Performance Analysis of Maximum Likelihood and Artificial Neural Network Classifiers for Training Sets with Mixed Pixels

Article *in* GIScience & Remote Sensing · July 2008

DOI: 10.2747/1548-1603.45.3.330

2 authors:

Taskin Kavzoglu
Gebze Technical University

Selçuk Reis
Aksaray Üniversitesi

Performance Analysis of Maximum Likelihood and Artificial Neural Network Classifiers for Training Sets with Mixed Pixels

Taskin Kavzoglu[1]
Department of Geodetic and Photogrammetric Engineering, Gebze Institute of Technology, 41400, Gebze-Kocaeli, Turkey

Selcuk Reis
Department of Geodesy and Photogrammetry, Aksaray University, 68100, Aksaray, Turkey

Abstract: This study evaluates the performance of an artificial neural network, specifically a multilayer perceptron, and a maximum likelihood algorithm to classify multitemporal Landsat ETM+ remote sensor data. The study area in Turkey is a mountainous region that contains many small scattered fields, usually 5–10 pixels in size. The classifiers were employed to identify eight land cover/use features covering the bulk of the study area using the same training and test datasets in order to avoid any difference resulting from sampling variations. Results show that the neural network approach performed better in extracting land cover information from multispectral and multitemporal images with training data sets including a large amount of mixed and atypical pixels. The maximum likelihood classifier was found to be ineffective, particularly in classifying spectrally similar categories and classes having subclasses.

INTRODUCTION

The last several decades have witnessed heavy industrialization, urbanization, and environmental degradation throughout the world. With the utilization of accurate, up-to-date, and repetitive remote sensor data at various scales, these issues can be investigated effectively and proper monitoring and planning activities can be carried out. Satellite images and extracted thematic maps provide top-level information for the inventory, monitoring, and management of natural resources. Given the diversity and heterogeneity of the natural and human-altered landscape, it is obvious that the time-honored and laborious method of ground inventory is inappropriate for mapping land use and land cover over large areas (Civco, 1993). Therefore, the use of remotely sensed images is essential particularly for regional- or global-scale studies.

[1]Email: kavzoglu@gyte.edu.tr

GIScience & Remote Sensing, 2008, **45**, No. 3, p. 1–13. DOI: 10.2747/1548-1603.45.3.1

One of the most significant recent developments in the field of land cover classification using remotely sensed data has been the introduction of artificial neural network models. They can be thought of as forms of models imitating complicated brain processing in a very simple way. They have been recently employed in a wide range of classification and pattern recognition problems ranging from signal recognition to image compression. In the remote sensing arena, they have been applied to many applications, but the most popular application in remote sensing is the classification of land cover information (Paola and Schowengerdt, 1995; Gopal and Woodcock, 1996; Erbek et al., 2004).

While the information obtained from the training data varies from one algorithm to another in the statistical supervised classification methods, supervised neural network models do not use any statistical information to identify unknown pixels present in an image. Instead, they use all patterns present in the training data. This is the characteristic that makes supervised neural network models more powerful than their statistical counterparts. However, the effect of any incorrect definition of training pixels is very important when using neural networks. In other words, they are more open to influence as they consider every single pixel as a pattern in the learning process. Therefore, this characteristic of artificial neural networks requires special attention when preparing training data sets.

In this study, the performances of two classification methods—maximum likelihood and artificial neural networks—were tested for the classification of eight land cover/use classes. The study area selected is the province of Trabzon in Turkey, which can be described as rugged terrain featuring small fields or parcels, usually less than 10 pixels. In addition to their small acreages, agricultural lands are generally surrounded by forests consisting of hazelnut or green tea trees. This particular characteristic certainly complicates the use of intermediate- and low-resolution satellite images for the study area, as the number of mixed pixels increases with pixel size considering the land cover types and land use activities in the region.

TEST SITE AND DATA

This study was undertaken in Trabzon Province, situated between 39º15'–40º15' E. Long and 41º8'–40º 30' N. Lat. in the Black Sea region of Turkey (Fig. 1). The total area of the province is about 466,000 ha˒ with a population of 975,137 estimated in 2000 (www.tuik.gov.tr). The main commercial agricultural products are hazelnuts and green tea. Proximity to the sea results in a temperate climate where summers are generally warm and winters mild, with an overall annual average temperature of 14.5° C and annual average precipitation of 838.4 mm (Reis and Yomralioglu, 2006).

