CLASSIFICATION OF MULTI-SPECTRAL, MULTI-TEMPORAL AND MULTI-SENSOR IMAGES USING PRINCIPAL COMPONENTS ANALYSIS AND ARTIFICIAL NEURAL NETWORKS: BEYKOZ CASE

M. Cetin [a], T. Kavzoglu [a, *], N. Musaoglu [b]

[a] Dept. of Geodetic and Photogrammetric Engineering, Gebze Institute of Technology, 41400 Gebze-Kocaeli, Turkey - (mcetin, kavzoglu)@gyte.edu.tr
[b] Istanbul Technical University, Civil Engineering Faculty, 34469 Ayazaga, Istanbul, Turkey - nmusaoglu@ins.itu.edu.tr

KEY WORDS: Land Cover, Classification, Artificial Neural Networks, Principal Components, Maximum Likelihood, PCA

ABSTRACT:

The thematic maps derived from remotely-sensed images are invaluable sources of information for various investigations since they provide spatial and temporal information about the nature of Earth surface materials and objects. The robustness of classification techniques used to produce these thematic maps can be crucial especially for complex classification problems. This study aims to determine the level of contributions of multi-temporal and multi-sensor data together with their principal components for Maximum Likelihood and Artificial Neural Network classifiers. The performance of a multi-layer perceptron that learns the characteristics of the data using backpropagation algorithm is compared to that of Maximum Likelihood classifier in identifying major land cover classes present in the study area, Beykoz district of Istanbul, Turkey. The image data available for the study are from Landsat ETM+ and Terra ASTER images. Image band combinations are inputted to the neural network for training and the success of the classification is tested using test data sets. Results show that the neural network approach is an attractive and effective way of extracting land cover information using multi-spectral, multi-temporal and multi-sensor satellite images. It is also observed that the level of contribution of principal components to the results is much less than the contribution of multi-temporal data in terms of the classification accuracy.

1. INTRODUCTION

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-honoured 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 for regional or global scale studies. With the launch of recent satellites, much image data are available from different sources regarding the same area. Naturally, each data type represents different characteristics of the area. Using all these data may improve the accuracy of classification significantly but introduces redundancy and requires more training data. Therefore, new data should be added only if they contribute to an improved classification. In the literature, a single date satellite image has been usually used for classification problems. However, it is well known that it is necessary to consider the reflectance characteristics of surface objects varying with the seasonal conditions to improve the accuracy of the classification.

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 (ANN) models. They can be thought of as forms of models imitating the complicated brain processing in a very simple way. This method has been recently used 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 recently applied to many applications, but the most popular application of the method in remote sensing is the classification of land cover information (Paola and Schowengerdt, 1995; Gopal and Woodcock, 1996; Sunar Erbek et al. 2004). The use of ANN method has become popular due to the unique advantages of the method over conventional statistical methods. Perhaps the most important characteristic of ANN is that there is no underlying assumption about the distribution of data. Furthermore, it is possible to employ data from different sources to improve the accuracy of the classification.

For this study the performances of classification methods, the Maximum Likelihood and ANN classifiers, were tested for the inclusion of multi-temporal, multi-sensor and principal component images in classification processes. In order to meet the objective of this research Landsat ETM+ and Terra ASTER images and their first three principal components accounting for the highest variances were employed in land cover classification and the results for the two classifiers were compared.

2. STUDY AREA AND DATA

The study area, Beykoz, located along the north-east side of the Bosphorous on the Asian side of Istanbul, is one of the least dense regions mainly due to its 60% forest coverage. The elevation of Beykoz ranges from 0 to 400 metres. Most

---

* Corresponding author

transportation in the region relies on a two-lane highway that parallels with the coastline. Almost all other roads are quite narrow, which causes traffic congestion. The construction of the two bridges in 1973 and 1989, one of which is located in lower left part of the study area (Figure 1), led to better highway transportation network, but it brought the rapid urbanization problem around the recently constructed transportation areas within Beykoz district.

