UDK 528.835:528.854:53.088.3:519.226
Original scientific paper / Izvorni znanstveni članak

Evaluating Different Machine Learning Methods on RapidEye and PlanetScope Satellite Imagery

Nikola KRANJČIĆ – Varaždin[1], Damir MEDAK – Zagreb[2]

ABSTRACT. Since the first satellite imagery of RapidEye and PlanetScope became available, numerous studies have been conducted. However, only a few authors have focused on evaluating the accuracy of more than two machine learning methods in land cover classification. This paper evaluates the accuracy of four different machine learning methods, namely: support vector machine, artificial neural network, naive Bayes, and random forest. All analysis was conducted on cities in Croatia, Varaždin and Osijek. On Varaždin area on RapidEye satellite imagery support vector machine achieved overall kappa value 0.80, artificial neural network 0.37, naive Bayes 0.84 and random forest 0.76. On Varaždin area on PlanetScope satellite imagery support vector machine achieved overall kappa value 0.77, artificial neural network 0.38, naive Bayes 0.76 and random forest 0.75. On Osijek area on RapidEye satellite imagery support vector machine achieved overall kappa value 0.75, artificial neural network 0.36, naive Bayes 0.85 and random forest 0.76. On Osijek area on PlanetScope satellite imagery support vector machine achieved overall kappa value 0.64, artificial neural network 0.23, naive Bayes 0.72 and random forest 0.63. Performance time of each method is also evaluated. Naive Bayes and random forest have best performance time in every scenario.

Keywords: support vector machines, artificial neural network, naive Bayes, random forest, RapidEye, PlanetScope.

1. Introduction

First satellite called Sputnik 1 was launched in 1957 and it broadcasted radio signals. Purpose of Sputnik 1 was presentation of technological development and throughout the history purpose of satellite missions changed, from military to civil purposes. Today, satellite data is used for different purposes and in various case scenarios. Usage of satellite data is best presented in numbers. In December 2019, 2218 satellites were operable (URL 1) and in February 2020 1480000

[1] Nikola Kranjčić, mag. ing. geod. et geoinf., Faculty of Geotechnical Engineering, University of Zagreb, Hallerova aleja 7, HR-42000 Varaždin, Croatia, e-mail: nikola.kranjcic@gfv.unizg.hr,
[2] prof. dr. sc. Damir Medak, Faculty of Geodesy, University of Zagreb, Kačićeva 26, HR-10000 Zagreb, Croatia, e-mail: damir.medak@geof.unizg.hr.

scientific papers were published on "satellite imagery" topic (data from Google Scholar, URL 2). Landsat, Copernicus, IKONOS; GeoEye1, WorldView-2, WorldView-3, Quickbird, RapidEye and PlanetScope satellite imagery is usually used for scientific and professional purposes. Since large amount of data is available, it is important to have efficient way to process data and gain desired information. Machine learning is useful way to gain large amount of information based on available data. Machine learning is study of statistical methods that computers use to perform task relying on patterns. Supervised machine learning methods build mathematical model based on known sample data to make decision without being programmed to perform the task. There are many machine learning methods available, however, most popular are artificial neural networks, decision trees, Bayesian networks and support vector machines. Mentioned machine learning methods are well described in available literature and widely tested. However, mentioned methods are tested on different datasets and for different purposes, while for green urban infrastructure extraction are not thoroughly explored. Kranjčić et al. (2019a, 2019b) tested mentioned machine learning methods on Sentinel 2 imagery for mapping of green urban infrastructure in cities. Their results indicate that support vector machine and random forest outperformed other methods. This paper is extension of Kranjčić et al. (2019a, 2019b) where four machine learning methods were tested. Since their research was made on Sentinel 2 imagery this paper will evaluate methods on RapidEye and PlanetScope imagery, because many authors evaluated the potential of RapidEye and PlanetScope imagery for different classification purposes. For example, Adelabu et al. (2013) used support vector machine and random forest method in order to distinguish tree species on RapidEye imagery. They achieved high overall accuracy with 88.75% for support vector machine and 85% for random forest method. Tigges et al. (2013) exploited the benefits of multitemporal RapidEye satellite data in order to map and classify urban vegetation with support vector machine. Since they used data from spring till autumn their results are various. However, in one scenario they achieved high accuracy of 85.5%. Adam et al. (2014) evaluated the performance of support vector machine and random forest method on RapidEye imagery for land cover classification in a heterogeneous coastal landscape. Their research is like Adelabu et al. (2013) but they achieved even higher accuracy of 91.80% for support vector machine and 93.07% for random forest. Ustuner et al. (2014) used different vegetation indices and support vector machines to classify crop type on RapidEye imagery. Their overall results are various and range from 56.45% to 87.46%. Nitze et al. (2012) compared artificial neural network, random forest and support vector machine with radial basis function and polynomial kernel to maximum likelihood classifier for crop type classification on RapidEye imagery. They conclude that in general support vector machine outperformed other classifiers in overall accuracy from 50.0% to 97.1% for different crop types. Wicaksono and Lazuardi (2018) used PlanetScope data to map seagrass species using support vector machine and object-based image analysis and achieved overall accuracy for support vector machine from 40.00% to 45.76%. Wicaksono and Lazuardi (2019) used random forest method on PlanetScope imagery to map benthic habitat for two islands in Java Sea. They achieved overall classification results of 60.60% and 78.60%. Gašparović et al. (2018) evaluated RapidEye, PlanetScope and WorldView-2 for land cover classification with random forest classifier. For RapidEye overall accuracy was 83.80%, for PlanetScope 85.20% and for WorldView 2 90.10%. Evidently, many authors evaluated the potential of RapidEye and PlanetScope imagery for different classification purposes. However, majority of them used either support vector machine or random forest method for classification, and some of them used neural networks or maximum likelihood classifiers. There are lot of machine learning methods available and various authors (Civco 1993, Duro et al. 2012, Praveena et al. 2013, Kranjčić et al. 2019b) tested the most of them on different imagery. Their results were different depending on satellite imagery, methods used and application area.

The main objective of this paper is to test four different machine learning methods on RapidEye and PlanetScope imagery with emphasise on classification of green urban areas in cities. Based on available literature and previous research done, support vector machine and random forest method should provide highest overall accuracy.

