WALAILAK JOURNAL
http://wjst.wu.ac.th Engineering and Physical Sciences

Land Use Classification using Support Vector Machine and Maximum Likelihood Algorithms by Landsat 5 TM Images

Abbas TAATI*, Fereydoon SARMADIAN, Amin MOUSAVI, Chamran Taghati Hossien POUR and Amir Hossein Esmaile SHAHIR

Department of Soil Science Engineering, University of Tehran, Karaj, Iran

(*Corresponding author's e-mail: taatyabbas@yahoo.com)

Received: 9 August 2014, Revised: 27 November 2014, Accepted: 28 December 2014

Abstract

Nowadays, remote sensing images have been identified and exploited as the latest information to study land cover and land uses. These digital images are of significant importance, since they can present timely information, and capable of providing land use maps. The aim of this study is to create land use classification using a support vector machine (SVM) and maximum likelihood classifier (MLC) in Qazvin, Iran, by TM images of the Landsat 5 satellite. In the pre-processing stage, the necessary corrections were applied to the images. In order to evaluate the accuracy of the 2 algorithms, the overall accuracy and kappa coefficient were used. The evaluation results verified that the SVM algorithm with an overall accuracy of 86.67 % and a kappa coefficient of 0.82 has a higher accuracy than the MLC algorithm in land use mapping. Therefore, this algorithm has been suggested to be applied as an optimal classifier for extraction of land use maps due to its higher accuracy and better consistency within the study area.

Keywords: Remote sensing, satellite, overall accuracy, kappa coefficient

Introduction

Remote sensing technologies are of the capability of monitoring the Earth's surface with different spatial, spectral, and temporal resolutions. These technologies have many advantages in terms of time and cost compared to land surveying. In view of all these features, remote sensing has an important data source to extract land use/land cover information. With the developments of remote sensing technology, remotely sensed data have been widely applied to classify the land cover, which provide the capability of updating maps more frequently and on a near real-time basis. It is worth mentioning that the Landsat satellite data are the most widely used data for land use/land cover mapping, because of their 35-year data records and the relatively high spatial resolution [1,2]. The classification of remotely sensed images is an important phase in the determination of land use/land cover information. In general, classification techniques fall into 2 broad categories: parametric and non-parametric classifiers. Parametric classifiers assume that the data for individual classes are distributed normally [3]. The most widely used parametric classifier is the MLC, which creates decision surfaces based on the mean and covariance of each class [4]. On the other hand, non-parametric techniques, such as the SVM classification, make no assumptions about the statistical nature of the data and are the most recent additions to the existing group of image classification techniques.

SVM is a non-parametric classifier, which includes a set of related learning algorithms that are used for classification and regression [5-7]. The theory of SVM was originally proposed by Vapnik and Chervonenkis [8] and later discussed in detail by Vapnik [9]. In addition, the SVM has been explored in remote sensing applications [10,11]. Huang *et al*. [12] implemented the SVM classification for a spatially degraded Landsat Thematic Mapper (TM) data. The SVM classification accuracy was superior to that obtained using a maximum likelihood algorithm and a decision tree algorithm. Yousefi *et al*. [13] investigated the nine supervised classification algorithms (SVM, Neural network, Mahalanobis distance, Maximum likelihood, Minimum distance, Spectral angle mapper, Spectral information divergence, parallelepiped and binary code) in land use mapping in Mazandaran province, Iran and used Landsat ETM$^+$ images. Their results confirmed that the SVM classification by a kappa coefficient of 0.9503 and an overall accuracy of 90.94 % is better than the other methods. The objective of the present study was to create land use classification using SVM and MLC algorithms in Qazvin, Iran, by TM images of the Landsat 5 satellite.

Materials and methods

Study area and satellite data

The study area is located in the Qazvin area, Iran, with an area of 16,618 ha. It is positioned between the latitudes of 36° 00' and 36° 11' N and between the longitudes of 50° 16' and 50° 20' E (**Figure 1**). It is worth noting that the climate of this area is semi-arid with an average annual rainfall of 258 mm, and a minimum and maximum relative humidity of 52 and 82 percent, respectively. The mean annual temperature is 14.1 °C. Two images were used in this study area, i.e. the Landsat 5 TM images for 2011. These images are located on the satellite path 165 and row 35, which have been acquired from https://www.usgs.gov.

