

REMOTE SENSING MACHINE LEARNING ALGORITHMS IN ENVIRONMENTAL STRESS DETECTION - CASE STUDY OF PAN-EUROPEAN SOUTH SECTION OF CORRIDOR 10 IN SERBIA

IVAN POTIĆ^{1*}, MILICA POTIĆ²

¹Faculty of Geography, University of Belgrade, Belgrade, Serbia

³Independent researcher, Belgrade, Serbia

ABSTRACT

The construction of the Pan-European Corridor 10 is one of the major projects in the Republic of Serbia, and it enters the final phase. A vast natural area suffered a significant change to complete the project and therefore is the existence of a need to monitor those changes. Nature requires adequate and accurate detection of environmental stresses which inevitably arise after implementation of such large construction projects. Conversely to traditional field monitoring of the environment, this paper will present the remote sensing method which includes usage of European Space Agency's Sentinel 2A optical satellite data processed with different Machine Learning algorithms. An accuracy assessment is performed on land cover map results, and change detection carried out with best resulting data.

Keywords: Environment Monitoring, Gaussian Mixture Model, Random Forest, K-Nearest Neighbors, Confusion Matrix.

INTRODUCTION

One of the major projects of the Republic of Serbia funded by the World Bank (WB), European Investments Bank (EIB), Hellenic Plan for the Economic Reconstruction of the Balkans (HiPERB) and the Republic of Serbia, is the construction of the main branch of Pan-European Corridor 10. The corridor connects Salzburg in Austria and Thessaloniki in Greece through Ljubljana in Slovenia, Zagreb in Croatia, Belgrade, and Niš in Serbia, Skopje, and Veles in Macedonia (Figure 1). In Serbia, the south part of Corridor 10 is called the "Highway E75 – project SOUTH" and it is presented and constructed as the motor road at this point (Koridori Srbije, 2017).



Figure 1. Pan-European corridors in Serbia.
Source: belgradenet.com

The Highway E75 – project SOUTH extends for 74 km, from Grabovnica to Levosoje (Figure 2). There are five sections to complete in this area: Grabovnica – Grdelica (L=5.6 km), Grdelica – Caričina Dolina (L= 11.8 km), Caričina Dolina – Vladičin Han (L= 14.3 km), Vladičin Han – Donji Neradovac (L= 26.3 km), and Donji Neradovac - Levosoje (L= 16 km) (Figure 2) (Koridori Srbije, 2017).

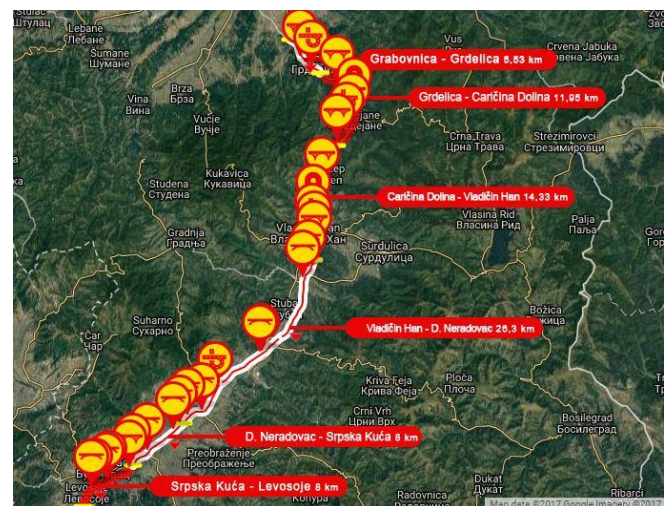


Figure 2. Corridor 10 South project in Serbia.
Source: www.koridor10.rs printscreen

The construction zone of this scale indubitably has a significant impact on the environment. A proper monitoring is crucial to conserve the nature and mitigate the environmental stress. Considering that technology has advanced, we are going to use the achievements of remote sensing and its methods to monitor the changes that have occurred during the construction of Corridor 10. Further, the change detection of the land cover

* Corresponding author: ipotic@gmail.com

will be performed to present the changes for the monitored period. Area of interest is selected within the area that is under active construction and covers 1.095,4 sq. km (Figure 3).

