COMPARISON OF MACHINE LEARNING ALGORITHMS IN DETERMINATION OF SHALLOW WETLAND AREAS

N.Yagmur [a,*], A.Dervisoglu [a], B. B. Bilgilioglu [a,b]

[a] ITU, Civil Engineering Faculty, Geomatics Engineering Department, 34000 Maslak Istanbul, Turkey - (adervisoglu, bilgilioglu16, yagmurn)@itu.edu.tr

[b] Gumushane University, Faculty of Engineering and Natural Sciences, Geomatics Engineering Department, Gumushane, Turkey- (bahabilgilioglu@gumushane.edu.tr)

KEY WORDS: Wetland, Remote Sensing, Machine learning, MLC, SVM, RF and NN

ABSTRACT:

Wetlands which have been emerged as a result of natural processes, are considered as the most important genetic reservoirs of the earth with its rich plant and animal species. In the literature, it is seen that using remote sensing data and methods is an effective and important way in determining, analysing and monitoring the change in the wetlands from past to present years. In this study, Duden (Kulu) Lake which is located within the Konya Closed Basin (KCB), is chosen as the study area. In order to obtain the effects of climate change and negative human activities on the wetland surface area, multi-temporal analyses are widely used with remote sensing techniques. Google Earth Engine (GEE) is quite usable in especially multi-temporal analysis. With this study, the water surface area of the Duden Lake was investigated temporally over the years 1984 and 2018 by using the GEE, and the change of the water surface area was obtained approximately as 50%. Landsat satellite has provided medium spatial resolution images from the 1970s and the collected data are commonly used to get multi-temporal analysis. The usage of machine learning algorithms has been widely increased in the remote sensing field over the recent years. In this study, Sentinel 2 satellite images, with having high spatial resolution, were used to analyse the performance of different machine learning algorithms in terms of detecting shallow water area. The classification results obtained through the usage of machine learning algorithms were examined with accuracy assessment.

1. INTRODUCTION

Wetlands provide breeding, growing and feeding opportunities to many fish and wildlife species and are shallow waters which are considered as the bridge between terrestrial and water systems (Cowardin et al., 1979). Wetlands and lakes have an important position for the Earth's surface system such as atmosphere, hydrosphere etc, and are affected from environmental changes and human activities (Liu et al., 2019). Because of the wildlife and structure of wetlands, accessing wetlands is often difficult; so, monitoring wetlands via remote sensing is the optimal solution (Zomer et al., 2009).

There are numerous methods to detect surface water area such as spectral indices and classification. In spectral indices, especially Normalized Difference Water Index (NDWI) has been widely used in change detection. Liu et al. (2019) improved a system which automatically identifies the lake area using NDWI enabling the determination of spatio-temporal change area.

Another commonly used method is classification. Unsupervised classification and supervised classification methods are used for mapping of wetlands with using different satellite images. However, classification of wetlands is difficult due to the different characteristics of the wetlands (Ozesmi & Bauer, 2002). Tian et al. (2016), applied random forest classification method to detect the wetland landcover. Amani et al. (2017) compared the five different machine learning classifiers (K-nearest neighbour, Maximum Likelihood, Support vector machine, Classification and regression trees, and Random forest) in terms of performance in the wetland classification and the results showed that the overall accuracies of classifications were nearly similar, except RF classifier. According to Ozesmi and Bauer (2002), maximum likelihood classification method is commonly used supervised classification method.

In this study, two approaches will be evaluated. In the first approach, spatio-temporal change of the Duden (Kulu) Lake was determined with NDWI. In the second approach, four machine learning classifiers were evaluated according to the classification performance in the shallow wetland area. These classifiers are Random Forest (RF), Support Vector Machine (SVM), Maximum Likelihood Classifier (MLC) and Neural Networks (NN). Landsat satellite data were used for temporal change detection due to the its large image archive from the 1970's, whereas Sentinel 2 data were used for classification because of having higher spatial resolution.

