6th Intercontinental Geoinformation Days

igd.mersin.edu.tr

Creation of land use and land cover maps of the Khachmaz-Shabran region of Azerbaijan using machine learning methods

Nariman Imranli*[1], Raziye Hale Topaloglu [2], Elif Sertel [3]

[1] *Turkish National Defence University, Atatürk Strategic Studies and Graduate Institute, Faculty of Satellite Technologies, Istanbul, Türkiye*
[2] *Yildiz Technical University, Faculty of Civil Engineering, Department of Geomatics Engineering, Istanbul, Türkiye*
[3] *Istanbul Technical University, Faculty of Civil Engineering, Department of Geomatics Engineering, Istanbul, Türkiye*

Keywords
Remote sensing
LULC
SVM
RF

Abstract

This study aims to produce Land Use/ Land Cover (LU/LC) maps of the Azerbaijan Khachmaz-Shabran region using machine learning (ML) methods and remotely sensed data. We used two common ML classification algorithms, Random Forest (RF) and Support Vector Machine (SVM). We generated two different LU/LC maps using an Azersky satellite image with a spatial resolution of 1.5 and a Sentinel-2 image with 20 m spectral bands. Both images were acquired on 1 August 2020. We implemented LU/LC class definitions of Level 2 of the CORINE nomenclature. After the classification step, an error matrix was created using the same reference points for Azersky and Sentinel-2. We compared the results of two classifications to determine the better-performing approach in obtaining the region's LU/LC maps.

1. Introduction

Remote sensing is used to map landscapes and infrastructures worldwide, manage natural resources, and study environmental change. With the emergence of high and very high-resolution satellite images with the developing technology, examining the current land use situation and the dynamic changes that have occurred over the years has become more accessible. Therefore, high and very high-resolution satellite data, which provides fast and reliable information, has become one of the practical information sources (Sertel et al., 2018; Topaloglu et al. 2022).

Information on Land Use/Land Cover (LU/LC) is essential for proper planning, management, and utilization of natural resources, including agricultural land and environmental protection (Sertel et al., 2018). Assessment of changes in LU/LC helps better understand the interactions between people and the environment, leading to better management of natural resources and sustainable development. Many methods and tools exist to classify and evaluate LU/LC and changes (Alipbeki et al. 2019; Topaloglu et al. 2022).

Machine learning (ML) techniques for LU/LC classification enable accurate and rapid analysis in many areas, such as land use planning, environmental management, monitoring of natural resources, and agricultural management (Algancı et al. 2015; Albayrak et al. 2018; Dey et al. 2021). Specifically, using ML algorithms such as Support Vector Machine (SVM), Random Forest (RF), and Neural Networks (NN), detailed LU/LC classes can be detected (Mountrakis et al. 2011; Algancı et al. 2015; Lui et al. 2016).

The RF classification algorithm is an ML method based on decision trees (Breiman 2001). Decision trees analyze the classes of training data and determine which class the test data belongs to according to the rules extracted from the training data (Algancı et al. 2015). The SVM is one of the most widely used classifiers for multispectral images. The method uses an optimal hyperplane with a maximum distance between the closest points. The reference vectors lie at the boundaries of the training samples, giving the maximum margin between the two classes and placing a separating linear hyperplane between them. An advantage of SVM is that it requires a small training sample size compared to traditional classifiers (Huang et al. 2002; Algancı et al. 2015).

The primary purpose of this study is to determine LU/LC maps of the Azerbaijan Khachmaz-Shabran region using two different ML techniques and to compare the accuracies of LU/LC maps created from Azersky and Sentinel-2 satellite data. RF and SVM were applied to classify 13 different land categories accurately, and

* **Corresponding Author**

*(imranli.nariman.94@gmail.com) ORCID ID 0009-0009-8711-9701
(haletopaloglu48@gmail.com) ORCI) ORCID ID 0000-0001-9706-8068
(sertele@itu.edu.tr) ORCID) ORCID ID 0000-0003-4854-494X

Cite this study

Imranlı, N., Topaloğlu, H., & Sertel, E. (2023). Creation of land use and land cover maps of the Khachmaz-Shabran region of Azerbaijan using machine learning methods. Intercontinental Geoinformation Days (IGD), 6, 199-203, Baku, Azerbaijan

results were examined to evaluate the performance of two algorithms and multi-sensor satellite images.

