COMPARISON OF SUPPORT VECTOR MACHINES, RANDOM FOREST AND DECISION TREE METHODS FOR CLASSIFICATION OF SENTINEL - 2A IMAGE USING DIFFERENT BAND COMBINATIONS

Taskin Kavzoglu (1), Furkan Bilucan (1), Alihan Teke (1)

[1] Gebze Technical University, Department of Geomatics Engineering, Gebze, Kocaeli, 41400, Turkey
Email: kavzoglu@gtu.edu.tr; f.bilucan2020@gtu.edu.tr; a.teke2020@gtu.edu.tr

KEY WORDS: Classification, Sentinel-2A, Support Vector Machines, Random Forest, Decision Tree Algorithm

ABSTRACT: Classification of remotely sensed images is a prerequisite for many earth observation studies including change detection, yield forecast and water quality analysis. Recent studies showed that machine learning algorithms used for the classification of satellite image high accurate results. In this study, three popular machine learning algorithms namely, random forest (RF), support vector machines (SVM) and decision tree (DT) classifiers were utilized considering three datasets that comprise different band combinations of a Sentinel-2A image. These datasets consist of five, seven and eleven bands containing an image of normalized difference vegetation index (NDVI). In the classification process, six land use/cover classes covering the bulk of the study area were determined as forest, grass, asphalt road, soil and bare area, urban and water. In the classification stage, 700 pixels for training and 300 pixels for testing were selected for each class to avoid possible bias among the classes. Classification resulted revealed that SVM classifier produced the best accuracy results for all three datasets. The highest accuracy (95.17%) was achieved with SVM classifier using the 11-band combination dataset. The combinations containing high spatial resolution bands provided higher accuracies. Moreover, McNemar's test was applied to analyze statistical significance of classifier performances for the datasets. Also F-score test was applied for all class to evaluate classification accuracy results. The results indicated that the differences between the performances were statistically significant except for SVM and RF using 7-band and 11-band combinations. To sum up, the efficiency of the machine learning algorithms applied in this study were all found effective in classification of Sentinel-2A imagery.

1. INTRODUCTION

In remote sensing, land use and land cover (LULC) maps are demanded hugely for land management and monitoring natural resources (DeFries et al., 2004). It is important to emphasize that thematic maps enable analysis of earth cover visually (Foody, 2002). The most significant way to produce LULC maps is by through classifying remotely sensed images. LULC maps are one of the vital instruments that supply information for several studies such as evaluation land use policy, urban planning, agricultural planning and ecosystem services (Guidici and Clark, 2017).

Remote sensing and image processing studies have gained acceleration with the developing satellite technologies, as in the case of Sentinel-2 mission by European Space Agency. It is designed for land and coastal studies. The mission has two platforms, namely Sentinel-2A and Sentinel-2B**.** In addition, it provides free access and high spatial resolution (10-60 m) imagery to users. It is extremely vital to use the correct classification method together with selecting the appropriate satellite image in order to produce useful LULC maps (Lu and Weng, 2007).

Until present, various algorithms, classifiers and strategies have been developed and applied to produce LULC maps (Kavzoglu and Colkesen, 2013; Xie et al., 2008). Machine learning algorithms such as random forest (RF), support vector machines (SVM) and decision trees (DT) are examples of the developed popular classifier methods. In recent years, machine learning algorithms have taken great attention in remote sensing applications (Kavzoglu et al., 2018; Thanh Noi and Kappas, 2017). They provide superior performance compared to traditional methods, especially in remote sensing applications (Gislason et all., 2006).

One of the most applied machine learning algorithms in studies the literature is random forest (RF) (Belgiu, and Dragut, 2016)**.** The widespread usage of RF is owing to the fact that it can be used classification and regression problems (Abdi, 2020). This algorithm generates more than one decision tree during the classification process to increase the accuracy of resulting thematic maps. Support vector machines (SVM), is another popular machine learning algorithm used for classification purposes. During the SVM process, data is transformed into a higher dimension. A hyperplane is created, and it is used to separate the two classes (Kavzoglu and Colkesen, 2010). Decision tree (DT) algorithm is a supervised classification method that has been successfully used in many areas with their structures similar to flow charts. The non-parametric structure of the method and its speed in problem solving have made the use of this algorithm widespread (Pal and Mather, 2003).