Two Landsat ETM+ images acquired on 2.6.2001 and 11.10.2002 were used to classify eight land cover/use classes, which are green tea, hazelnut tree, deciduous forest, coniferous forest, pasture, rocky land, agriculture, and urban. Image bands except for thermal bands were stacked to create a new multi-layer image including the study area. The images were registered to the UTM coordinate system using topographic maps scaled to 1:25,000 produced by the General Command of Mapping. A first-order polynomial transformation was applied and RMSE values estimated for image transformations were about 0.7 pixel. After the registration process, all images

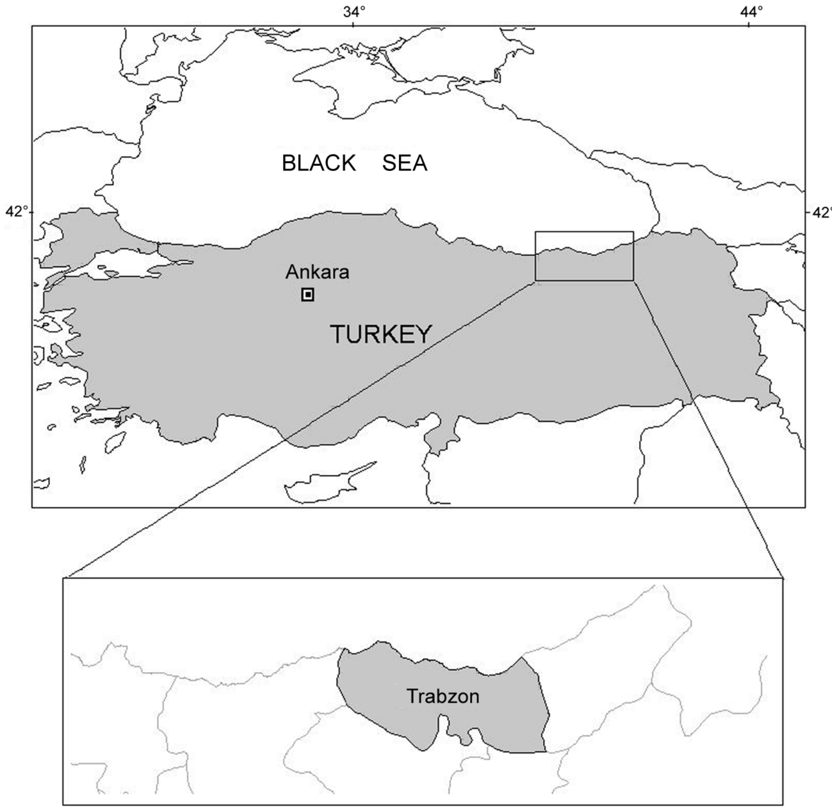

Fig. 1. Location of the study area, province of Trabzon in Turkey.

were resampled at a spatial resolution of 30 m, and 4,358 pixel × 2,529 pixel portions of the images covering the study area were extracted for subsequent analysis.

In the Eastern Black Sea Region, the principal challenges in generating land cover types are rough topographic structure, and small and scattered agricultural lands. Therefore, it was difficult to collect a large number of "pure" training pixels to delineate the characteristics of the land cover/use classes. As pointed out by Lillesand and Kiefer (2000), obtaining an accurate classification depends on determining training areas homogeneously on land and defining them in appropriate sizes.

In the preparation of ground reference data, it was difficult to collect samples for some particular classes due to the small parcel sizes and low spatial resolution for satellite images. In agricultural fields except for hazelnut and green tea areas, cultivation of maize, tobacco, beans, and potatoes are restricted due to the topographic structure of the region. These crops are only grown for domestic consumption. Because these small cultivated areas were mostly scattered around hazelnut or green tea fields and situated near the urban areas, it was difficult to acquire training data. In order to resolve this difficulty, all classes representing agricultural crops were grouped under a unique agricultural crop class. Also, it was difficult to determine homogeneous hazelnut and green tea areas due to the fact that fields are relatively small and include a variety of plant species. For example, it is common to see deciduous tree types inside and around the hazelnut fields. According to Akyol and Sesli (2000), 70–80% of the parcels in Trabzon have an area of lower than 0.5 ha, corresponding to approximately six pixels in Landsat ETM+ imagery. It is obvious that collecting training data in hazelnut and green tea areas in the region is challenging. On the other hand, forested lands were divided into two groups; deciduous and coniferous forests. Because the deciduous tree types (brushwood, hornbeam, chestnut, etc.) have a complex distribution in areas near the shore line, some difficulties were also encountered in determining the training areas for this particular class. However, collecting training data was relatively easy in highlands because deciduous species (e.g. hornbeam, oak) mostly cover large areas in those regions.