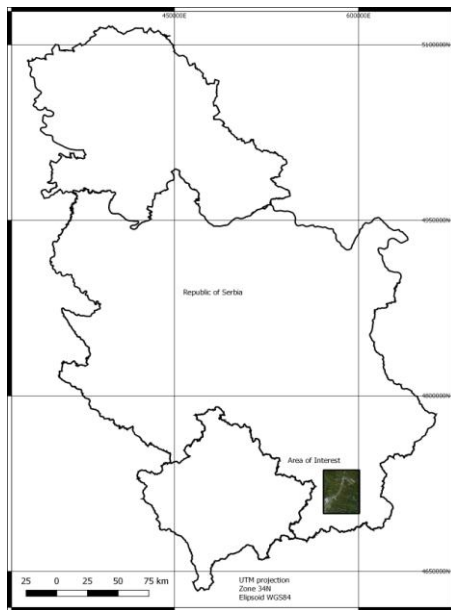


Figure 3. Area of Interest - part of Corridor 10 working zone.

EXPERIMENTAL

Materials and methods

Remote sensing technology is employed to achieve the goal of this paper with the contemporary methodology that employs the *Machine Learning* (ML) algorithms (Canziani et al., 2008; Mas & Flores 2008; Jensen et al., 2009; Duro et al., 2012; Lary et al., 2016).

Sentinel 2 satellite imagery was obtained using Copernicus Sci Hub (Copernicus Open Access Hub, 2017) as starting data for the analysis. Sentinel 2 product consists of the granules that represent the particular region. The granule comes with 13 different bands where three different ground resolution bands are present: 10 m, 20 m, and 60 m. 10 m bands are: visible Blue (B), Green (G), Red (R), and Near InfraRed (NIR). 20 m bands are three Vegetation Red Edge bands, Narrow NIR and two Short Wave InfraRed (SWIR) bands. 60 m bands are Coastal Aerosol, Water, Vapour and SWIR Cirrus band (Sentinel 2 MSI, 2017).

Two different Sentinel 2 products Level-2A were downloaded for 2017. Since there were cloudy parts in the research area, the mosaic was made using two different granules T34TEN date from 01.07. – 31.07.2017. Remote sensing/ raster processing plugin for QGis was applied to perform the mosaicking tasks.

To perform the change detection for the research area, the same images from August 2016 were downloaded from the Copernicus Sci Hub, and sub-scene created. The image was cloud-free, and there was no need for mosaicking. The product was Level-1C, so the data was processed to Level-2A using

SNAP (Sentinel Application Platform) toolbox software (ESA STEP, 2017), which took more than 13 hours to complete. Sentinel 2 products have multiple processing phases:

- *Level-0* and *Level-1A&B* products are in preprocessing phase and not available to users;
- *Level-1C* processing uses the *Level-1B* product and applies radiometric and geometric corrections (including orthorectification and spatial registration);
- Atmospheric correction is applied to Top-Of-Atmosphere (TOA) *Level-1C* orthoimage products, and a scene classification is presented as the *Level-2A* product. Bottom-Of-Atmosphere (BOA) corrected reflectance product is Level-2A with main output as an orthoimage. Additional outputs are Aerosol Optical Thickness (AOT) map, a Water Vapour (WV) map and a Scene Classification Map (SCM) together with Quality Indicators (QI) for cloud and snow probabilities at 60 m resolution (Sentinel 2 MSI, 2017).

Sentinel 2 bands used to complete the analysis are Red, Green, Blue and Near Infra-Red bands with 10m ground resolution.

Pixel-based *Machine Learning* (ML) algorithms were used to produce the land cover map of the area of interest. The most common three ML tasks are *Regression*, *Classification*, and *Clustering*.

Regression is employed as supervised learning task for modeling and predicting variables, where we have numeric *true* ground values for the research area. There are different regression algorithms, such as:

- *Linear Regression* (works when there are linear relationships between dataset variables);
- *Regression Tree* or *Decision Trees* repeatedly splits the dataset into separate branches and maximize the information gain. This allows the algorithm to learn nonlinear relationships;
- *Deep Learning* algorithm applies to multi-layer neural networks to learn extremely complex patterns using convulsions and drop-out mechanisms, and others;
- *Honorable Mention (Nearest Neighbors)* save each training observation. Further, they make predictions for new observations as they search for similar training observations and join the values (Elite Data Science, 2017).