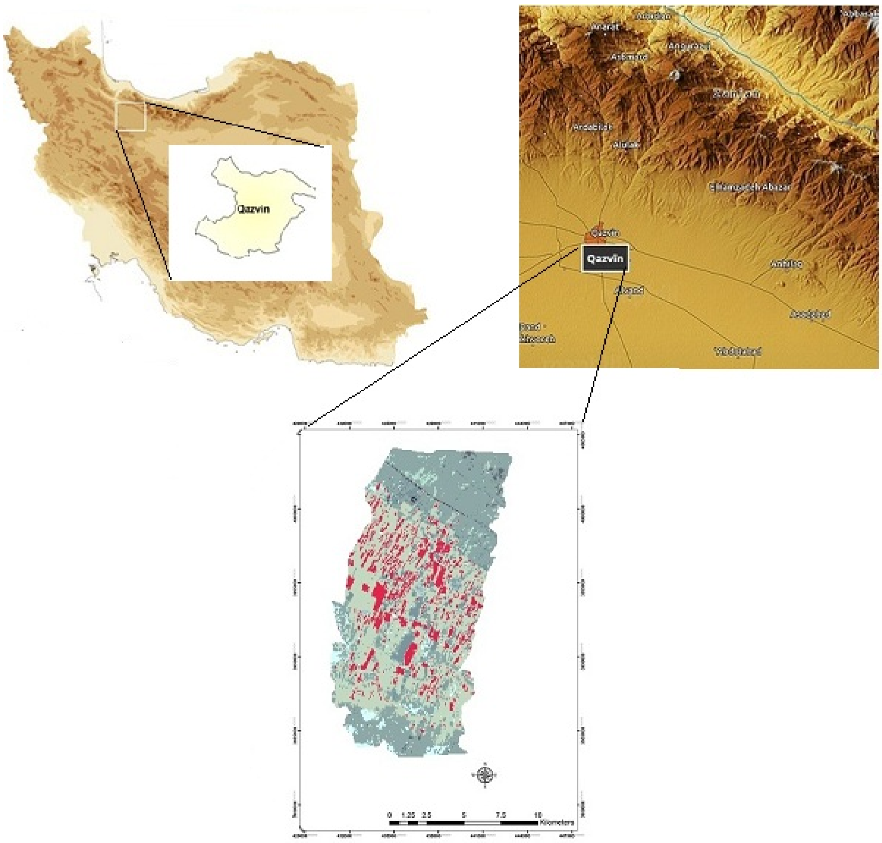


Figure 1 Location of the study area.

Radiometric and geometric corrections

Atmospheric correction is not required for the remote sensing applications, such as the classification of the same date images. The reason is that atmospheric correction of a single-date image would often mean to subtract a constant value from all pixels in the spectral band [3]. The geometric correction of TM images is performed using the image to image method by the ERDAS IMAGINE software. In addition, the nearest neighbor method is used for re-sampling of uncorrected pixel values. Finally, the root-mean-square error (RMSE) images were obtained as less than 0.4 pixels, which are acceptable [14].

Image classification

Image classification was carried out by using the MLC and SVM algorithms. In the following subsections, a brief explanation of the 2 algorithms is provided.

Maximum likelihood classification

A maximum likelihood classification algorithm is one of the well-known parametric classifiers used for supervised classification. According to Erdas [15] the algorithm for computing the weighted distance or likelihood **D** of an unknown measurement vector **X** belonging to one of the known classes $\mathbf{M_c}$ is based on the Bayesian equation.

$$D = \ln(ac) - [0.5 \ln(|Covc|)] - [0.5(X - Mc)T(Covc - 1)(X - Mc)] \quad (1)$$

c a particular class, **ac** percent probability that any candidate pixel being a member of class c, $\mathbf{cov_c}$ the covariance matrix of the pixels in the sample of class c, **|** $\mathbf{cov_c}$**|** determinant of cov_c , $\mathbf{cov_c}$**-1** inverse of cov_c, **In** natural logarithm function, **T** transposition function. The unknown measurement vector is assigned to the class in which it has the highest probability of belonging. The advantage of the MLC as a parametric classifier is that it takes into account the variance-covariance within the class distributions and for normally distributed data, the MLC performs better than the other known parametric classifiers [15]. However, for data with a non-normal distribution, the results may be unsatisfactory.

