Machine learning methods for classification of green infrastructure in city areas

Nikola Kranjčić 1, Damir Medak 2, Robert Župan 3, Milan Rezo 1…

1 Faculty of Geotechnical Engineering, University of Zagreb, Hallerova aleja 7, 42000 Varaždin, Republic of Croatia

2 Faculty of Geodesy, University of Zagreb, Kačićeva 26, 10000 Zagreb, Republic of Croatia

[nkranjcic@gfv.hr](mailto:nkranjcic@gfv.hr); nikola.kranjcic@gmail.com

**Abstract**. Reducing green urban infrastructure pose a huge threat to cities sustainability. It is important to monitor and track health of vegetation. For efficient planning of urban development and maintenance of green urban infrastructure key is to have up to date maps. Using satellite imagery is the easiest way to cover large city areas. In order to map vegetation, there are many possible solutions, however most effective way is using machine learning methods. Machine learning is divided into supervised and unsupervised classification and each can be divided into several different methods. Many authors have considered different methods and they use them to access accuracy on satellite image information extraction. They have used different satellite images and naturally higher resolution imagery results in better classification. However, there is still lack of comprehensive analysis of more supervised machine learning methods on similar cities. This paper aims to provide analysis of four different machine learning methods: support vector machine, artificial neural network, naïve Bayes and random forest. The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points where hyperplanes are decision boundaries that help classify the data points and support vectors data points that are closer to the hyperplane and influence the position and orientation of the hyperplane. Artificial neural network are brain-inspired systems which are intended to replicate the way that humans learn. Neural networks consist of input and output layers, and hidden layers. They are excellent tools for finding patterns which are far too complex or numerous for a human programmer to extract and teach the machine to recognize. A Naive Bayes classifier is a probabilistic machine learning model that’s used for classification task and the crux of the classifier is based on the Bayes theorem. Random Forest creates a forest and makes it random and is an ensemble of Decision Trees, most of the time trained with the bagging method which general idea is that a combination of learning models increases the overall result. All of the mentioned methods will be tested on Sentinel-2A imagery. Sentinel-2A multispectral imager has thirteen sensors which is useful in vegetation extraction. Methods will be compared using error matrix and kappa statistics.

1. Introduction

Since more and more people are living in cities it became extremely important to plan and manage green urban areas. Rapid urbanisation results in upgrade of interest in green urban areas and how can green space benefit cities and their residents [1]. The importance of green urban areas in order to improve quality of life has been discussed in [2] and [3] discussed appropriate sizes of green areas in order to improve local climatic conditions. How combination of trees can optimize thermal comfort in extreme summer conditions has been discussed in [4] and authors concluded that urban planning of green areas is important in improving quality of life of cities residents. Green infrastructure has been widely explored and investigated in order to improve or even gain sustainability. [5] discussed how sustainability could be achieved with smart green urban planning and pointed out that green infrastructure plans are developed at different planning scales. As seen from this short introduction many authors have considered the issue of green urban infrastructure planning and they all came to the similar conclusion that it is important to have quality green urban spaces in cities. In order to successfully plan the development it is necessary to have good and up to date maps. [6] found that is difficult to describe landscapes using traditional land use and land cover classification techniques but with the use of multispectral imagery mapping could be improved.

There has been many attempts and published papers which considers different techniques in order to extract information from satellite imagery especially using different machine learning methods. However there has not been full comprehensive analysis of available machine learning methods on one or two locations. In this paper four machine learning methods will be examined as it follows: support vector machines, artificial neural networks, naïve Bayes and random forest. This paper is the follow-up of paper published by [7]. Due to page limitations only introduction to each method will be presented.

Proposed methods will be tested on two cities in Croatia, Varaždin and Osijek.

1. Machine learning methods
   1. *Support Vector Machine*

Support Vector Machines were presented by [8] and since that day it has been widely explored. [7] have showed in their work that with the right combination of parameters and kernels classification can achieve high accuracy. Their optimal parameters are shown in table 1 with the use of Radial Basis Function kernel (RBF).