After the formation of ground reference data, the data sets required for training and testing were created. For each of the eight land cover classes, 250 pixels were randomly selected for training the network, 50 pixels were selected as a validation data set to control the training process, and 200 pixels were randomly selected for testing the performance of the trained networks.

CLASSIFICATION METHODS

Classification is a process of identification that is addressed and used in all scientific disciplines as a way of comprehending and ordering a mass of data. Classification of land cover features from remotely sensed image data has been one of the main applications in the remote sensing field. It is an important and difficult task, inasmuch as satellite images are highly dimensional and complex in nature. As the number of categories and the amount of data involved increase, so does the complexity of the classification problem; then it becomes more difficult to determine the characteristics of the categories and allocate a pixel to one of the categories. Statistical classification methods have been mostly used in the classification of remotely sensed images. These methods assume that similar cover types have similar spectral properties so they can be discriminated from each other using some estimated statistical measures. Although this assumption is generally valid, it is not true for classes composed of several subclasses. For example, a class labeled forest may include deciduous and coniferous forest types that have significantly different spectral behavior. Due to the drawbacks of frequency distribution of the input data and data types, non-statistical (non-parametric) techniques have been introduced. Artificial neural networks that have recently become popular in the scientific community are of this kind. It should be pointed out that the accuracy of a supervised classification method depends on the representativeness (i.e., objectiveness) and size of the samples selected, and the degree of departure from the assumption upon which the classification technique is based (Mather, 1999).

Maximum Likelihood Classifier

The Maximum Likelihood (ML) classifier is one of the statistical classifiers that rely on multivariate normal (i.e., Gaussian) distribution of the data in each class. Swain and Davis (1978) suggest that if the assumption of a normal distribution for each class is correct, then the classification has a minimum overall probability of error and the maximum likelihood classifier is the optimal choice. It is based on the idea that the geometrical shape in feature space of the pattern of pixels belonging to a given class can be represented by an ellipsoid. The locations, shapes, and sizes of these ellipsoids are derived from the mean vectors and variance-covariance matrices of the individual classes. A series of concentric ellipses centered on the mean vector of a given class is used to evaluate pixels to be classified in terms of likelihood probabilities. These concentric ellipses represent the probability of membership of a class with contours in such a way that the probability declines away from the mean center. Distance is not the only criterion for deciding whether a pixel belongs to one class or another. The shape of the probability contours depends on relative dimensions of the axes of the ellipse as well as on its orientation. In essence, the maximum likelihood function describes ellipsoidal "equi-probability contours," which can be viewed as decision boundaries (Tso and Mather, 2001). The resulting classification might be expected to be more accurate than other statistical ones because the training sample data are being used to provide estimates of the shapes of distribution of membership of each class in the n-dimensional feature space as well as of the location of the center point of each class. However, it should be noted that the reliability of the results declines when the distribution of the data departs from the normality. Also, the violation of multivariate normality can result in rejection of the null hypothesis by significance tests, leading generally to the adoption of excessively complex models. In order to hold the normality condition, modified versions of maximum likelihood estimators for various distribution types are suggested in the literature (e.g., Vaughan and Tiku, 2000). Another way is to use a transformation model (such as the Box-Cox transformation, the logarithm of the likelihood function) to convert the data to a normal distribution (URL-1, 2008).

Artificial Neural Networks

Recently, artificial neural networks (ANNs), theoretically more sophisticated and robust methods of image classification, have been employed in classification applications. ANNs are heuristic algorithms, in that they can learn from experience via samples and can subsequently be applied to recognize new data. These systems are intended, in an extremely simple way, to imitate the behavior of the network of neurons in the human brain. ANNs can also provide superior classification results compared to the conventional methods because they require less training data to delineate the characteristics of the classes as reported by Blamire (1996) and Foody (1995). Despite their significant advantages, they have the main drawback of having a poorly interpretable nature. Therefore, they are often called black-box methods.