The main aim of this study is to apply machine learning algorithms to different band combinations of Sentinel-2A satellite images and compare the results. Not only the classification methods, but also selected data is important to produce LULC maps. In accordance with the purpose of study, training and test pixels were selected for each class and pixel-based image classification method was used to produce thematic maps. Three datasets were created to examine the effect of datasets on accuracies. The results were evaluated using McNemar's test to verify whether the results are statistically significant or not. Also, F-score values were calculated for each class to analyze accuracy assessment.

2. STUDY AREA AND DATASETS

The study site is located on the European and Asian sides of Istanbul in Turkey covering an area of approximately 234 km^2 land (Figure 1). It is surrounded by the Black Sea and Bosphorus, also includes newly constructed Yavuz Sultan Selim Bridge. One of the most significant conditions for a successful classification process is to select the appropriate dataset with high representativeness (Kavzoglu, 2009). Therefore, three datasets obtained from Sentinel-2A imagery were used as data sources with the intension of produce thematic maps of study field as described in Table 1. The study area mainly covers six main LULC classes; namely forest, grass, asphalt road, soil and bare area, urban and water. In compliance with the aim of this study, all datasets were constructed with various band combinations. The first dataset called Dataset-I includes 10 m spatial resolution bands and NDVI of Sentinel-2A image (i.e. band 2,3,4,8 and NDVI). NDVI for Dataset-I was calculated by the ratio of bands 8 (NIR) and 4 (RED). Dataset-II is the second dataset that consists of 20 m spatial resolution bands and NDVI image (i.e. 5,6,7,8A,11,12 and NDVI). It should be noted that, NDVI image for Dataset-II was computed from band 8A and band 5 (Fernández-Manso et al., 2016). Dataset-III is the last dataset which contains 10 m spatial resolution bands, pan-sharpened bands and NDVI (totally eleven bands). For pan-sharpening process, Gram-Schmidt sharpening method was employed to resample the 20 m resolution bands to 10 m resolution.

