The use of Back Propagating Artificial Neural Networks in Rare Vegetation Communities Classification from High-Resolution Satellite Imagery

Hassanov, H. G., Gambarov, A. Y., and Gambarova, Y. M.,
Institute for Aerospace Informatics, National Aerospace Agency, Baku, Azerbaijan
E-mail: hhassanov@yahoo.com, AGambarov@sahil.info, yelena.gambarova@integris.az

Abstract
This paper has presented artificial neural networks (ANNs) for rare vegetation communities' classification using remotely sensed data. Three variants of training of the Multi Layer Perceptron (MLP) based on three different classification schemes are used. At first 12 types of rare vegetation communities were defined and the main classification scheme was designed on that basis. After preliminary statistical tests for training samples, two modification algorithms of the classification scheme were defined: the first one led to creating a scheme, which consisted of 7 classes, and the second one led us to creating of 5 classes' scheme. Testing results show that the use of ANNs of 5 classes' scheme can produce higher classification accuracies than other alternative. The training procedures of these classifiers are described in details along with analysis and post processing products using Geoinformation Technologies. Ancillary geospatial data: DTM and its derivable (DEM, Slope, Aspect), as well as topographical, hydrological data and land use maps were created in order to support post classification operations. This result demonstrates that a level of classification accuracy achieved by artificial neural networks is higher than those generated by the statistical classifiers.

1. Introduction

The interpretation of remotely sensed data uses techniques from a number of disciplines including remote sensing, pattern recognition, artificial intelligence, computer vision, image processing and statistical analysis. The move towards automated analysis of remotely sensed data is encouraged by the ever increasing volumes of data as well as by the high cost of ground surveying. The new generation of satellite-borne instruments is providing higher spatial and spectral resolution data, leading to the wider application of remotely sensed products and further emphasizing the need for more automated forms of analysis. A number of methodologies have been developed and employed for image classification from remotely sensed data within the past 20 years. Statistical image classification techniques are ideally suited for data in which the distribution of the data within each of the classes can be assumed to follow a theoretical model. The most commonly used statistical classification methodology is based on maximum likelihood, a pixel-based probabilistic classification method which assumes that spectral classes can be described by a normal probability distribution in multispectral space (Swain and Davis, 1978). This traditional approach to classification is found to have some limitations in resolving interclass confusion if the data used are not normally distributed.

As a result, in recent years, and following advances in computer technology, alternative classification strategies have been proposed. In most instances, human beings are good pattern recognizers. This observation led researchers in the field of pattern recognition to consider whether computer systems based on a simplified model of the human brain can be more effective than the standard statistical and knowledge-based classification methods. Research in this field led to the adoption of artificial neural networks (ANN), which have been used in remote sensing over the past ten years, mainly for image classification. An important characteristic of ANNs is their non-parametric nature, which assumes no *a priori* knowledge, particularly of the frequency distribution of the data. Because of their adaptability and their ability to produce results with classification accuracies that are higher than those generated by statistical classifiers, the use of ANNs has spread in the scientific community at large, leading to an increasing amount of research in the remote sensing field (Paola and Schowengerdt, 1995b and Atkinson and Tatnall, 1997). The network comprises a large number of simple processing elements linked to each other by weighted connections according to a specified architecture. These networks learn from the training data by adjusting the connection weights.

ISSN 1686-6576 / © Geoinformatics International

There is a range of artificial neural network architectures designed and used in various fields, including pattern recognition (Bishop, 1995). In remote sensing applications, the multi-layered feed-forward network, also called the multi-layer perceptron, is generally used (Benediktsson et al., 1990, Zhang and Scofield, 1994 and Foody, 1995a). The structure of the neural-network makes it very easy to incorporate ancillary data or spatial information. Determination of the 'best' bands that are assigned to the input neurons of an artificial neural network (ANN) is one of the critical steps in designing the ANN for a particular problem. Methods used to select the optimum inputs are known as *feature selection techniques*. Their use in the context of artificial neural networks was investigated in study (Kavzoglu and Mather, 2002). A method for improving artificial neural network performance by using multi-temporal, multi-spectral and multi-source remotely-sensed data as features (optimum inputs) for classifying agricultural crops is described in another paper (Oliveira et al., 2003). ANNs models of multi-layer feed-forward perceptron were designed using NDVI (Normalized Difference Vegetation Index), EVI (Enhanced Vegetation Index), red (RED) and near infrared (NIR) reflectance values as input patterns, respectively by Bocco et al., (2007) and Walthall et al., (2004).