2. Methods and datasets

In this paper four different machine learning methods are evaluated: support vector machine with polynomial, radial basis function and sigmoid kernel, artificial neural network, naive Bayes and random forest. All the evaluation is done on RapidEye and PlanetScope imagery on Varaždin and Osijek areas. Methods mentioned in introduction are well explained in available literature (Vapnik 1995, Kavzoglu and Colkesen 2009, Civco 1993, Langley and Sage 1994, Breinman 2001). Support vector machine is abstract machine learning method first mentioned in Vapnik (1995). Usually it can be improved with the use of kernels which are functions that simulates projecting the original data into space with higher dimension where data is considered linearly segregative. Usually in image processing radial basis function, polynomial and sigmoid kernel are used (Yekkehkhany et al. 2014). Kernels used in this paper are (URL 3):

- Polynomial kernel: $K(u',v) = (\gamma \times \langle u',v \rangle + coef\ 0)^d$
- Radial basis function (RBF) kernel: $K(u,v) = exp^{-\gamma \times |u-v|^2}$
- Sigmoid kernel: $K(u',v) = \tanh(g \times \langle u',v \rangle + coef\ 0)$

where $coef\ 0$ is coefficient, γ is gamma, and d is degree. The factor C parameter eliminates training errors and model complexity. A small value for C will increase the number of training errors, while a large C will lead to behaviors like the hard-margin SVM. γ is the free parameter of the radial basis function. If γ is large the variance is small, implying that the support vector does not have a broad propagation effect. Basically, a large γ leads to large biases and small variance models, and vice versa. Different authors (Yekkehkhany et al. 2014, Kranjčić et al. 2019a) tested different parameters and kernels on different satellite imagery and scenes. Their research resulted in parameters which provide highest classification accuracy and this parameters are selected for each kernel function and tested on RapidEye and PlanetScope imagery. Parameters used are presented in Table 1.

Table 1. *Support Vector Machine classification parameters for each kernel function.*

Kernel	**Polynomial**	**Radial Basis Function**	**Sigmoid**
γ	1	1	0
C	1	28	1

Two neuroscientists McCulloch and Pitts (1943) proposed neural networks and how they should work, but significant breakthrough of neural networks is linked to development of computer processing power. Sun et al. (2016) evaluated how the depth of neural network reflects in results. However, they did not get precise results, and further work should be conducted in order to get parameters to achieve highest accuracy. If the outputs of layer n is x_j and output of layer $n + 1$ y_i artificial neural network are calculated as follows:

$$u_i = \sum_j \left(w_{i,j}^{n+1} \cdot x_j \right) + w_{i,bias}^{n+1}$$

$$y_i = f(u_i)$$

where w is the weight of each input layer and f is a function. The weights are computed by the training algorithm. There are three different functions, identity function, symmetrical sigmoid, and the Gaussian function. For this research symmetrical sigmoid function is used, which can be calculated by expression:

$$f(x) = \beta \cdot (1 - e^{-\alpha x}) / (1 + e^{-\alpha x})$$

where for α and β default values are 1. α and β value was not changed in this paper, therefore how changing them affect the results is not tested. In this study number of neurons and number of iterations were changed and examined. Combination of parameters used for artificial neural network is presented in Table 2. Usually, potential network flexibility is higher if the network is bigger network, therefore accuracy can be higher. Parameters shown in Table 2 are proven to provide highest classification results among other tested parameter combinations (Kranjčić et al. 2019b).

Table 2. *Artificial Neural Network classification parameters.*

	Parameter combination 1	**Parameter combination 2**	**Parameter combination 3**
Number of Layers	5	5	5
Number of Neurons	3	5	15
Number of Iterations	100	300	1500

Third used machine learning method is naive Bayes. Langley and Sage (1994) declared that naive Bayes method is most simple and widely used probabilistic induction method where each sample is associated with the value that presents the probability that the sample will be evaluated in machine learning. Typical for naive Bayes method is that learns conditional probability and relationships from training datasets (Park and Stenstrom 2006). Among other tested methods this is the simplest one. User doesn't need to define or optimize parameters and results only depends on selection of training datasets and input satellite imagery.

Last tested method is one from decision tree family of machine learning methods called random forest. According to Breinman (2001) and Rodriguez-Galiano et al. (2015) random forest is regression technique where numerous decision trees are combined in order to predict or classify the value of a variable. In random forest term bagging is a process where the diversity of trees is created by growing them form various subsets of training data. Association between different trees is avoided with the use of bagging, which results in higher classification accuracy due to smaller correlation between trees. In this paper two different parameters were tested. First is the depth of the tree, where low value results in underfitting and high value will result in overfitting. Cross-validation can be used in order to define optimal value for tree depth. Second parameter tested is minimum number of samples mandatory for a leaf node to be split. Usual value is low percentage of the total number of samples. Selection of parameters in this paper is shown in Table 3 and it is accordant to Kranjčić et al. (2019b) who tested larger number of parameters, and these three combinations have proven to achieve highest accuracy.

Table 3. *Random Forest classification parameters.*

	Parameter combination 1	**Parameter combination 2**	**Parameter combination 3**
Maximum tree depth	20	30	50
Maximum sample count	4	6	10
Maximum number of categories	5	5	5

Four described machine learning methods are tested on RapidEye and PlanetScope satellite imagery. RapidEye sensor is operational since 2008 and it consist of five spectral channels and orthorectified pixel size is 5 meters. PlanetScope sensor consist of four spectral channels and orthorectified pixel size is 3 meters. For the purpose of this research RapidEye and PlanetScope imagery was downloaded within GEMINI (Geospatial monitoring of green infrastructure by means of terrestrial, airborne and satellite imagery) project. Since focus in this paper is on classification accuracy of green urban areas in cities, dates of downloaded imagery are accordant to most abundant vegetation. Also, satellite imagery is downloaded for dates when cloud coverage of the study area is minimal. Dates and product levels of downloaded imagery are presented in Table 4.

Table 4. *Dates and product levels of downloaded imagery.*

	Varaždin		**Osijek**	
	Date	**Product level**	**Date**	**Product level**
RapidEye	01.09.2016.	Level 3A	14.08.2016.	Level 3A
PlanetScope	29.08.2016.	Level 3B	19.08.2016.	Level 3B

RapidEye level 3A means that radiometric and sensor corrections have been applied to the data and the imagery is orthorectified using an elevation model and rational polynomial coefficients (Planet Labs Inc 2019). PlanetScope level 3B means that imagery is orthorectified and suitable for analytic and visual application. Imagery is projected to a cartographic projection. (Planet Labs Inc 2019). After downloading imagery, it was clipped with administrative borders of cities which were downloaded from (URL 4). Training area is presented on Figure 1. Training samples were defined using red-green-blue (RGB) composition of colours and infrared-green-blue (IRGB) composition of colours since vegetation is more visible in infrared spectre due to spectral characteristics of vegetation. 69 training samples and 26 control samples were selected for Varaždin, and 41 training and 26 control samples for Osijek study area. Lower training samples – control samples ratios in Osijek could result in lower classification accuracies for all methods on all imagery, especially on PlanetScope imagery which has higher resolution than RapidEye.

