NEURAL AND ANT COLONY OPTIMIZATION VERSUS STATISTICAL MODELS FOR SUPERVISED CLASSIFICATION OF MULTISPECTRAL REMOTE-SENSING IMAGERY

Elena-Cătălina NEGHINĂ[1], Victor-Emil NEAGOE[2], Radu-Mihai STOICA[3], Adrian Dumitru CIOTEC[4]

This paper comparatively evaluates performances of Computational Intelligence and statistical approaches for the task of classification in multispectral satellite images. The considered Computational Intelligence approaches are represented by Artificial Neural Networks (Multilayer Perceptron - MLP and Radial Basis Function - RBF) and the Swarm Intelligence technique (Ant Colony Optimization - ACO). For benchmarking, statistical classifiers have been considered: Nearest Neighbour (NN) and K-means. The considered techniques have been evaluated using two Landsat images: one image of the city of Bucharest (with 4 classes) and the second image for the city of Kosice, Slovakia (with 7 classes). The best recognition score for the Bucharest test image was obtained with RBF (84.78%) and the best recognition score for Kosice test image was obtained with MLP (94.41%).

Keywords: multispectral imaging, geoscience and remote sensing, pattern recognition, supervised learning, Artificial Neural Networks (ANN), Multilayer Perceptron (MLP), Radial Basis Function networks (RBF), Ant Colony Optimization (ACO), Nearest Neighbour (NN), K-means.

1. Introduction

Modern environmental remote sensing satellite imagery, due to their large volume of high-resolution data, offers greater challenges for automated image analysis. Classification, which extracts useful information from remote-sensing data, has been one of the key topics in remote-sensing studies. Multispectral

[1] PhD student, Depart. of Electronics, Telecommunications and Information Technology, University POLITEHNICA of Bucharest, Romania, e-mail: neghina.catalina@gmail.com
[2] Prof., Depart. of Electronics, Telecommunications and Information Technology, University POLITEHNICA of Bucharest, Romania
[3] PhD student, Depart. of Electronics, Telecommunications and Information Technology, University POLITEHNICA of Bucharest, Romania
[4] PhD student, Depart. of Electronics, Telecommunications and Information Technology, University POLITEHNICA of Bucharest, Romania

imagery classification involves the grouping of image data into a finite number of discrete classes. Hence, the output from a multispectral image classification system is a thematic map in which each *n*-dimensional pixel in the original image has been classified into one of *M* classes. The algorithms are based on the fact that each class of materials, in accordance to its molecular composition, has its own spectral signature. For example, classification is frequently carried out to obtain land use/cover information.

Applications are needed both for remote sensing of urban/suburban infrastructure and socio-economic attributes as well as to detect and monitor land-cover and land-use changes. Conventionally, pattern recognition in remote sensing imagery has been mainly based on classical statistical methods and decision theory. Recently, several artificial intelligence approaches have been used with promising degrees of success in remote sensing image analysis [3][5][6][7][10]. By observing and studying natural systems, new algorithmic models able to solve increasingly complex problems have been developed. Enormous success has been achieved during the last decade through modelling of biological and natural intelligence, resulting in so-called „intelligent systems". These *nature-inspired intelligent techniques* are included under *Computational Intelligence (CI)* [4]. This paper compares the results of *Artificial Neural Networks (ANN)* and *Ant Colony Optimization (ACO) model* for the task of classification, also including statistical techniques for benchmarking.

2. Algorithm description

As shown in *Fig. 1*, we further consider the supervised classification algorithms for multispectral image analysis belonging to one of three categories presented below:

- **Artificial Neural Networks (ANN)**, consisting of either Multilayer Perceptron (MLP) or Radial Basis Function neural network (RBF).
- **Swarm Intelligence (SI)** represented by Ant Colony Optimization (ACO) classifier.

Statistical classifiers, consisting of one of the following two algorithms: Nearest Neighbour (NN) and k-Means.