The basic element of ANNs is the processing node that corresponds to the neuron of the human brain. Each processing node receives and sums a set of input values, and passes this sum through an activation function providing the output value of the node. The structure of a multi-layer perceptron includes one input layer, at least one hidden layer, and one output layer (Fig. 2). The input layer representing input features such as a spectral band introduces the distribution of the data for each class to the network. Hidden layers are used for computations and the values associated with each node are estimated from the sum of the multiplications between input node values and weights

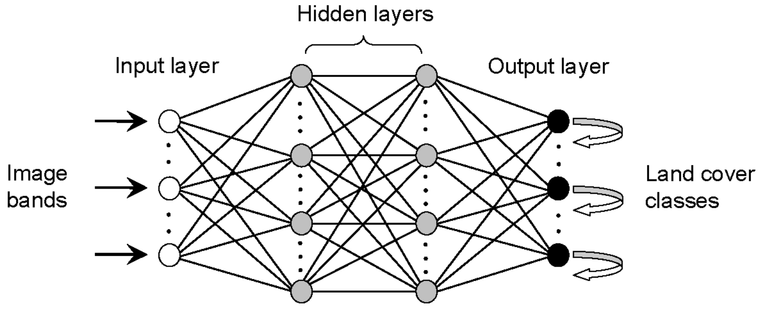

Fig. 2. A simple four-layer feed-forward neural network architecture.

of the links connected to that node. The output layer is the final processing layer that has a set of codes to represent the classes to be recognized. Processing nodes make up a set of fully interconnected layers, except that there are no interconnections between nodes within the same layer in the multilayer perceptron. All inter-node connections have associated weights, which are usually initially randomized. When a value passes through an inter-connection, it is multiplied by the weight associated with that inter-connection.

In order to learn the characteristics of the data sets, a learning algorithm is required. The learning algorithm defines how network weights are adjusted between successive training cycles or epochs. Although a number of learning strategies have been developed, the most popular is the back propagation learning algorithm, also called the generalized delta rule, introduced by Rumelhart et al. (1986). The method is based on iterative gradient descent training. The process is repeated until the error is reduced to an acceptable level in terms of a predetermined number of times, or a specified threshold value set for the error estimated for the training data set, or a combination of training and validation data sets.

Two crucial stages in the application of neural networks are the design of a network and setting the parameters for the learning algorithm selected by the analyst. The specification of the number and size of the hidden layer(s) is critical for the network's capability to learn the characteristics of the training data sets and recognize the pixels that are new to the network. It should be noted that the number of nodes in the hidden layers defines the complexity and the power of the neural network model to describe underlying relationships and structures inherent in a data set (Kavzoglu and Mather, 2003). The problem of determining the optimum number of hidden layer nodes is mainly dependent on the numbers of input and output units, the number of training pixels, the complexity of the classification problem, and the level of noise in the data. Setting the network training parameters (i.e., initial weights, learning rate, and momentum term) also has a major influence on the performance of the learning algorithm and the trained network. A trial-and-error strategy is usually applied in the literature for the determination of appropriate values of the learning parameters. In this study, all experiments were carried out considering the guidelines suggested by Kavzoglu and Mather (2003) for the design and application of the neural networks.

RESULTS AND DISCUSSION

A major task in designing a neural network is to determine the number of hidden layers and the number of nodes in those layers. Essentially, the number of nodes in the hidden layers defines the complexity and power of the neural network model to delineate underlying relationships and structures inherent in a data set. In this study, the number of hidden layer nodes was estimated using the following expression of Garson (1998):

$$N_H = N_p/[r \cdot (N_i + N_o)], \tag{1}$$

where N_i, N_o, and N_H indicate the numbers of input, output, and hidden neurons, respectively; N_p shows the number of training samples (or patterns); and the symbol r is a constant that is related to the noise level of the data. Typically, r is ranges from 5 to 10 although it can be as low as 2 considering the difficulty of the problem under consideration.

It is known that a considerable number of mixed and atypical pixels exist in the Landsat ETM+ images of the study area, particularly in the parts where relatively small parcel sizes of land cover types exist. In our case, this was specifically observed for lands covered by tea, hazelnut, and agricultural crops. Considering these points, r is set to 4. For the training data set including 2,000 pixels, N_H is estimated as 25. Thus, the network of 12–25–8 was employed in the classification process. In light of the guidelines suggested by Kavzoglu and Mather (2003), weights in the network were randomly initialized in the range of [–0.25, 0.25], learning rate and momentum term were set to 0.3 and 0.6, respectively, and later reduced to 0.2 and 0.5, and lastly to 0.1 and 0.4.

