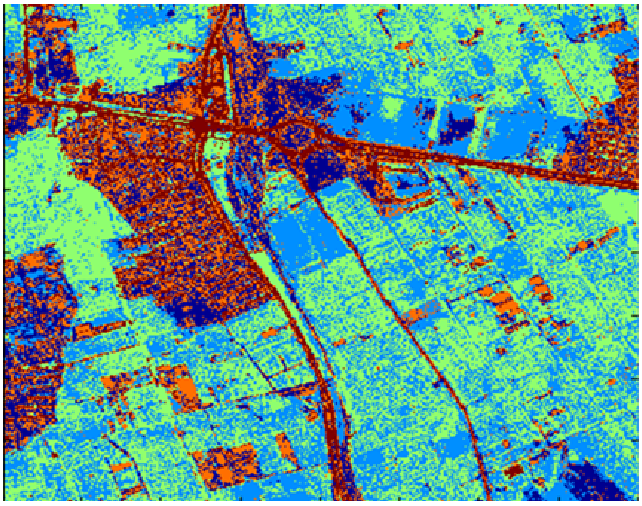

Fig. 11. *The classified image using MLC*

At the end, three algorithms are compared with each other regarding time duration for image classification. Fig.12demonstrates that ANN algorithm has the maximum time duration for image classification (11.58 seconds), because the learning network step needs more time to be trained, whereas SVM and MLC take less time for image classification, 2.3 and 9.43 seconds respectively.

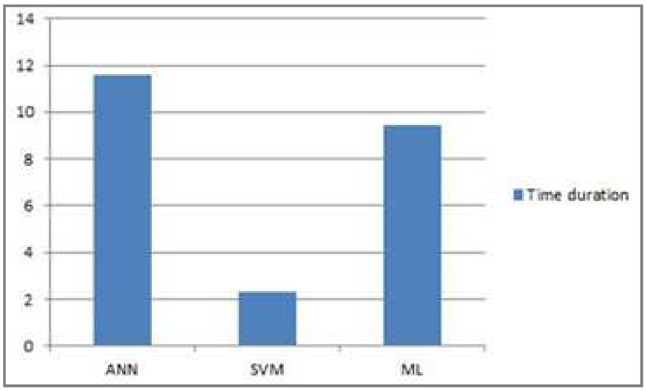

Fig. 12. *Time duration need for image classification*

But ANN provides more accuracy in comparison with other methods. Fig 13.

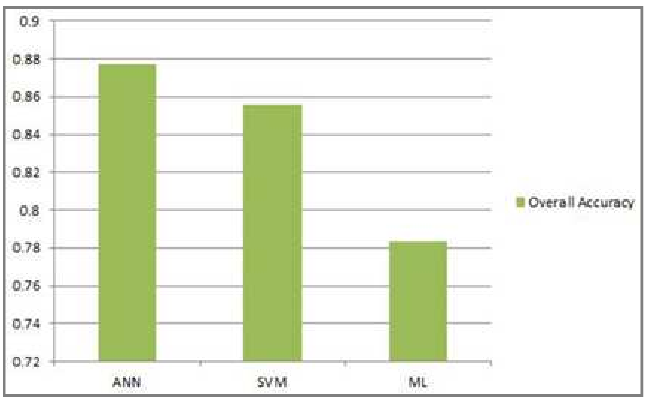

Fig. 13. *Overall classification accuracy for three methods*

7. Conclusion

The purpose of this paper was evaluating the performance of the neural network method to classify high spatial resolution image and compare with SVM and maximum likelihood statistical approaches. According to the results, the ANN classifier in the best case can achieve higher accuracy than both the ML or SVM, classifier, but it needs more time for image classification compared to the other methods. Classification accuracy of the ANN and SVM methods was close to each other; this indicates that the nonparametric approach like machine learning methods can provide more satisfactory accuracy than the parametric approach like maximum likelihood classifier. This is due to the fact that they are not sensitive to form of the underlying probability density function.

On the other hand, the ANN and SVM approaches require choice of a suitable architecture and parameter. In ANN method, it depends to inputs, outputs, nature of the problem, network training methods and in SVM approach there is no established heuristics and theoretical method to determine a kernel function and its parameters, which frequently leads to a trial-and-error approach in both methods.

Another advantage of machine learning is that ancillary information like GIS or texture information can be easily integrated with this algorithm to increase accuracy of image classification in comparison with maximum likelihood classification.

References

[1] T. Lillesand, "M. and Kiefer, R., W., 1994. Remote Sensing and Image Interpretation," *New York*.

[2] T. M. Lillesand, R. W. Kiefer, and J. W. Chipman, *Remote sensing and image interpretation*: John Wiley & Sons Ltd, 2004.M. Caetano, " IMAGE CLASSFICATION ", ed. ESA ADVANCES TRAINING COURSE LAND REMOTE SENSING, 2009.

[3] M. Caetano, " IMAGE CLASSFICATION ", ed. ESA ADVANCES TRAINING COURSE LAND REMOTE SENSING, 2009.

[4] J. F. Mas and J. J. Flores, "The application of artificial neural networks to the analysis of remotely sensed data," *International Journal of Remote Sensing,* vol. 29, pp. 617-663, 2008/02/01 2007.

[5] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: A review," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 66, pp. 247-259, 2011.