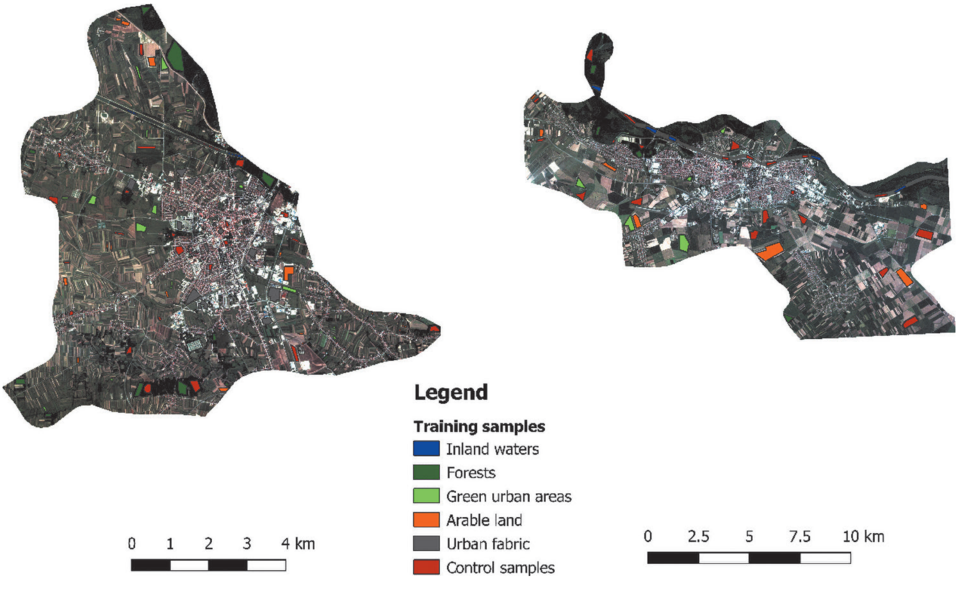

Fig. 1. *Training area with training samples; Varaždin (left) and Osijek (right).*

All of analysis is performed in SAGA GIS 7.5.0. on Intel ® Core ™ machine with i7-8550u, 16GB RAM, 64bit on Windows 10 OS. This is important because performance time, which presents elapsed time, of each machine learning method is evaluated. Elapsed time was measured only once in SAGA GIS. Classification accuracy is assessed with the kappa value, error matrix, commission and omission. Error statistics report that contains error matrix, commission, omission and kappa values is calculated in GRASS GIS with *r.kappa* analysis tool. Kappa value is statistical measure of agreement between two different classifiers of the same data, or how well the training dataset agreed with resulting classification. If the kappa value range is from 0.41 till 0.60 moderate classification accuracy is considered. If the range is from 0.61 till 0.80 high classification accuracy is considered and very high classification accuracy is achieved when kappa value is higher than 0.81 (modified from Viera and Garrett 2005). The commission percent shows what percentage of each class was confused with another class. The omission percent shows what percentage of each class was mistakenly classified as the wrong class. The difference between commission and omission is that commission percent reports how many pixels were placed into its class incorrectly, while omission percent shows how many pixels were not placed into its class correctly. Commission and omission values are shown in percentage.

After the classification process was done for each parameter defined in Table 1, 2 and 3 the error statistics report for each classification result was calculated. In results and discussion chapter statistics report is shown only for best overall accuracies for each method.

3. Results and discussion

Visual presentation of support vector machine with radial basis function on Varaždin is shown on Figure 2. From Table 5 omission, commission and estimated kappa per class can be seen. For class inland waters estimated kappa is 1.00 on RapidEye and 0.99 on PlanetScope imagery, while for class forest estimated kappa is 0.98 and 0.99 on RapidEye and PlanetScope respectively. On RapidEye imagery for class green urban areas 25% of pixels is misplaced or classified as other class, with estimated kappa of 0.72. On PlanetScope commission and omission is under 10% with high estimated kappa value of 0.91. For class arable land estimated kappa values are 0.45 and 0.30 on RapidEye and PlanetScope respectively, which are lowest values among other classes. Figure 2 shows that on PlanetScope imagery majority of arable land class is classified into green urban areas class, especially on west part of study area. If comparing RGB composition of colours on Figure 1 and classification on RapidEye imagery on Figure 2 it is seen that west part of study area is populated with more arable land than it is classified on PlanetScope imagery. For class arable land commission is 46.53% and 61.90%, while omission is 20.16% and 22.38% on RapidEye and PlanetScope respectively. This indicates that over 45% of pixels was mixed with other classes, and over 20% of pixels was not classified correctly. For urban fabric class values of commission and omission are opposite to arable land class, where commission percentage is lower than omission percentage. Estimated kappa value for urban fabric class is 0.88 on RapidEye and 0.83 PlanetScope imagery. Overall kappa value on RapidEye imagery is 0.80, while on PlanetScope imagery is 0.77 which is categorised as high classification accuracy.

Table 5. *Statistics report for support vector machine with radial basis function on Varaždin study area.*

Class no.	RapidEye			PlanetScope		
	Commission	**Omission**	**Estimated Kappa**	**Commission**	**Omission**	**Estimated Kappa**
1	0.00	0.00	1.00	0.40	0.00	0.99
2	0.99	0.07	0.98	0.04	0.94	0.99
3	25.36	24.50	0.72	8.31	9.31	0.91
4	46.53	20.16	0.45	61.90	22.38	0.30
5	8.80	34.07	0.88	11.93	42.66	0.83
Overall Kappa	0.796850			0.770525		

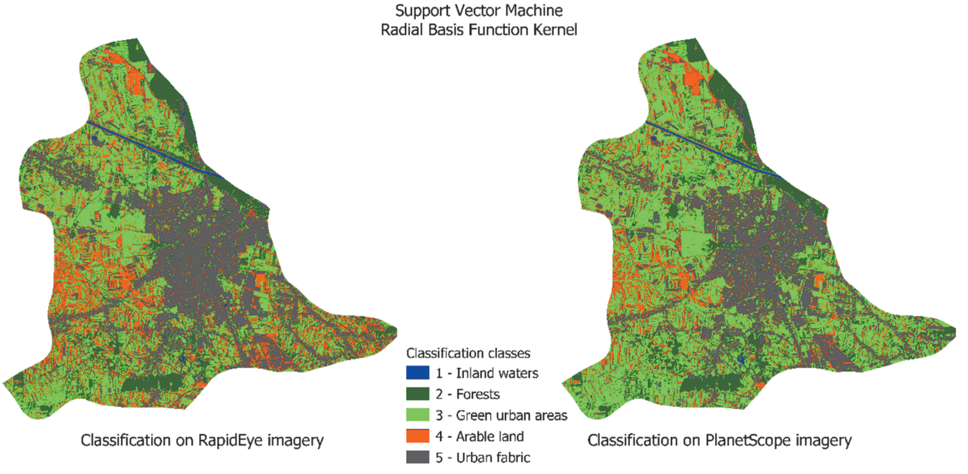

Fig. 2. *Support vector machine with radial basis function kernel classification result on Varaždin.*