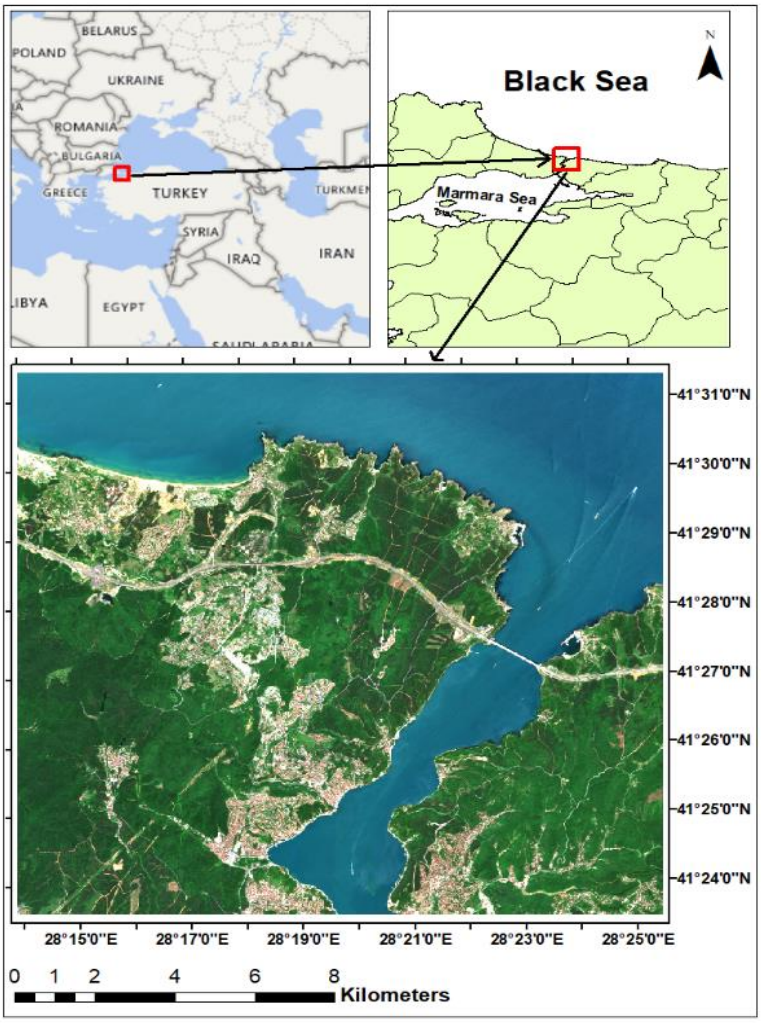

Figure 1: The location of the study area

Table 1: Bands of Sentinel 2-A sensor and datasets used in this study

Band	Band Names	Spatial Resolution (m)	Dataset-I	Dataset-II	Dataset-III
1	Coastal Aerosol	60	-	-	-
2	Blue	10	+	-	+
3	Green	10	+	-	+
4	Red	10	+	-	+
5	Vegetation Red Edge	20	-	+	+
6	Vegetation Red Edge	20	-	+	+
7	Vegetation Red Edge	20	-	+	+
8	NIR	10	+	-	+
8A	Narrow NIR	20	-	+	+
9	Water vapour	60	-	-	-
10	SWIR - Cirrus	60	-	-	-
11	SWIR	20	-	+	+
12	SWIR	20	-	+	+
Added	NDVI	10, 20	+	+	+

3. METHODOLOGY

3.1 Random Forest (RF)

Random forest (RF) is based on the principle of using decision trees as the basic classifier and creating a collective learning model by combining multiple decision trees (Breiman, 2001). RF classifier outperforms most classifiers because of robust against overfitting, easy to parametrize and speed (Kavzoglu, 2017). The main purpose of the RF classifier is to create multiple decision trees using bootstrapped sampling method. The training data set used to create tree models in the decision forest is randomly selected from the original training data set. Approximately, 2/3 of the randomly sampled data set is used to create the decision tree structure and the remaining part is used to test the validity of the created decision tree model. The class label of an uncertain sample is determined using the estimated majority voting principle of each tree model in the decision forest.

3.2 Support Vector Machines (SVM)

Support vector machines (SVM) is a nonparametric classifier method suggested to solve classification problems in data sets where patterns between variables are unknown. SVM is based on statistical learning theory. Although mathematical algorithms are designed to classify data that are linear and have two classes, it is generalized to classify nonlinear and data multi-class data. The working principle of SVM classifier is based on the method of defining the hyperplane that distinguishes the two classes optimally (Vapnik, 1995). Distance between support vectors are maximized and optimal decision function is created thanks to the obtained hyperplane.

3.3 Decision Tree (DT)

Decision trees method (DT) is a classification method that is widely applied in the literature since tree structures has simple rules used to create. In this classification method, the relationship between the data set and the classes are handled in stages (Colkesen, 2017). A simple tree structure consists of three basic parts namely, knots, branches and leaves. In the tree structure, each attribute is represented by a node (Friedl and Brodley, 1997). The basic principle to create a tree structure by using the attribute information of the training data can be expressed as asking questions to the data and reaching the results as soon as possible according to the obtained answers. The most significant processing step in creating DT is the criteria by which the branching in the tree will be made.

4. RESULT

In order to compare the performances of different machine learning methods for various datasets RF, SVM and DT classification algorithms were implemented for the three datasets. 700 training pixels and 300 testing pixels were chosen per class (i.e. six LULC classes) for the three datasets. Standard confusion matrices were used to calculate classification accuracies. The predicted kappa and overall accuracies values of the SVM, RF, DT classifiers for all datasets were given in Table 2**.** Moreover, F-score is computed from user and producer accuracies. The predicted overall accuracies and F-score values for all datasets, methods and classes as described in Table-3. It was observed that better results were obtained using SVM classification method compared to other classifiers considering the three datasets. It should be noted that the highest overall accuracy was computed as 95.17% when the SVM classifier was applied to Dataset-III including all bands at 10-meter resolution. On the other hand, the highest classification accuracy for RF classifier method was predicted as 94.00% also for Dataset-III. The highest accuracy for the DT method was acquired as 85.44% for Dataset-I. Furthermore, the lowest overall accuracy was predicted with DT classifier method for Dataset-1 as 78.56%. Results noticeably indicated that SVM classifier outperformed the RF and DT methods considering all three datasets. Comparing the results for the datasets overall accuracies estimated for the RF, SVM and DT classifiers varied about 6, 7 and 8%. Corresponding thematic maps were shown in Figure 2 from which visual interpretation can be performed.