Classification, as supervised learning task, is used in this paper to model and predict land cover categories as the ML algorithms can predict a class. Different classifications were used in this article to obtain the best possible accuracy of the data:

- *Classification Trees* is employed in Random Forest;
- *Gaussian Mixture Model* (GMM) take on that data points are generated from a mixture of a limited number of Gaussian distributions with unfamiliar parameters (Scikit learn, 2014).

K-Neighbors Classifier where the learning is based on the **k** nearest neighbors of each query point. **k** is an integer value specified by the user (Scikit learn, 2014).

The creation of a land cover map from BOA processed Sentinel 2 data required a ground training samples. To obtain such areas and create necessary vector file as training material, historical google maps were employed using different sources and plugins for QGIS. Seven different classes recognized for both 2016 and 2017 and consist of 175 and 164 polygons respectively. Two attributes created, as integer and text. Further, prepared subsene for each year was processed using dzetsaka ML plugin for QGIS.

The accuracy assessment was performed using training sample polygons in dzetsaka and SCP plugin for QGIS.

Confusion matrix was created and presents overall accuracy and kappa hat.

The land cover change was performed using SCP plugin in QGIS.

NUMERICAL RESULTS

After applying the algorithms, three different land cover maps for each year were created (Figure 4).

Accuracy assessment for created land cover maps is presented in Tables 1-3. As it can be seen, ML algorithms gave very decent results where Random Forests goes up to 100% of accuracy.