2. Study area and data

The Khachmaz-Shabran region, part of Azerbaijan's economic zone, is located in the northeast of the Greater Caucasus (Fig.1). Its primary industries are light and food industries; at the same time, the soils have a wide variety of crops and are very fertile. Various elevation ranges are observed in the region, with sea level in the coastal zone, lower elevation values in the plains adjacent to the coast, and higher elevations towards the western part, where forests and mountains are available. The region is rich in oil and gas, gravel, sand, clay, and other minerals (Bedelova and Valehov, 2019).

Figure 1. Study area

We used two remote sensing images; one was obtained from the Azersky and the other from the Sentinel-2 satellite. Both images were acquired on 01.08.2020 and with zero cloudiness. The Azersky and Sentinel-2 satellite data were obtained in level 2A and level 1 Ortho formats, respectively. Azersky image was pan-sharpened, including Red, Green, Blue, and Near-infrared spectral bands. We used the same spectral bands of Sentinel-2, but the spatial resolution is 20m in this case.

3. Methodology

We employed RF and SVM classification approaches to generate LU/LC maps. Figure 2 shows the whole flow chart of this study.

In this study, thirteen LU/LC classes based on Level-2 of CORINE nomenclature were mapped. These classes are:

- urban fabric
- industrial, comercial, and transport units
- mine, dump, and construction sites
- artificial, non-agricultural vegetated areas
- arable land
- pastures
- heterogeneous agricultural areas
- forests
- shrub and/or herbaceous vegetation associations
- open spaces with little or no vegetation
- coastal wetlands
- inland waters
- marine waters

Training sites were created for each class in ArcMap and ArcGIS Pro software. At least 20 samples were collected for each LU/LC class, and 300 samples were used to classify satellite images. The same training sites were utilized for Azersky and Sentinel-2 data sets and both SVM and RF classification algorithms.

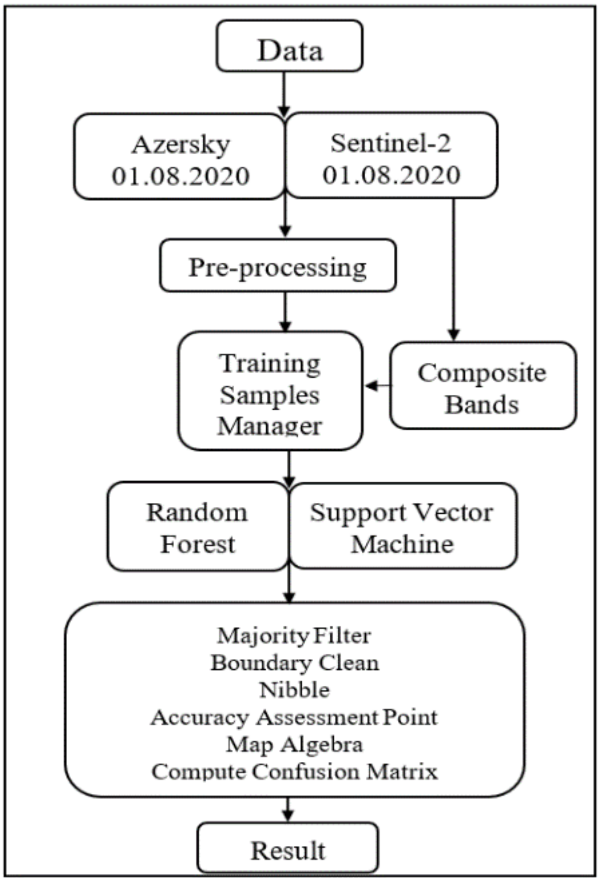

Figure 2. Flowchart of the study

3.1 Support vector machines

Support Vector Machine (SVM), a supervised and non-parametric statistical learning technique aims to find a hyperplane that divides training samples into a certain number of classes (Kavzoglu and Colkesen, 2009; Mountrakis et al. 2011; Du et al. 2019;). SVMs are classifiers that, in their most basic form, assign a binary test sample into one of two possible classes. By utilizing a kernel to translate feature space samples to a higher dimensional feature space, the SVM technique is extended to classes that cannot be separated linearly. SVMs are particularly useful in remote sensing because they can handle less training and achieve higher classification accuracy than conventional techniques (Chen and Guevara 2015).