2. Study Area

The Study Area (Sangachal) is located on the coast of the Caspian Sea, 45 kilometers (km) south of Baku, Azerbaijan with an area of around 100 km^2 (Figure 1). It is positioned with latitude 40°10'10"N and longitude 49°27'45"E.

3. Data used and Methodology

Two IKONOS images acquired in July 2005 and June 2006 were used for the delineation of 12 rare vegetation communities and soil types. The images being used were pan-sharpened multi-spectral images at resolution of 1m. It has well defined test site for 12 classes on the both images where training and test samples were gathered from. The Multilayer Perceptron (MLP), a feed-forward artificial neural network model is used in rare vegetation community's classification using remotely sensed data. And back propagation learning algorithm, also called the generalized delta rule, was an iterative gradient descent training procedure.

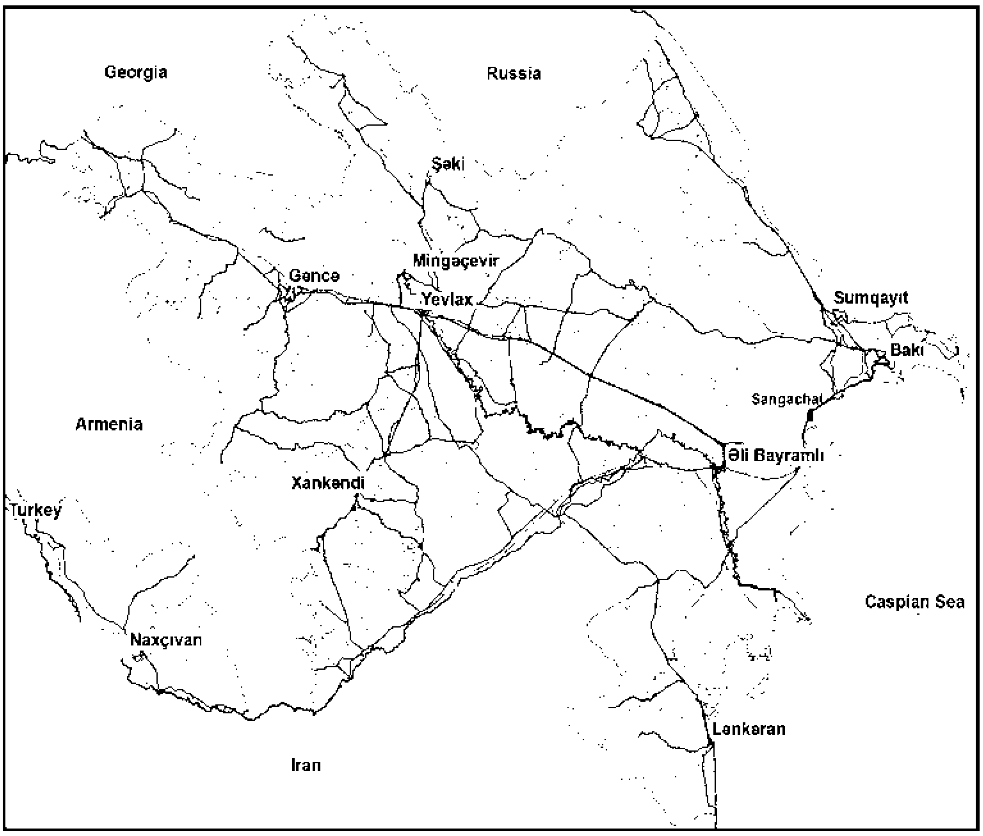


Figure 1: Map of the study area

A Specialized GIS was used as software environment for performing of workflow comprising of jobs connected with collecting of samples, hosting of classifier training and producing software as well as classification results analysis. We used ArcGIS Desktop 9.2 with its extensions (Spatial Analysis and 3D Analysis) as base (core) GIS. Using this software, we created Geographical Data Base consisting of relevant spatial data (Orthorectified satellite multi-spectral data, ancillary data: various spectral Indexes, DEM, Slope, Aspect as well as vector Topographical data) and Map template. We also developed and built our tools in Core GIS and UI for managing/automating all processes during the project lifetime: collecting/gathering of training and test sets of samples, designing, and learning, testing of ANNs and performing of classification on real data. Combination of these data and software we called Specialized GIS for solving applied tasks using RS data and ANNs.