Support vector machine

SVM has been successfully used for data classification in the remote sensing arena [12]. A SVM aims to fit an optimal separating hyper plane or set of hyper planes in a high or infinite dimensional space, to locate the optimal boundaries between classes. SVM is a set of related learning algorithms used for classification and regression. The SVM is also a non-parametric classifier. The theory of the SVM was originally proposed by Vapnik and Chervonenkis [8] and later discussed in detail by Vapnik [9]. The success of the SVM depends on how well the process is trained. The easiest way to train the SVM is by using linearly separable classes. According to Osuna *et al*. [16] if the training data with **k** number of samples is represented as $\{X_i,Y_i\}$, i =1,2,...,k where X € Rn is an n-dimensional space and y € {−1, +1} is a class label, then these classes are considered linearly separable if there exists a vector **W** perpendicular to the linear hyper-plane (which determines the direction of the discriminating plane) and **a** scalar **b** showing the offset of the discriminating hyper-plane from the origin. For the 2 classes, i.e. class 1 represented as −1 and class 2 represented as +1, 2 hyper-planes can be used to discriminate the data points in the respective classes. These are expressed as;

$$\mathrm{WXi} + \mathrm{b} \geq +1 \text{ for all } \mathrm{y} = +1, \text{i.e. a member of class 1} \quad (2)$$

$$\mathrm{WXi} + \mathrm{b} \leq -1 \text{ for all } \mathrm{y} = -1, \text{i.e. a member of class 2} \quad (3)$$

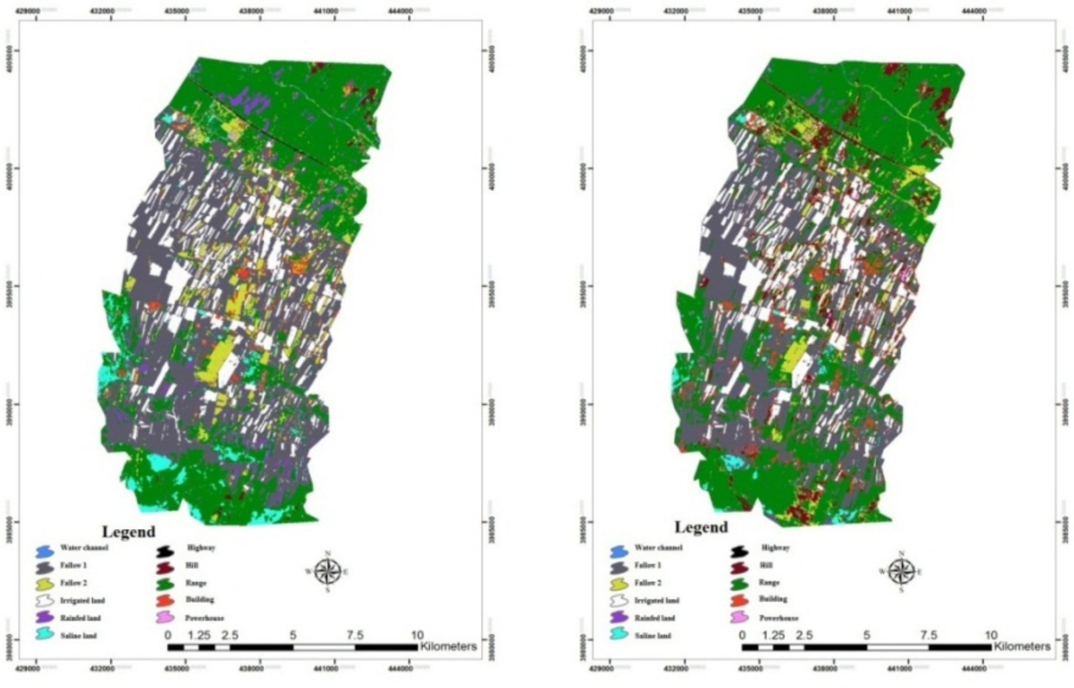
In some cases, the classes might not be linearly separable [17]. Kernel representations offer a solution in locating complex decision boundaries between classes. The SVM classifier provides four types of kernels: linear, polynomial, radial basis function (RBF), and sigmoid.

Accuracy assessment

The final stage of the image classification process usually involves an accuracy assessment step [18]. Accuracy estimation is in fact the quantification of mapping with the aid of remote sensing data to group classification conditions, which is useful in evaluation of classification algorithms and also in determination of the error level that might be contributed by the image. The accuracy of each classification is expressed in the form of an error matrix (also known as a confusion matrix) [19]. In this study, in order to evaluate the classification results with the ground truth, a total of 60 samples (collected by GPS) were randomly chosen for the accuracy assessment. Many methods for accuracy assessment have been discussed in remote sensing. In the current paper, 2 accuracy measures have been tested, namely the overall accuracy and the kappa coefficient [20,21].

