Intercontinental Geoinformation Days

igd.mersin.edu.tr

Supervised machine learning classification in Google Earth Engine: Time series analysis in Akkuyu, Turkey

Muzaffer Can İban*[1], Şafak Bozduman [2], Ezgi Şahin [2]

[1]Mersin University, Faculty of Engineering, Department of Geomatics Engineering, Mersin, Turkey
[2]Mersin University, Institute of Science, Department of Remote Sensing and Geographic Information Systems, Mersin, Turkey

Keywords
Supervised Classification
Machine Learning
Remote Sensing
Land Use and Land Cover
Google Earth Engine
Nuclear Power Plants

ABSTRACT

This paper tests the potential of the Google Earth Engine platform and Sentinel-2 data to classify and monitor land use and land cover nearby a nuclear power plant construction which is being eroded in Akkuyu district of Mersin Province in Turkey. After classification study, the performance of different supervised machine learning classification algorithms are compared. According to the accuracy assessment results, the random forest classifier has performed slightly better classification than other classifiers. According to the numerical results, Akkuyu NPP project has doubled the footprint of built-up areas. The project has brought a significant deforestation and filling areas on the seashore.

1. INTRODUCTION

The classification of land use and land cover (LULC) presents crucial information to comprehend the relationships between humankind and environment. Remote sensing (RS) and geographic information systems (GIS) have been regularly used for mapping and monitoring LULC, thanks to a wide range of free data and software. Nowadays, many researchers are able to implement more precise and accurate LULC change analysis with the help of high-resolution RS products. Lambin et al. 2001; Mas, 1999)

RS and geospatial big data approach can help to understand the continuous change on earth's surface. Such big data requires time-series composite images, which brings higher variability in the spectra of different LULC classes. Conventional approach to implement multi-temporal big data processing has required to satellite-data searching, downloading and storing. Besides, computational processing capacities need to be powerful enough to manage all data and to run different complex classifiers (Tamiminia et al. 2020).

European Space Agency's Copernicus Program has launched several satellite missions including Sentinel-1,2,3 and 5 satellite constellations. The scientific community, governments and private sectors have used Sentinel-2 data for land use and land cover monitoring. Sentinel-2 offers improved data compared to other low to medium spatial resolution satellite images, especially in temporal and spatial resolution (Gascon et al 2017).

To handle satellite imagery there are two different system architecture: on one hand cluster-based high performance computing where a single image is processed and stored by the co-operation of several computers. On the other hand, there are cloud-based platforms where the capabilities of supercomputers can be achieved without owning a good storage and processing system.

The Google Earth Engine (GEE) platform is a free cloud-based tool of Google where the users can access and process petabyte scales of remotely sensed data by using Javascript and Python coding on its own user interface. GEE takes advantage of Google's computational infrastructure to reduce operational time and provide a repository for script storing and sharing. The GEE data catalogue allows access to multiple satellite data and satellite-derived products (Gorelick et al. 2017).

This paper was conceived with two main objectives: (1) to test the potential of the GEE platform and Sentinel-2 data to classify and monitor land use and land cover nearby a nuclear power plant (NPP) construction which is being eroded in Akkuyu district of Mersin Province in Turkey; (2) to compare and assess the performance of different supervised machine learning (ML)

* **Corresponding Author**

*(caniban@mersin.edu.tr) ORCID ID 0000 – 0002 – 3341 - 1338
(safak.bozduman@gmail.com) ORCID ID 0000 – 0003 – 0359 – 9090
(ezgisahin.9600@gmail.com) ORCID ID 0000 – 0002 – 7455 – 8141

Cite this study

İban M C, Bozduman Ş & Şahin E (2021). Supervised machine learning classification in Google Earth Engine: Time series analysis in Akkuyu, Turkey. 2nd Intercontinental Geoinformation Days (IGD), 238-240, Mersin, Turkey

classification algorithms in terms of the obtained classification accuracy.

2. MATERIALS AND METHODS

2.1. Study Area

Akkuyu NPP is under construction on Mediterranean Sea coast in Mersin Province, Turkey. It will be Turkey's first nuclear power plant. Therefore, it is crucial to monitor the impact of such a mega-project on LULC of the surrounding area. As seen in "Fig. 1", the image classification frame is selected as a 25 x 15 km rectangle around the NPP construction site. However, LULC change analysis is implemented in 1-km and 5-km buffers centered on the site.

Figure 1. Image Classification Frame

2.2. Image Classification

Via GEE platform, the least cloudy Sentinel-2 MSI Level 2-A images were obtained from scihub for 2017, 2019 and 2021 in order to visualize the LULC change in biannual time series. Each LULC class (water bodies, built-in areas, agricultural areas, forests and barren land) are trained in pixel-wise approach. For each class, a set of 100 points are selected and marked on images. 75 of them are used for training and 25 points are used for testing.

GEE platform has built-in supervised machine learning image classification functions. These are classification and regression trees (CART)(Lawrance and Wright 2001), random forest (RF) (Pal 2005) and support vector machines (SVM) (Mountrakis et al. 2011). All these classification algorithms were executed on three different images with the same training and testing set of points. The hyper-parameters of ML algorithms are kept as default and they are not tuned on GEE platform.

After obtaining the classified images, the images are clipped to aforementioned buffer frames.

3. RESULTS

Table 1 shows the overall accuracy and kappa coefficients for the classifications of three different images with SVM, CART and RF classifiers.

Table 1. Accuracy Assessment

Years	ML Algorithm	Overall Accuracy	Kappa Coefficient
2017	SVM	0.74	0.67
	CART	0.81	0.75
	RF	0.86	0.83
2019	SVM	0.69	0.61
	CART	0.85	0.81
	RF	0.89	0.87
2021	SVM	0.81	0.77
	CART	0.84	0.79
	RF	0.89	0.87

"Fig. 2" and "Fig.3" shows the LULC classification maps of the region in 2017, 2019 and 2021 images with three classifying ML algorithms.