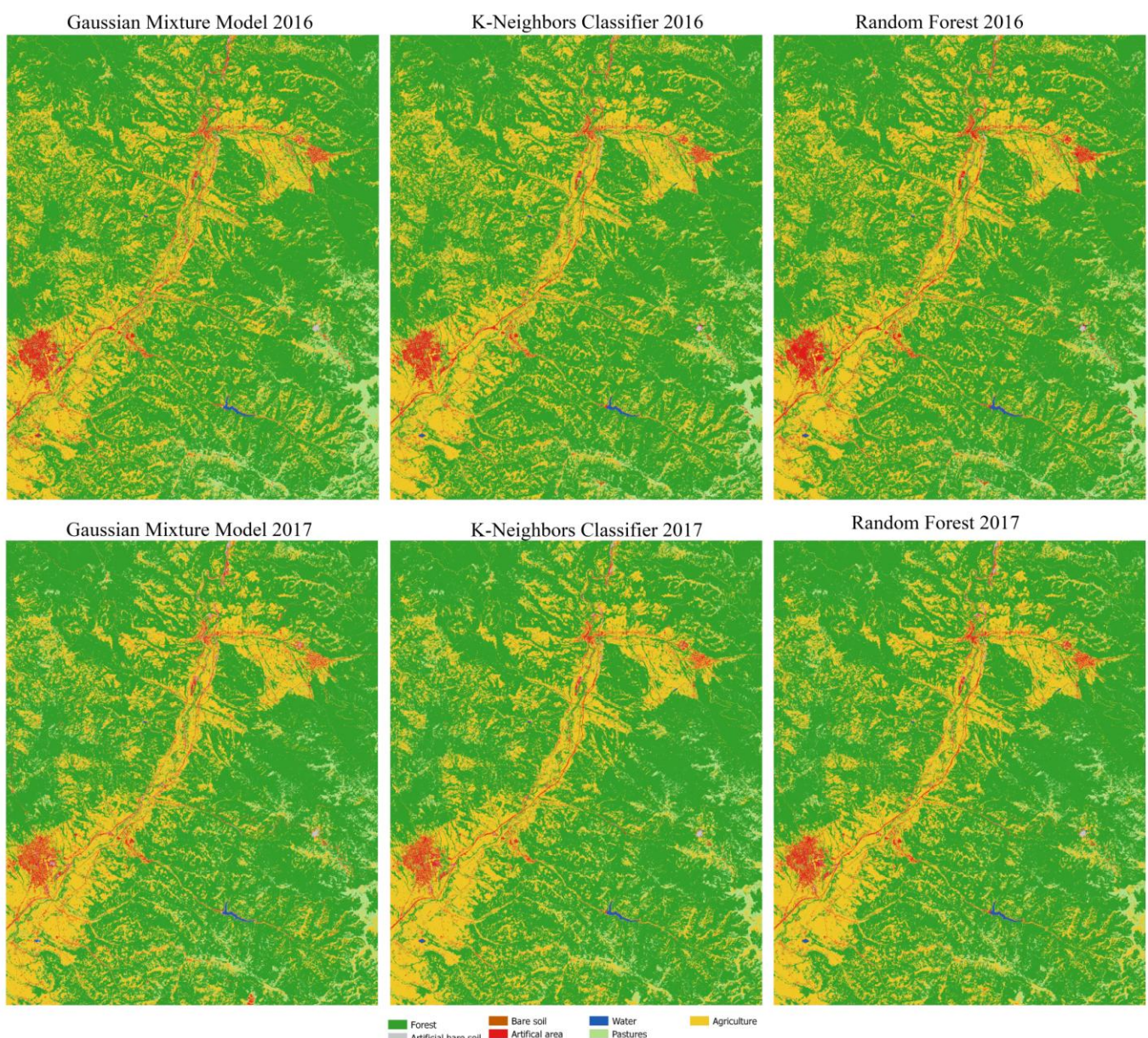


Figure 4. Land cover maps for the area of interest created using different ML algorithms.

Table 1. Confusion matrix for K-Neighbors Classifier.

Class 2016	Forest	Artificial bare soil	Bare soil	Artificial area	Water	Pastures	Agriculture	Class 2017	Forest	Artificial bare soil	Bare soil	Artificial area	Water	Pastures	Agriculture
Forest	115307	0	5	39	0	263	1601	Forest	116280	0	3	11	0	153	515
Artificial bare soil	0	2059	0	286	0	0	1	Artificial bare soil	0	2548	0	296	0	0	0
Bare soil	0	0	239	39	0	35	16	Bare soil	0	0	308	56	0	0	92
Artificial area	7	499	20	11002	0	9	313	Artificial area	0	227	33	5135	0	0	162
Water	0	0	0	8	2256	0	0	Water	0	0	0	1	2259	0	0
Pastures	212	1	26	363	0	13726	621	Pastures	180	0	22	0	0	4922	515
Agriculture	1139	58	1147	2085	0	1902	56578	Agriculture	240	19	1078	2168	0	458	56406
Kappa	91.63%							Kappa	94.09%						
Overall	94.95%							Overall	96.79%						

Table 2. Confusion matrix for Gaussian Mixture Model Classifier.

Class 2016	Forest	Artificial bare soil	Bare soil	Artificial area	Water	Pastures	Agriculture	Class 2017	Forest	Artificial bare soil	Bare soil	Artificial area	Water	Pastures	Agriculture
Forest	114235	2	5	17	51	489	2879	Forest	115280	0	3	1	7	748	511
Artificial bare soil	0	2083	0	1059	57	0	0	Artificial bare soil	0	2401	0	727	108	0	0
Bare soil	0	3	0	102	0	23	99	Bare soil	78	0	55	119	0	0	185
Artificial area	42	498	24	10270	3	3	1533	Artificial area	116	383	174	4509	0	3	665
Water	0	0	0	0	2145	0	0	Water	0	0	0	0	2144	0	0
Pastures	478	0	232	368	0	13085	922	Pastures	227	0	24	2	0	11892	869
Agriculture	1910	31	1176	2006	0	2335	53697	Agriculture	999	10	1188	2309	0	2553	56905
Kappa	87.23%							Kappa	89.93%						
Overall	92.28%							Overall	94.15%						

Table 3. Confusion matrix for Random Forest Classifier.

Class 2016	Forest	Artificial bare soil	Bare soil	Artificial area	Water	Pastures	Agriculture	Class 2017	Forest	Artificial bare soil	Bare soil	Artificial area	Water	Pastures	Agriculture
Forest	116665	0	0	0	0	0	2	Forest	116453	0	2	8	0	485	298
Artificial bare soil	0	2617	0	0	0	0	0	Artificial bare soil	0	2647	0	150	1	0	2
Bare soil	0	0	1437	0	0	0	0	Bare soil	0	0	825	13	0	0	42
Artificial area	0	0	0	13821	0	0	0	Artificial area	0	138	59	6526	0	0	187
Water	0	0	0	0	2256	0	0	Water	0	0	0	0	2258	0	0
Pastures	0	0	0	0	0	15933	0	Pastures	93	0	10	0	0	10611	213
Agriculture	0	0	0	1	0	2	59128	Agriculture	154	9	548	970	0	4100	58393
Kappa	100%							Kappa	93.71%						
Overall	100%							Overall	96.35%						

Accuracy assessment results demonstrate how those ML algorithms execute the classification. The best result is given by the Random Forest algorithm with perfect accuracy of 100% for 2016 and 96.35% for 2017. In next part of this research, Random Forest land cover map will be used for the final analysis. Classification results are presented in Table 4:

Table 4. Classification results for RF land cover maps.