2. STUDY AREA

Konya Closed Basin (KCB) is one of the 25 hydrological basins in Turkey and covers 7% of the country. There are numerous wetlands and natural lakes in KCB and one of them is Duden (Kulu) Lake. The Duden Lake, which was declared as a protected area in 1992, is the most important growth point of endangered white-headed duck worldwide. The lake is an important accommodation point for coastal birds during their migration periods. Duden Lake is an enclosed shallow wetland that contains 185 bird species, most of which are flamingos. Duden Lake was chosen as the study area because of having shallow water. Figure 1 shows the study area.

* Corresponding author.

Figure 1. Study area a) location of the Duden Lake and KCB b) Duden Lake on Landsat 8 False Color (5-3-2)

3. DATA AND METHODOLOGY

3.1 Data

In this study, two different satellite images were used for detecting the water area; which are Landsat and Sentinel 2 images, both providing free data access. Landsat has provided satellite images from 1970s, so it has a wide range of usage for temporal analysis. Landsat 5 TM and 8 OLI images were used with a 10-year interval. Landsat data was used in the first approach, and Sentinel data with its higher spatial resolution was used in the second approach. The information of these satellite images are given in Table 1.

Satellite	Landsat -5 TM	Landsat -8 OLI	Sentinel 2 MSI
Spatial Resolution	Red, Green, Blue, NIR, SWIR 1-2: 30m TIR: 120 m	Pan: 15 m Coastal, Cirrus, Red, Green, Blue, NIR, SWIR 1-2: 30m TIR 1-2: 100 m	Coastal, Water vapour, SWIR Cirrus: 60m Red, Green, Blue, NIR: 10 m Vegetation Red Edge, SWIR 1-2 : 20 m
Radiometric Resolution	8 bit	12 bit	12 bit
Temporal Resolution	16 days	16 days	5 days
Spectral Resolution	Red :0,63-0.69 µm Green :0,52-0.60 µm Blue :0,45-0.52 µm NIR :0,76-0.90 µm SWIR 1:1.55-1.75 µm SWIR 2:2.08-2.35 µm	Red : 0,636-0.673 µm Green :0,533-0.590 µm Blue :0,452-0.512 µm NIR :0,851-0.879 µm SWIR 1:1.566-1.651 µm SWIR 2:2.107-2.294 µm	Red :0,665 µm Green :0,560 µm Blue :0,490 µm NIR :0,842 µm SWIR 1:1.610 µm SWIR 2:2.190 µm
Used Date	May 6, 1984 May18, 1994 May13, 2004	June 26, 2014 May 4, 2018	May 17, 2018

Table 1. Landsat and Sentinel 2 satellite image properties

Kulu Meteorological Station is the closest station in the Duden Lake area; so the data obtained from Kulu Meteorological Station was used in determining the reasons of the temporal change with the evaluation of the precipitation and temperature parameters.

3.2 Methods

In this study, Duden Lake was evaluated with two different approaches. In the first approach, temporal change of Duden Lake was determined by using NDWI on Landsat images. NDWI uses the green and NIR bands of the satellite images. Results change between -1 and 1, and values which are greater than zero shows the water area. In the second approach, the performance of the four different machine learning classification methods was evaluated in detection of the shallow water area by using Sentinel 2 image. Flowchart of the study is given in the Figure 2.