**Table 1.** Combination of SVM parameters to achieve high classification accuracy

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| City | Kernel | γ | C | Kappa |
| **Varaždin** | RBF | 1 | 28 | **0.87** |
| **Osijek** | RBF | 1 | 28 | **0.89** |

* 1. *Artificial Neural Networks*

First mention of neural networks was by McCulloch and Pitts in 1943 when they proposed a paper on how neurons might work. Afterward there were not many researches on artificial networks due to limitation in computation power. Breakthrough for artificial neural network was in 1990 when they became more and more explored. [9] considered how depth of neural network affects the result. They have concluded that there is still work to do in order to find specific parameters to gain maximum accuracy. [10] have shown that the highest overall accuracy of classification of high-resolution imagery is produced when there are 15 hidden neurons and number of iterations is 1500. There are parameters such as function, training methods and number of iterations that could affect the satellite imagery classification. In this paper different parameters will be tested and preliminary results will be shown.

* 1. *Naïve Bayes*

In order to represent knowledge based on Bayes theorem [11] proposed Bayesian networks or Naïve Bayes. A Bayesian network is a graph showing information flow as directed links between nodes without any loops consisting of nodes and directed links [11,12]. Each node of the network corresponds to a variable and the variables have discrete values [13]. It is typical for the relationships and the conditional probability table to be learned from a training data sets [14]. In this paper the preliminary results of satellite image classification using Naïve Bayes will be shown. Also, the results will be compared to the other machine learning methods.

* 1. *Random Forest*

Random forest is a regression technique that combines the performance of numerous decision trees algorithms to classify or predict the value of a variable [15,16]. To avoid the correlation of the different trees, random forest increases the diversity of the trees by growing them from different training data subsets created through a procedure called bagging [16]. In order to gain stability some data may be used more than once while others might never be used. Bagging technique is used for training data by resampling randomly the original datasets with replacements [16]. To reduce generalisation error, random forest uses the best feature/split point within a subset of evidential features which has been selected randomly form the total set of input features [15,16]. In this paper different parameters will be tested and preliminary results shown, and compared with other methods as well.

