Neural Network Approach for Training Models for Multispectral Images

Dragan Stevic
University of Pristina, Kneza Miloša 7
38220 Kosovska Mitrovica, Serbia

Igor Hut
University of Belgrade,Kraljice Marije 16, 11120
Belgrade, Serbia
ihut@mas.bg.ac.rs

Nikola Dojcinovic
MySkin, Inc, Kosovska 17, 11000
Belgrade, Serbia
nikoladojcinovic@gmail.com

Jugoslav Jokovic
University of Nis, Aleksandra Medvedeva 14
18000 Niš, Serbia
jugoslav.jokovic@elfak.ni.ac.rs

ABSTRACT: *In this paper we have presented the neural network-based approach for LANDSAT vegetation classification with multispectral image set. We have given many training models which came out with the measures such as accuracy and a good framework.*

Keywords: Remote Sensing, Neural Networks, Multispectral Images

Received: 29 Janaury 2022, Revised 8 April 2022, Accepted 20 April 2022

DOI: 10.6025/isej/2022/9/1/15-20

1. Introduction

Vegetation classification is an important component in the management and planning of natural resources [1]. Recently, there were efforts to characterize vegetation culture by its reflectance [2,3]. It is shown that vegetation cultures can be distinguished by their reflectance. Remote sensing with multispectral or/and hyperspectral data derived from various satellites in combination with topographic variables is valuable tool in vegetation classification. Based on this requirement, we propose method for automatic vegetation culture recognition based on its reflectance data from multispectral Landsat 7 satellite image sets, using neural networks.

Landsat 7 satellite is equipped with Enhanced Thematic Mapper Plus (ETM+) imaging system. In addition to Thematic Mapper, system that operates on Landsat 4 and Landsat 5, ETM+ is capable of acquiring panchromatic band with 15m spatial resolutions, enhanced radiometric calibration and a thermal IR channel with 60m spatial resolution. Altogether, ETM+ acquires images of 7 wavelength bands, plus panchromatic channel[4, 5], as is presented in Table I. Each channel is represented with corresponding image and it covers 185km width and length of a land surface. Therefore, maximal resolution of usable rectified image area is (12333×12333)pixels for panchromatic, or (6166×6166)pixels for channels 1, 2, 3, 4, 5 and 7. The IR channel 6 usable area resolution is (3083×3083)pixels. Each pixel is represented with 8 bit values, ranging from 0 to 255. Channel images are stored in uncompressed image formats.

Band Number	Spectral range (nm)	Ground resolution
1	450 - 515	30
2	525 - 605	30
3	630 - 690	30
4	750 – 900	30
5	1550 - 1750	30
6	10400 - 12500	60
7	2090 - 2350	30
Pan	520 - 900	15

Table 1. Landsat 7 Channels Properties

2. Methodology

In order to characterize reflectance of vegetation culture represented on an image, construction of a multi-spectral descriptor is proposed. The descriptor is constructed as vector of acquired reflectance values by wavelength bands. Due unavailability of 6^{th} band for civil use, descriptor consists of values of averaged reflected light intensity in period of acquisition for bands 1, 2, 3, 4, 5 and 7. This approach to the descriptor construction quantifies reflected light intensities, quantizing it to non-overlapping and non-continuous bands, yielding vector of six elements with values ranging from 0 to 255.

However, due to information loss caused by the quantization, no information between bands boundaries and averaging within bands, mapping from reflectance to descriptor is not injective, and therefore not invertible. Let's denote mapping *f* from continuous reflectance value R_λ to discrete descriptor D_Λ as:

$$f: R_\lambda \rightarrow D_\lambda,\ f(r_\lambda) = d_\lambda$$

wherer $r_\lambda \in R_\lambda$ and $d_\Lambda \in D_\Lambda$. Due non-invertibility of mapping *f*, it is not possible to define analytical invers mapping f^1 from *f*.