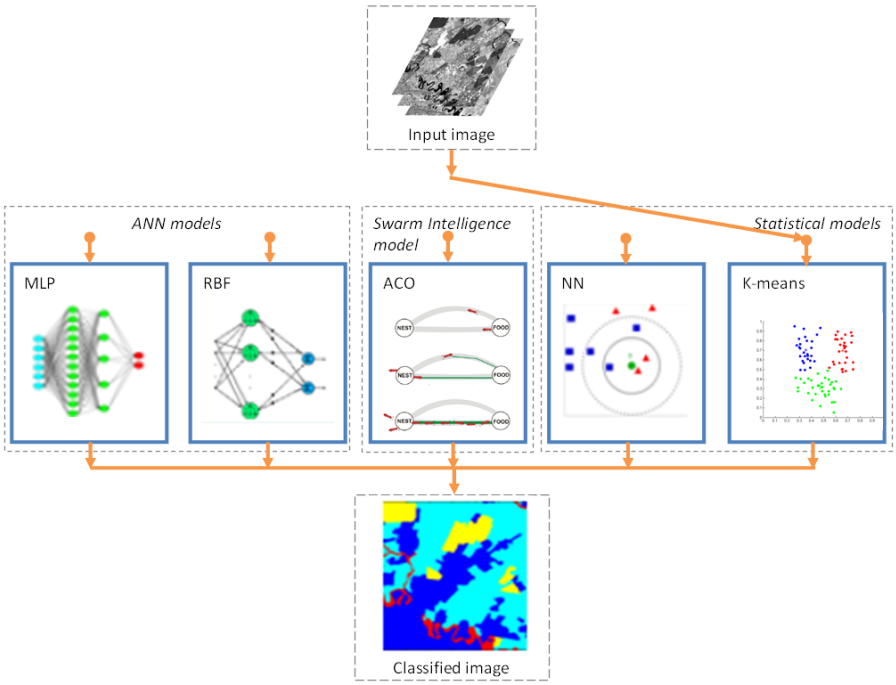


Fig. 1. Flowchart of the considered supervised classification model for multispectral remote-sensing imagery

2.1. Artificial Neural Networks (ANN)

The advantages of applying ANN for classification of satellite images are the following [2][3]:

- neural classifiers do not require initial hypotheses on the data distribution and they are able to learn non-linear and discontinuous input data;
- architecture of neural networks is very flexible, so it can be easily adapted for improving the performances of a particular application;
- the neural classifiers are generally more accurate than the statistical ones.

Multilayer Perceptron (MLP) is a supervised feed-forward artificial neural network, which can distinguish data that is not linearly separable. Most of the MLP networks use back-propagation as learning algorithm, such that the error of the output layer (compared to the desired outputs) is propagated backwards through the network and influences the weights of the various intermediary layers. The MLP is a powerful neural classifier with very good performances for a wide range of data mining applications

Radial Basis Function (RBF) networks have a supervised two-layer neural architecture similar to that of MLP, with the important difference that each RBF unit implements a radial-activated function and the output units implement a weighted sum of the hidden unit outputs. Due to their non-linear approximation

properties, RBF networks are able to model complex mappings, which MLP networks can only model by means of multiple intermediary layers. In pattern classification applications, the Gaussian function is preferred, and even mixtures of Gaussians have been considered. Generally, the RBF nets lead to better pattern recognition performances than those of MLP.

2.2. Swarm Intelligence (SI)

In a colony of social insects, such as ants, bees, wasps and termites, each insect usually performs its own tasks independently from other members of the colony. However, the tasks performed by different insects are related to each other in such a way that the colony, as a whole, is capable of solving complex problems through cooperation. Swarm Intelligence (SI) techniques [1] belong to the field of computational intelligence [4]; they are essentially complex multi-agent systems where low-level interactions between individual agents result in complex behaviour of the whole. SI models have been used to solve complex pattern recognition problems; since then they promise to be very good candidates for knowledge discovery in remote-sensing image databases.

Ant Colony Optimization (ACO) as a main component of SI is inspired by the behaviour of real ants, namely by the fact that they are capable of finding the shortest path between a food source and the nest (adapting to changes in the environment) without the use of visual information [1][5][8][11]. The main idea is the indirect communication between the ants by means of chemical pheromone trails, which enables them to find short paths between their nest and food. In nature, ants usually wander randomly, and upon finding food return to their nest while laying down pheromone trails. If other ants find such a path (pheromone trail), they are likely not to keep travelling at random, but to instead follow the trail, returning and reinforcing it if they eventually find food. However, as time passes, the pheromone starts to evaporate. The more time it takes for an ant to travel down the path and back again, the more time the pheromone has to evaporate (and the path to become less prominent). A shorter path, in comparison will be visited by more ants (can be described as a loop of positive feedback) and thus the pheromone density remains high for a longer time. Ant Colony Optimization (ACO) is implemented as a team of intelligent agents which simulate the ants' behaviour, walking around the graph representing the problem to solve using mechanisms of cooperation and adaptation.