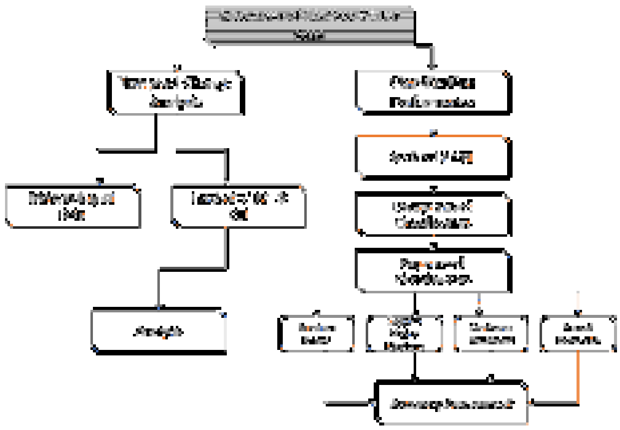

Figure 2. Flowchart of the study

Temporal change analysis was made with Google Earth Engine (GEE). GEE platform is a cloud-based platform; enables to use, monitor and process the satellite images without the need of downloading and provide earth images more than forty years ago (URL 1). Temporal change results were analysed with meteorological parameters of the Kulu Meteorological Station.
For evaluation of the classification performance, firstly ISODATA (Iterative Self Organizing DATA) unsupervised classification method was applied for determining training datasets. ISODATA calculates the class means iteratively using minimum distance (Abbas et al., 2016).
Four different machine learning algorithms were applied on Sentinel 2 satellite image to classify shallow water. The first one is Random Forest (RF) classification algorithm. RF classifier is an ensemble classifier which generates multiple decision trees using a randomly selected set of training samples and variables (Belgiu & Drăguţ, 2016). Support Vector Machine (SVM) and Neural Networks (NN) are non-parametric supervised classifiers. The hyper-planes (support vectors) are selected in SVM. These support vectors maximize the distance between the given classes (Otukei & Blaschke, 2010). In the NN, there are processing nodes which are referred as neurons. It uses weights of neurons instead of the algorithm to regulate the network connection (Jiang et al., 2011). Unlike SVM and NN, Maximum Likelihood Classifier (MLC) is a parametric supervised classifier. It depends on the statistics of a Gaussian probability density function (pdf) model for each class and computes the likelihood of unknown measurement based on the Bayesian equation (Paola & Schowengerdt, 1995; Otukei & Blaschke, 2010).

4. RESULTS AND DISCUSSION

4.1 Temporal Change Analysis

Temporal change of Duden Lake was determined with NDWI from 1984 to 2018 with ten years of interval. NDWI was applied on Landsat images. Satellite images were selected from May, when the water was the most abundant. Because of the images being cloudy, satellite image was selected from June in 2014. Results are shown in the Figure 3.

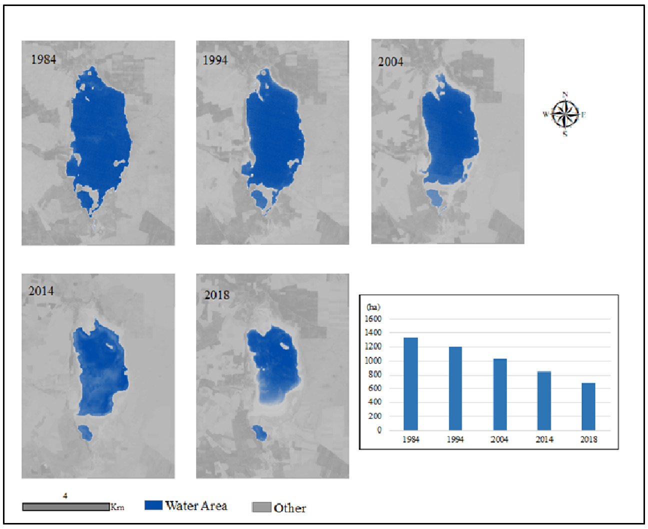

Figure 3. Changes of water surface area by using NDWI between 1984 and 2018

According to NDWI results, Duden Lake had 1340 ha water surface area in 1984, but it has decreased to 686 ha in 2018. It has lost approximately 50% of the water surface area. Precipitation and temperature changes from 1984 to 2017 were shown which is shown in the Figure 4.