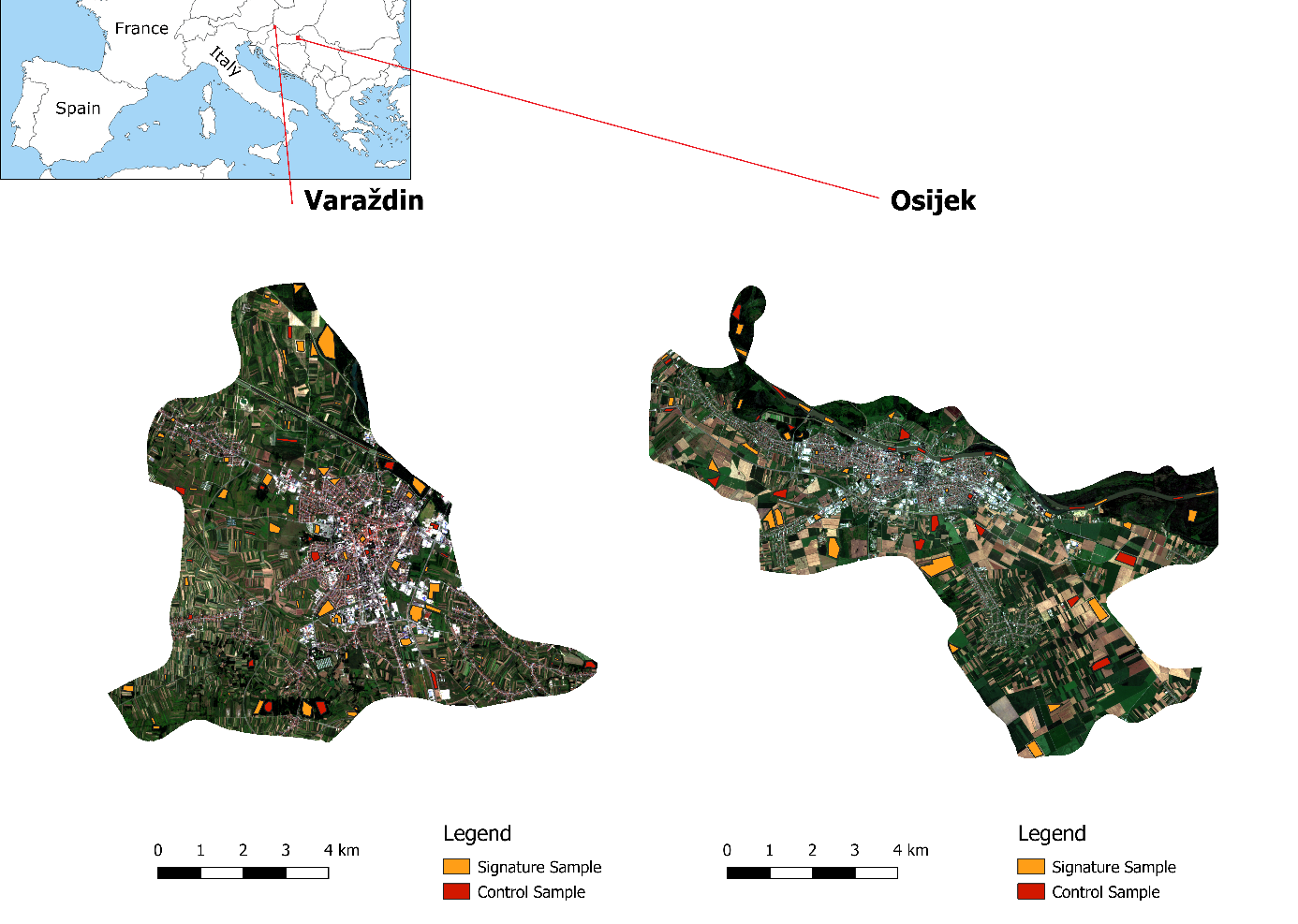
1. Study area

The study area is located in two towns in Croatia: Varaždin and Osijek and they have similar land cover use. Both are located on the River Drava, and the central area is populated with an urban fabric while smaller areas are filled with green urban areas. The wider city centre is populated with arable land, forests and inland waters. The dominant types of vegetation in Varaždin are forests populated with beech, oak, and chestnut and larger parks with grass and shrubbery. The dominant types of vegetation in Osijek are also parks, but in Osijek there are many swamp plants and linden and oak trees dominate forests. The coordinates for the centre of the study areas of Varaždin and Osijek are shown in Table 2. Study areas with a signature and control samples are presented in Figure 1.

**Table 2.** The study area coordinates (projection EPSG 3765 | WGS84).

|  |  |  |
| --- | --- | --- |
|  | **E (m) | φ (d°m’ s’’)** | **N (m) | λ (d°m’ s’’)** |
| Varaždin | 487,550.00 | 46°18’35’’ | 5,130,000.00 | 16°20’18’’ |
| Osijek | 671,000.00 | 45°33’03’’ | 5,048,000.00 | 18°41’24’’ |

The combination of parameters tested in this paper are presented in Table 3. and in Table 4. In Saga-GIS Normal Bayes doesn’t have parameters that could be changed. All of the machine learning methods are based on OpenCV library.

****

**Figure 1.** The study area: Varaždin (left); Osijek (right).

**Table 3.** Combination of parameters for Artificial Neural Networks.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **#1** | **#2** | **#3** | **#4** | **#5** |
| Number of Layers | 5 | 5 | 5 | 5 | 5 |
| Number of Neurons | 3 | 5 | 7 | 9 | 15 |
| Number of Iterations | 100 | 300 | 500 | 700 | 1500 |

**Table 4.** Combination of parameters for Random Forest.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **#1** | **#2** | **#3** | **#4** | **#5** |
| Maximum tree depth | 5 | 10 | 20 | 30 | 50 |
| Maximum sample count | 1 | 2 | 4 | 6 | 10 |
| Maximum categories | 5 | 5 | 5 | 5 | 5 |

1. Results and discussions

In Table 5. and Table 6. the results will be shown. All the results shown are presented in Kappa indices.

