Intercontinental Geoinformation Days

igd.mersin.edu.tr

The effect of auxiliary data (slope, aspect and elevation) on classification accuracy of Sentinel – 2A image using random forest classifier

Furkan Bilucan [*1], **Taskin Kavzoglu** [1]

[1]Gebze Technical University, Department of Geomatics Engineering, Kocaeli, Turkey

Keywords
Remote sensing
Random Forest Classifier
Auxiliary data
DEM
Sentinel-2A

ABSTRACT

Land use/land cover (LULC) maps provide irreplaceable information about the earth's surface, and investigation of their accuracy has always been an attractive research topic. Furthermore, it is known that accuracy of LULC maps could be improved using auxiliary data. One of the prevalent auxiliary data is the digital elevation model (DEM) because the surface landscape is affected directly by topography. Slope, aspect and elevation which are the main characteristics of the land surface are extracted from DEM. In this study, the effects of slope, aspect and elevation on classification accuracy were analyzed. For this purpose, Sentinel-2A satellite image together with the DEM data from the ALOS PALSAR satellite was assessed as auxiliary data for the classification process. Seven LULC classes covering the bulk of the study area were specified as urban, road, forest, water, bare and soil lands, cultivated and non-cultivated land in the classification process. To avoid possible bias among the determined classes, 700 pixels for training and 300 pixels for testing were chosen for each class. Classification results revealed that the highest accuracy (96.19%) were obtained when spectral bands were used together with elevation, slope, aspect data.

1. INTRODUCTION

Remote sensing is one of the most powerful tools for monitoring the Earth's cover or natural resources. In addition, satellite images that are the main product of remote sensing technologies, are a significant source to obtain information about the Earth. Land use/land cover (LULC) maps can be produced in accordance with the purpose of working with the analysis of remotely sensed images. The most commonly used method for producing LULC maps is the classification of satellite images. Extraction of accurate information from remotely sensed imagery has been a significant exploration issue for researchers (Rwanga and Ndambuki, 2017). For a successful classification process, it is vital to select a dataset with high representativeness (Kavzoglu, 2009). Moreover, it is well known fact that LULC maps play fundamental role in many studies including estimating forest supplies, observing climate changes, crop production forecasting and assessing water quality. Therefore, auxiliary data are commonly applied to increase the classification accuracy of thematic maps (Sang et al. 2021; Hurskainen et al. 2019). It has been reported that auxiliary data such as digital elevation model (DEM) and its derivatives can improve the classification accuracy (Zhu et al. 2016; Nguyen and Pham, 2015). Land characteristics can have a considerable influence the accuracy of the classification result in areas with large topographic differences. Additionally, high topographic changes could cause pixels belonging to the same class to have different spectral values and pixels belonging to different classes to have similar spectral values (Fahsi et al. 2000). In this case, it brings about decrease in accuracy and errors to occur during the classification process. In light of this information, the main purpose of this study is to investigate DEM, which has a positive effect on thematic maps classification accuracy. Furthermore, five different datasets were created using Sentinel-2A image and DEM data. In accordance with the aim of study, 700 train and 300 test pixels were selected per class to produce thematic maps considering pixel-based image classification method. Random forest algorithm was applied to five datasets. Overall accuracy, F-score values and Kappa coefficient values were calculated to compare the performances on the constructed five datasets.

* **Corresponding Author**

*(f.bilucan2020@gtu.edu.tr) ORCID ID 0000 – 0001 – 7920 – 6914
(kavzoglu@gtu.edu.tr) ORCID ID 0000 – 0002 – 9779 – 3443

Cite this study

Bilucan F & Kavzoglu T (2021). The Effect of Auxiliary Data (Slope, Aspect and Elevation) on Land Use/Land Cover Classification Accuracy of Sentinel–2A Image Using Random Forest Classifier. 2nd Intercontinental Geoinformation Days (IGD), 143-146, Mersin, Turkey

2. METHOD

2.1. Study Area and Datasets

The study area covering an area approximately 537 km^2 land is located on Kocaeli province in Turkey (Figure 1). The study site basically covers seven essential LULC classes namely forest, bare and soil lands, road, water, urban, cultivated, and non-cultivated. Five datasets were created to evaluate the effect of elevation, slope and aspect on classification accuracy.