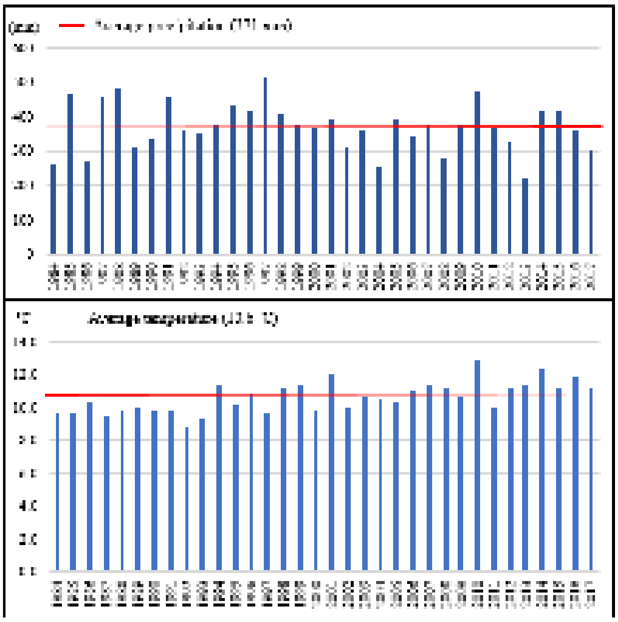

Figure 4. Precipitation and temperature of Kulu Meteorological Station between 1984 and 2017

Average precipitation rate in Turkey has been detected as 571 mm between 1980 and 2017. According to meteorological data, average of the precipitation rate in Kulu is 371 mm and average of the temperature is 10.6 ºC. Average of the precipitation rate in Kulu Meteorological Station is quite lower than Turkey's average precipitation. It is observed that temperature has increased approximately 1.5 ºC from 1984 to 2017.

4.2 Classification Performance

In the second approach, shallow water of the Duden Lake was classified as four classes according to spectral signatures of each class. Firstly, ISODATA, unsupervised classification method, was applied with 300 clusters to detect the water clusters. Training data was prepared with these clusters and according to their spectral signature of the classes; clusters were grouped into 5 classes: 4 water classes and the others. Spectral signatures of four water classes are shown in the Figure 5.

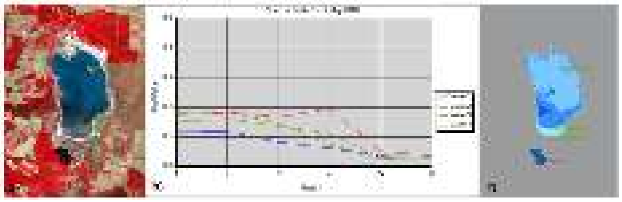

Figure 5. a) False colour Sentinel 2 image b) Spectral signatures of 4 water classes c) Unsupervised classification result with 4 water classes

Water was classified according to their different spectral reflectance. Water classes are respectively classified as water1 (deeper), water2 (deep), water3 (shallow) and water4 (shallower). Different machine learning classification methods which are mentioned before were applied to the satellite image. Classification results are shown in the Figure 6.

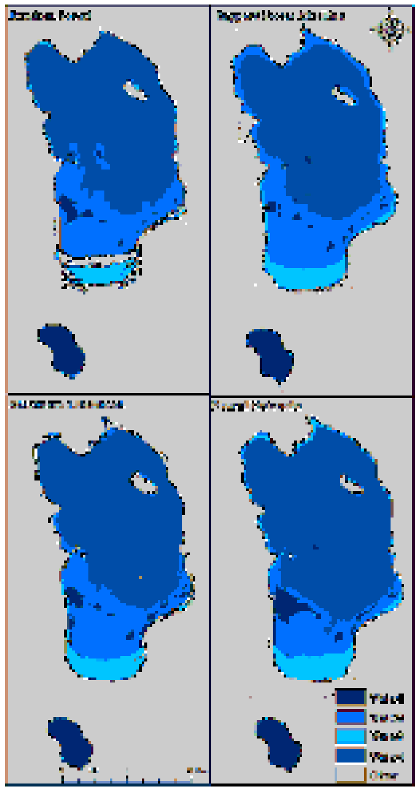

Figure 6. Classification results

Surface area of each class was calculated for different classification methods. Accuracy of the classification results was evaluated with 200 random selected points. Overall accuracy and kappa values were determined and are shown in the Table 2.