Training processes for all network structures were controlled by taking the error level for the validation data into consideration, which is commonly known as cross-validation. In other words, the learning process is stopped when the error on the validation set starts to rise. The generalization capabilities of the trained networks were tested using the test pattern file. It should be noted that a 50% threshold is set for a pixel to be assigned to one of the classes. The error matrices for maximum likelihood and artificial neural network classifications produced from the same training and test data sets are presented in Tables 1 and 2, respectively. The classification accuracies were also estimated in terms of the Kappa coefficient, which is a more realistic statistical measure of accuracy than overall accuracy because it incorporates the off-diagonal elements using row and column totals (i.e., omission and commission errors) in addition to the diagonal elements of the error matrix. The term "unrecognized" in the tables indicates the number of pixels having the highest membership value of less than 50%. It should be noted that several trials were also carried out to determine the effect of using randomly selected training and test data sets from total samples collected with reference to the ground truth image. It was found that classification accuracies estimated with these data sets differed at a level of about 2%. It should be also mentioned that some researchers (e.g., Saerens et al., 2002; Brenning et al., 2006) suggest using randomly selected data sets in the training stage to be independent of a particular partitioning. Estimated accuracies are then averaged over all test data sets.

Table 1. Contingency Matrix for the Maximum Likelihood Classification[a]

Class	1	2	3	4	5	6	7	8	Total
1	**162**	2	20	0	1	0	2	0	187
2	1	**168**	13	0	0	0	8	0	190
3	6	4	**176**	0	0	0	7	0	193
4	0	0	4	**189**	0	0	4	0	197
5	0	1	0	0	**194**	1	3	0	199
6	0	0	0	0	0	**196**	0	3	199
7	1	7	2	0	2	1	**177**	3	193
8	0	0	0	0	0	1	2	**196**	199
Total	170	182	215	189	197	199	203	202	**1557**

[a]Overall accuracy = 92.11%; Kappa = 0.9115; 43 pixels unrecognized. Key to classes: 1 = green Tea; 2 = hazelnut; 3 = deciduous; 4 = coniferous; 5 = pasture; 6 = rock; 7 = agriculture; 8 = urban. A 50% threshold was applied to assign a pixel to one of the classes.

Table 2. Contingency Matrix for Artificial Neural Network Classification[a]

Class	1	2	3	4	5	6	7	8	Total
1	**174**	5	12	0	1	0	0	0	192
2	1	**175**	7	0	1	0	8	0	192
3	4	6	**180**	1	0	0	3	0	194
4	0	0	2	**198**	0	0	0	0	200
5	0	1	0	0	**195**	0	0	0	196
6	0	0	0	0	0	**196**	0	3	199
7	0	5	2	0	2	0	**182**	1	192
8	0	0	0	0	0	0	2	**198**	200
Total	179	192	203	199	199	196	195	202	**1565**

[a]Overall accuracy = 93.62%; Kappa = 0.9263; 34 pixels unrecognized. Key to classes: 1 = green Tea; 2 = hazelnut; 3 = deciduous; 4 = coniferous; 5 = pasture; 6 = rock; 7 = agriculture; 8 = urban. A 50% threshold was applied to assign a pixel to one of the classes.

When the overall accuracy and Kappa coefficient values estimated for the classifiers were compared, they appeared to be close to each other. However, it should be pointed out that slight differences in the performances result from the definition of class boundaries for spectrally close or similar classes. In order to better compare the performances, User's and producer's accuracies estimated for each class by the two classification approaches are presented side by side in Table 3. The table clearly conveys the robustness of neural networks for class separation. Attention should be

Table 3. Comparison of Producer's and User's Accuracies for Maximum Likelihood and Neural Network Classifications[a]

	Producer's accuracy		User's accuracy	
Class	ML	ANN	ML	ANN
1	86.6	90.6	95.3	97.2
2	88.4	91.1	92.3	91.1
3	91.2	92.8	81.9	88.7
4	95.9	99	100	99.5
5	97.5	99.5	98.5	98
6	98.5	98.5	98.5	100
7	91.7	94.8	87.2	93.3
8	98.5	99	97	98

[a]Key to classes: 1 = green tea; 2 = hazelnut; 3 = deciduous; 4 = coniferous; 5 = pasture; 6 = rock; 7 = agriculture; 8 = urban.

paid to the accuracies estimated for the green tea (1), hazelnut (2), deciduous forest (3), and agriculture (7) classes for which ANN produced better results up to 3% in terms of producer's accuracy. Spectral closeness of these classes and resulting misclassification can be easily comprehended from the incorrectly classified pixels in Tables 1 and 2. For example, in ML classification 20 pixels of green tea were incorrectly classified as deciduous forest pixels, whereas 13 pixels of hazelnut were mistakenly classified as deciduous and 8 other pixels as agriculture. The level of confusion is much less for the ANN classifier, as it correctly classified most of the pixels that were incorrectly classified by the ML technique.