Description of five constructed dataset used in this study were given in Table 1. Dataset-I is the main dataset that includes Sentinel-2A image bands (i.e. band 2, 3, 4, 5, 6, 7, 8, 8A, 11, 12). Dataset-II was created by adding the slope feature to the Dataset-I (totally eleven bands). Dataset-III is the third dataset obtained by adding the aspect land features to the Dataset-I (totally eleven bands). Another dataset called Dataset-IV was produced using Dataset-I including spectral bands and elevation map. The last dataset named as Dataset-V was created using a combination of Sentinal-2A image bands together with topographical features of slope, elevation and aspect. DEM data at 12,5 m spatial resolution was obtained from the ALOS PALSAR satellite and employed in this study.

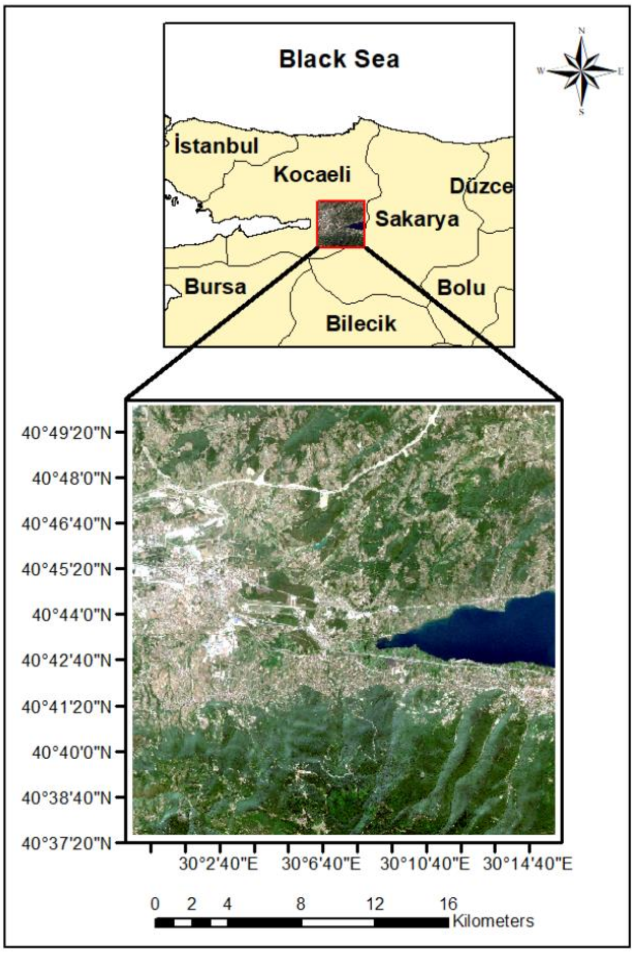

Figure 1. The study area, Kocaeli province of Turkey

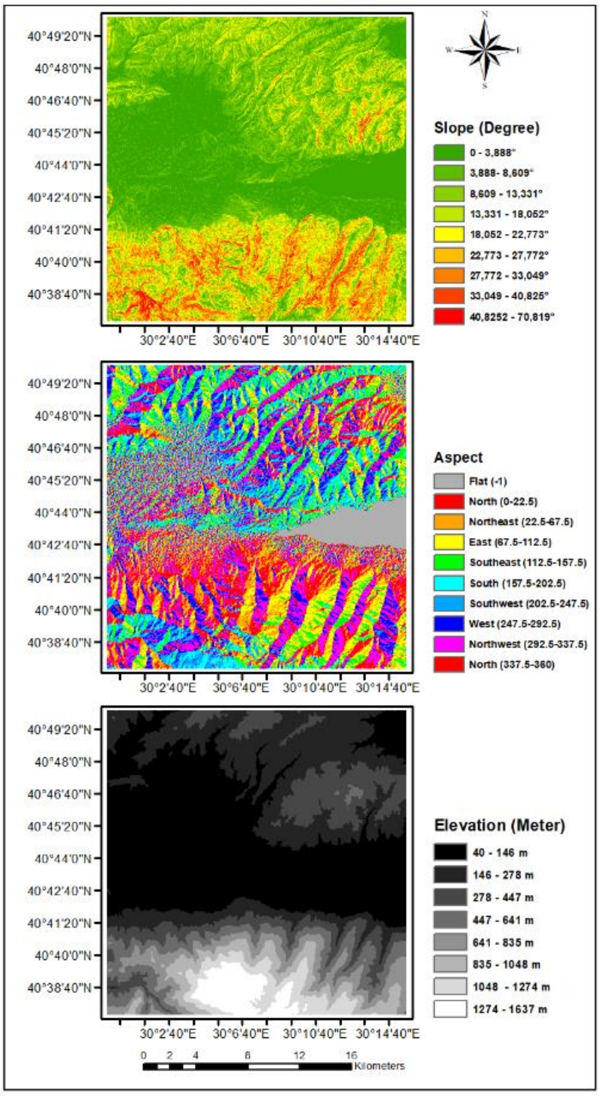

Figure 2. Slope, aspect and elevation feature maps

Table 1. Datasets used in this study

Datasets	Sentinel-2A bands	Slope	Aspect	Elevation
Dataset-I	√	-	-	-
Dataset-II	√	√	-	-
Dataset-III	√	-	√	-
Dataset-IV	√	-	-	√
Dataset-V	√	√	√	√