The graph associated to ACO model can be described by the following relations. From its starting node, an ant iteratively moves from one node to another. When being at a node, an ant chooses to go to an unvisited node at time *t* with a probability given by:

$$p_{i,j}^{k}(t) = \frac{[\tau_{i,j}(t)]^{\alpha}[\eta_{i,j}(t)]^{\beta}}{\sum_{l \in N_i^k}[\tau_{i,l}(t)]^{\alpha}[\eta_{i,l}(t)]^{\beta}} \quad (1)$$

- N_i^k is the feasible neighbourhood of node *i* for ant_k
- $\tau_{i,j}(t)$ is the pheromone value on the edge *(i, j)* at the time *t*
- α is the weight of pheromone
- $\eta_{i,j}(t)$ is a priori available heuristic information on the edge *(i, j)* at the time *t*
- β is the weight of heuristic information
- $\tau_{i,j}(t)$ is updated according to equation (2)

$$\tau_{i,j}(t) = \rho \cdot \tau_{i,j}(t-1) + \sum_{k=1}^{n} \Delta\tau_{i,j}^k(t) \tag{2}$$

where $\Delta\tau_{i,j}^k(t)$ is the pheromone quantity left by ant_k on arc *(i, j)*, at moment *t*, and is computed according to equation (3).

$$\Delta\tau_{i,j}^k(t) = \begin{cases} \dfrac{Q}{L_k(t)}, \textit{ if arc } (i, j) \textit{ is chosen by } ant_k \\ 0, \textit{ otherwise} \end{cases} \tag{3}$$

- ρ is the pheromone trail evaporation rate ($0 < \rho < 1$)
- *n* is the number of ants
- *Q* is a constant for pheromone updating
- L_k represents the length of the path chosen by the ant_k

Parpinelli et al. [11] were the first to propose ACO for discovering classification rules using a system called Ant-Miner. Their study demonstrates that Ant-Miner produces better accuracy and simpler rules than some decision tree methods. Ant Miner [11] has been designed to be used with loose discrete sets of values. The satellite imagery field offers a very dense input space for each dimension and, in order to adapt the algorithm to the density of the input space, Liu et al. [5] has proceeded with a discretization process through which discrete intervals [like (0-13),(14-25),(26-41),...,(240-255)] are defined for each band by determining break points based on entropy. These intervals are then given as discrete options to the ants in order to form the terms of the rules. $Term_i$ could be expressed as "the value along dimension *i* falls in the respective interval".

The present paper proposes an ACO-based classifier for satellite images, with several distinct particularities compared to the [11] and Liu [5] algorithms. These specific differences are pointed out in the concluding remarks.

The ACO algorithm for classification builds a solution to the classification task iteratively. The solution is a set of rules, each rule being associated with an output class. The ACO algorithm for classification attempts to find a set of rules of the following type:

if ($term_1$ *and* $term_2$ *and... and* $term_N$) *then* ($class_k$).

The algorithm is a sequence of steps for each dimension to discover the corresponding rules, by discovering both thresholds *a* and *A* of each term.

2.3. Statistical techniques

We have considered for comparison two well-known statistical algorithms: Nearest Neighbour (NN) and K-means.

3. Experimental Results

For experiments we have used 2 databases: a Landsat image *of the city of Bucharest (with 4 classes) and a Landsat image of the city of Kosice (with 7 classes).*

3.1. Experimental results for Landsat image of Bucharest

The data for this study consists of a Landsat 7 ETM+ (Landsat **E**nhance **T**hematic **M**apper) *image of the city of Bucharest (Romania) and its environs. The whole image consists of* 116.963 pixels (343 x 341 pixels).

A pseudo-colour representation of the satellite image selection is shown in *Fig. 2*. The performances of the considered techniques have been evaluated using as a reference the CORINE maps. These maps have been created by the European Environmental Agency (EEA) under the CORINE Land Cover (CLC2000) project [12]. The CORINE maps are generated based on the visual interpretation of experts, sometimes helped by aerial photographs, topographic maps and other additional information. The map labelled in 4 classes (*class1: artificial surfaces, class2: agricultural areas, class3: forests, class4: water*) using CLC2000 is shown in *Fig. 3*.