Results and discussion

After processing the satellite images, the land use map of the area was analyzed using the SVM and MLC algorithms and land use classes were extracted: irrigated agricultural lands (including the irrigated land, fallow 1, and fallow 2), highway, hill, rainfed land, water channel, range, building, powerhouse, and saline land. **Figure 2** shows the land use classification map of the study area by 2 algorithms of SVM and MLC. Moreover, the results of the accuracy assessment are shown in **Table 1**. The overall accuracy for the MLC and SVM was 80 and 86.67 %, respectively. Added to them, the kappa coefficient for the MLC and SVM was 0.72 and 0.82, respectively. The area of each land use class is displayed in **Figure 3** based on the SVM algorithm.



(a) (b)

Figure 2 Land use classification map of the study area by (a) the SVM and (b) the MLC algorithms.

Table 1 Classification accuracy assessment report.

Class	SVM		MLC	
	Producers Accuracy (%)	**Users Accuracy (%)**	**Producers Accuracy (%)**	**Users Accuracy (%)**
Water channel	100	100	100	100
Fallow 1	100	81.82	100	100
Fallow 2	71.43	83.3	50	100
Irrigated and	33.33	100	100	100
Rainfed land	100	100	100	50
Saline land	100	100	66.67	100
Highway	0	0	0	0
Range	100	75.81	100	61.54
Hill	100	100	100	100
Building	0	0	0	0
Powerhouse	0	0	0	0
Overall accuracy (%)	86.67		80	
Kappa coefficient	0.82		0.72	

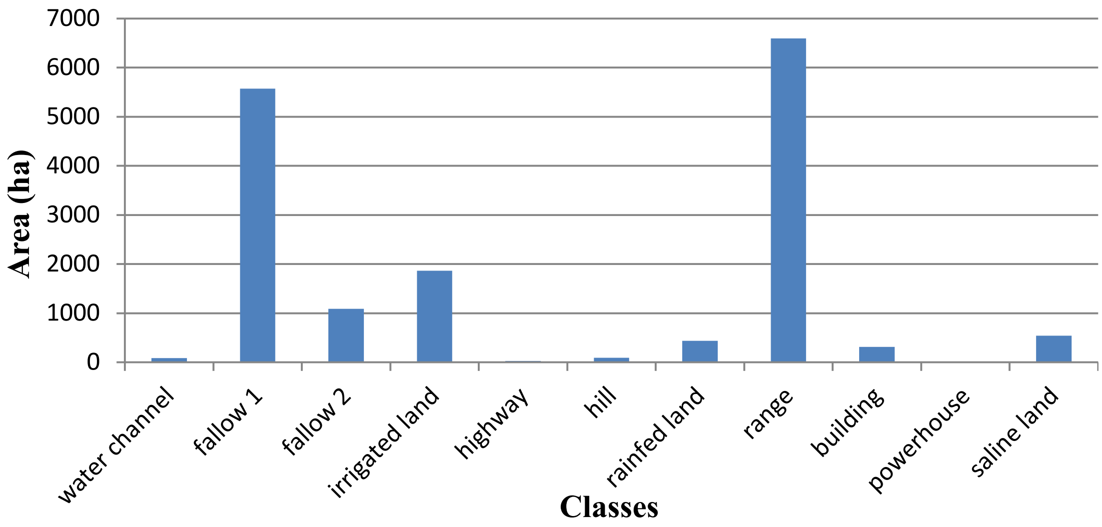


Figure 3 Area of each land use classes based on the SVM algorithm.

According to the above figure, the total study area includes 16,618 ha and the irrigated agricultural lands, including the irrigated land, fallow 1, and fallow 2 had the highest level of the area. In contrast, the highway had the lowest level of the area. Furthermore, Otukei and Blaschke [22] and Huang *et al*. [12] evaluated various algorithms for classification in land use mapping, and concluded that the SVM algorithm in comparison with the MLC algorithms and decision trees has a higher accuracy in the preparation of land use maps. Deilmai *et al*. [23] studied the comparison of 2 classification methods (MLC and SVM) to extract land use and land cover in Johor Malaysia. The results showed that the SVM classification based on kappa coefficient 0.86 was the most accurate method.

Conclusions

Today, applying land use maps is one of the critical issues in generating information for macro and micro planning. Using satellite images has been recommended as an appropriate strategy for preparing land use maps. By using the images of different backgrounds, the capabilities and limitations of these data can be recognized. The purpose of this study has been to land use classification using SVM and MLC algorithms in Qazvin, Iran, by TM images of the Landsat 5 satellite. After the necessary corrections and pre-processing of images, 2 different algorithms were applied in order to classify the images. The evaluation results demonstrated that the SVM algorithm with an overall accuracy of 86.67 % and a kappa coefficient of 0.82 has a higher accuracy in comparison with the MLC algorithm in land use mapping. This algorithm has been suggested as an optimal classifier for the extraction of land use maps because of its higher accuracy and better consistency with the study area. This study confirmed that the proper implementation of the SVM algorithm and the proposed land use map can facilitate the management in line with sustainable development.