Slope, elevation and aspect features were derived from the DEM. The spatial resolution of the Sentinel-2A image bands was converted to 12.5 m spatial resolution by resampling method to combine with the topographical features. Slope is the angle of between any topographic surface with the horizontal plane and aspect is the position of the slopes against the sun. Additionally, vertical distance between sea level and any point on the land surface is defined as elevation. The slope in the study area varies from 0 and 70 degrees. Moreover, the study site has elevation values ranging from 40 to 1637 meters. In the aspect map produced for the study area, flat regions were represented by -1° and aspect map delineates nine geographical directions as shown in Figure 2.

2.2. Random Forest Classifier

In random forest algorithm (RF), decision trees are used as the basic classifier; thus, a collective learning model is created by combining multiple decision trees (Breiman, 2001). The RF technique that is one of the collective learning algorithms has become popular in remote sensing due to the high accuracy it provides (Belgiu and Drâgut 2016; Kavzoglu 2017).

Random subsets are obtained from the original training dataset for training decision trees located in the forest. Almost two thirds of the generated subsets are used to form the decision tree structure, and the remaining part is used to test the validity of the model (Kavzoglu et al. 2020). Each decision tree receives one vote as a result of classification; hence, the structure of the tree is determined by number of votes received. If there are enough trees in the forest for a RF model, the probability of overfitting problem is reduced. It is also used not only in classification applications, but also in the regression applications.

3. RESULTS

In order to examine the effect of the auxiliary data on classification accuracy, RF classifier was applied to the constructed datasets. All classification processes were performed using sklearn library in Python programming language. The parameter n_estimators which is the number of trees in forest was selected as 100. The verbose and n_jobs parameters were selected as 1. Overall accuracies and Kappa coefficient values were calculated using standard confusion matrices (Table 2). The calculated F-score values for all LULC class were shown in Table 3 where the highest F-score values were estimated for the water class. It was observed that the highest classification accuracy was estimated with 96.19% overall accuracy using Sentinel-2A image bands, slope, elevation and aspect.

The thematic map was produced with the best performing Dataset-V (Figure 3). It was found that elevation is the most informative or the contributing one compared to other topographical features considering differences in classification accuracy. The second highest overall accuracy result was estimated the with 96.10% using Sentinel-2A image bands and elevation land feature (Dataset-IV).

Table 2. Accuracy assessment results (OA: Overall accuracy)

Datasets	OA (%)	Kappa(%)
Dataset-I	94.86	94.00
Dataset-II	95.67	94.94
Dataset-III	94.76	93.89
Dataset-IV	96.10	95.44
Dataset-V	96.19	95.56

Table 3. F-score values of the datasets (DS: Dataset)

Classes	DS-I (%)	DS-II (%)	DS-III (%)	DS-IV (%)	DS-V (%)
Urban	89.68	91.19	89.61	90.60	91.06
Road	94.58	95.07	94.60	95.38	95.05
Forest	97.68	99.34	97.68	99.17	99.34
Water	100	100	100	100	100
Bare soil	92.78	93.93	92.57	95.25	95.40
Cultivated	96.36	97.68	96.52	98.68	98.84
Non-cultivated	92.79	92.38	92.21	93.46	93.60

Results noticeably indicated that accuracy increased when elevation and slope features were used as auxiliary data. It should be highlighted that the aspect feature did not contribute to the solution by slightly decreasing the classification accuracy. To sum up, the worst classification result (94.76%) was obtained using Sentinel-2A image bands and aspect (Dataset-III).