Therefore, mapping *g* is defined as inverse mapping:

$$g: D_\lambda \rightarrow R_\lambda,\ g(d_\lambda) = r'_\lambda$$

where $r'_\lambda \in R_\lambda$, satisficing condition

$$|r'_\lambda - r_\lambda| = 0 \quad (1)$$

In order to compensate nonlinearities and unknown complexity of inverse mapping function, our method employes artificial neural network (ANN) to estimate inverse mapping *g* in order to minimize the difference in (1).

The additional simplification can be achieved with discretization of vegetation cultures to classes. Using finite number of classes, N_{out}, a recognition problem is reduced to a classification problem. A number of classes yields number of neurons in output layer to be equal as a number of vegetation classes. The input layer of the neural network is defined with the descriptor vector length. In order to compensate nonlinearities, the network in the proposed solution has one hidden layer with 3 times more neurons than in the input layer. All neurons are activated with the sigmoid function. If network weights are denoted with Θ, than corresponding cost function of network can be presented as:

$$J(\Theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{k=1}^{10} y_k^{(i)} log\left(h_\Theta(x^{(i)})\right) + (1 - y_k^{(i)}) log\left(h_\Theta(x^{(i)})\right)\right] + \frac{\lambda}{2m}\sum_{l=1}^{3}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}\left(\Theta_{ij}^{(l)}\right)^2 \qquad (2)$$

where m is dimension of training set, h_Θ is hypothesis calculated of x (i), input value of i-th training sample, and y_k is correct response on k-th output, for i-th training sample, and sl number of neurons in l-th layer. Neural network is trained to minimize the cost function $J(\Theta)$ on training set.

It is assumed that each pixel corresponds to a single class. A pixel purity problem is not treated. A classification of such pixels requires additional physical information. The availability of data representing reliable ground truth is often a problem in a remote sensing pattern classification and the problem would be further complicated with mixed pixels. Therefore, special attention was paid to proper training set forming, omitting location positioned on boundaries between classes.

Training set consists of the descriptor vector for a particular pixel as input data and matching vegetation culture for that geographic position, as desired response. The vegetation data was organized in geo-referent map covering same area as Landsat 7 image set. For easier visualization hybrid image was created by placing of translucent geo-referent vegetation map over geo-referenced Landsat 7 image set image of one of channels.

Process of training set formation was semi-automatized with software written for this purpose, presented in Fig. 1. It allows observer/network_trainer to navigate through hybrid image in search for desired vegetation class for training set labeling. Once suitable class is found, observer chooses one location in particular class region. Selecting location by clicking mouse positioned over desired location dialog window pops-up, allowing observer to specify vegetation class present in that particular position. As temporary data pair of an image location and corresponding class are stored in memory. After process of labeling