Table 2: Classification overall accuracy and kappa values of datasets (OA: Overall Accuracy)

	Dataset - I		**Dataset - II**		**Dataset - III**	
	OA (%)	**Kappa (%)**	**OA (%)**	**Kappa (%)**	**OA (%)**	**Kappa (%)**
SVM	93.22	91.87	89.44	87.33	95.17	94.20
RF	92.39	90.87	88.56	86.27	94.00	93.80
DT	85.44	82.53	78.56	74.27	83.56	80.27

Table 3: Classification overall accuracies and F-score values of datasets (OA: Overall Accuracy, DS: Dataset)

LULC Classes	**SVM**			**RF**			**DT**		
	F-Score (%)			**F-Score (%)**			**F-Score (%)**		
	DS-I	**DS-II**	**DS-III**	**DS-I**	**DS-II**	**DS-III**	**DS-I**	**DS-II**	**DS-III**
Forest	93.70	93.13	93.52	93.93	93.73	94.16	91.19	91.85	91.34
Grass	93.53	87.02	93.53	93.42	87.46	93.85	87.85	77.73	87.85
Asphalt	94.12	81.39	95.39	93.35	80.42	93.83	81.34	64.29	74.83
Bare - Soil	89.43	85.76	94,76	86.25	83.01	90.69	77.58	64.66	68.98
Urban	88.45	89.77	93,76	87.19	87.52	91.40	82.15	71.90	78.25
Water	100	99.50	100	100	99.33	100	99.33	99.67	99.50
OA (%)	93.22	89.44	95.17	92.39	88.56	94.00	85.44	78.56	83.56

McNemar's non-parametric test was used to evaluate the statistical significance of the results obtained by classification algorithms (Foody, 2009; Demsar 2006). McNemar's test is based on χ^2 distribution, and it is applied to compare the errors of the two classifiers. If the chi-squared table value that 3.84 for 95% confidence interval is smaller than estimated value, it means that difference for two classification result are statistically significant. Values less than 3.84 are shown in bold in Table 4. McNemar's test verified that differences in classification success were statistically significant for the pairwise comparison of except SVM and RF classifiers for Dataset-I and Dataset-II combinations.

Table 4: McNemar's statistical test comparison table of classifier methods

	McNemar's Test		
	Dataset - I	**Dataset - II**	**Dataset - III**
SVM - RF	7.93	**3.51**	**2.68**
SVM - DT	186.08	109.24	155.24
DT - RF	168.50	92.84	140.00