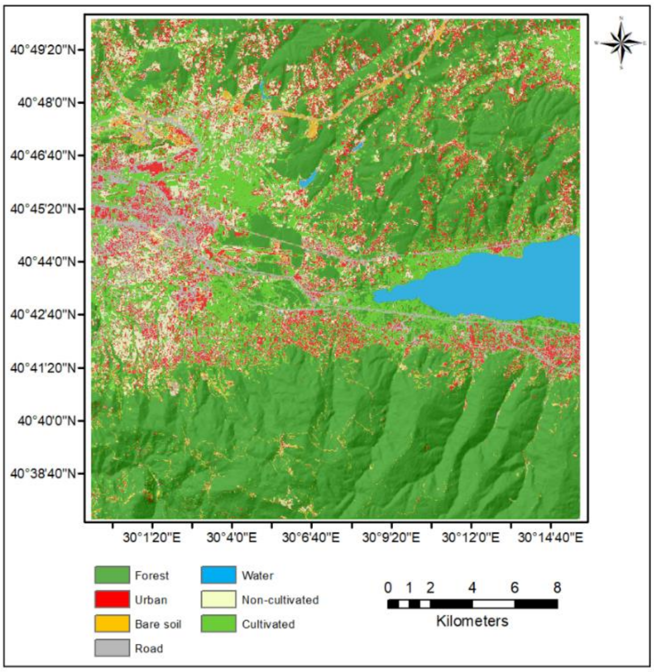

Figure 3. Thematic map produced with full dataset regression applications.

4. CONCLUSION

It is widely accepted that using auxiliary data is influential to enhance classification performance in remote sensing applications. In accordance with the aim of this study, the slope, aspect and elevation data were employed as auxiliary data to investigate their effect on classification result. Few researches focus on the impact of the slope, elevation and aspect on the classification result. In this study, pixel-based classification method was conducted using all datasets and some important results were reached. Firstly, Dataset-V was the most informative and contributing one compared to the other datasets since the highest accuracy was achieved with the use of this dataset. It should be emphasized that the second highest classification accuracy was achieved with the Dataset-VI that elevation used as auxiliary data. The reason for this result is due to fact that the land elevation change is large in study area. It should be also noted that when slope, elevation and aspect were evaluated together in classification processes, overall accuracy increased significantly. To be more specific, improvement in overall accuracy was approximately 1.5%. This finding indicates that elevation was more informative in classification processes compared to the slope and aspect. Even though the elevation and slope features have positive influence on classification accuracy, aspect caused a slight as specified in Table 2.

As a result, the main purpose of this study is to evaluate the contribution and quantify the effectiveness of DEM derivatives in improving LULC classification using Sentinel-2A image. Results pointed out that slope, aspect and elevation extracted from DEM, could be used as auxiliary data to obtain more accurate thematic maps. It is clearly revealed that high topographic changes in land cover considerably influence the LULC map accuracy. To sum up, the use of slope and elevation information should be utilized for areas with high topographic changes to enhance classification performance.

REFERENCES

Belgiu M & Drâgut L (2016). Random Forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing, 114, 24-31.

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

Fahsi A, Tsegaye T, Tadesse W & Coleman T (2000). Incorporation of digital elevation models with Landsat-TM data to improve land cover classification accuracy. Forest Ecology and Management, 128(1-2), 57-69.

Hurskainen P, Adhikari H, Siljander M, Pellikka P K E & Hemp A (2019). Auxiliary datasets improve accuracy of object-based land use/land cover classification in heterogeneous savanna landscapes. Remote Sensing of Environment, 233.

Kavzoglu T (2009). Increasing the accuracy of neural network classification using refined training data. Environmental Modelling & Software, 24(7), 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 & Balas V E, Amsterdam: Elsevier, Chapter 33, 607-619.

Kavzoglu T, Bilucan F & Teke A (2020). Comparison of Support Vector Machines, Random Forest and Decision Tree Methods for Classification of Sentinel-2A Image. 41st Asian Conference on Remote Sensing.

Nguyen T T H & Pham T T T (2016). Incorporating ancillary data into Landsat 8 image classification process: a case study in Hoa Binh, Vietnam. Environmental Earth Science, 75(5), 1-8.

Rwanga S S & Ndambuki M J (2017). Accuracy Assesment of Land Use/Land Cover Classification Using Remote Sensing and GIS. International Journal of Geosciences, 8(4), 611-622.

Sang X, Guo Q, Wu X, Xie T, He C, Zang J, Qiao Y, Wu H & Li Y (2021). The Effect of DEM on the Land Use/Cover Classification Accuracy of Landsat OLI Images. Journal of the Indian Society of Remote Sensing, 5.

Zhu Z, Gallant A L, Woodcock C E, Pengra B, Olofsson P, Loveland T R, Jin S, Dahal D, Yang L & Auch R F (2016). Optimizing selection of training and auxiliary data for operational land cover classification for the LCMAP initiative. ISPRS Journal of Photogrammetry and Remote Sensing, 122, 206-221.