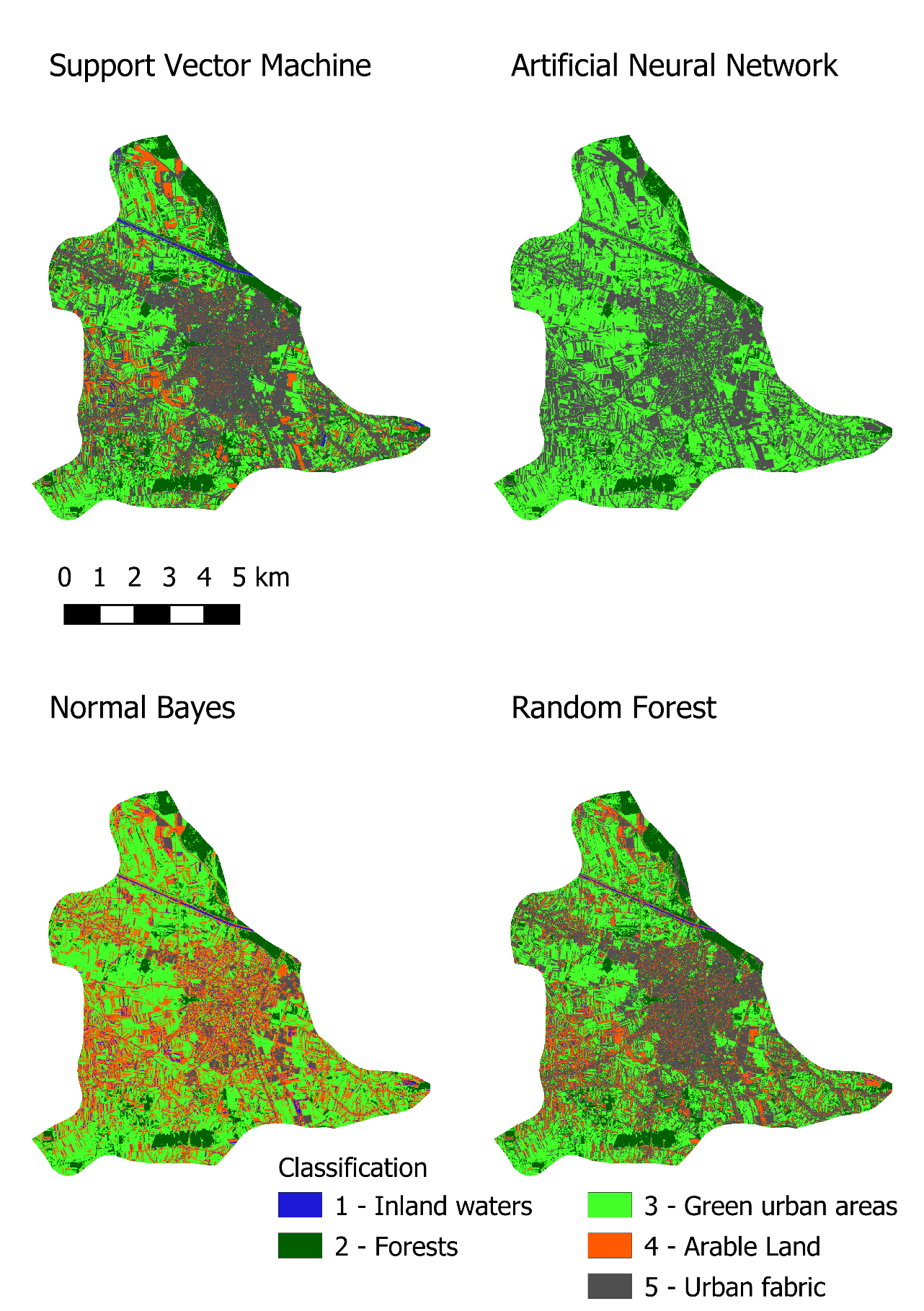
**Table 5.** Classification accuracy for city of Varaždin.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **#1** | **#2** | **#3** | **#4** | **#5** |
| Artificial Neural Network | 52,19% | 49,37% | 0,00% | 33,99% | 43,59% |
| Normal Bayes | 64,25% | | | | |
| Random Forest | 63,82% | 72,65% | 77,02% | 77,87% | 78,36% |