4. Definition of the Classification Scheme

Initially 12 types of rare vegetation communities and soil types were defined that—according to ecologists' opinion—are indicators of anthropogenic impact on environment in the region being studied. Below the names of them including their Latin analogs (in parenthesis) are presented (Table 1). Ecologists presented the geographical coordinates of these vegetation communities. At first, these sites were geolocated, then using GIS procedures the areas of location of these vegetation communities were determined for extraction of samples for the classifier training and testing. For the MLP classifier training, sizes of training and test sets of samples for the 12 classes scheme were defined. The training and test sets of samples were tested for representativeness and separability based on their calculated statistical parameters. These tests have shown that: Class 1 and Class 4 have heavily overlapped each other; Class 6 has completely contained Class 5; Class 7 and Class 8 have heavily overlapped each other; Class 10 has completely contained Class 9 and Class 3. These tests pointed out to a direction of possible modification of "Initial classification scheme". The algorithm of this modification is presented (Table 2). Using this algorithm, we created new training and test sets, which were combination of one or more classes from the Initial sets of samples. Having received the new sets, we performed the same statistical tests of representativeness and separability. It led us to defining a new classification scheme consisted of 5 classes. The algorithm of this modification is presented (Table 3).

Table 1: The Initial classification scheme - 12 types of vegetation communities and soil

Class number	Full name of vegetation communities and soil types
Class 1	Chal meadow / Reedbed wetland
Class 2	Chal meadow / Tamarix scrub (Tamarix)
Class 3	Coastal zone semi desert
Class 4	Phragmites australis reedbed wetland (Phragmaties australis)
Class 5	Salsola ericoides
Class 6	Salsola nodulosa
Class 7	Salsola nodulosa / Artemesia lerchiana
Class 8	Salsola Nodulosa /Grasses
Class 9	Semi desert vegetation, Kalidium caspicum (Kalidium caspicum)
Class 10	Semi desert scrub alhagi dominated (Alhagi pseudoalhagi)
Class 11	Bare ground
Class 12	Salsola nodulosa / Bare ground

Table 2: Modified classification scheme (7 classes) of vegetation communities and soil types

Class number	Full name of vegetation communities and soil types
Class 7_1	Init. Class1 + Init. Class 4 - Chal meadow / reedbed wetland + Phragmites australis reedbed wetland
Class 7_2	Init. Class 2 - Chal meadow / Tamarix scrub
Class 7_3	Init. Class 3 + Init. Class 9 + Init.Class 10 - Coastal zone semi desert + Semi desert vegetation, Kalidium caspicum + Semi desert scrub alhagi dominated
Class 7_4	Init. Class 5 + Init. Class 6 - Salsola ericoides + Salsola nodulosa
Class 7_5	Init. Class 7 + Init. Class 8 - Salsola nodulosa / Artemesia lerchiana + Salsola Nodulosa / Grasses
Class 7_6	Init. Class 11 - Bare ground
Class 7_7	Init. Class 12 - Salsola nodulosa / Bare ground

Table 3: "Optimized scheme" (5 classes) of vegetation communities and soil types

Class number	Full name of vegetation communities and soil types
Class 1	Init. Class 1 + Init. Class 4 - Chal meadow / reedbed wetland + Phragmites australis reedbed wetland
Class 2	Init. Class 2 - Chal meadow / Tamarix scrub
Class 3	Init. Class 5 + Init. Class 6 + Init. Class 7 + Init. Class 8 - Salsola ericoides + Salsola nodulosa + Salsola nodulosa / Artemesia lerchiana + Salsola Nodulosa /Grasses
Class 4	Init. Class 11 - Bare ground
Class 5	Init. Class 12 - Salsola nodulosa / Bare ground

We called this scheme as "Optimized classification scheme." Having received new sets, we performed the same statistical tests for representativeness and separability, which show the advances have come using new optimization of classification scheme. There still was a small overlapping between Class 5_1 and Class 5_2, Class 5_4 and Class 5_5. However, it was a high probability that after neural classifier would have been trained, it would be able to overcome these uncertainties and clearly recognize thin transitions between objects from these classes. Subsequent testing of quality of collected samples was going to be performed during classifier training procedure.