3.2 Random forest

A group of binary decision trees makes up the ensemble method known as Random Forest (RF) classifier. By assembling a linear mixture of simpler models, it improves model performance. RF selects observations randomly for "small" binary trees. The performance of the method is enhanced by the comprehensive solution of this substantial forest (set) of trees. Based on the majority of labels from each tree, the final result is obtained (Li et al. 2020). When determining the specific classes, each classifier typically gives one vote, and classification is done by considering votes for all classifiers (Jing et al. 2020).

After the classification, we implemented a post-classification process to enhance the classification results. We employed smoothing with "Majority Filter", "Boundary Clean" and "Nibble" functions (Fig. 3), (Esri 2023).

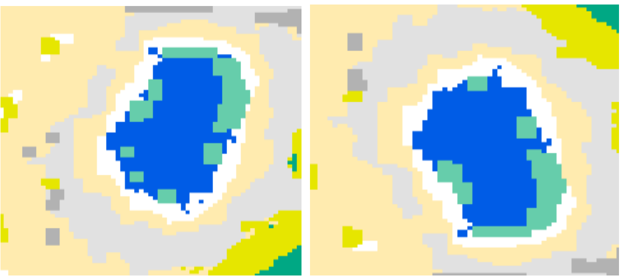

Figure 3. The difference before and after smoothing

To determine the accuracy of the final LU/LC maps, we conducted an accuracy assessment by generating an error matrix and calculating overall (OA), producer's (PA), and user's accuracy (UA) values and kappa statistics.

4. Results and Discussion

We determined 13 different LU/LC classes in the region both from Sentinel-2 and Azersky satellite images. Selected regions from LU/LC maps of Sentinel-2 and Azersky images using RF and SVM methods are shown in Figure 4. Each column represents a small area within the region. For the first region, more natural grassland areas were classified with RF-based Sentinel-2 classification. Inland water surfaces could be identified with both images and both methods. However, some parts of the shorelines of the lake were classified as the sea with the Azersky image.

We randomly generated approximately 400 reference pixels to conduct an accuracy assessment. The error matrix results are illustrated in Table 1 and Table 2.

We presented the accuracy metrics of Sentinel-2 LULC classification results in Table 1. We obtained a better OA value of 0.88 with the RF method compared to SVM which is 0.83. Specifically, for urban (1.1) and industrial, commercial, and transport units (1.2) classes, RF produced higher UA than the SVM method for Sentinel-2 classification. When we checked the confusion matrix of the Sentinel-2 RF LULC map, we observed that urban (1.1) and coastal wetlands (4.2) were mixed with the open spaces (3.3) class. In Sentinel-2 SVM, we found out that the industrial, commercial, and transport units (1.2), besides the urban (1.1) and coastal wetlands (4.2) classes, were mixed with open spaces (3.3).