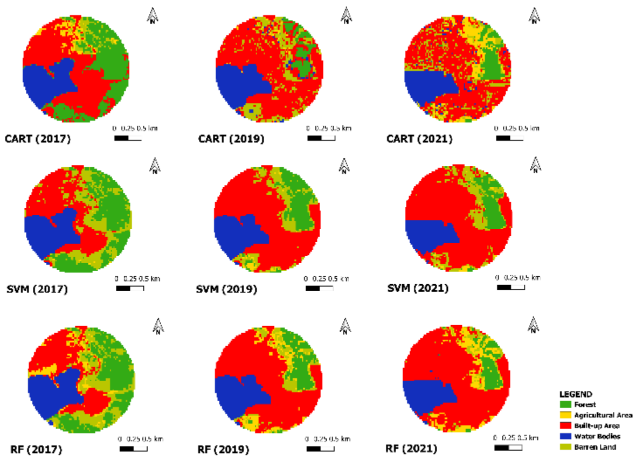

Figure 2. LULC Classification Maps in 1-km Buffer Zone

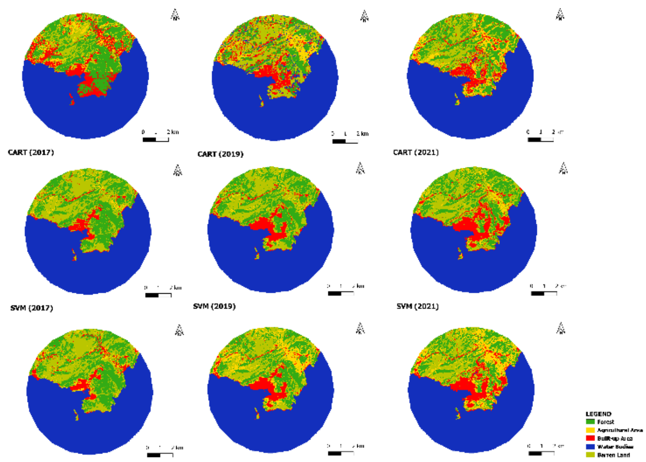

Figure 3. LULC Classification Maps in 5-km Buffer Zone

Table 2 shows the percentage of each LULC classes in the classification maps. By this table, it is possible to investigate the LULC change between 2017 and 2021 during the construction phase

Table 2. LULC Analysis

Buffer Zones	Year	ML	Forest (%)	Agriculture (%)	Built-up (%)	Water Bodies (%)	Barren Land (%)
1 km	2017	SVM	26.66	0.23	31.81	19.73	21.56
		CART	29.25	4.39	43.08	18.65	4.62
		RF	25.93	7.52	26.50	19.85	20.20
	2019	SVM	13.88	1.43	54.40	17.57	12.71
		CART	6.81	1.62	55.67	19.62	16.28
		RF	11.49	8.72	54.19	17.76	7.85
	2021	SVM	10.05	1.53	60.96	14.35	13.11
		CART	8.50	10.45	49.45	16.77	14.82
		RF	7.99	10.88	61.57	14.99	4.58
5 km	2017	SVM	18.71	1.52	4.07	54.70	21.01
		CART	20.02	5.02	10.30	54.43	10.23
		RF	18.09	2.44	4.35	54.73	20.38
	2019	SVM	18.05	2.07	5.79	54.48	19.61
		CART	10.50	5.55	8.82	57.51	17.61
		RF	15.38	9.16	5.79	54.64	15.02
	2021	SVM	16.00	3.60	8.43	54.06	17.90
		CART	16.62	7.57	7.23	54.57	14.00
		RF	13.40	11.35	8.46	54.30	12.48

4. DISCUSSION AND CONCLUSIONS

According to the accuracy assessment results, the RF classifier has performed slightly better classification than SVM and CART classifiers. Even though CART classifier gives a better result than SVM, LULC classification maps derived from CART classifier seems to be over-trained. In contrast, RF and SVM classifiers enable to obtain a descent distribution of LULC classes.

According to the numerical results of LULC change analysis between 2017 and 2021, Akkuyu NPP project has doubled the footprint of built-up areas. The project has brought a significant deforestation and filling areas on the seashore. The operating NPP is expected to absorb more population, which may lead to new development areas, and more loss in forest and agricultural lands.

REFERENCES

Gascon, F., Bouzinac, C., Thépaut, O., Jung, M., Francesconi, B., Louis, J., ... & Fernandez, V. (2017). Copernicus Sentinel-2A calibration and products validation status. Remote Sensing, 9(6), 584.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote sensing of Environment, 202, 18-27.

Lambin, E. F., Turner, B. L., Geist, H. J., Agbola, S. B., Angelsen, A., Bruce, J. W., ... & Xu, J. (2001). The causes of land-use and land-cover change: moving beyond the myths. Global environmental change, 11(4), 261-269.

Lawrence, R. L., & Wright, A. (2001). Rule-based classification systems using classification and regression tree (CART) analysis. Photogrammetric engineering and remote sensing, 67(10), 1137-1142.

Mas, J. F. (1999). Monitoring land-cover changes: a comparison of change detection techniques. International journal of remote sensing, 20(1), 139-152.

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.

Pal, M. (2005). Random forest classifier for remote sensing classification. International journal of remote sensing, 26(1), 217-222.

Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S., & Brisco, B. (2020). Google Earth Engine for geo-big data applications: A meta-analysis and systematic review. ISPRS Journal of Photogrammetry and Remote Sensing, 164, 152-170.