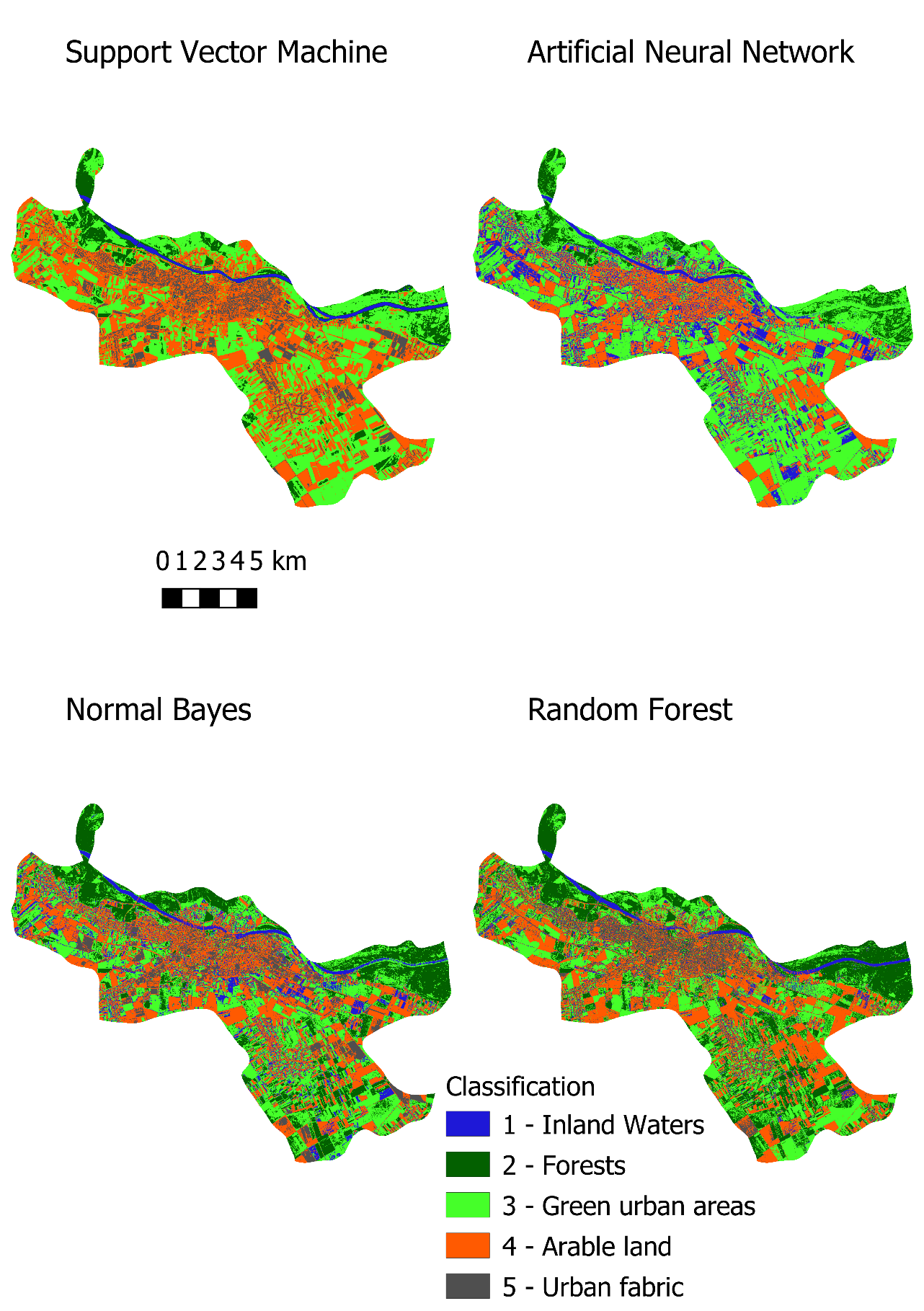
**Table 6.** Classification accuracy for city of Osijek.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **#1** | **#2** | **#3** | **#4** | **#5** |
| Artificial Neural Network | 29,05% | 67,61% | 46,06% | 36,21% | 36,76% |
| Normal Bayes | 66,30% | | | | |
| Random Forest | 73,34% | 78,84% | 89,29% | 89,57% | 89,18% |

Since it is mentioned in introduction that the support vector machine method has been tested by [7] their results will be also shown. On Figure 2 and Figure 3 there are shown results that are provided with the best classification accuracy in each method. Form Table 1, Table 5 and Table 6 it can be seen that the support vector machine results in highest overall accuracy compared to artificial neural networks, normal Bayes and random forest. Second highest accuracy has random forest. Random forest method also uses less memory and it is less time consuming than artificial neural networks. Normal Bayes provides high accuracy, but there are not many parameters that could affect the results of classification. Artificial neural networks provide the less accurate classification quality and it is most time consuming.