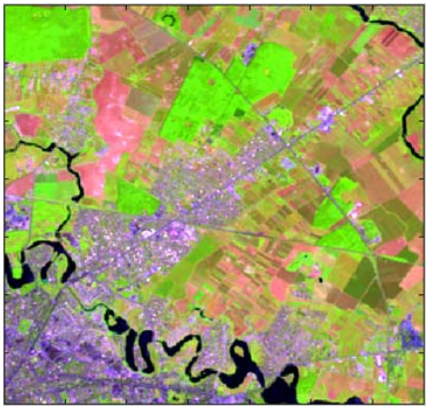


Fig. 2. Landsat ETM+ image
Represented in pseudo-color space
(Red = Band 5, Green = Band 4, Blue = Band 1)

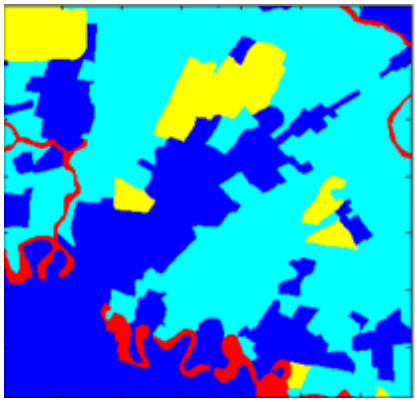


Fig. 3. Reference map of the Landsat image
using CLC classes

The pixel distribution of the four classes, according to the reference map shown in *Fig. 3*, is described in *Table 1*.

Table 1

Percentage of pixels assigned to classes

Class	Pixels number	Percentage
[1] Artificial surfaces	41697	35.65 %
[2] Agriculture area	59468	50.84 %
[3] Forest	10735	9.18 %
[4] Water	5063	4.33 %
	TOTAL = 116963	**100 %**

The image pixels have been divided in a training set and a test set. The training set contains 800 pixels from each class, adding up to a total of 3.200 pixels. Their positions are shown in black in the CLC reference map, in *Fig. 4*. The test set contains 116.963 pixels (those 3.200 pixels used for training are part of the 116963 pixels used for test).

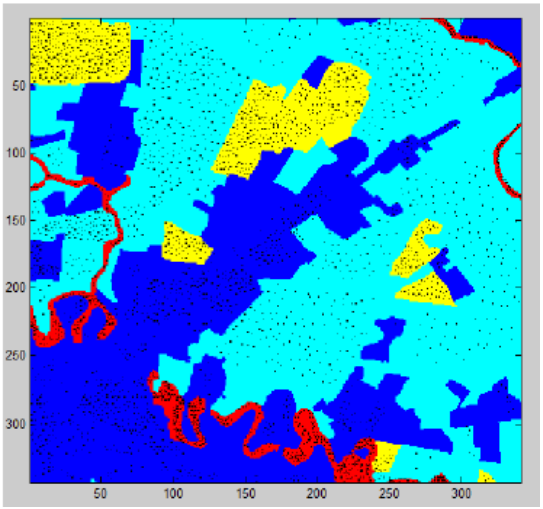


Fig.4. Selected pixels (black) in the CLC reference map

3.1.1. Results obtained with ACO

For ACO, we have considered the following parameters: N = number of epochs; P = raster quantum; M = maximum number of unclassified vectors; L = minimum number of vectors belonging to a rectangular hyper-parallelepiped.

Experiment 1. For the first set of experiments, we set N = 25, M = 25, L = 25 and P having the values {0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.15, 0.20, 0.33}. The resulting classification scores are represented in *Fig. 5*.

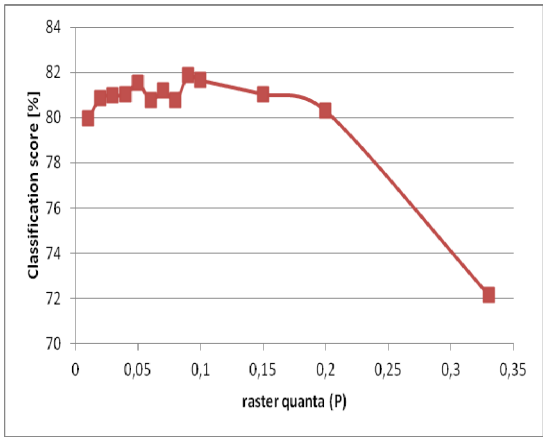


Fig.5. Classification score for test set, for different values of P (M = L = 25, N = 25)

Experiment 2. For the second set of experiments we kept the raster quanta constant (P = 0.01, is the raster quanta for which the best classification score in *Experiment1* was obtained), M = 25, L = 25 and the number of epochs (N) has been varied between 5 and 40. The classification scores are shown in *Fig. 6.*