Like on Varaždin study area similar results are on Osijek study area for support vector machine paired with radial basis function. Statistics report for support vector machine paired with radial basis function on Osijek study area is presented in Table 6. Class inland waters achieved highest estimated kappa of 0.99 with negligible commission and omission percentage on both satellite imagery. Class forests has high estimated kappa of 0.73 on RapidEye and 0.76 on PlanetScope. Green urban areas class achieved estimated kappa 0.75 on RapidEye which is similar result to result achieved on Varaždin training area. However, same class achieved estimated kappa of 0.56 on PlanetScope on Osijek area, while on Varaždin estimated kappa is 0.91. On RapidEye classification of class arable land resulted in lowest estimated kappa among other classes of 0.62, which is higher than on Varaždin study area. On PlanetScope class arable land estimated kappa is also lowest among other classes with value 0.38. Urban fabric class estimated kappa on RapidEye is 0.77 and on PlanetScope 0.53. In general, estimated kappa values per class are lower on Osijek study area than on Varaždin study area. However, on RapidEye imagery classification achieved high accuracy with kappa value 0.75. Based on classification accuracy ranking on PlanetScope imagery classification also resulted in high accuracy, however, with significantly lower kappa value of 0.64. Lower overall accuracy on Osijek area can be result of different class distribution, or different training to control samples ratios. Visual presentation of classification is shown on Figure 3. While for Varaždin training area difference in results are easily visible on Figure 2, on Figure 3 classification differences between RapidEye and PlanetScope imagery are not easily visible. Therefore, it should be discreet while evaluating classification accuracy based on visual interpretation. In Table 5 and Table 6 only results for support vector machine paired with radial basis function is shown, because support vector machine paired with polynomial kernel didn't produce results due to long performance time. Support vector machine paired with sigmoid kernel achieved overall kappa of 0.33 and 0.41 on Varaždin area for RapidEye and PlanetScope imagery respectively. On Osijek area achieved overall kappa is 0.45 and 0.49 for RapidEye and PlanetScope imagery.

Table 6. *Statistics report for support vector machine with radial basis function on Osijek study area.*

Class no.	RapidEye			PlanetScope		
	Commission	Omission	Estimated Kappa	Commission	Omission	Estimated Kappa
1	0.04	0.01	0.99	0.02	0.00	0.99
2	21.68	0.08	0.73	19.24	0.00	0.76
3	17.63	19.47	0.75	33.06	23.79	0.56
4	28.18	22.24	0.62	46.99	41.47	0.38
5	20.40	78.90	0.77	40.94	89.05	0.53
Overall Kappa	0.753408			0.635084		

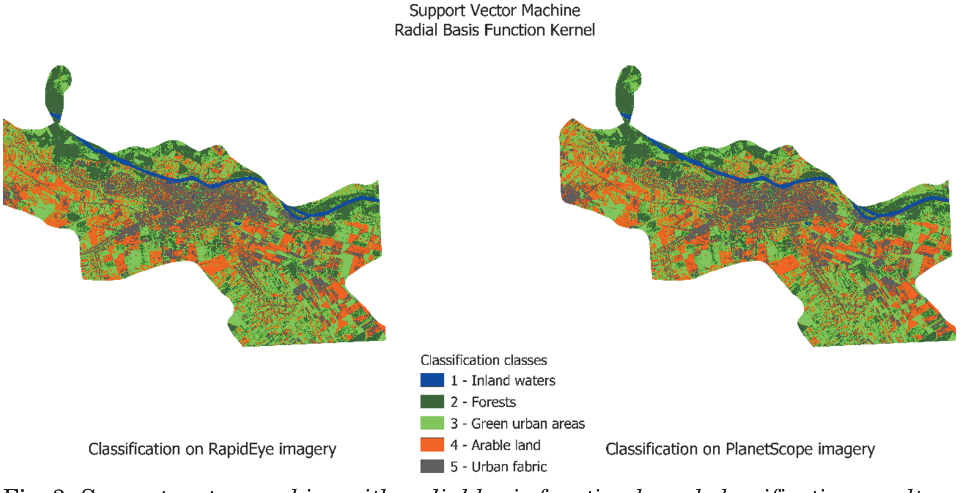

Fig. 3. *Support vector machine with radial basis function kernel classification result on Osijek.*

Artificial neural network didn't achieve high classification accuracy since it was not optimized and therefore on Varaždin overall kappa value on RapidEye is 0.37 and on PlanetScope 0.38. On Osijek overall kappa on RapidEye is 0.36 and on PlanetScope 0.23. Since the results for artificial neural network method are low, visual presentation won't be shown here. Application of random forest method resulted in high classification accuracy. On Figure 4 random forest classification for Varaždin is shown. Differences between classification on RapidEye and classification on PlanetScope imagery can be seen on Figure 4, especially on distribution of class arable land and urban fabric. Biggest differences are seen on south-east part of study area and on west or north-west part of study area. On Figure 4 on RapidEye imagery there are more pixels classified into class arable land and urban fabric, than on PlanetScope. In Table 7 is shown that like support vector machine, random forest method also achieved highest estimated kappa for inland waters class, on RapidEye 1.00 and on PlanetScope 0.99 on Varaždin study areas. Second highest estimated kappa value is for class forests, with 0.99 for both satellite imagery. Results achieved on Varaždin area with random forest are like support vector machine paired with radial basis function. Urban fabric class achieved very high classification accuracy with estimated kappa of 0.85 on RapidEye and 0.83 on PlanetScope imagery. On RapidEye class green urban areas achieved estimated kappa of 0.71, but with over 25% of commission and omission. On PlanetScope class green urban areas estimated kappa is 0.85, which is higher than on RapidEye like on classification with support vector machine. Lowest values for estimated kappa are for class arable land on both imageries, on RapidEye 0.35 and 0.25 on PlanetScope. Overall classification accuracy is high, with overall kappa 0.76 for RapidEye and 0.75 for PlanetScope imagery.