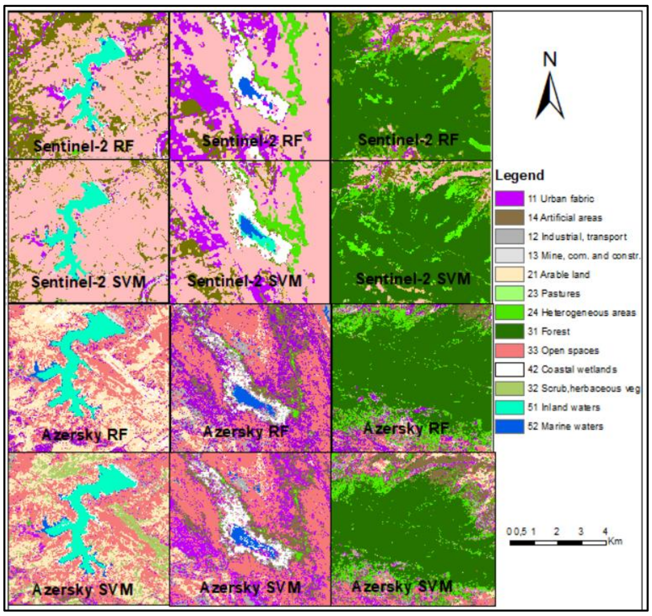

Figure 4. Classification results

On the other hand, we showed the accuracy metrics of the Azersky LULC classification results in Table 2. We acquired a better OA value of 0.78 with the SVM method. The RF method exhibited superior performance in terms of UA compared to the SVM method for the forest (3.1) and marine waters (5.2) classes in the Azersky classification. When we analyzed the RF confusion matrix of the Azersky satellite image based LULC map, we observed that the highly mixed class was the open spaces class (3.3), confused with the urban (1.1), arable land (2.1), coastal wetlands (4.2), mine, dump and construction sites (1.3) classes. In SVM based LULC mapf from the Azesky satellite image, urban (1.1), industrial, commercial, and transport units (1.2) and mine, dump, and construction sites (1.3) were mixed with the open spaces (3.3) class.

Table 1. Accuracy assessment for Sentinel-2 LULC maps

	Random Forest		Support Vector Machine	
	User's Accuracy	Producer's Accuracy	User's Accuracy	Producer's Accuracy
11 urban fabric	0.84	1	0.58	1
12 industrial, commercial and transport units	0.90	0.90	0.55	0.81
13 mine, dump and construction sites	0.87	0.96	0.87	0.90
14 artificial, non-agricultural vegetated areas	0.65	0.91	0.77	1
21 arable land	0.71	0.69	0.84	0.93
23 pastures	0.90	0.82	0.97	0.97
24 heterogeneous agricultural areas	0.94	1	0.94	1
31 forests	1	0.91	1	0.89
32 shrub and/or herbaceous vegetation associations	0.97	0.97	0.97	0.91
33 open spaces with little or no vegetation	0.97	0.65	0.97	0.40
42 coastal wetlands	0.77	0.92	0.35	0.92
51 inland waters	1	1	1	1
52 marine waters	1	0.94	1	0.86
overall accuracy	0.88		0.83	
kappa	0.81		0.82	

Table 2. Accuracy assessment for Azersky LULC maps

	Random Forest		Support Vector Machine	
	User's Accuracy	Producer's Accuracy	User's Accuracy	Producer's Accuracy
11 urban fabric	0.58	0.86	0.61	0.56
12 industrial, commercial and transport units	0.73	0.88	0.63	0.95
13 mine, dump and construction sites	0.58	0.90	0.58	0.90
14 artificial, non-agricultural vegetated areas	0.61	0.90	0.55	0.89
21 arable land	0.74	0.61	0.87	0.64
23 pastures	0.58	0.75	0.94	0.78
24 heterogeneous agricultural areas	0.94	0.97	1	1
31 forests	0.97	0.88	0.94	1
32 shrub and/or herbaceous vegetation associations	0.87	0.64	0.94	0.81
33 open spaces with little or no vegetation	0.84	0.45	0.94	0.45
42 coastal wetlands	0.42	0.38	0.48	0.50
51 inland waters	0.87	1	1	0.89
52 marine waters	0.97	0.83	1	0.90
overall accuracy	0.75		0.78	
kappa	0.73		0.75	

5. Conclusion

Our results show that the highest OA of 0.88 was obtained with the RF based classification of the Sentinel-2 image. The Coastal Wetlands, Urban, and Green Urban classes exhibited lower UA and PA values for both satellite images. The Coastal Wetlands class is mostly mixed with Open Spaces and Arable land. Some Urban areas were misclassified as the Open Spaces and some Road, Industry areas were misclassified as the Open Spaces. Inland waters, marine waters, and forest areas were easily identified from both satellite images with both classification methods.