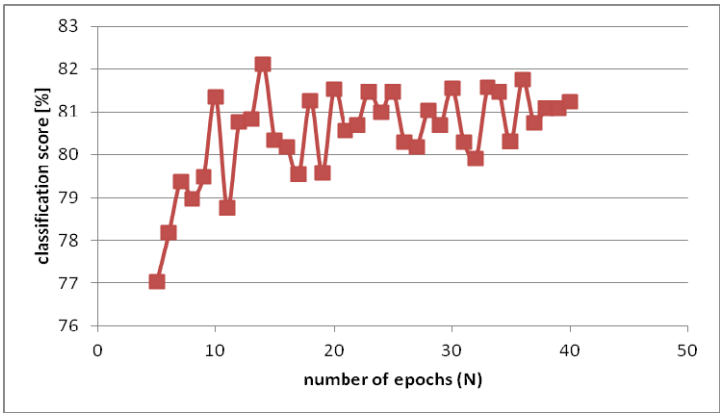


Fig. 6. Classification score for test set, for different values of N (M = L = 25, P = 0.01)

Experiment 3. For the third set of experiments we kept constant the raster quanta (P = 0.01, is the raster quanta for which the best classification score in *Experiment1* was obtained) and the number of epochs (N = 35 is the number of epochs for which was obtained the best classification score in *Experiment2*) and the variables M (the number of unclassified vectors used for the last rule) and L (the minimum number of vectors covered by a rule) have been varied between 7 and 41 (with step 2). The classification scores have improved, as shown in *Fig. 7.*

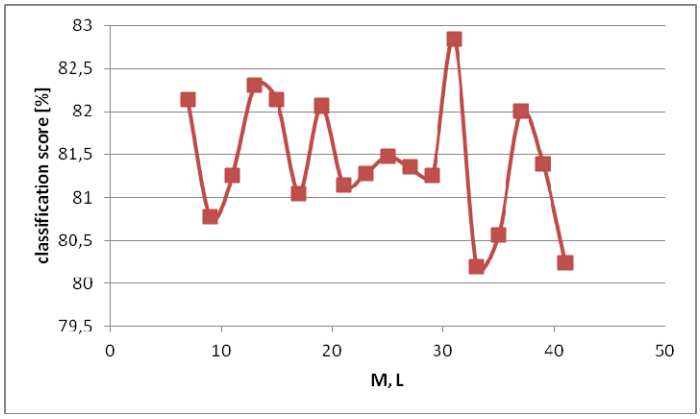


Fig.7. Classification score for test set, for different values of M and L (P = 0.01, N = 35)

From these three experiments, the best classification score obtained for the test set was **82.84%** for *Experiment 3* (P = 0.01, N = 35, M = L = 31).

3.1.2. Results obtained with MLP

A MultiLayer Perceptron network with one hidden layer was used in the experiment. The number of neurons from hidden layer has been varied between 2 and 50 (with step 3). The resulting classification scores are represented in *Fig. 8.* The best result obtained with MLP was **84.75%**, for 23 neurons on the hidden layer.

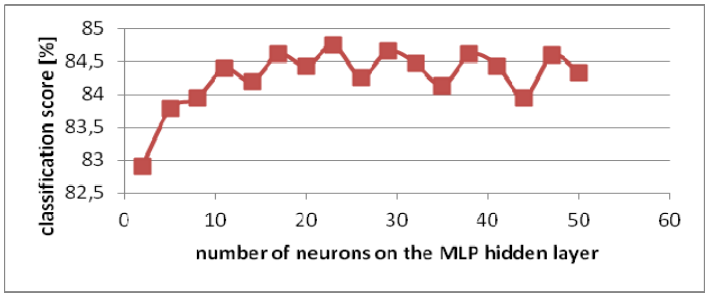


Fig. 8. Classification score for test set, using MLP, for different numbers of neurons on hidden layer

3.1.3. Results obtained with RBF

For the classification using RBF, the *spread* parameter has been varied between 0.5...100, with step 0.5. The resulting classification scores are represented in *Fig. 9.* The best result obtained with RBF was **84.78%**, for *spread* =10.5.