In this study, multisensor data including Landsat ETM+ and Terra ASTER imagery, acquired in May 2001 and October 2002 respectively, were used for the delineation of eight main land-cover classes. These classes are namely coniferous forest, deciduous forest, urban, inland water, grassland, bare soil, road and sea. Due to the short time difference between the acquisition dates it is assumed that there was no dramatic change on the types of ground cover classes. The area under the analysis (Figure 1) was about a 430 km² region covering Beykoz district. Borders of Beykoz district are also depicted on Figure 1.
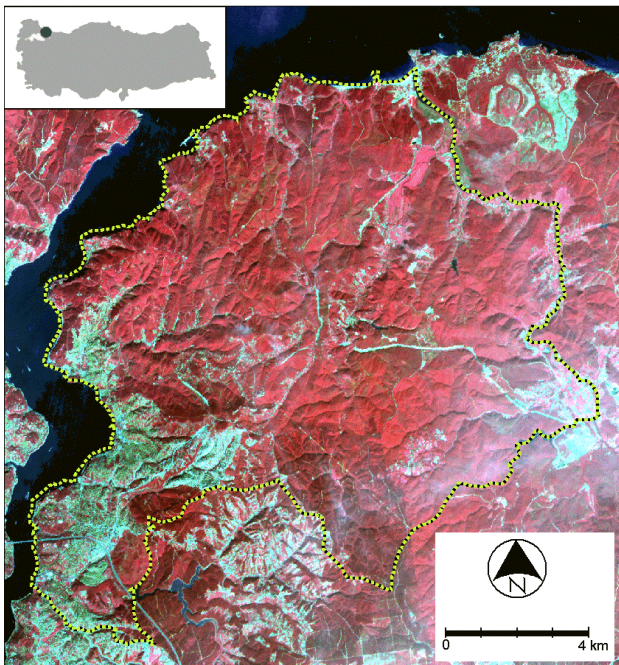

Figure 1. The location of study area, Beykoz district, Istanbul

The Landsat ETM+ image was rectified to the UTM projection using several 1:25000 scale topographic maps. In the geometric correction process, a total of 22 Ground Control Points (GCP) were used, which resulted in a Root Mean Square Error (RMSE) of less than 0.5 pixels. The Terra ASTER image of the study area was later rectified to the geometrically corrected Landsat ETM+ image using 29 GCPs giving a RMSE of around 0.5 pixels. The two corrected images were resampled at a spatial resolution of 30m using the nearest neighbour algorithm in order to prevent the formation of new pixel values. Subimages of 714 by 668 pixels covering the study area were extracted and used in subsequent analyses.

3. METHODS

3.1 Principal Components Analysis

Adjacent bands in a multispectral remotely-sensed image are generally correlated. Presence of the correlation among the bands of the multispectral image implies that there is redundancy in the data. In other words, some information is being repeated. It is the repetition of the information between the bands that is reflected in their inter-correlations (Mather, 1999). Principal components analysis (often called PCA or Karhumen–Loeve analysis) has proved to be of value in the analysis of multispectral remotely-sensed data. The transformation of the raw remote sensing data using PCA can result in new component images that may be more interpretable than the original data (Ashutosh, 2002). Additionally, this technique reduces contributions of noise and error. PCA can be used to reduce the information included in the raw data into two or three bands without losing significant information (Monger, 2002). The principal components analysis can be used for effective classification of land use, colour representation or visual interpretation with multi-band data and change detection with multi-temporal data (Sunar, 1998).

PCA is used in this study to improve the quality of the classification. It is applied to the satellite images to obtain uncorrelated (i.e. statistically independent) principal components lying on orthogonal axes that the original data are reprojected. The results of PCA for both images including eigenvalues and variances of each component are given in Tables 1 and 2. Whilst the first three components for the Landsat ETM+ image represent 99% of the image data, those for the Terra ASTER image account for 97% of the image data. First three components of PCA analyses for both images were used to form three-layer images, which are later used in classification processes.

Component	Eigenvalue	Variance (%)	Total (%)
1	2445.38	71.39	71.39
2	884.21	25.81	97.20
3	72.65	2.12	99.32
4	11.98	0.35	99.67
5	8.88	0.26	99.93
6	2.27	0.07	100.00

Table 1. PCA for Landsat ETM+ image

Component	Eigenvalue	Variance (%)	Total (%)
1	16094.89	82.07	82.07
2	2118.43	10.80	92.87
3	883.73	4.51	97.37
4	191.99	0.98	98.35
5	109.97	0.56	98.91
6	85.03	0.43	99.35
7	51.58	0.26	99.61
8	47.79	0.24	99.85
9	28.81	0.15	100.00