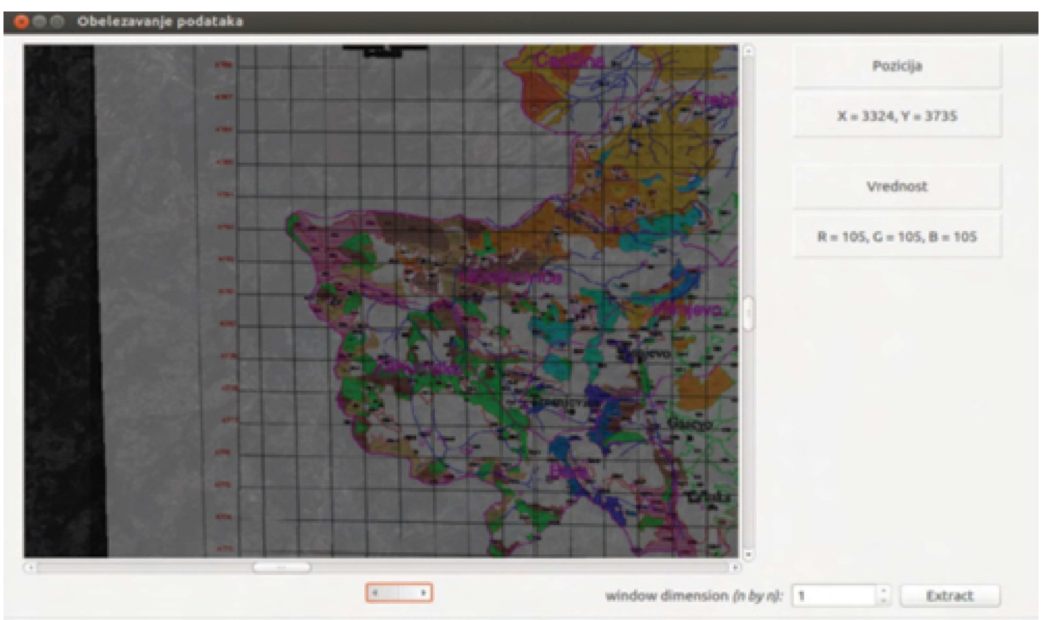
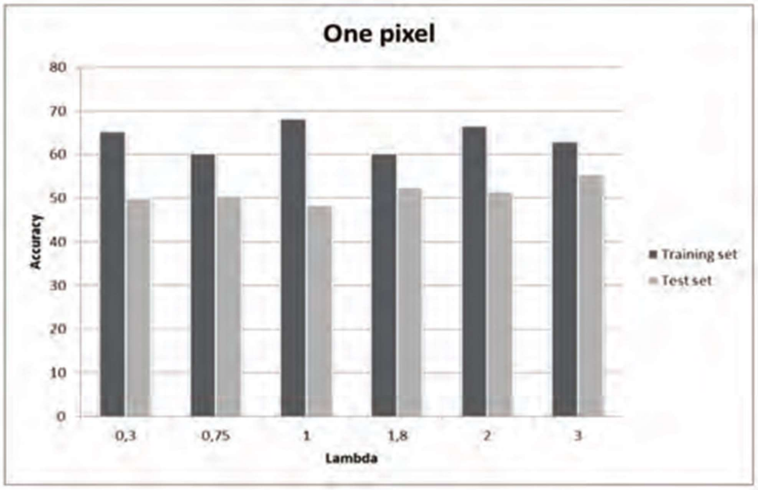
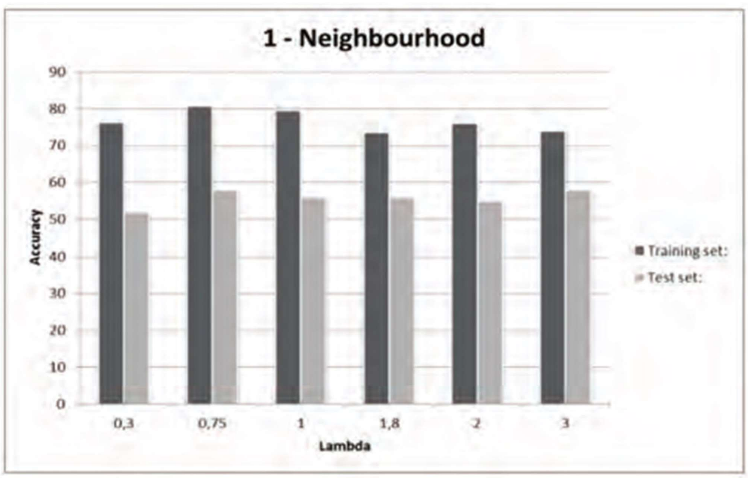
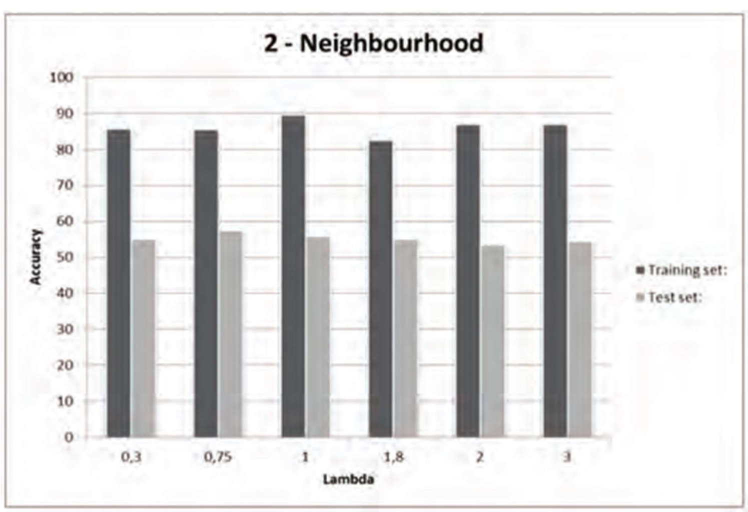