Table 7. *Statistics report for random forest on Varaždin study area.*

Class no.	RapidEye			PlanetScope		
	Commission	**Omission**	**Estimated Kappa**	**Commission**	**Omission**	**Estimated Kappa**
1	0.00	0.00	1.00	0.50	0.00	0.99
2	0.78	0.11	0.99	0.05	1.37	0.99
3	25.90	25.94	0.71	13.90	8.69	0.85
4	56.11	25.19	0.35	67.15	26.87	0.25
5	10.40	38.77	0.85	11.39	44.93	0.83
Overall Kappa	0.764300			0.748076		

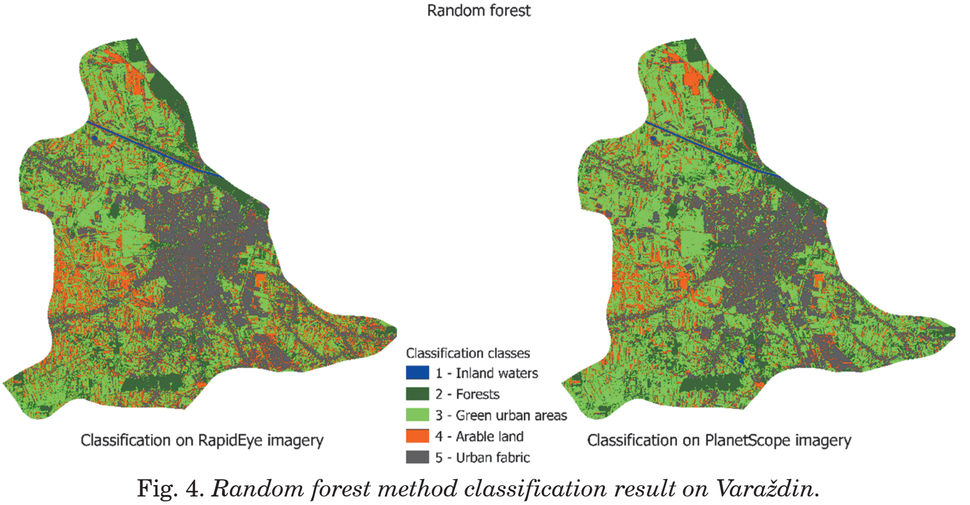

Fig. 4. *Random forest method classification result on Varaždin.*

On Osijek area results are in accordance with those obtained on Varaždin training area and statistics report is shown in Table 8. Highest estimated kappa is for class inland waters with value of 0.99 for both imageries. Second highest value is for class forest with values for estimated kappa 0.96 on RapidEye and 0.82 on PlanetScope. Other three classes on both imageries achieved medium, or high classification accuracy with high percentage of commission of omission. Green urban areas class achieved estimated kappa 0.66 on RapidEye and 0.50 on PlanetScope. Arable land class estimated kappa on RapidEye is 0.52 and on PlanetScope is 0.33, which is the lowest estimated kappa among other classes. For class urban fabric estimated kappa is higher than for class arable land with values of 0.74 on RapidEye and 0.54 on PlanetScope. However, on RapidEye omission percentage is over 80% and on PlanetScope near 90% for class urban fabric, which presents that almost every pixel from class was mistaken with other classes. In general, results are better on RapidEye than on PlanetScope imagery. On Osijek area on RapidEye imagery overall kappa is 0.76 and on PlanetScope 0.62. Figure 5 presents random forest classification on Osijek area and the differences on pixel distribution per class can be seen. Main differences are between classes green urban areas, arable land and urban fabric. On RapidEye imagery less pixels are distributed as green urban areas than on PlanetScope, which is seen on south-east part and west part of imagery. However, on RapidEye imagery, more pixels are redistributed as class urban fabric than on PlanetScope which is visible on Figure 5 from central part of training area.

Table 8. *Statistics report for random forest on Osijek study area.*

Class no.	**RapidEye**			**PlanetScope**		
	Commission	**Omission**	**Estimated Kappa**	**Commission**	**Omission**	**Estimated Kappa**
1	0.05	0.05	0.99	0.22	0.00	0.99
2	2.97	0.10	0.96	14.30	0.05	0.82
3	26.06	3.71	0.66	38.02	22.46	0.50
4	35.87	31.24	0.52	50.78	45.94	0.33
5	23.15	83.60	0.74	39.14	89.42	0.54
Overall Kappa	0.758897			0.621349		

Last tested method was naive Bayes and it achieved highest overall accuracy among tested methods. On Varaždin area estimated kappa for class inland waters is 1.00 on RapidEye and 0.99 on PlanetScope, which is result accordant to other methods. Also, very high classification accuracy is for class forests, with estimated kappa 0.95 on RapidEye and 0.99 on PlanetScope. Class green urban areas also achieved very high classification accuracy with estimated kappa 0.94 on RapidEye and 0.96 on PlanetScope. Contrary to support vector machine and random forest, class arable land achieved very high accuracy on RapidEye with estimated kappa 0.83 and low classification accuracy on PlanetScope with estimated kappa 0.49. Class urban fabric achieved low estimated kappa 0.57 on RapidEye and 0.49 on

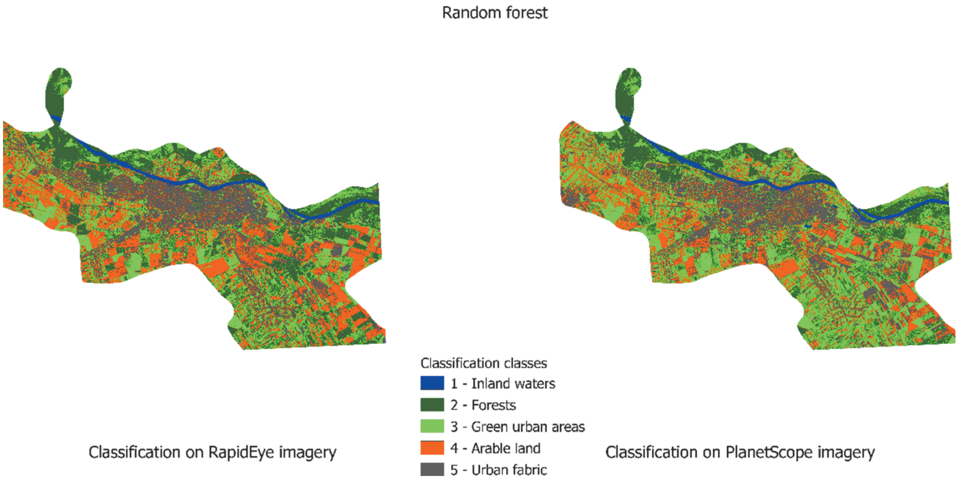

Fig. 5. *Random forest method classification result on Osijek.*

PlanetScope. Visual presentation of results can be seen on Figure 6. There are no obvious differences between two satellite imageries and classification on them. Main differences can be seen in distribution of classes which achieved lowest accuracy, arable land and urban fabric. Most obvious difference is on south-east area of training set. Overall kappa for classification on RapidEye imagery is 0.84 and on PlanetScope is 0.76.