Table 2. PCA for Terra ASTER image

3.2 Classification

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, since such images are high-dimensional and complex in nature. As the number of categories and the amount of data involved increases, so does the complexity of the classification problem because it becomes more difficult to determine the characteristics of the categories and allocate a pixel to one of the categories. Digital image processing of satellite data provides tools for analysing the image through different classification algorithms. The two data sets used in this research were independently classified using two strategies, Maximum Likelihood and Artificial Neural Network approaches. Each of these algorithms is based on a different mathematical theory that is described in following sections. Information obtained from fieldwork, maps and aerial photographs helped in identifying sample pixels for each land cover category on the imagery. The classifiers use the properties of such sample pixels, known as training pixels, to work out parameters for land cover classes to which each pixel on the imagery would be assigned. The classification process generates thematic maps on which different land cover categories are generally presented in different colours.

3.2.1 Maximum Likelihood Classifier

The Maximum Likelihood (ML) classifier is one of the statistical classifiers that rely on normal distribution of the data in each class. The geometrical shape of a number of pixels belonging to a class is represented by an ellipsoid. The locations, shapes and sizes of the ellipsoids are derived from the means and variance-covariance matrices of classes.

These ellipses represent contours of probability of membership, the values of which decline with distance from the mean centre. Distance from centre 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. 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 centre point of each class (Mather, 1999).

It is important to take into consideration that a ML classifier gives good results if the frequency distribution of the data is in the multivariate normal distribution. Unsupervised classification methods can be used to find out whether training data represents the assumption of normal distribution. After estimating the probabilities of each pixel being a member of the classes, the most likely class having the highest probability value is assigned to the pixel with a class label. If the highest probability value of a pixel is lower than a threshold to be set by an analyst, then the pixel is labelled as unclassified.

3.2.2 Artificial Neural Networks

The basic element of artificial neural networks (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, which in turn forms one of the inputs to a processing node in the next layer of ANNs. Whilst activation functions, also known as transfer functions, are employed to decrease the number of iterations, they introduce non-linearity into the network and thus improve its performance. The structure of a feed-forward artificial neural network (i.e. multi-layer perceptron) includes input, hidden and output layers (Figure 2). The input layer introduces the distribution of the data for each class to the network. Each input layer node represents one of the input features, such as a Landsat ETM+ band. The output layer is the final processing layer that has a set of values (or codes) to represent the classes to be recognised. The layers between the input and output layers are called hidden layers. Through these layers that the internal representations of the input patterns can be produced.

Processing nodes make up a set of fully interconnected layers, except that there are no interconnections between nodes within the same layer in the standard multiplayer perceptrons. All inter-node connections have associated weights, which are usually initially randomised. When a value passes through an inter-connection, it is multiplied by the weight associated with that inter-connection. The weights in the network determine class boundaries in the feature space.
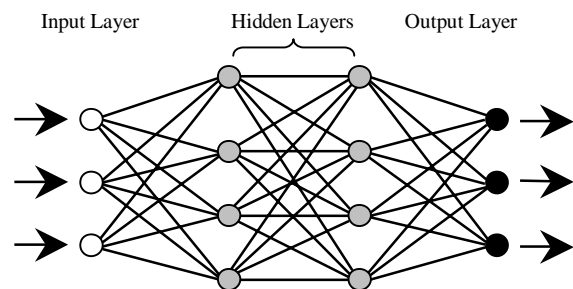

Figure 2. A simple four layer feed-forward neural network

The most interesting characteristic of ANNs is their ability to learn. The learning algorithm used for this purpose defines the way the network weights are adjusted between successive training cycles or epochs. Although many learning strategies have been developed, the most widely used procedure is the backpropagation learning algorithm, also known as the generalised delta rule. The algorithm operates by searching an error surface, defined as a function of the weights, using a gradient descent technique to locate the point (i.e. weight combination) with minimum error. Each iteration in the backpropagation algorithm has two basic movements: forward and backward. The forward propagation cycle starts with the presentation of a set of input patterns to the network. The backward error correction starts at the output layer and error is fed backward through the intermediate layer towards the input layer in order to adjust the weights and reduce the error. The process continues iteratively until the error is reduced to an acceptable level.