2016 Class	Pixel Sum	Percentage (%)	Area (km ²)
Forest	7209286	65.89	720.93
Artificial bare soil	10320	0.09	1.03
Bare soil	7955	0.07	0.80
Artificial area	267578	2.45	26.76
Water	9883	0.09	0.99
Pastures	428576	3.92	42.86
Agriculture	3008299	27.49	300.83
2017 Class	Pixel Sum	Percentage (%)	Area (km ²)
Forest	7419839	67.85	741.98
Artificial bare soil	17226	0.16	1.72
Bare soil	4049	0.04	0.40
Artificial area	192010	1.76	19.20
Water	9860	0.09	0.99
Pastures	543802	4.97	54.38
Agriculture	2748450	25.13	274.85

The results show that two classes are dominant with more than 90% of the research area: Forest with 65.9% in 2016 and 67.6% in 2017 and Agriculture with 27.9% and 25.1% respectively. Percentage of change is presented in Table 5.

Table 5. Change in classes. The positive values represent the increase of pixels in 2017 while negative values present decrease in 2017.

Class	Area (km ²)	Percentage (%)
Forest	21.06	2.92
Artificial bare soil	0.69	66.92
Bare soil	-0.39	-49.10
Artificial area	-7.56	-28.24
Water	0.00	-0.23
Pastures	11.52	26.89
Agriculture	-25.98	-8.64

Change detection data in table 5 confirms the table 4 data and presents how much each class has changed. The highest increase has the Artificial bare soil (where our primary goal of this work belongs – Corridor 10 under construction), and Pasture classes versus the Bare Soil, Agriculture, and Artificial classes which decrease in area percentage cover. Figure 5 shows the difference in the northern part of the research area where the construction of Corridor 10 is in its full swing.