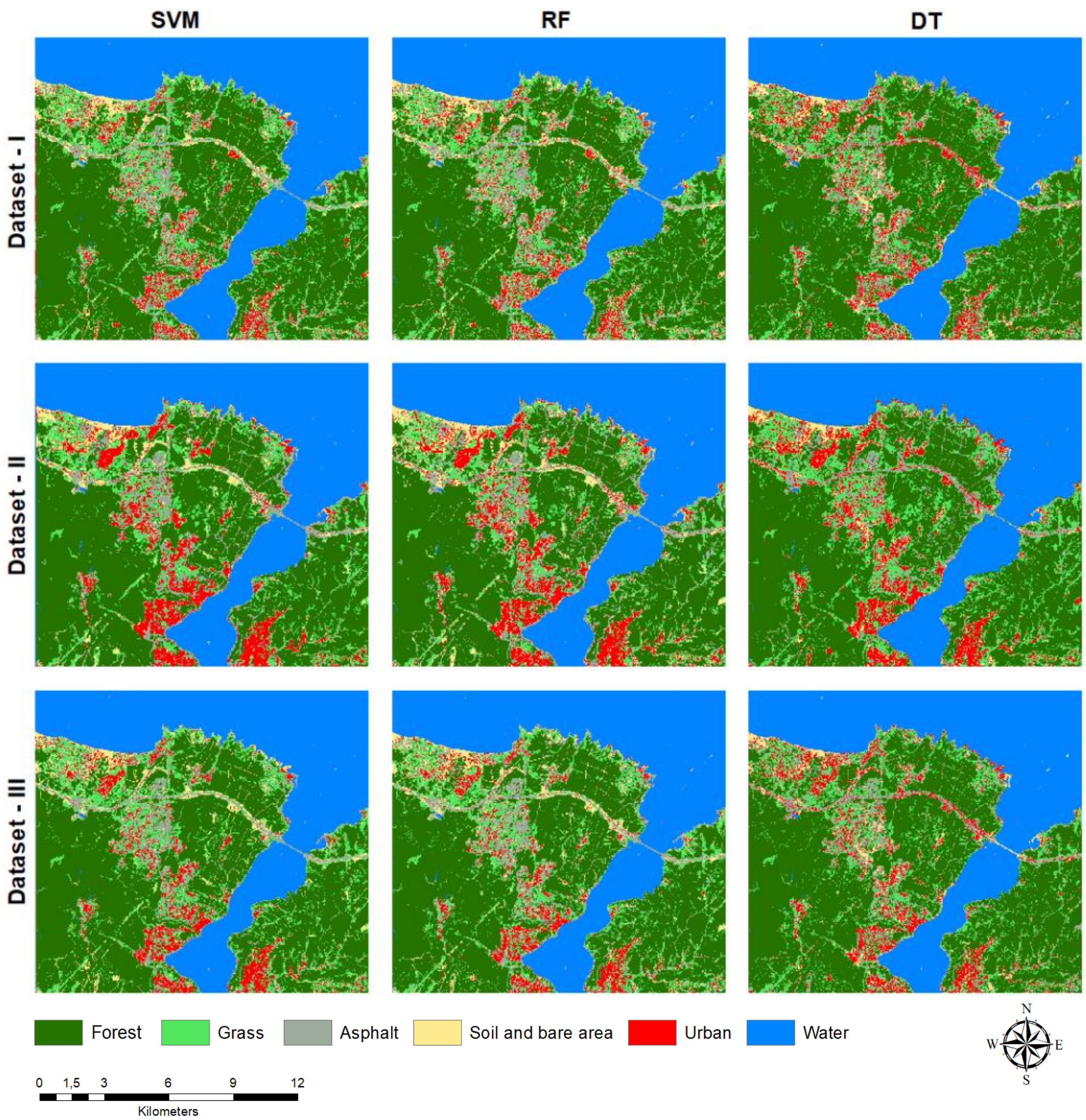

Figure 2: Produced thematic maps

5. CONCLUSION

The purpose of this study was to compare the efficiency of three machine learning classifiers, namely RF, SVM and DT using freely available Sentinel-2A imagery covering part of Istanbul, Turkey. In this context, pixel-based classification process was applied to the datasets consisting of different band combinations to delineate the effect of spectral bands. Some important findings were revealed using different datasets and methods. Firstly, it should be noted that Dataset-III was more informative and sufficient for classification, compared with Dataset-I and Dataset-II. Secondly, the result noticeably showed that SVM classifier was found more effective in terms of classification performances compared to RF and DT. To be more specific, the best classification accuracy was achieved as 95.17% for SVM classifier using Dataset-III. In addition, for RF classifier the highest classification accuracy was estimated as 94.00% using Dataset-III. Despite obtained better results for RF and SVM classifiers using Dataset-III, it was attained that better classification accuracy result as 85.44% when applied DT classifier for Dataset-I. Therefore, it could be recommended to use SVM classifier in classification process. Results point out that added bands influence positively the classification accuracy and delineation of the thematic maps for SVM and RF classifiers. For further assessment predicted results McNemar's statistical test was used and it was revealed that results are statistically significant except bold values in Table 4. Additionally, F-score values were calculated to evaluate overall accuracies for each class thus. Thirdly, the results showed that estimated overall accuracies for the RF, SVM and DT classifiers increased approximately 6, 7 and 8% respectively. Finally, the efficiency of SVM, RF and DT classifier methods in classification of satellite images for three datasets were examined with this study and SVM classifier outperformed compared other classifiers for all datasets. In conclusion, the SVM classifier and pan-sharpened bands together can be used to obtain more accurate LULC maps.