Classification maps of the study area were produced using the maximum likelihood (ML) and artificial neural network (ANN) classifiers (Fig. 3). For ANN classification, the entire image was input to the trained neural network to produce the thematic map of the study area. The robustness of ANN classification over ML classification can be easily observed from the figures, especially for tea and hazelnut fields. The major differences in the thematic maps and classified images are highlighted by circles. It is straightforward to comprehend the differences in the results and the robustness of the ANN method.

Detailed analysis of the results in terms of total areas covered by the classes is provided in Figure 4. In the ML classification, fewer pixels were identified for the green tea, hazelnut, deciduous, and coniferous classes. On the other hand, more pixels were classified as pasture, rock, and agriculture classes in comparison to the ANN classification. The reason for this could be that the samples collected for classes are known to be mostly mixed pixels that encompass a substantial region in the feature space compared to the others. Due to the spectral similarity and the superposition of spectral regions of several classes, the ML algorithm relying solely on statistical estimates wrongly identified many pixels in the resulting thematic image. This can be easily seen from the contingency matrices given in Tables 1 and 2. In the ML

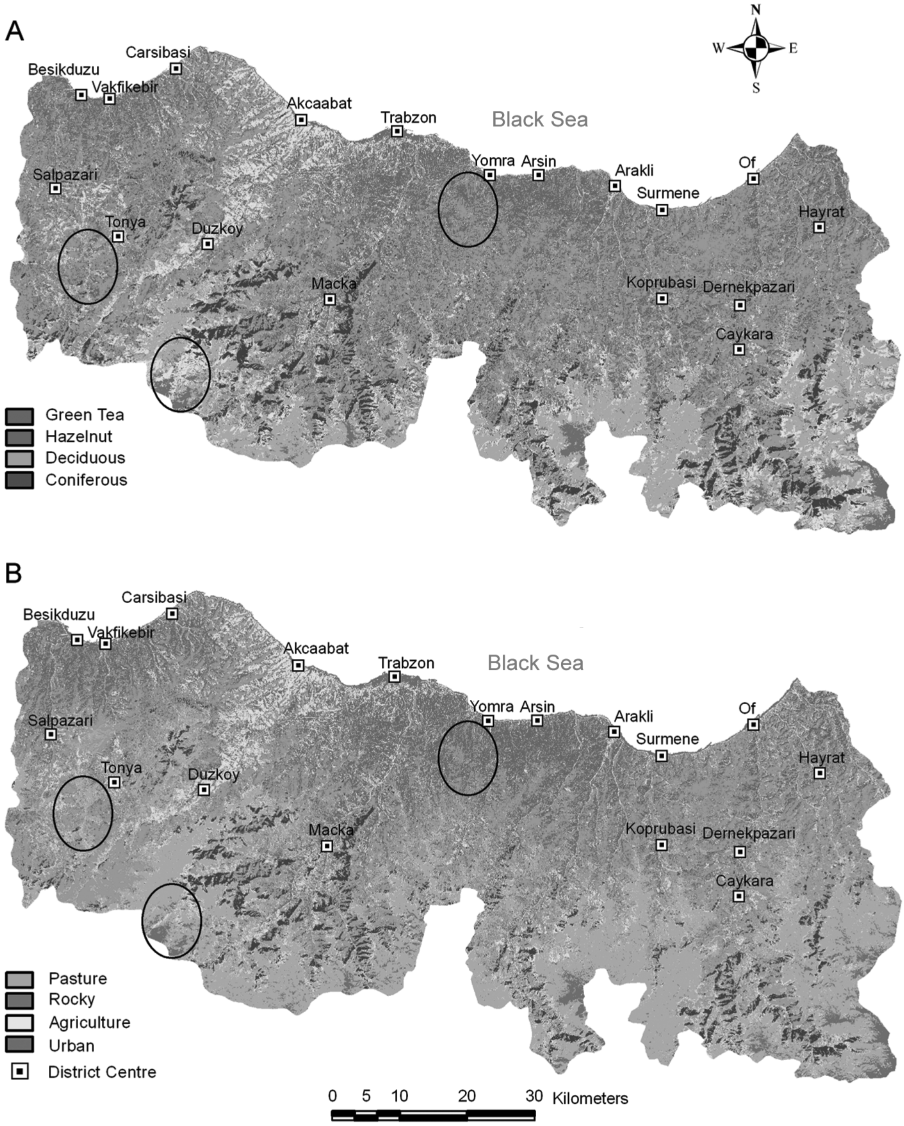

Fig. 3. Classification results produced by artificial neural networks (A) and the maximum likelihood classifier (B). The major classification problems are highlighted for the areas shown in the circles.