Table 9. *Statistics report for naive Bayes on Varaždin study area.*

Class no.	RapidEye			PlanetScope		
	Commission	Omission	Estimated Kappa	Commission	Omission	Estimated Kappa
1	0.00	0.00	1.00	1.15	0.00	0.99
2	2.96	0.00	0.95	0.40	0.94	0.99
3	5.47	20.27	0.94	3.81	19.21	0.96
4	12.60	25.74	0.83	40.79	33.25	0.49
5	36.28	16.70	0.57	40.10	41.33	0.49
Overall Kappa	0.841415			0.760630		

As expected on Osijek area for class inland waters estimated kappa is highest among other classes with values 0.99 on both satellite imagery. Second highest estimated kappa is for class forest, which on RapidEye is 0.99 and on PlanetScope is 0.82. As seen from Table 10 for class forests on PlanetScope imagery 14% of pixels is misplaced for other classes. Class green urban areas on RapidEye achieved estimated kappa 0.71, but on PlanetScope is 0.47 which is lowest estimated kappa among other tested methods. Class arable land achieved high classification accuracy, with estimated kappa 0.79 on RapidEye and 0.72 on PlanetScope. Class

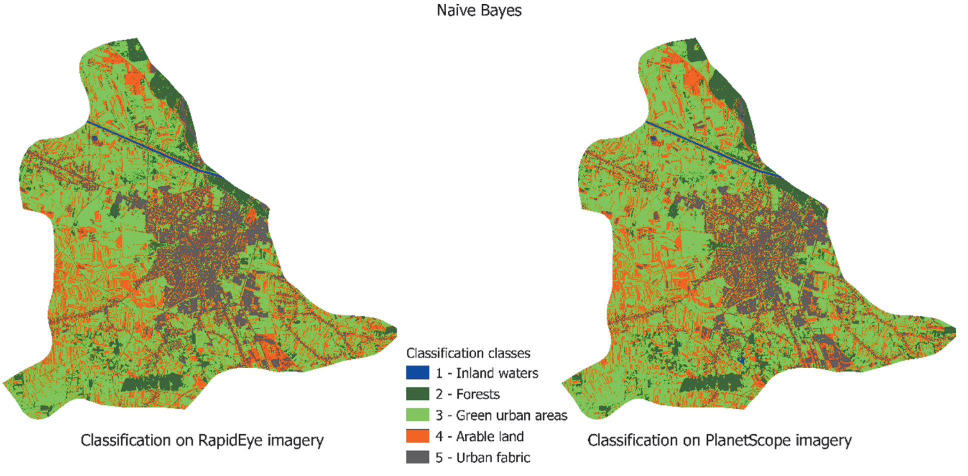

Fig. 6. *Naive Bayes classification result on Varaždin.*

urban fabric on RapidEye achieved estimated kappa 0.82 and on PlanetScope 0.66. Visual presentation of results is shown on Figure 7. Differences are mainly seen on central and south-east part of training area, especially on distribution of classes arable land and urban fabric, which is expected due to high percentage of omission. However, on Osijek area on RapidEye imagery very high classification accuracy is achieved with overall kappa 0.85 and on PlanetScope imagery high classification accuracy with overall kappa is 0.72.

Table 10. *Statistics report for naive Bayes on Osijek study area.*

Class no.	**RapidEye**			**PlanetScope**		
	Commission	**Omission**	**Estimated Kappa**	**Commission**	**Omission**	**Estimated Kappa**
1	1.09	0.00	0.99	0.25	0.00	0.99
2	1.06	0.05	0.99	14.01	0.00	0.82
3	21.88	3.76	0.71	40.56	25.65	0.47
4	14.81	22.07	0.79	18.30	35.89	0.72
5	17.37	64.61	0.82	32.29	62.67	0.66
Overall Kappa	0.851732			0.719083		

In SAGA GIS all of machine learning methods used in this paper are based on OpenCV library. For naive Bayes method change of parameters is not intended. In earlier analysis of machine learning methods on Sentinel 2 imagery (Kranjčić 2019b) naive Bayes method resulted in worse results than support vector machine or random forest method. This could indicate, that for achieving higher classification accuracy with support vector machine and random forest method on RapidEye and PlanetScope imagery, selection of classification parameters should be

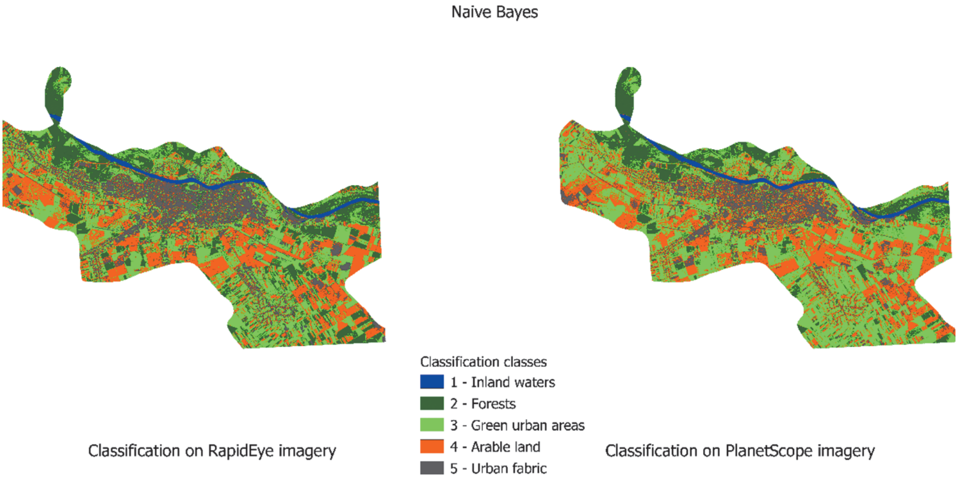

Fig. 7. *Naive Bayes classification result on Osijek.*

exploited in detail. However, with currently presented choice of parameters classification accuracy can be ranked as high or very high depending on method and satellite sensor. As mentioned earlier, support vector machine and random forest could result in similar classification accuracy, which was confirmed, and it could be seen in Tables 5, 6, 7 and 8.

As mentioned in chapter Methods and datasets performance time is measured. Graphical presentation of results is shown on Figure 8. RES shortcut on Figure 8 stands for RapidEye and PS for PlanetScope. General conclusion is that land cover classification performed on PlanetScope imagery has worse performance time that classification performed on RapidEye imagery, which was expected since PlanetScope imagery has higher spatial resolution than RapidEye imagery. The longest performance time had support vector machine paired with sigmoid kernel. Second worst execution time had artificial neural network and execution time increased with the increasement of iterations. Naive Bayes method and random forest method had quickest response time.