This study showed that both Sentinel-2 and Azersky satellite images can be used to create up-to-date LU/LC maps. Although some classes suffered from mixture problems, most of the classes were accurately identified. To overcome this problem, some of the similar characteristic classes could be merged, and LULC maps could be generalized if it fits the purpose of the study.

References

Albayrak, S., Koc-San, D., & Sivrikaya, F. (2018). Land cover classification using machine learning algorithms. Journal of Environmental Management, 207, 205-214.

Alganci, U., Ozdogan, M., Sertel, E., & Ormeci, C. (2015). Estimating maize and cotton yield in southeastern Turkey with integrated use of satellite images, meteorological data and digital photographs. Field Crops Research, 157, 8-19.

Alipbeki, O. A., Alipbekova, Ch. A., & Amangaliev, K. S. (2019). Assessment of The Accuracy of Land Use Classification. International Scientific journal "Bulletin of Science" No. 12

Bedelova, A. N., & Valehov, N. S. (2019). GİS əsasında Xaçmaz rayonunun torpaq və meşə örtüyünün yamacların meylliyindən asılı olaraq ekoloji qiymətləndirilməsi.

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

Chen, J., & Guevara, M. (2015). Support vector machine for land cover classification: a study from the central Amazon region, Brazil. Environmental Monitoring and Assessment, 187(3), 1-13. https://doi.org/10.1007/s10661-015-4416-4.

Esri, (2023). https://desktop.arcgis.com/en/arcmap/, accessed in May 2023.

Dey, P., Das, M., & Maiti, S. K. (2021). Land use and land cover classification of a mining region using machine learning algorithms: a case study in India. Environmental Earth Sciences, 80(2), 1-17.

Du, P., Feng, X., Li, Y., & Wan, J. (2019). Mapping land use and land cover using machine learning techniques: A comparative study of random forest and support vector machine algorithms. Remote Sensing, 11(23), 2829.

Huang, C., Davis L. S., & Townshend J. R. G. (2002). An assessment of support vector machines for land cover classification, International Journal of Remote Sensing, 23(4), 725-749.

Jing, L., Zhang, Z., Yang, X., & Wang, H. (2020). Comparison of Random Forest and SVM in the classification of land cover using Sentinel-2 imagery. Peer Journal, 8, e10105.

Kavzoglu, T., Colkesen, I. (2009). A kernel functions analysis for support vector machines for land cover classification, International Journal of Applied Earth Observation and Geoinformation, 11, 5, 352-359.

Li, W., Wu, C., Zhou, W., Li, W., Huang, X., & Zhang, W. (2020). An improved random forest algorithm for the classification of Corine land cover using Sentinel-2A data. International Journal of Applied Earth Observation and Geoinformation, 93, 102218. https://doi.org/10.1016/j.jag.2020.102218.

Lui, W., Wang, J., & Qiu, S. (2016). A random forest based approach for land use and land cover classification using Landsat 8 imagery. Remote Sensing, 8(9), 786.

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing, 66(3), 247-259.

Sertel, E., Topaloğlu, R. H., Şallı, B., Algan, I. Y., & Aksu, G. A., (2018). Comparison of Landscape Metrics for Three Different Level Land Cover/Land Use Maps. ISPRS Internatıonal Journal Of Geo-Informatıon ,7,10.

Sertel, E., Musaoğlu, N., Alp, G., Yay, I., Kaya, Ş., Yüksel, B., & Yılmaz, A. (2018). 1:25000 ölçekli ulusal Arazi Örtüsü/Kullanımı Sınıflandırma Sistemi ile HGK TOPO Veritabanının Karşılaştırılması. Harita Dergisi , 160, 34-46.

Topaloglu, R. H., Aksu, G. A., Ghale, Y. A. G., & Sertel, E., (2022). High-resolution land use and land cover change analysis using GEOBIA and landscape metrics: A case of Istanbul, Turkey. Geocarto International, 37, (25), 9071-9097.