Figure 1. Software for semi-automatic data labelling

feature computing cost for neural network training progressively increases, choosing "one pixel" model over the remaining three would be adequate.A relatively low accuracy is, probably, primarily due to low separability of chosen classes. Precision of terrain map is questionable, spatial resolution in visible bands is 30m, therefore one pixel can encompass area with mixed vegetation [6].







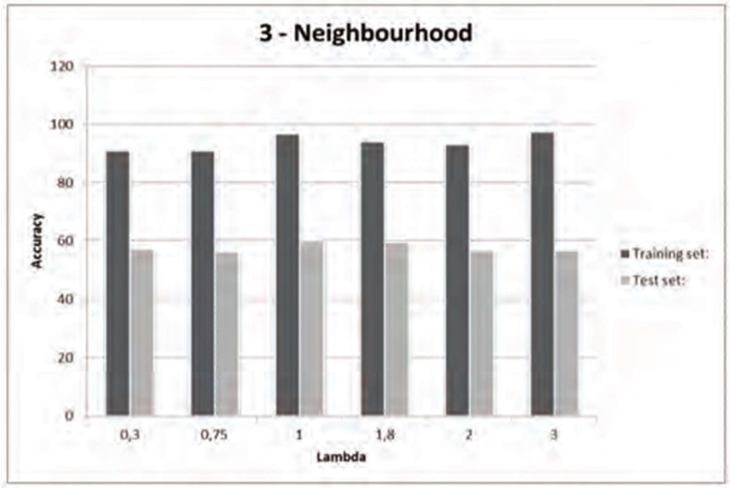


Figure 2.Comparison of model accuracy for the training set and the test set, as a function of regularization parameter λ

1 x 1	Lambda					
	0,3	0,75	1	1,8	2	3
Training set:	65,25	60,25	68,25	60,25	66,5	63
Test set:	50	50,5	48,5	52,5	51,5	55,5

3 x 3	Lambda					
	0,3	0,75	1	1,8	2	3
Training set:	76,25	80,75	79,5	73,5	76	74
Test set:	52	58	56	56	55	58

5 x 5	Lambda					
	0,3	0,75	1	1,8	2	3
Training set:	85,75	85,5	89,5	82,5	87	87
Test set:	55	57,5	56	55	53,5	54,5

7 x 7	Lambda					
	0,3	0,75	1	1,8	2	3
Training set:	91	91	96,75	94	93,25	97,5
Test set:	57,5	56,5	60	59,5	57	57

Table 2. Comparison of model accuracy for the training set and the test set, as a function of regularization parameter λ: “One pixel”, “1-”, “2-” and “3” -neighbourhood

is finished, training set is exported by creating training pair that consists of the descriptor created for location stored in temporary data set, and corresponding vegetation class.