As mentioned in Introduction Adelabu et al. (2013) achieved very high accuracy of 0.89 for support vector machine and 0.85 for random forest method on RapidEye imagery. In this research paper, on RapidEye imagery on Varaždin study area support vector machine achieved overall accuracy 0.80 and on Osijek area 0.75, which are lower results than Adelabu et al. (2013) achieved. Random forest results are also lower with overall accuracy of 0.76 for both training areas. Tigges et al. (2013) focused on classification of urban vegetation with support vector machine on RapidEye imagery. Their results are in general lower that results achieved in this paper. However, in one scenario their results are higher than in this paper, with overall accuracy of 0.86. Much higher results achieved Adam et al. (2014) for land cover classification on RapidEye imagery with overall accuracy 0.92 for support vector machine and 0.93 for random forest. On RapidEye imagery Ustuner et al. (2014) tested support vector machines to classify crop types. Their highest result is 0.87 which is higher than result achieved in this paper. However, their results are various and generally lower than achieved in this paper. For different

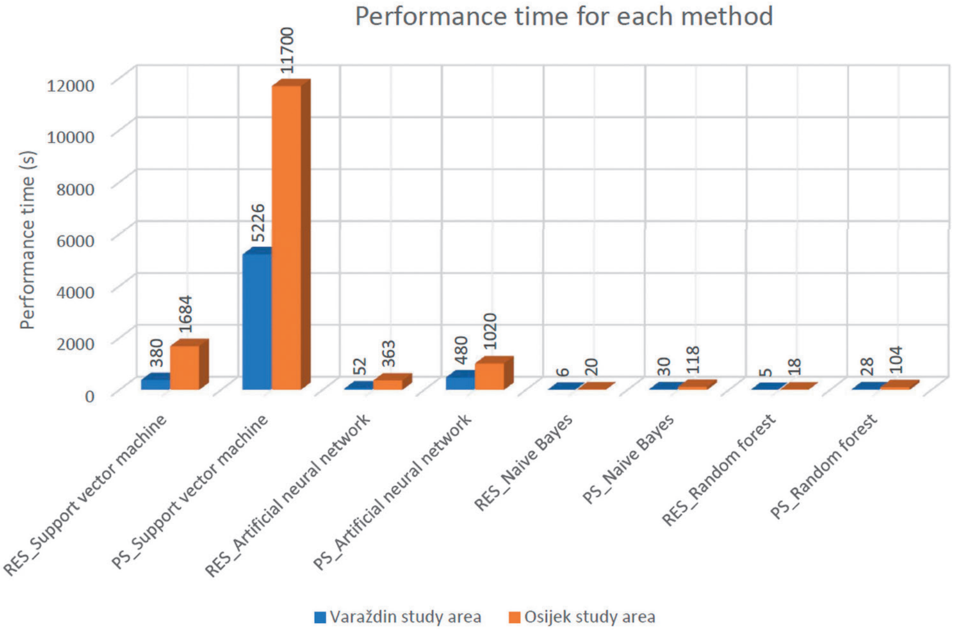

Fig. 8. *Performance time for each method (in seconds).*

crop types Nitze et al. (2012) with support vector machine achieved highest overall accuracy of 0.97, but their results vary and usually are lower than achieved in this paper. Wicaksono and Lazuardi (2018, 2019) did their research on PlanetScope imagery. In one research they used support vector machine and achieve overall accuracy from 0.40 to 0.46. In other research they used random forest and achieved overall results from 0.61 to 0.79. In this paper support vector machine used on PlanetScope imagery achieved overall accuracy 0.77 on Varaždin area and 0.64 on Osijek are, which is higher than previous mentioned research. Random forest method used on PlanetScope imagery achieved overall accuracy 0.75 on Varaždin area and 0.62 on Osijek area which is like results obtained by Wicaksono and Lazuardi (2019). Comparing results to Gašparović et al. (2018) results obtained in this paper are lower. Gašparović et al. (2018) evaluated random forest and achieved overall accuracy 0.84 and 0.85 on RapidEye and PlanetScope respectively.

4. Conclusion

After analysis of four different machine learning methods on satellite imagery with different resolution, results can be summarized and concluded. For RapidEye imagery on Varaždin area naive Bayes classifier has achieved highest overall accuracy with kappa value 0.84. Second highest accuracy is achieved with support vector machine paired with radial basis function kernel with overall kappa 0.80, random forest method results in overall kappa 0.76 and these methods are ranked as high accuracy classification. Artificial neural network method resulted in poor classification accuracy and with poor performance time. If artificial neural network is not optimized, it should be avoided for land cover classification. For RapidEye imagery on Osijek area naive Bayes resulted in highest accuracy with overall kappa 0.85. Second best result is achieved with random forest and overall kappa value 0.76 and third result is from support vector machine method combined with radial basis function kernel with overall kappa 0.75. Like Varaždin, on Osijek area artificial neural network method provided poor result with overall kappa 0.36. For PlanetScope imagery on Varaždin area support vector machine with radial basis function outperformed other methods with overall kappa 0.77. Naive Bayes achieved overall kappa 0.76, random forest 0.75 and artificial neural network 0.38. However due to noticeably increased performance time of support vector machine it could be stated that random forest and naive Bayes method outperformed support vector machine. For PlanetScope imagery on Osijek area naive Bayes with overall kappa 0.72 outperformed other machine learning methods. Support vector machine paired with radial basis function achieved overall kappa 0.64, random forest 0.62 and artificial neural network 0.23.

RapidEye and PlanetScope satellite imagery can be used for land cover classification and with the use of machine learning methods very high classification accuracy can be achieved. In this paper only naive Bayes method achieved very high classification accuracy. Support vector machine and random forest method achieved high classification accuracy which is an indication that the parameters used in this paper for classification should be thoroughly examined, evaluated and improved.

ACKNOWLEDGMENT. This research is performed on RapidEye and PlanetScope satellite imagery obtained with the support of Croatian science foundation project "Geospatial Monitoring of Green Infrastructure by Means of Terrestrial, Airborne and Satellite Imagery", grant number: IP-2016-06-5621.

References

Adam, E., Mutanga, O., Odindi, J. et al. (2014): Land-use / cover classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the performance of random forest and support vector machines classifiers, Int J Remote Sens, 35, 3440–3458.

Adelabu, S., Adam, E., Cho, M. A. (2013): Exploiting machine learning algorithms for tree species classification in a semiarid woodland using RapidEye image, J Appl Remote Sens, 7.