classification, overall accuracy was 92.11% with individual class accuracies (i.e., producer's accuracy) ranging from 86.6 to 97.5%. The highest accuracies (98.5%) were estimated for the rock and urban classes, which are the most distinct classes of all. On the other hand, classes of coniferous, pasture, rock, and urban were all classified with classification accuracies equal or above 98.5% in the ANN classification. The ML algorithm produced the lowest classification accuracies for green tea and hazelnut

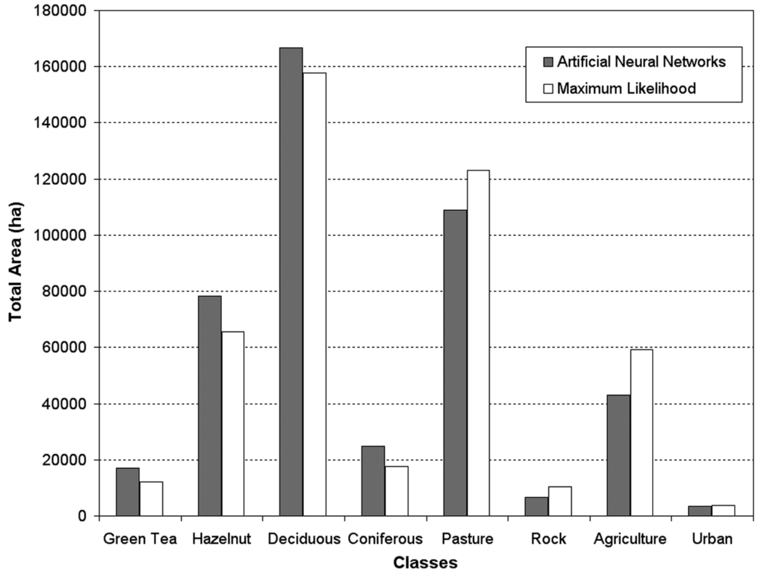

Fig. 4. Analysis of classification results in terms of total area coverage of the land cover/use classes.

classes, with 86.6% and 88.4%, respectively. These two classes were the most difficult classes for which to collect ground reference data, due to relatively small size of the parcels.

It should be also pointed out that more pixels were unrecognized by the ML classification algorithm (i.e., less than 50% membership for a class). The results clearly show the effectiveness of the ANN method for the classification of land cover/use classes with limited training data. The limitation reflects not only the number of samples but also the difficulty of the problem—that is, a large amount of mixed pixels due to the spatial resolution of the Landsat ETM+ sensor and the parcel sizes in the study area.

CONCLUSIONS

Some characteristics, such as flexible decision region capability, the ability to use data from different sources, and its non-parametric nature, have made artificial neural networks (ANNs) a new standard tool in the analysis of remotely sensed images. For years the most widely used method has been the maximum likelihood classifier (ML), a conventional statistical classifier. The performance of a multilayer perceptron learning the characteristics of the training data using a back propagation algorithm is compared to that of maximum likelihood classifier in identifying major land cover/use classes present in the study area, the city of Trabzon, Turkey. The study area selected for this research has unique characteristics in terms of its terrain features, parcel sizes, and vegetation variability.