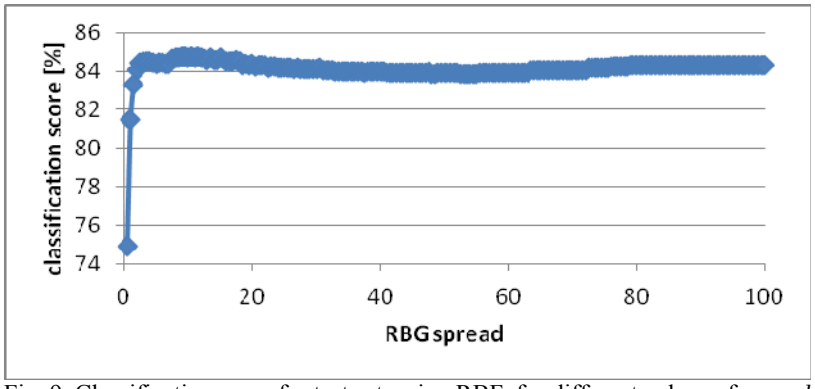


Fig. 9. Classification score for test set, using RBF, for different values of *spread*

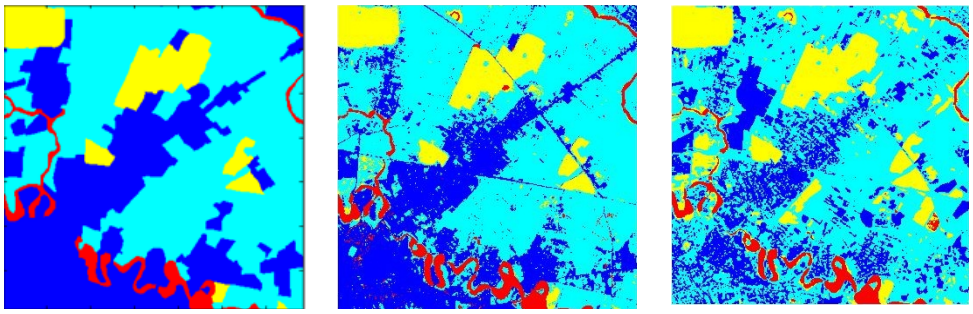
3.1.4. Comparative results

The comparative results are given in *Table 2* and *Fig.10.*

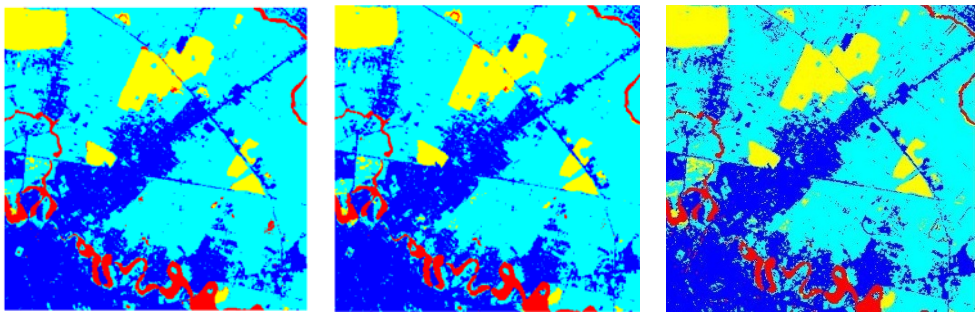
Table 2

Recognition score for Bucharest Landsat dataset

Classifier	Recognition score for test set [%]	Specifications
ACO	82.84	P = 0.01, N = 35, M = L = 31
MLP	84.75	23 neurons in the hidden layer
RBF	**84.78**	*spread* = 10.5
1-NN	83.61	
k-Means	65.91	



a) CLC 2000 referance map b) Classification with 1-NN c) Classification with k-Means



d) Classification with MLP e) Classification with RBF f) Classification with ACO

Fig. 10. Classification for Bucharest Landsat image

The best result was obtained with RBF (84.78%). The confusion matrix obtained for RBF is presented in *Table 3.*

Table 3

Confusion matrix for RBF (Bucharest LANDSAT dataset)