5. The Main Results of the Classifiers' Learning

Training of a neural network requires that the user specifies the network structure and sets the learning parameters. Heuristics proposed by a number of researchers to determine the optimum values of network parameters were compared (Kanellopoulos and Wilkinson, 1997, Kavzoglu, 2001 and Kavzoglu and Mather, 2003). The input layer consisted of 4 neurons, corresponding to four spectral channels of IKONOS satellite data: we used the red, green, blue, and near infrared (NIR) channel. The hidden layer had 25 neurons and the output layer had 12 neurons for the "Initial set", 7 for "Modified set" and 5 for "Optimized set", respectively. An activation function was hyperbolic tangent.

The back propagation algorithm was used for neural network training. A network structure of 4-25-12 was trained with the learning parameters listed in Table 4. For evaluation of training process, we used the following quality parameters:

- The Mean Square Error **(MSE);** Having reached *threshold* set for the **MSE** level of 0.01 the training process was stopped;
- The Correlation coefficient **(*r*)**, which reflects the degree of correlation between directions of changing of real and desired outputs of the neural network;
- **% Error** - Error per element of the neural network.

The accuracy of classification has traditionally been measured by the overall accuracy by generating a confusion matrix and determining accuracy levels by dividing the total number of correctly classified pixels (sum of major diagonal of confusion matrix, also called *actual agreement*) by the total number of reference pixels. The accuracy of classification has traditionally been measured by the overall accuracy by generating a confusion matrix and determining accuracy levels by dividing the total number of correctly classified pixels (sum of major diagonal of confusion matrix, also called *actual agreement*) by the total number of reference pixels.

Table 4: Optimum setting of network structure and learning parameters

Parameters	Choice
Initial weight range	[0, 0.05]
Number of input nodes	4
Number of hidden layers	1
Number of hidden nodes	25
Learning rate between input and hidden layers	0.5
Momentum term between input and hidden layers	0.7
Learning rate between hidden and output layers	0.25
Momentum term between hidden and output layers	0.7
Type of activation function	hyperbolic tangent
Error threshold er_{thresh}	0.01

When we reached the maximal number of iterations and the training process was stopped, we compiled confusion matrix with the results of recognition of samples from training sets (so called self-testing procedure). We estimated the common degree of correctness CDC (the level of classification accuracy) by the following formula:

$$CDC = 100\% \times (N_{CCS} / N_{total})$$

Equation 1

Where, N_{CCS} - number of correctly classified samples; N_{total} - total number of samples. ANNs results obtained from the configuration given in Table 4 for training and test samples are presented in Table 5 and Table 6, respectively.

5.1 Training Results - the Initial Classification Scheme (12 Classes)

To analyze the results, confusion matrices were generated using the three classification schemes. The confusion matrices computed on the results of the classifier training of "Initial classification scheme" have shown that: **CDC** = 91.63% (for training samples) (Table 5);

CDC = 81.39% (for test samples) (Table 6).

CDC was quite high; however, review that is more careful showed that this result was based on high results of recognition of samples from three densely populated Classes with numbers: 1, 2 and 6 from the Initial classification scheme (see Table 1). For Class 3 and Class 10 classification results were found as unpredictable. Class 7 shielded Class 8. Results for Class 11 and Class 12 were found as high as for Class 1, Class 2 and Class 6. Though the maximum number of iterations was reached, the **MSE** value was far from its *threshold* set of 0.01 and the correlation coefficient **r** value was also far from optimal (of 1), and the **% Error** value also was high enough. All these factors pointed out that the training procedure being performed using this classification scheme was not successful. Furthermore we tested this classifier for samples of test sets, which were not used in the training procedure. The results were worsened as we had expected. The value of **CDC** was equal to 81.39 and the main character of shortcomings remained the same (Table 6). These results proved conclusions about inadequacy of "Initial classification scheme" which were based on the results of preliminary tests.