Although ANN-based classification methods are more robust than conventional statistical approaches, difficulties in their use relate to their long training time requirements and to the determination of an efficient network structure, both of which have a direct effect on training time and classification accuracy. Large networks take a long time to learn, and tend to

give accurate classification results for training data but not for unknown test data. Network structure, therefore, has a direct impact on the generalisation capabilities of networks, that is, their ability to recognise patterns that are not present within the training set (Kavzoglu and Mather, 1999).

4. RESULTS AND DISCUSSION

As a result of a field study and visual interpretation of aerial photography (1: 5000 scale) of the study area, eight major land cover classes were decided, and fields were selected to collect representative pixels for the classes to be used in classification processes. Thus, a ground truth image containing a total of 10477 pixels was created. Table 3 shows total number of pixels selected for each land cover type. As can be seen from the table, it was difficult to collect sample pixels for some classes such as grassland, inland water and bare soil classes. It should be also noted that the selection of pixels for road class was tedious and problematic as the width of the roads in the study area is mainly smaller than the pixel size of 30 metres, suggesting that these pixels are mostly mixed in nature.

Class	Number of Pixels
Coniferous Forest	1667
Deciduous Forest	3633
Urban	1144
Inland Water	383
Grassland	389
Bare Soil	590
Road	1029
Sea	1642

Table 3. Number of pixels used for training and testing

In the formation of training, validation and test pattern files, an in-house software developed by the second author of this paper was employed. The program randomly selects pixels from the images by taking the ground truth image into account. It also allows the user to decide minimum and maximum number of pixels for each pattern file. For minimum 380 pixels were selected whilst 1000 pixels for maximum. For all band combinations considered in this study training files included 200 pixels for each class (1600 pixels in total), validation files contained 40 pixels for each class, and testing files comprised 3550 pixels.

In order to test the effectiveness of addition of multi-temporal and principal component bands, several combinations of the image data were produced by stacking image layers. In addition to the single date Landsat ETM+ and Terra ASTER images, combination of both images together with the principal components were prepared and employed in the classification stage. These combinations of the images are described in Table 4. In the table, abbreviations of PCA1 and PCA2 represent the images having the three principal components estimated for Landsat ETM+ and Terra ASTER images, respectively.

Training, validation and test pattern files were produced for the combinations given in Table 4. In order to estimate the number of training samples, set the optimum rates for the learning parameters and define the network structure (i.e. number of hidden layer neurons), the guidelines suggested by Kavzoglu and Mather (2003) were used.

Combination	Image (band)
C1	Landsat ETM+ (6)
C2	Terra ASTER (9)
C3	Landsat ETM+ (6) + PCA1 (3)
C4	Terra ASTER (9) + PCA2 (3)
C5	Landsat ETM+ (6) + Terra ASTER (9)
C6	Landsat ETM+ (6) + Terra ASTER (9) + PCA1 (3) + PCA2 (3)

Table 4. Image band combinations used in classification

According to these guidelines, weights in the network were randomly initialised in the range of [-0.25, 0.25], learning rate and momentum term were set to 0.2 and 0.5 respectively, and the number of hidden layer nodes were estimated using the following expression of Garson (1998);

$$N_p / [r \cdot (N_i + N_o)] \qquad (1)$$

where numbers of input and output layer nodes are represented by N_i and N_o respectively, and the number of training samples (or patterns) is represented by N_p. The symbol r is a constant set by the noise level of the data. Typically, r is in the range from 5 to 10. It should be noted that the Stuttgart Neural Network Simulator (SNNS) developed at the Institute for Parallel and Distributed High Performance Systems at the Stuttgart University was chosen to implement the neural network models created for each image combination. Training processes for all network structures were controlled by taking the error level for the validation data into consideration, which is known as cross-validation − a robust stopping criterion for training process. In other words, learning process is stopped when the error on the validation set starts to rise. The generalisation capabilities of the trained networks were tested using the test pattern file. The results for all combinations including individual class accuracies were shown in Table 5. The table also includes the results of the Maximum Likelihood classification that is performed with exactly the same training and test pixels. The classification accuracies were estimated in terms of Kappa coefficient, which is a more realistic statistical measure of accuracy than overall accuracy since 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. Network column in the table shows the network structures established for the corresponding combination. For instance, 6-16-8 indicates 6 input nodes, 16 hidden nodes and 8 output nodes.