References

[1] J Knorn, A Rabe, V Radeloff, T Kuemmerle, J Kozak and P Hostert. Land cover mapping of large areas using chain classification of neighboring Landsat satellite images. *Remote Sens. Environ.* 2009; **113**, 957-64.

[2] R Borges, A Hernandez and L Nykjaer. Analysis of sea surface temperature time series of the south-eastern North Atlantic. *Int. J. Remote Sens.* 2004; **5**, 869-91.

[3] P Srivastava, D Han, M Rico-Ramirez, M Bray and T Islam. Selection of classification techniques for land use/land cover change investigation. *Adv. Space Res.* 2012; **50**, 1250-65.

[4] J Richards and X Jia. *Remote Sensing Digital Image Analysis*. 4th ed. Heidelberg, Springer, Germany, 2006, p. 359-88.

[5] M Bray and D Han. Identification of support vector machines for runoff modelling. *J. Hydroinform*. 2004; **6**, 265-80.

[6] A Moghaddamnia, M Gosheh, M Nuraie, M Mansuri, D Han and E Schmitter. Performance evaluation of LLR, SVM,CGNN and BFGSNN models to evaporation estimation estimation. *Water Geosci.* 2010; **9**, 108-13.

[7] D Han, L Chan and N Zhu. Flood forecasting using support vector machines. *J. Hydroinform*. 2007; **9**, 267-76.

[8] N Vapnik and A Chervonenkis. On the uniform convergence of the relative frequencies of events to their probabilities. *Theor. Probab. Appl*. 1971; **16**, 264-80.

[9] N Vapnik. An overview of statistical learning theory. *IEEE Trans. Neural Network*. 1999; **10**, 988-99.

[10] B Dixon and N Candade. Multispectral land use classification using neural networks and support vector machines. *Int. J. Remote Sens.* 2008; **29**, 1185-206.

[11] G Foody and A Mathur. Toward intelligent training of supervised image classifications: Directing training data acquisition for SVM classification. *Remote Sens. Environ.* 2004; **93**, 107-17.

[12] C Huang, LS Davis and JRG Townshend. An assessment of support vector machines for land cover classification. *Int. J. Remote Sens.* 2002; **23**, 725-49.

[13] S Yousefi, M Tazeh, S Mirzaee, HR Moradi and SH Tavangar. Comparison of different classification algorithms in satellite imagery to produce land use maps (Case study: Noor city). *J. Appl. RS GIS Tech. Nat. Res. Sci.* 2011; **2**, 15-25.
[14] CJ Tucker, DM Grant and JD Dykstra. NASA's global orthorectified landsat data set. *Photogramm. Eng. Rem. Sens*. 2004; **70**, 313-22.
[15] Erdas Inc. *Erdas Field Guide*. Erdas Inc., Atlanta, Georgia, 1999.
[16] EE Osuna and R Freud. *Support Vector Machines: Training and Applications*. Massachusetts Institute of Technology Cambridge, USA, 1977, p. 1-144.
[17] A Erener. Classification method, spectral diversity, band combination and accuracy assessment evaluation for urban feature detection. *Int. J. Appl. Earth Observ. Geoinform.* 2013; **21**, 397-408.
[18] BP Manisha, G Chitra and N Umrikar. Image classification tool for land use/ land cover analysis: A comparative study of maximum likelihood and minimum distance method. *Int. J. Geo. Earth. Environ*. 2012; **6**, 189-96.
[19] RG Congalton. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.*1991; **37**, 35-46.
[20] D Lu, P Mausel, E Brondizio and E Moran. Change detection techniques. *Int. J. Remote Sens.* 2004; **25**, 2365-407.
[21] K Sundarakumar, M Harika, SK Aspita Begum, S Yamini and K Balakrishna. Land use and land cover change detection and urban sprawl analysis of Vijayawada city using multitemporal landsat data. *Int. J. Eng. Sci. Tech.* 2012; **4**, 170-8.
[22] JR Otukei and T Blaschke. Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *Int. J. Appl. Earth Observ. Geoinform.* 2010; **12**, 527-31.
[23] B Deilmai, B Ahmad and H Zabihi. Comparison of two classification methods (MLC and SVM) to extract land use and land cover in Johor Malaysia. *IOP Conf. Ser. Earth Environ. Sci*. 2014; **20**, 1-6.