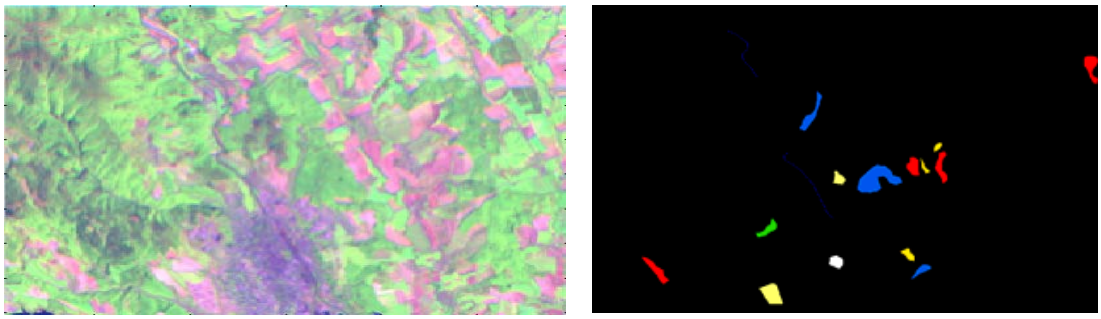
Estimated class	Real class			
	1	**2**	**3**	**4**
1	**71.90**	4.89	1.71	4.95
2	26.06	**93.26**	8.61	12.34
3	0.85	1.25	**89.38**	1.26
4	1.19	0.60	0.30	**81.45**

As shown in the confusion matrix for the overall best classification result, in *Table 4*, the class with the greatest percentage of correctly classified pixels is *Agricultural Areas* (93.26%), while the worst percentage of correctly classified pixels is for class *Artificial Surfaces* (71.90%).

3.2. Experimental results for Kosice Landsat image

The data for this study consists of a Landsat Thematic Mapper (TM) image of the city of Kosice (located in Eastern Slovakia) and its environs [9]. The whole image consists of 368.125 7-dimensional pixels, out of which 6.331 pixels were classified by an expert into seven thematic categories (classes): *1-urban area; 2-barren fields, 3-bushes, 4-agricultural fields, 5-meadows, 6-woods, 7-water.*

In *Figure 11*, there are represented the Landsat image in pseudo-color space and the calibrated pixels used in tests (6.331 pixels).



a) Landsat image represented in pseudo-color space h) Calibration image

Fig. 11. Kosice Landsat image and labeled pixels

The 6.331 labeled pixels have been split in a training set (210 pixels equaly balanced in 7 classes) and test set (the remaining 6.121 pixels). The distribution of all 6.331 vectors in classes is represented in *Figure 12*.

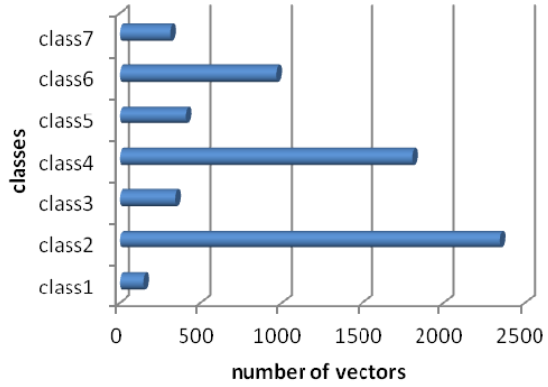


Fig. 12. Distribution of vectors in classes (Kosice Landsat dataset)

The best results obtained with ACO, MLP, RBF, 1-NN and k-Means are in *Table 4*.

Table 4

Recognition score for Kosice Landsat dataset

Classifier	Recognition score for test set [%]	Specifications
ACO	90.07	P = 0.05, N = 35, M = L = 3
MLP	**94.41**	17 neurons in the hidden layer
RBF	93.78	*spread* = 34
1-NN	92.17	
k-Means	87.74	

The best result was obtained with MLP (94.41%). The confusion matrix obtained for MLP is presented in *Table 5.*

Table 5

Confusion matrix for MLP (Kosice LANDSAT dataset)

Estimated class	Real class						
	1	**2**	**3**	**4**	**5**	**6**	**7**
1	**91.53**	0.26	1.29	0.12	0	0.22	0.71
2	0	**99.66**	0	0.05	0	0	0
3	1.69	0	**80.39**	0.46	1.84	1.49	1.43
4	0.85	0	2.25	**95.20**	0	14.73	2.49
5	0	0.08	0	0	**98.16**	0	0
6	0	0	7.07	4.17	0	**83.14**	0
7	5.93	0	9.00	0	0	0.42	**95.37**

As shown in the confusion matrix for the overall best classification result, in *Table 6*, the class with the greatest percentage of correctly classified pixels is *Barren Fields* (99.66%), while the worst percentage of correctly classified pixels is for class *Bushes* (80.39%).