Table 5: ANN results for three classification schemes for training samples

Quality parameters	**Initial scheme (12 classes)**	**Modified scheme (7 classes)**	**Optimized scheme (5 classes)**
Iteration	25000	25000	15000
Mean Square Error (MSE)	0.0382	0.0417	0.0361
Correlation coefficient (r)	0.7822	0.8898	0.9632
Error per element of the neural network (% error)	1.2823	1.4983	1.5031
Common degree of correctness (CDC), %	91.63	94.61	96.37

Table 6: ANN results for three classification schemes for test samples

Quality parameters	**Initial scheme (12 classes)**	**Modified scheme (7 classes)**	**Optimized scheme (5 classes)**
Mean Square Error (MSE)	0.0795	0.0983	0.0745
Correlation coefficient (r)	0.6960	0.7734	0.9107
Error per element of the neural network (% error)	1.9938	2.5686	2.0283
Common degree of correctness (CDC), %	81.39	85.8	91.45

5.2 Training Results - the Modified Classification Scheme (7 Classes)

The confusion matrices computed on the results of the classifier training of "Modified classification scheme" have shown that:

CDC = 94.61% (for training samples) (Table 5);

CDC = 85.8% (for test samples) (Table 6).

As it was expected, the value of **CDC** increased but it still was much less, than desired *threshold* set of 90%. Besides, it was revealed that:

- The **MSE** value increased (in comparison with analogous self-testing of the classifier for Initial classification scheme);
- Learning procedure was not robust as the correlation coefficient (r) value diminished to a small value;
- The Error per element (**% Error**) value increased too.

In the whole, the results of testing of the classifier trained on "Modified classification scheme" approved the statistical tests results about the necessity for transferring it to 5 class's classification scheme called as "Optimized classification scheme".

5.3 Training Results. The Optimized Classification Scheme (5 Classes)

The confusion matrices computed on the results of the classifier training of "Optimized classification scheme" (5 classes) have shown that:

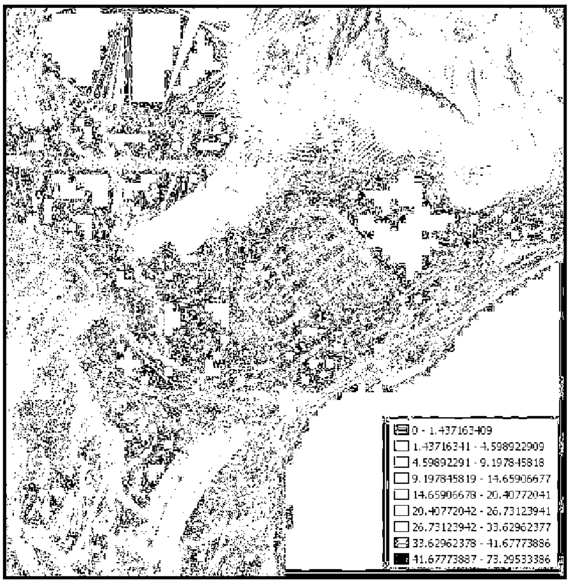
CDC = 96.37% (for training samples) (Table 5);

CDC = 91.45% (for test samples) (Table 6)

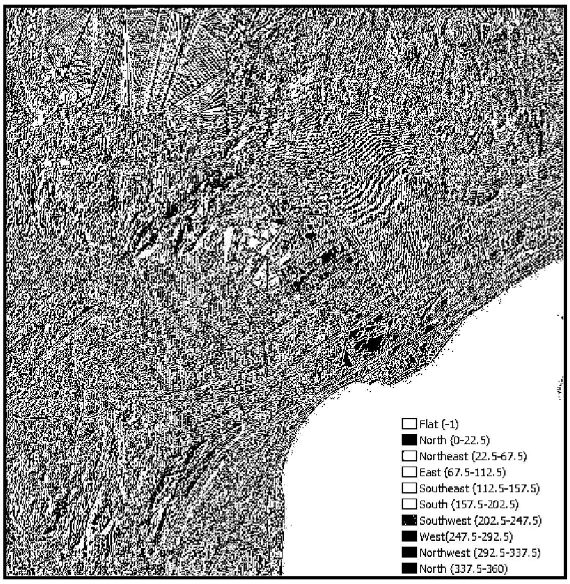
These results are a good enough. These results have shown a little overlapping between Class 1 and Class 2. However, it is in an acceptable extent. There are still some shortcomings concerning the quality parameters. The value of **MSE** is remarkably higher than the minimum threshold value of 0.01. However, the high value of correlation coefficient **(r)** reflects the fact that the training process has had robust character at this stage. This fact points out that there have been realized certain regularities in the learning process and adapting of the classifier to these regularities is going on successfully. Summarizing the classifier testing results, we can conclude that beside some realized uncertainties in recognizing of test samples and non-ideal values of quality parameters being received, the "Optimized" neural classifier has been trained properly. Results produced by other two classifiers ("Initial" and "Modified") on test datasets were found lower than 90% threshold of **CDC** (Table 6.). In comparison, a *Maximum Likelihood classifier (MLC)* using the same datasets produced an overall classification accuracy (common degree of correctness) Table 7: This result demonstrates that levels of classification accuracy achieved by artificial neural networks is higher than those generated by the statistical classifiers. The artificial neural networks designed using the guidelines recommended in this paper could identify the rare vegetation with around 3% better accuracy than the *Maximum Likelihood Classifiers (MLC)*. In case of arising difficulties in production period, additional geospatial data (DTM and its derivable, other topographical, hydrological data as well as land use information and etc.) could be involved into the process for ANNs classification and the problems of recognition of objects would be solved (Figure 2). There are another methods which determinate the quality of classification products. Among them a visual analysis of the results of the classification is very important.