Classification Method	RF	SVM	MLC	NN
Water1 (ha)	46.58	42.95	50.09	58.99
Water2 (ha)	432.66	413.25	430.53	481.18
Water3 (ha)	153.29	201.9	149.83	103.76
Water4 (ha)	46.18	74.86	51.89	80.98
Total (ha)	678.71	732.96	682.34	724.91
Overall Accuracy	91.5	94.0	96.0	89.5
Kappa	0.89	0.92	0.95	0.87

Table 2. Surface area of each class and accuracy assessments of classification methods

According to Table 2, overall accuracy and kappa values of MLC and SVM methods have slightly higher than other methods and MLC showed the best performance in determination of the shallow water area. SVM showed the better performance than RF and NN, but it determined the water surface area more than other algorithms.

5. CONCLUSION

Machine learning algorithms have been widely used in studies about wetlands. In this study, firstly it was observed that the surface water area of Duden Lake has decreased approximately 50%. Secondly, it was determined that machine learning classification methods are usable for detecting shallow water areas. MLC method showed the better performance than others according to the accuracy assessment results. In future studies, different classification methods might be applicable to detect shallow water area. Classification performance can be more clearly observed with very high resolution satellite images.

REFERENCES

Abbas, A. W., Minallh, N., Ahmad, N., Abid, S. A. R., & Khan, M. A. A., 2016. K-Means and ISODATA clustering algorithms for landcover classification using remote sensing. *Sindh University Research Journal-SURJ* (Science Series), 48(2).

Amani, M., Salehi, B., Mahdavi, S., Granger, J. E., Brisco, B., & Hanson, A., 2017. Wetland classification using multi-source and multi-temporal optical remote sensing data in Newfoundland and Labrador, Canada. *Canadian Journal of Remote Sensing*, 43(4), pp. 360-373.

Belgiu, M., & 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.

Cowardin, L. M., Carter, V., Golet, F. C., & LaRoe, E. T., 1979. Classification of wetlands and deepwater habitats of the United States. US Department of the Interior, US Fish and Wildlife Service.

Jiang, X., Lin, M., & Zhao, J., 2011. Woodland cover change assessment using decision trees, support vector machines and artificial neural networks classification algorithms. In 2011 Fourth International Conference on Intelligent Computation Technology and Automation (Vol. 2, pp. 312-315). IEEE.

Liu, Z., Yao, Z., & Wang, R., 2019. Automatic identification of the lake area at Qinghai–Tibetan Plateau using remote sensing images. *Quaternary International*, 503, pp. 136-145.

Otukei, J. R., & Blaschke, T., 2010. Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *International Journal of Applied Earth Observation and Geoinformation*, 12, pp. 27-31.

Ozesmi, S. L., & Bauer, M. E., 2002. Satellite remote sensing of wetlands. *Wetlands Ecology and Management*, 10(5), pp. 381-402.

Paola, J. D., & Schowengerdt, R. A., 1995. A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification. *IEEE Transactions on Geoscience and remote sensing*, 33(4), pp. 981-996.

Tian, S., Zhang, X., Tian, J., & Sun, Q., 2016. Random forest classification of wetland landcovers from multi-sensor data in the arid region of Xinjiang, China. *Remote Sensing*, 8(11), 954.

URL 1: https://earthengine.google.com/

Zomer, R. J., Trabucco, A., & Ustin, S. L., 2009. Building spectral libraries for wetlands land cover classification and hyperspectral remote sensing. *Journal of Environmental Management*, 90(7), pp. 2170-2177.