Combination	Network	ANN	ML
C1	6-16-8	0.88958	0.84851
C2	9-20-8	0.92553	0.89012
C3	9-20-8	0.88837	0.84209
C4	12-16-8	0.91715	0.87410
C5	15-26-8	0.94943	0.90516
C6	21-25-8	0.95876	0.92051

Table 5. ANN and ML results for image band combinations

It is clear from Table 5 that classification accuracies produced for Landsat ETM+ and Terra ASTER images (C1 and C2) are higher than those produced by these images together with their principal components (C3 and C4). This shows the ineffectiveness of PCA bands for land cover type delineation. While the PCA bands do not introduce different distinctive characteristics, they increase the dimensions and the complexity of the data. Therefore, they do not make any significant contribution to the results produced by the two classifiers. On the other hand, the use of the combination of the two multi-temporal images (C5) in classification resulted in slightly better performances. However, it should be borne in mind that it is difficult to increase the level of classification accuracy after a certain point, mainly depending on the complexity of the data set, the number of output classes to be recognised and the classification technique used.

After the accuracy assessment stage, subset images were fed into the trained networks to produce the thematic maps of the study area including eight land cover classes. As an example, the ML and ANN classification results for C6 combination are given in Figure 3. The robustness of ANN classification over ML classification can be easily observed from the figure, especially for road class. In the ML classification many pixels were identified as road since the samples collected for this particular class are known to be mostly mixed pixels that encompass a substantial region in the feature space compared to the others. This forced the ML, which relies on statistical estimates, to wrongly identify many pixels as road in the resulting thematic image. The weakness of the ML classification can be also observed for urban pixels near the upper right corner of the image. Except for these problems, the ML classification method performed well for other classes.

When the result of ANN classifier is analysed from Figure 3, two points draw attention. The first is related to larger number of bare soil pixels in the image compared to the ML result. The analysis of ground truth data showed that ANN classifier produced more realistic results for this class, as well. The second problem is about the incorrect classification of sea pixels as inland water for pixels along the seashore. This is most likely resulted from the ground truth image due to the possibility of selected pixels of inland water and water reservoirs being sand or rock. Therefore, the pixels assigned to inland water class along the seashore could be representing sand or rock.

5. CONCLUSIONS

The degree of contributions of multi-temporal image data and their principal components in the production of a classified thematic map is investigated in this study. The classification methods considered are the Maximum Likelihood (ML) and Artificial Neural Network (ANN) classifiers, sophisticated methods that have been recently used for many investigations. The image data available for the study were from Landsat ETM+ and Terra ASTER images acquired in 2001 and 2002, respectively. The main objective was to use these images to determine the effect of principal components in the results in terms of the classification accuracy. ANNs were trained to learn the characteristics of the training sets prepared for combinations of images together with their principal components. Trained networks were later used in accuracy assessment and in the classification of subset image combinations. Several important results can be derived from the results. Firstly, the ANN classifier yielded more accurate results than the ML classifier. The robustness of the ANN can be easily seen from Table 5 and from the classified image
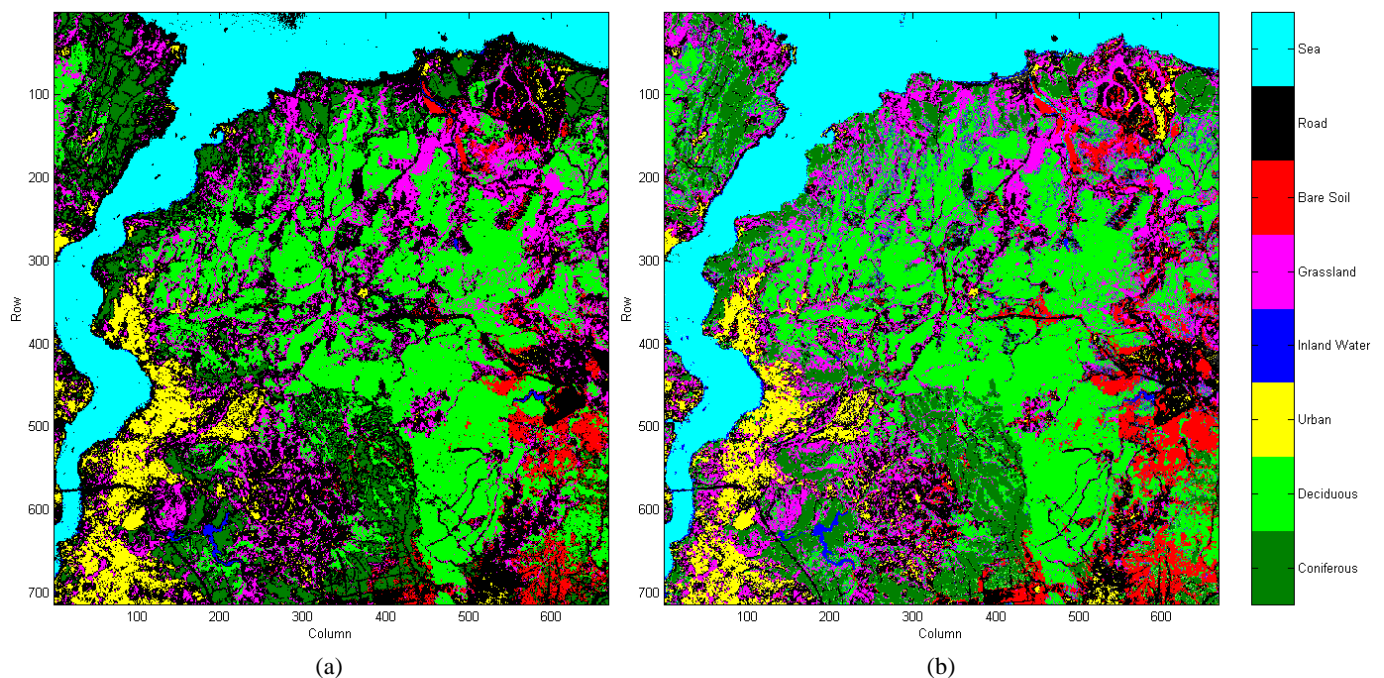