Table 7: MLC results for three classification schemes for test samples

Quality parameter	Initial scheme (12 classes)	Modified scheme (7 classes)	Optimized scheme (5 classes)
Common degree of correctness (CDC), %	78.51	83.64	88.93



a) Ancillary data: Slope

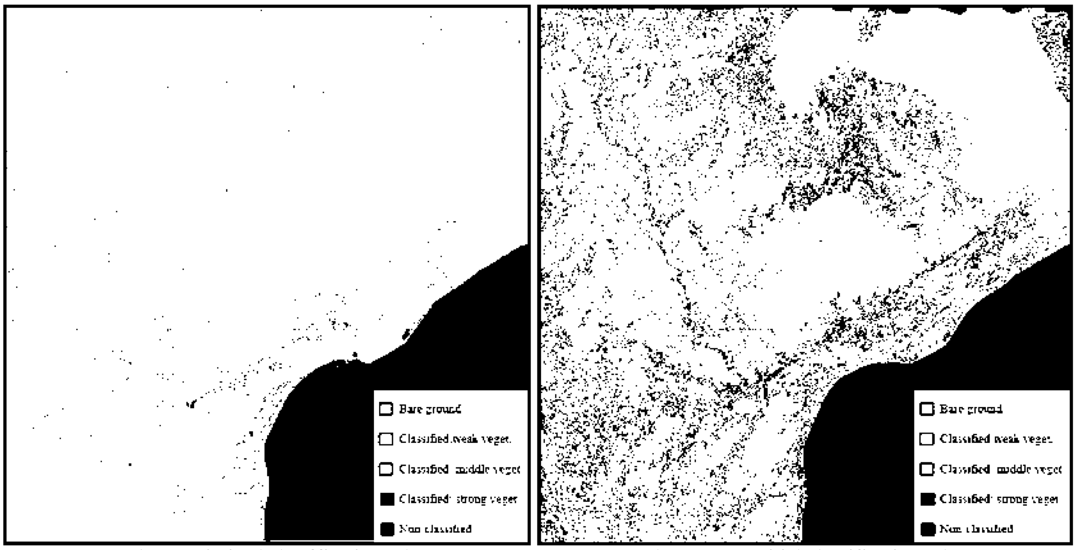


b) Ancillary data: Aspect

Figure 2: Ancillary data

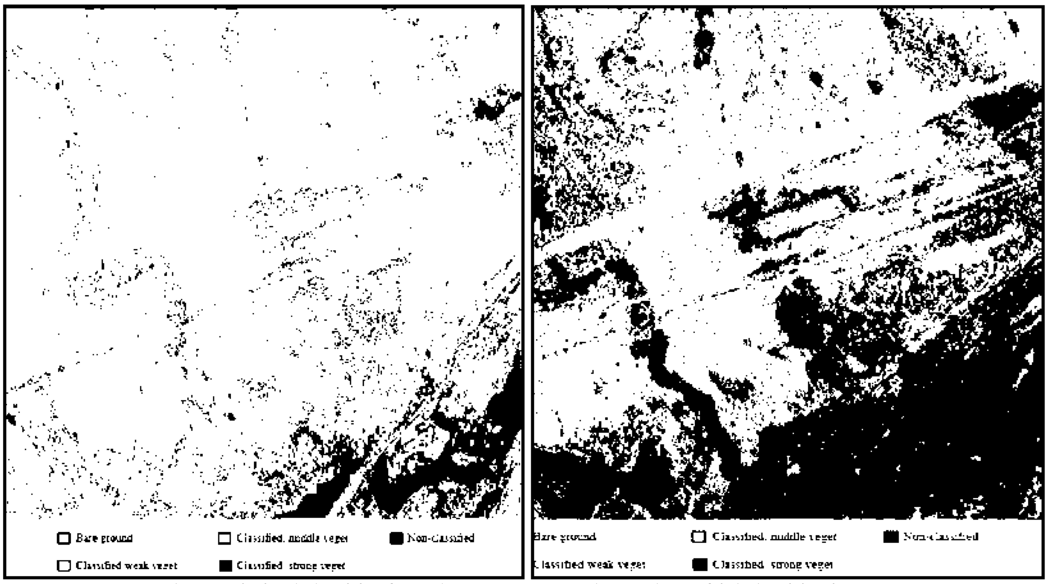
6. Visual Analysis of Classification Results

A full scene of IKONOS satellite image covering area about 100 km^2 was used in these investigations. After completion of training procedures, a datasets covering this scene were classified by neural network according to all three classification schemes described above. The resulting images representing themselves as thematic rasters were analyzed. Using the GeoInformation technologies, amount of non-classified and not exactly classified pixels was estimated. Below the results of this analysis are presented. The various aspects are highlighted by presenting pairs of images: the first one is classification results of the classifier trained on the Optimized classification scheme and the second is a product of the classifier trained on the Initial classification scheme. A general view of the products of these classifiers on overall scene is presented (Figure 3). The classified pixels are presented in the light grey color and the non-classified pixels are presented in the black color. It is easy to notice that the thematic raster representing results of the classification on the Initial scheme has been heavily distorted. There is one difference that is more characteristic: thin transitions between high-density vegetation communities (Figure 4).



a) On the "Optimized classification scheme" b) On the "Initial classification scheme"

Figure 3: Classification results: non-classified pixels are represented in the black color. Fragment 1