A large number of mixed pixels exist in the images due to the small sizes of the fields in the study area, considering the 30 m resolution of the Landsat ETM+ imagery. The effect of mixed pixels in training data was found effective in the performances of the classifiers. Due to the spectral similarities and superposition of clusters, the ML algorithm produced relatively poorer classification accuracy. Both classifiers were trained and tested with the same data sets (as is the general practice in the literature) to eliminate the bias of using data sets with different characteristics, including their sizes and distributions. In addition to error matrices of the test data set, total numbers of pixels for the entire image were estimated for each class. The ANN produced better results in terms of both overall accuracy and individual class accuracies. Particularly, the poor classification performance for green tea and hazelnut classes was improved with the use of the ANN classifier. It is observed that the ML produces lower classification accuracies for classes having similar spectral characteristics, or including several subclasses, as in the case of the deciduous tree class. It should be also pointed out that the ANN classifier identifies more pixels under the same conditions (i.e., training/test data sets, threshold value).

ACKNOWLEDGMENTS

The authors are grateful to The Prime Ministry State Planning Organization (DPT) and Karadeniz Technical University for their support and for supplying remote sensing data.

REFERENCES

Akyol, N. and A. F. Sesli, 2000, *Tarım Arazilerinin Parçalanması Trabzon İli Örneği*. Trabzon, Turkey: Karadeniz Technical University, Technical Report 2000/4 (in Turkish).

Blamire, P. A., 1996, "The Influence of Relative Sample Size in Training Artificial Neural Networks," *International Journal of Remote Sensing*, 17(1):223–230.

Brenning, A., Kaden, K. and S. Itzeroft, 2006, "Comparing Classifiers for Crop Identification Based on Multitemporal Landsat TM/ETM Data," in *Proceedings of the Second Workshop of the EARSeL SIG on Land Use & Land Cover*, Bonn, Germany, 64–71.

Civco, D. L., 1993, "Artificial Neural Networks for Land-Cover Classification and Mapping," *International Journal of Geographical Information Systems*, 7(2): 173–186.

Erbek, F. S., Özkan, C. and M. Taberner, 2004, "Comparison of Maximum Likelihood Classification Method with Supervised Artificial Neural Network Algorithms for Land Use Activities," *International Journal of Remote Sensing*, 25(9):1733–1748.

Foody, G.M., 1995, "Using Prior Knowledge in Artificial Neural Network Classification with a Minimal Training Set," *International Journal of Remote Sensing*, 16(2):301–312.

Garson, G. D., 1998, *Neural Networks: An Introductory Guide for Social Scientists*, London, UK: Sage.

Gopal, S. and C. Woodcock, 1996, "Remote Sensing of Forest Change Using Artificial Neural Networks," *IEEE Transactions on Geoscience and Remote Sensing*, 34(2):398–403.

Kavzoglu, T. and P. M. Mather, 2003, "The Use of Back Propagating Artificial Neural Networks in Land Cover Classification," *International Journal of Remote Sensing*, 24(23):4907–4938.

Lillesand, T. M. and R. W. Kiefer, 2000, *Remote Sensing and Image Interpretation*, 4th ed., New York, NY: The Lehigh Press.

Mather, P. M., 1999, *Computer Processing of Remotely-Sensed Images: An Introduction*, Chichester, UK: John Wiley.

Paola, J. D. and R. A. Schowengerdt, 1995, "A Detailed Comparison of Back Propagation Neural Network and Maximum Likelihood Classifiers for Urban Land Use Classification," *IEEE Transactions on Geoscience and Remote Sensing*, 33(4):981–996.

Reis, S. and T. Yomralioglu, 2006, "Detection of Current and Potential Hazelnut (*Corylus*) Plantation Areas in Trabzon, North East Turkey Using GIS&RS," *Journal of Environmental Biology*, 27(4):653–659.

Rumelhart, D. E., McClelland, J. L., and PDP Group, 1986, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*, London, UK: The MIT Press.

Saerens, M., Latinne, P. and C. Decaestecker, 2002, "Adjusting the Outputs of a Classifier to New *a priori* Probabilities: A Simple Procedure," *Neural Computation*, 14:21–41.

Swain, P. H. and S. M. Davis, 1978, *Remote Sensing: The Quantitative Approach*, New York, NY: McGraw-Hill.

Tso, B. and P. M. Mather, 2001, *Classification Methods for Remotely Sensed Data*, London, UK: Taylor & Francis.

URL-1, 2008, "What to Do When Data are Non-normal, in *NIST/SEMATECH e-Handbook of Statistical Methods*, December 5 [http://www.itl.nist.gov/div898/handbook/].

Vaughan D. C. and M. L. Tiku, 2000, "Estimation and Hypothesis Testing for a Non-normal Bivariate Distribution with Applications," *Mathematical and Computer Modelling*, 32:53–67.