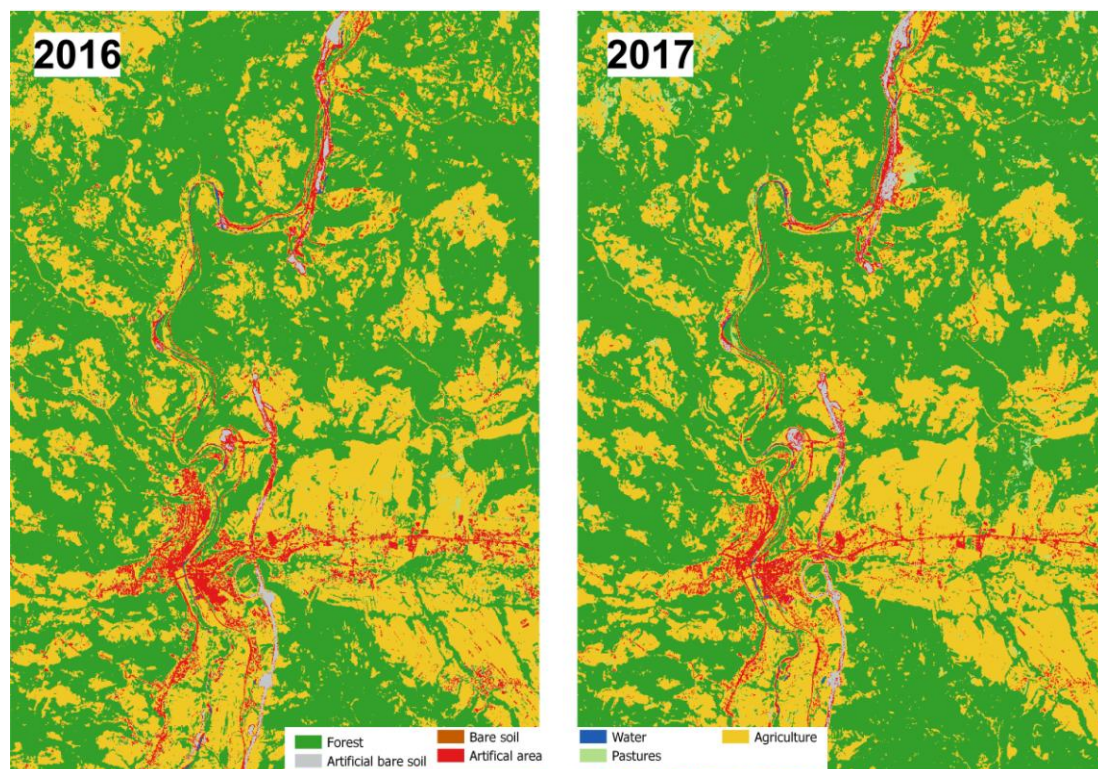


Figure 5. Northern part of the research area - Corridor 10 ongoing construction site.

CONCLUSION

As table 5 is presenting, the class of interest in this research is within Artificial bare soil which presents the construction area of new Corridor 10. It can be seen that there is an increase of the area covered by this class which indicates that in one year there were changes in the environment. Since the land cover is still presented with same class and did not change into an Artificial area where constructed – paved highway belongs, we can conclude that the motorway is still under construction. This data acquired using remote sensing analysis of Sentinel 2 satellite imagery can be of great help in monitoring changes of the environment and big construction projects. Since the satellite data are widely accessible and have satisfying ground resolution with low, or no cost, we cannot exclude the remote sensing techniques from the environmental research, but we must expand the knowledge and capabilities provided. Random Forest machine learning algorithm used in this paper confirms that the classifying algorithms have advanced to the level when they can be of great help to the environment analysts. High accuracy of classified data obtained using Classification Tree algorithm gives new perspective to remote sensing. Furthermore, different machine learning algorithms (Random Forest, Gaussian Mixture Model, K-Neighbors Classifier, and other) along with the Artificial Neural Networks and Object Based Image Analysis (OBIA) classification are in the focus of remote sensing professionals and researchers, while rapid development and improvement of the algorithms is in progress.

With this methodology, it is possible to perform a broad spectrum of analysis, such as environmental stress detection (landslides, wildfires, flooding, etc.) or land cover map creation and other, with the very high percentage of accuracy while we save time and money in the process that used to last much longer.

REFERENCES

- Belgrade Net. (2017). Retrieved from <http://www.belgradenet.com/business/>
- Canziani, G., Ferrati, R., Marinelli, C., & Dukatz, F. 2008. Artificial Neural Networks and Remote Sensing in the Analysis of the Highly Variable Pampean Shallow Lakes. *Mathematical Biosciences and Engineering*, 5(4). doi:10.3934/mbe.2008.5.691
- Copernicus Open Access Hub. (2017). Retrieved from <https://scihub.copernicus.eu/dhus/#/home>
- 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 Sensing of Environment*, 118, pp. 259-272. doi:10.1016/j.rse.2011.11.020
- Elite Data Science. (2017). Retrieved from <https://elitedatascience.com/machine-learning-algorithms>
- ESA STEP. (2017). Retrieved from <http://step.esa.int/main/toolboxes/snap/>
- ESA. 2017. Sentinel 2 MSI. Retrieved from <https://earth.esa.int/web/sentinel/user-guides>
- Jensen, R.R., Hardin, P.J., & Yu, G. 2009. Artificial Neural Networks and Remote Sensing. *Geography Compass*, 3, pp. 630-646. doi:10.1111/j.1749-8198.2008.00215.x
- Koridori Srbije. (2017). Retrieved from <http://www.koridor10.rs>
- Lary, D.J., Alavi, A.H., Gandomi, A.H., & Walker, A.L. 2016. Machine Learning in Geosciences and Remote Sensing. *Geoscience Frontiers*, 7(1), pp. 3-10. doi:10.1016/j.gsf.2015.07.003
- Mas, J.F., & Flores, J.J. 2008. The Application of Artificial Neural Networks to the Analysis of Remotely Sensed Data. *International Journal of Remote Sensing*, Iss., 29(3). doi:10.1080/01431160701352154
- Scikit learn. (2014). Retrieved from <http://scikit-learn.org/0.15/modules/>