a) On the "Optimized classification scheme" b) On the "Initial classification scheme

Figure 4: Classification results: non-classified pixels are represented in the black color (Fragment 2)

Kanellopoulos, I., and Wilkinson, G., 1997, Strategies and Best Practice for Neural Network Image Classification. *International Journal of Remote Sensing*, 18, 711–725.

Kavzoglu, T., 2001, An Investigation of the Design and use of Feed-Forward Artificial Neural Networks in the Classification of Remotely Sensed Images. *PhD thesis*, School of Geography, The University of Nottingham, UK 2001.

Kavzoglu, T., and Mather P., 2002, The Role of Feature Selection in Artificial Neural Network Application. *International Journal of Remote Sensing*, 23, 2919–2937.

Kavzoglu, T., and Mather P., 2003, The use of Back propagating Artificial Neural Networks in Land Cover Classification, *International Journal of Remote Sensing*, 24, 4907–4938.

Oliveira, C., Mather, P., and Aplin P. P., 2003, Improving Artificial Neural Network Performance by Temporal-Spectral Features for Agricultural Crop Classification". *17th European Simulation multiconference*, Nottingham Trent University, Nottingham, England.

Paola, J. D., and Schowengerdt, R. A., 1995b, A Review And Analysis Of Back Propagation Neural Networks For Classification Of Remotely-Sensed Multi-Spectral Imagery. *International Journal of Remote Sensing*, 16, 3033–3058.

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

Zhang, M., and Scofield, R. A., 1994, Artificial Neural Network Techniques For Estimating Heavy Convective Rainfall and Recognizing Cloud Mergers. *International Journal of Remote Sensing*, 15, 3241-3261.

Walthall, C., Dulaney, W., Anderson, M., Norman, J., Fang, H., and Liang, S., 2004, A Comparison of Empirical and Neural Network Approaches for Estimating Corn and Soybean Leaf Area Index from Landsat ETM+ Imagery. *Remote Sensing Environment,* 92, 465-474.