REFERENCES

Abdi, A.M., 2020. Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. GIScience & Remote Sensing, 57(1), pp. 1-20.

Belgiu, M., and Drăguţ, L., 2016. Random Forest in Remote Sensing: A Review of Applications and Future Directions. ISPRS Journal of Photogrammetry and Remote Sensing, 114, pp. 24-31.

Breiman, L., 2001. Random forests. Machine Learning, 45(1), pp. 5-32.

Colkesen, I., 2017. Object-Based Classification with Functional Trees: A Case Study of Worldwiev-2 Imagery. Harita Dergisi, 157 (1), pp. 9-21 (in Turkish).

DeFries, R.S., Foley, J.A., and Asner, G.P., 2004. Land-use choices: balancing human needs and ecosystem function. Frontiers in Ecology and the Environment, 2(5), pp. 249-257.

Demsar, J., 2006. Statistical Comparisons of Classifiers over Multiple Data Sets. Journal of Machine Learning Research, 7, pp. 1-30.

Fernández-Manso, A., Fernández-Manso O., and Quintano C., 2016. SENTİNEL-2A red-edge spectral indices suitability for discriminating burn severity. International Journal of Applied Earth Observation and Geoinformation, 50, pp. 170-175.

Foody, G.M., 2009. Classification accuracy comparison: Hypothesis test and the use of confidence intervals in evaluations of difference, equivalence and non-inferiority. Remote Sensing of Environment, 113(8), pp. 1658-1663.

Foody, G.M., 2002. Status of land cover classification accuracy assessment. Remote Sensing of Environment, 80(1), pp. 185-201.

Friedl, M.A., and Brodley, CE., 1997. Decision tree classification of land cover from remotely sensed data. Remote Sensing of Environment, 61(3), pp. 399-409

Gislason, P.O., Benediktsson J.A., and Sveinson J.R., 2006. Random Forest for Land Cover Classification. Pattern Recognation Letters, 27(4), pp. 294-300.

Guidici, D., and Clark, M., 2017. One-Dimensional Convolutional Neural Network Land-Cover

Classification of Multi-Seasonal Hyperspectral Imagery in the San Francisco Bay Area, California. Remote Sensing, 9(6), pp. 629.

Kavzoglu, T., 2009. Increasing the accuracy of neural network classification using refined training data. Environmental Modelling & Software, 24(7), pp. 850-858.

Kavzoglu, T., 2017. Object-Oriented Random Forest for High Resolution Land Cover Mapping Using Quickbird-2 Imagery. In: Handbook of Neural Computation, edited by Samui, P., Roy, S.S., and Balas, V.E., Amsterdam: Elseiver, Chapter 33, pp. 607-619.

Kavzoglu, T., and Colkesen, I., 2010. Investigation of the Effects of Kernel Functions in Satellite Image Classification Using Support Vector Machines. Harita Dergisi, 144(2), pp. 73-82 (in Turkish).

Kavzoglu, T., and Colkesen, I., 2013. An assessment of the effectiveness of a rotation forest ensemble for land-use and land-cover mapping. International Journal of Remote Sensing, 34(12), pp. 4224-4241.

Kavzoglu, T., Tonbul, H., Yildiz Erdemir, M., and Colkesen, I., 2018. Dimensionality reduction and classification of hyperspectral images using object-based image analysis. Journal of the Indian Society of Remote Sensing, 46(8), pp. 1297-1306.

Lu, D., and Weng, Q., 2007. A survey of image classification methods and techniques for improving classification performance. International Journal of Remote Sensing, 28(5), pp. 823-870.

Pal, M., and Mather, P.M., 2003. An assessment of the effectiveness of decision tree methods for land cover classification. Remote Sensing of Environment, 86(4), pp. 554-565.

Thanh Noi, P., and Kappas, M., 2018. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. Sensors, 18(1), pp. 18.

Vapnik, V.N., 1995. The Nature of Statistical Theory. Springer-Verlag, New York.

Xie, Y., Sha, Z., and Yu, M., 2008. Remote sensing imagery in vegetation mapping: a review. Journal of Plant Ecology, 1(1), pp. 9-23