Four descriptor models were used for training. The first model takes into account only descriptor of selected position, while other three models takes into account local neighborhood of dimensions 1 pixel (3×3window), 2 pixels (5×5window) and 3 pixels (7×7window). The descriptor is than formed by sequencing single pixel descriptor from upper left corner of local window, row by row. Therefore, descriptors for models with 0, 1, 2 and 3 neighborhood consists of 6, 54, 150 and 294 features, respectively.

3. Results and Discussion

We tracked the change of training set error and test set error with the respect to four different models and 6 chosen values for the regularization parameter λ. The training error is likely to be lower than the actual generalization error, which is the case for our models as well. Figure 2. and Table 2 present comparison of accuracy for the training set and the test set, in the cases of descriptors for models with one pixel and 1-, 2- and 3- neighborhood, respectively.

As can be seen from the data presented Table 2 as well as charts in Figure 2, performance, in the terms of accuracy, of all four neural networks varies significantly for the data regarding training set. The accuracy in this case is defined as a relative number of accurate predictions ((#accurate predictions)/(#all predictions)*100%), that neural network outputs for the training set or the test set as an input. As one would expect, the model with the largest number of features (3-neighbourhood, 294 input neurons) displays the best performance in fitting the training data set (97.5% for λ=3), whilst the simplest, one-pixel, model has the lowest accuracy in fitting the training set (68, 25% for λ=1). Situation is rather different if accuracy of our models for fitting the test data set is compared. For the test set, accuracy of all four models is between 48.5 and 60%. Taking into account that with every added

4. Conclusion

In this paper one possible approach for automated vegetation classification for Landsat 7 multispectral images is proposed and early stage research results are presented. Generally, automated vegetation classification based on Landsat 7 multispectral images can be helpful for human interpretation of remote sensing images as the classification results could be used along with the original image data.In order to clearly identify types of vegetation on land surface and achieve an accurate classification based on satellite imagery, it is of crucial importance to determine the number of discrete vegetation categories (number of classes) and to choose distinctive characteristics (based on reflectance power distribution spectra) of these that are considered most suitable for the vegetation type categorization of the study area. More specifically, the first step of the classification procedure is the careful selection of the number of classes that represent sufficiently every discrete vegetation type. The good knowledge of the study area should help us to choose more representative classes depicted at the satellite images. Further, larger training and test set has to be made based on carefully collected pixels for every class from the study area. Recently, Support-Vector Machines (SVM) with kernels has been used very successfully for classification in remote sensing applications[7], so it is possible to compare behavior of ANN and SVM for this concrete application.

References

[1] Walthall, C. (2004) A comparison of empirical and neural network approaches for estimating corn and soybean leaf area index from Landsat ETM+ imagery. *Remote Sensing of Environment*, 92, 465–474 [DOI: 10.1016/j.rse.2004.06.003].

[2] Govender, M., Chetty, K. & Bulcock, H. (2007) *A review of hyperspectral remote sensing and its application in vegetation and water resource studies.* Water SA, 33, 145–152 [DOI: 10.4314/wsa.v33i2.49049].

[3] Cochrane, M.A. (2000). *Using Vegetation Re Ectance Variability for Species Level Classication of Hyperspectral Data*, Vol. 21, pp. 2075–2087.

[4] No title [Online]. Available: geo.arc.nasa.gov/sge/landsat/l7.html.

[5] N Aeronautics. Landsat 7 Science Data Users Handbook Landsat 7 Science Data Users Handbook (1972).

[6] Chen, C.H. & Peter Ho, P.-G. (2008) Statistical pattern recognition in remote sensing. Pattern Recognition, 41, 2731–2741 [DOI: 10.1016/j.patcog.2008.04.013].

[7] Kolios, S. & Stylios, C.D. (2013) Identification of land cover/land use changes in the greater area of the Preveza peninsula in Greece using Landsat satellite data. *Applied Geography*, 40, 150–160 [DOI: 10.1016/j.apgeog.2013.02.005].