Breinman, L. (2001): Random forests, Mach Learn, 45, 1–33.

Civco, D. L. (1993): Artificial neural networks for land-cover classification and mapping, Int J Geogr Inf Syst, 7, 173–186.

Duro, D. C., Franklin, S. E., Dubé, M. G. (2012): A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery, Remote Sens Environ, 118, 259–272.

Gašparović, M., Dobrinić, D., Medak, D. (2018): Urban vegetation detection based on the land-cover classification of Planetscope, Rapideye and Wordview-2 satellite imagery, In: 18th International Multidisciplinary Scientific GeoConference SGEM2018 – Photogrametry and Remote Sensing.

Kavzoglu, T., Colkesen, I. (2009): A kernel functions analysis for support vector machines for land cover classification, Int J Appl Earth Obs Geoinf, 11, 352–359.

Kranjčić, N., Medak, D., Župan, R., Rezo, M. (2019a): Support Vector Machine Accuracy Assessment for Extracting Green Urban Areas in Towns, Remote Sens, 11, 13.

Kranjčić, N., Medak, D., Župan, R., Rezo, M. (2019b): Machine Learning Methods for Classification of the Green Infrastructure in City Areas, ISPRS Int J Geo-Information, 8, 1–15.

Langley, P., Sage, S. (1994): Induction of Selective Bayesian Classifiers, Morgan Kaufmann Publishers, Inc.

McCulloch, W. S., Pitts, W. H. (1943): A logical calculus of the ideas immanent in nervous activity, Bull Math Biophys, 5, 115–133.

Nitze, I., Schulthess, U., Asche, H. (2012): Comparison of machine learning algorithms random forest, artificial neuronal network and support vector machine to maximum likelihood for supervised crop type classification, Proc 4th Conf Geogr Object-Based Image Anal – GEOBIA, 35–40.

Park, M. H., Stenstrom, M. K. (2006): Using satellite imagery for stormwater pollution management with Bayesian networks, Water Res., 40, 3429–3438.

Planet Labs Inc (2019): Planet imagery product specifications.

Praveena, S., Singh, S. P., Muralikrishna, I. V. (2013): An Approach for the Segmentation of Satellite Images Using Moving KFCM and Naive Bayes Classifier, J Electron Eng, 3, 7–15.

Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., Chica-Rivas, M. (2015): Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines, Ore Geol Rev, 71, 804–818.

Sun, S., Chen, W., Wang, L., Liu, X., Liu, T.-Y. (2016): On the Depth of Deep Neural Networks: A Theoretical View, In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16), Phoenix, AZ, USA, 12–17, 2066–2072.

Tigges, J., Lakes, T., Hostert, P. (2013): Urban vegetation classification: Benefits of multitemporal RapidEye satellite data, Remote Sens Environ, 136, 66–75.

Ustuner, M., Sanli, F. B., Abdikan, S., Kurucu, Y. (2014): Crop type classifications using vegetation indices of RapidEye imagery, Int Arch Photogramm Remote Sens Spat Inf Sci, XL, 195–198.

Vapnik (1995): Support Vector Networks, Mach Learn.

Viera, A. J., Garrett, J. M. (2005): Understanding interobserver agreement: the kappa statistic, Fam Med, 37.

Wicaksono, P., Lazuardi, W. (2018): Assessment of PlanetScope images for benthic habitat and seagrass species mapping in a complex optically shallow water environment, Int J Remote Sens, 39, 5739–5765.

Wicaksono, P., Lazuardi, W. (2019): Random forest classification scenarios for benthic habitat mapping using PlanetScope image, IGARSS 2019 – 2019, IEEE Int Geosci Remote Sens Symp, 346, 8245–8248.

Yekkehkhany, B., Safari, A., Homayouni, S., Hasanlou, M. (2014): A comparison study of different kernel functions for SVM-based classification of multi-temporal polarimetry SAR data, Int Arch Photogramm Remote Sens Spat Inf Sci, XL-2/W3.

URLs

URL 1: USC Satellite Database,
https://www.ucsusa.org/resources/satellite-database, (24. 2. 2020).

URL 2: Google Scholar,
https://scholar.google.hr/, (24. 2. 2020).

URL 3: Saga-Gis,
http://www.saga-gis.org/saga_tool_doc/2.2.0/imagery_svm_0.html, (4. 3. 2020).

URL 4: DIVA-GIS,
http://www.diva-gis.org/, (3. 2. 2020).

Ocjena točnosti različitih metoda strojnog učenja na satelitskim snimkama RapidEye i PlanetScope

SAŽETAK. Otkako su prve satelitske snimke senzora RapidEye i PlanetScope postale dostupne, na njima su provedena brojna istraživanja. Međutim, samo se nekoliko autora usredotočilo na ocjenu točnosti više od dvije metode strojnog učenja pri klasifikaciji pokrova zemljišta. U ovom radu daje se ocjena točnosti četiri različite metode strojnog učenja: metode potpornih vektora, metode umjetnih neuronskih mreža, metode naivni Bayes i metode slučajnog šuma. Sve su analize provedene na gradovima u Hrvatskoj: Varaždinu i Osijeku. Na satelitskom snimku senzora RapidEye, za područje Varaždina, metoda potpornih vektora postigla je ukupnu kappa vrijednost 0,80, metoda umjetnih neuronskih mreža 0,37, metoda naivni Bayes 0,84 i metoda slučajnog šuma 0,76. Na satelitskom snimku senzora PlanetScope, za područje Varaždina, metoda potpornih vektora postigla je ukupnu kappa vrijednost 0,77, metoda umjetnih neuronskih mreža 0,38, metoda naivni Bayes 0,76 i metoda slučajnog šuma 0,75. Na satelitskom snimku senzora RapidEye, za područje Osijeka, metoda potpornih vektora postigla je ukupnu kappa vrijednost 0,75, metoda umjetnih neuronskih mreža 0,36, metoda naivni Bayes 0,85 i metoda slučajnog šuma 0,76. Na satelitskom snimku senzora PlanetScope, za područje Osijeka, metoda potpornih vektora postigla je ukupnu kappa vrijednost 0,64, metoda umjetnih neuronskih mreža 0,23, metoda naivni Bayes 0,72 i metoda slučajnog šuma 0,63. U radu se također mjeri i vrijeme izvedbe svake metode. Metoda naivni Bayes i metoda slučajnog šuma imaju najbolje vrijeme izvedbe u svim slučajevima.

Ključne riječi: *metoda potpornih vektora, metoda umjetnih neuronskih mreža, metoda naivni Bayes, metoda slučajnog šuma, RapidEye, PlanetScope.*

Received / Primljeno: 2020-02-10

Accepted / Prihvaćeno: 2020-03-11