4. Concluding Remarks

a) For satellite multispectral image classification, a variety of models have been considered in this paper: *statistical methods* (1-NN, K-means), *neural networks* (MLP, RBF) and *swarm intelligence methods* (ACO).

b) For *Bucharest Landsat image*, the best recognition score of 84.78% was obtained using the RBF model followed by the scores of 84.75% using MLP, 83.61 % using 1-NN, 82.84% using ACO and 65.91% using K-means.

c) For *Kosice Landsat image*, the best recognition score of 94.41% was obtained using the MLP model followed by the scores of 93.78% using RBF, 92.17% using 1-NN, 90.07% using ACO and 87.14% using K-means.

c) The advantage of nature-inspired intelligent models (ANN and ACO) consists not only in their classification score but also in their flexibility, learning capacity and adaptability to the specific task of remote-sensing image classification.

d) By comparison to existing literature, the proposed ACO classification algorithm introduces novelties regarding the generation of rule terms and the rule pruning section of the algorithm. It sets equidistant points according to the desired granularity; instead of selecting an interval, ants select the lower and upper bound of the interval among those points, in order to create terms of the rule. Regarding the rule pruning, all previous algorithms extend the subspace covered by a rule through elimination of terms from the rule. Rather than deleting terms, the proposed algorithm extends the covered subspace by extending intervals

gradually, from the points selected by the ant towards the absolute upper or lower bound (whichever extends the intervals). If the interval stretches from the absolute minimum to the absolute maximum along the respective dimension, it has the same effect as eliminating the terms regarding that dimension.

R E F E R E N C E S

[1] *A. Abraham, H. Guo, and H. Liu*, "Swarm Intelligence: Foundations, Perspectives and Applications," in Studies in Computational Intelligence (SCI), vol. 26, Berlin: Springer, 2006, pp. 3–25.

[2] *M. Bishop*, "Pattern Recognition and Machine Learning", Springer, New York, 2006.

[3] *F. Del Frate, F. Pacifici, G. Schiavon, and C. Solimini*, "Use of neural networks for automatic classification from high-resolution images," IEEE Trans. Geosci. Remote Sens., vol. 45, no. 4, pp. 800–809, Apr. 2007.

[4] *A. P. Engelbrecht*, "Computational Intelligence", Wiley, Hoboken, NJ, USA, 2002.

[5] *X. Liu, X. Li, L.Liu, J. He, and B. Ai*, "An Innovative Method to Classify Remote-Sensing Images Using Ant Colony Optimization," IEEE Trans. Geosci. Remote Sensing, vol. 46, no. 12, pp. 4198-4208, Dec. 2000.

[6] *V. Neagoe and A. Ropot*, "A New Neural Approach for Pattern Recognition in Space Imagery", in Harbour Protection through Data Fusion Technologies, NATO Science for Peace and Security Series-C: Environmental Security, pp. 283-289, Springer, 2009.

[7] *V. Neagoe and G. Strugaru*, "A Concurrent Neural Network Model for Pattern Recognition in Multispectral Satellite Imagery", in Proc. of the World Automation Congress, 2008 (WAC 2008), Internat. Symposium on Soft Computing in Industry (ISSCI'08), Sept. 28–Oct. 2, 2008, Hawaii, USA, ISBN:978-1-889335-38-4, IEEE Catalog No. 08EX2476.

[8] *V. Neagoe, C. Neghina and M. Neghina*, "Ant Colony Optimization for Logistic Regression and Its Application to Wine Quality Assessment," in Proc. Internat. IEEEAM Conf. Mathem. Models for Eng. Science (MMES'10), Tenerife (Spain), Nov.30-Dec.2, 2010, vol. I, pp. 195-200, ISBN: 978-960-474-252-3.

[9] V. E. Neagoe, and D. Ropot, " *Pattern Recognition in Multispectral Satellite Images Using Concurrent Self-Organizing Modular Neural Networks*", Emerging and Future Technologies for Space Based Operations Support to NATO Military Operations (pp. 7-1 - 7-10) Neuilly-sur-Seine, France: RTO, 2006.

[10] *J. D. Paola, R. A. Schowengerdt*, "A Detailed Comparison of Backpropagation Neural Network and Maximum-Likelihood Classifiers for Urban Land Use Classification," IEEE Trans. Geosci. Remote Sens., Vol. 33, No. 4, July 1995, pp. 981-996.

[11] *R. S. Parpinelli, H. S. Lopes, and A. A. Freitas*, "Data mining with an ant colony optimization algorithm," IEEE Trans. Evol. Comput., vol. 6, no. 4, pp. 321–332, Aug. 2002.

[12] CLC2006 technical guidelines, European Environment Agency, Technical Report No. 17/2007, ISSN 1725–2237, ISBN 978-92-9167-968-3.