(a) (b)

Figure 3. (a) Maximum Likelihood classification result, (b) Artificial neural network result

shown in Figure 3b. Secondly, both classifiers had difficulties to learn the characteristics of road class as it largely included mixed pixels due to the pixel resolution of the images (i.e. 30m). Thirdly, it is found that the addition of first three principal components to the classification process did not make any improvement. On the contrary, it reduced the classification accuracy probably because of the increased complexity and dimensionality of the data. This point obviously needs to be clarified with further research. Lastly, the classification methods, especially ML classifier, show sensitivity to classes depending on their spectral variability. ML algorithm dominated road class over the image whilst ANN classifier was slightly sensitive to inland water class.

REFERENCES

Ashutosh, S., 2002. Principal component-based algorithm on multispectral remote sensing data for spectral discrimination of tree cover from other vegetation types. *Current Science*, 82(1), pp. 67-69.

Civco, D. L., 1993. Artificial neural networks for land-cover classification and mapping. *International Journal of Geographical Information Systems*, 7(2), pp. 173-186.

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

Gopal, S. and Woodcock, C., 1996. Remote sensing of forest change using artificial neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 34(2), pp. 398-403.

Kavzoglu, T. and Mather, P.M., 1999. Pruning artificial neural networks: an example using land cover classification of multi-sensor images. *International Journal of Remote Sensing*, 20(14), pp. 2787-2803.

Kavzoglu, T. and Mather, P.M., 2003. The use of backpropagating artificial neural networks in land cover classification. *International Journal of Remote Sensing*, 24(23), pp. 4907-4938.

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

Monger, M., 2002. Soil classification in arid lands with Thematic Mapper data. *Terra*, 20(2), pp. 89-100.

Paola, J. D. and Schowengerdt, R. A., 1995. A detailed comparison of backpropagation neural network and maximum likelihood classifiers for urban land use classification. *IEEE Transactions on Geoscience and Remote Sensing*, 33(4), pp. 981-996.

Sunar, F., 1998. An analysis of changes in a multidate dat set: a case study in the Ikitelli area, Istanbul, Turkey. *International Journal of Remote Sensing*, 19(2), pp. 225-235.

Sunar Erbek, F., Özkan, C. and Taberner, M., 2004. Comparison of maximum likelihood classification method with supervised artificial neural network algorithms for land use activities. *International Journal of Remote Sensing*, 25(9), pp. 1733-1748.