[6] M. Pal and P. Mather, "Support vector machines for classification in remote sensing," *International Journal of Remote Sensing,* vol. 26, pp. 1007-1011, 2005.

[7] V. F. Rodriguez-Galiano and M. Chica-Rivas, "Evaluation of different machine learning methods for land cover mapping of a Mediterranean area using multi-seasonal Landsat images and Digital Terrain Models," *International Journal of Digital Earth,* pp. 1-18, 2012.

[8] J. Benediktsson, P. H. Swain, and O. K. Ersoy, "Neural network approaches versus statistical methods in classification of multisource remote sensing data," *IEEE Transactions on geoscience and remote sensing,* vol. 28, pp. 540-552, 1990.

[9] L. Bruuone, C. Conese, F. Maselli, and F. Roli, "Multisource classification of complex rural areas by statistical and neural-network approaches," *Photogrammetric Engineering & Remote Sensing,* vol. 63, pp. 523-533, 1997.

[10] J. Rogan, J. Franklin, D. Stow, J. Miller, C. Woodcock, and D. Roberts, "Mapping land-cover modifications over large areas: A comparison of machine learning algorithms," *Remote Sensing of Environment,* vol. 112, pp. 2272-2283, 2008.

[11] S. Haykin, *Neural Networks. A Comprehensive Foundation*: Prentice Hall, 1999,.

[12] R. M. Golden, "BOOK REVIEWExploring the Diversity of Artificial Neural Network Architectures Simon Haykin. Neural Networks: A Comprehensive Foundation.," *journal of mathematical psychology,* vol. 41, , pp. 287-292, 1997.

[13] A. Khotanzad and J.-H. Lu, "Classification of invariant image representations using a neural network," *Acoustics, Speech and Signal Processing, IEEE Transactions on,* vol. 38, pp. 1028-1038, 1990.

[14] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," DTIC Document1985.

[15] M. B. Howard Demuth, *Neural Network Toolbox For Use with MATLAB*.

[16] V. N. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.

[17] A. Tzotsos and D. Argialas, "Support vector machine classification for object-based image analysis," in *Object-Based Image Analysis*, ed: Springer, 2008, pp. 663-677.

[18] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *Geoscience and Remote Sensing, IEEE Transactions on,* vol. 42, pp. 1778-1790, 2004.

[19] C. Huang, L. Davis, and J. Townshend, "An assessment of support vector machines for land cover classification," *International Journal of Remote Sensing,* vol. 23, pp. 725-749, 2002.

[20] Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *Neural Networks, IEEE Transactions on,* vol. 13, pp. 415-425, 2002.

[21] J. A. Richards, *Remote Sensing Digital Image Analysis,*. Berlin: Springer-Verlag, 1999.

[22] J. A. Richards and X. Jia, *Remote sensing digital image analysis* vol. 3: Springer, 1999.

[23] S. Lawrence and C. L. Giles, "Overfitting and neural networks: conjugate gradient and backpropagation," in *Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on*, 2000, pp. 114-119.

[24] Y. Liu, J. A. Starzyk, and Z. Zhu, "Optimizing number of hidden neurons in neural networks," in *Artificial Intelligence and Applications*, 2007, pp. 138-143.

[25] D. Whitley, T. Starkweather, and C. Bogart, "Genetic algorithms and neural networks: Optimizing connections and connectivity," *Parallel computing,* vol. 14, pp. 347-361, 1990.

[26] G. Castellano, A. M. Fanelli, and M. Pelillo, "An iterative pruning algorithm for feedforward neural networks," *Neural Networks, IEEE Transactions on,* vol. 8, pp. 519-531, 1997.

[27] K. G. Sheela and S. Deepa, "Review on methods to fix number of hidden neurons in neural networks," *Mathematical Problems in Engineering,* vol. 2013, 2013.

[28] S. Roweis, "Levenberg-marquardt optimization," *Notes, University Of Toronto,* 1996.

[29] http://en.wikipedia.org/wiki/Cohen's_kappa.

[30] P. Gaspar, J. Carbonell, and J. L. Oliveira, "On the parameter optimization of Support Vector Machines for binary classification," *Journal of integrative bioinformatics,* vol. 9, p. 201, 2012.

[31] C.-L. Huang and C.-J. Wang, "A GA-based feature selection and parameters optimizationfor support vector machines," *Expert Systems with applications,* vol. 31, pp. 231-240, 2006.

[32] M. I. Lourakis and A. A. Argyros, "Is Levenberg-Marquardt the most efficient optimization algorithm for implementing bundle adjustment?," in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, 2005, pp. 1526-1531.

[33] J. Stastny and V. Skorpil, "Genetic algorithm and neural network," in *Proceedings of the 7th WSEAS International Conference on Applied Informatics and Communications*, 2007, pp. 345-350.

[34] T. Hastie, R. Tibshirani, J. Friedman, T. Hastie, J. Friedman, and R. Tibshirani, *The elements of statistical learning* vol. 2: Springer, 2009.

[35] G.-B. Huang, P. Saratchandran, and N. Sundararajan, "An efficient sequential learning algorithm for growing and pruning RBF (GAP-RBF) networks," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on,* vol. 34, pp. 2284-2292, 2004.

[36] G. Giacinto, F. Roli, and L. Bruzzone, "Combination of neural and statistical algorithms for supervised classification of remote-sensing images," *Pattern Recognition Letters,* vol. 21, pp. 385-397, 2000.