**Figure 2.** Results for city of Varaždin.



**Figure 3.** Results for city of Osijek.

1. Conclusions

Urban planning for and increasing the amount of green urban infrastructure are crucial for healthier living. To achieve this effectively, experts have combined machine learning modules with satellite imagery. The development of the Copernicus satellite missions represents a huge advance in land observation due to the increment in spatial resolution. Many analyses have been conducted to determine which of the parameters is most suitable for green urban areas extraction in satellite images. SVM is one of the most commonly used machine learning modules in image classification, and, according to different authors, the most common kernel is the radial basis function. The radial basis function is one of the kernels that produces classifications of the highest quality. In this paper presented four different machine learning methods and preliminary results. Test have shown that support vector machines provides highest accuracy among three other methods. However, there are still test necessary in order to fully understand and evaluate which machine learning method generate the best results.

References

1. Sandström, U.G. Green infrastructure planning in urban Sweden. *Plan. Pract. Res.* **2002**, *17*, 373–385, doi:10.1080/02697450216356.

2. Botkin, D.B.; Beveridge, C.E. Cities as environments. *Urban Ecosyst.* **1997**, *1*, 3–19, doi:10.1023/a:1014354923367.

3. Gómez, F.; Tamarit, N.; Jabaloyes, J. Green Zones in Urban Planning.Pdf. *Landsc. Urban Plan.* **2001**, *55*, 151–161.

4. Zheng, B.; Bedra, K.B.; Zheng, J.; Wang, G. Combination of tree configuration with street configuration for thermal comfort optimization under extreme summer conditions in the urban center of Shantou City, China. *Sustainability* **2018**, *10*, doi:10.3390/su10114192.

5. Mell, I.C. Can green infrastructure promote urban sustainability? *Eng. Sustain.* **2009**, *162*, 23–34, doi:10.1680/ensu.2009.162.1.23.

6. Dennis, M.; Barlow, D.; Cavan, G.; Cook, P.; Gilchrist, A.; Handley, J.; James, P.; Thompson, J.; Tzoulas, K.; Wheater, C.P.; Lindley, S. Mapping Urban Green Infrastructure: A Novel Landscape-Based Approach to Incorporating Land Use and Land Cover in the Mapping of Human-Dominated Systems. *Land* **2018**, *7*, 17, doi:10.3390/land7010017.

7. Kranjčić, N.; Medak, D.; Župan, R.; Rezo, M. Support Vector Machine Accuracy Assessment for Extracting Green Urban Areas in Towns. *Remote Sens.* **2019**, *11*, 655, doi:10.3390/rs11060655.

8. Vapnik, V.N. *The Nature of Statistical Learning Theory*; 1995; ISBN 0387987800.

9. Sun, S.; Chen, W.; Wang, L.; Liu, X.; Liu, T.-Y. On the Depth of Deep Neural Networks: A Theoretical View. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)*; 2015; pp. 2066–2072.

10. Mokhtarzade, M.; Zoej, M.J.V. Road detection from high-resolution satellite images using artificial neural networks. *Int. J. Appl. Earth Obs. Geoinf.* **2007**, *9*, 32–40, doi:10.1016/j.jag.2006.05.001.

11. Pearl, J. *Probabilistic reasoning in intelligent systems: networks of plausible inference*; Morgan Kaufmann Publishers Inc., 1998; ISBN 0-934613-73-7.

12. Neapolitan, R.E. *Probabilistic reasoning in expert systems: theory and algorithms*; John Wiley & Sons, Inc. Hoboken, New Jersey, 1990; ISBN 0-471-61840-3.

13. Mitchell, T.M. *Machine Learning*; McGraw-Hill Science/Engineering/Math, 1997;

14. Park, M.H.; Stenstrom, M.K. Using satellite imagery for stormwater pollution management with Bayesian networks. *Water Res.* **2006**, *40*, 3429–3438, doi:10.1016/j.watres.2006.06.041.

15. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32, doi:10.1023/A:1010933404324.

16. Rodriguez-Galiano, V.; Sanchez-Castillo, M.; Chica-Olmo, M.; Chica-Rivas, M. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geol. Rev.* **2015**, *71*, 804–818, doi:10.1016/j.oregeorev.2015.01.001.