Turkish Journal of Remote Sensing and GIS
Türk Uzaktan Algılama ve CBS Dergisi

Turk J Remote Sens GIS, March 2024, 5(1):125-137
Journal homepage: http://www.dergipark.org.tr/en/pub/rsgis

# Object Based Classification in Google Earth Engine Combining SNIC and Machine Learning Methods (Case Study: Lake Köyceğiz)

## *Google Earth Engine Platformunda SNIC ve Makine Öğrenimi Yöntemlerini Birleştiren Nesne Tabanlı Sınıflandırma (Örnek Olay: Köyceğiz Gölü)*

**Pınar Karakuş[1]***

[1]*Osmaniye Korkut Ata University, Faculty of Engineering and Natural Sciences, Geomatics Engineering, 80000, Osmaniye/Türkiye.*

***ORIGINAL PAPER***

***Corresponding author:**
Pınar Karakuş
pinarkarakus@osmaniye.edu.tr

**doi:** 10.48123/rsgis.1411380

**Article history**
Received: 28.12.2023
Accepted: 17.03.2024
Published: 28.03.2024

## *Abstract*

*Köyceğiz Lake is one of our country's most critical coastal barrier lakes, rich in sulfur, located at the western end of the Mediterranean Region. Köyceğiz Lake, connected to the Mediterranean via the Dalyan Strait, is one of the 7 lakes in the world with this feature. In this study, water change analysis of Köyceğiz Lake was carried out by integrating the Object-Based Image Classification method with CART (Classification and Regression Tree), RF (Random Forest), and SVM (Support Vector Machine) algorithms, which are machine learning algorithms. SNIC (Simple Non-iterative Clustering) segmentation method was used, which allows a detailed analysis at the object level by dividing the image into super pixels. Sentinel 2 Harmonized images of the study area were obtained from the Google Earth Engine (GEE) platform for 2019, 2020, 2021, and 2022, and all calculations were made in GEE. When the classification accuracies of four years were examined, it was seen that the classification accuracies(OA, UA, PA, and Kappa) of the lake water area were above 92%, F-score was above 0.98 for all methods using the object-based classification method obtained by the combination of the SNIC algorithm and CART, RF, and SVM machine learning algorithms. It has been determined that the SVM algorithm has higher evaluation metrics in determining the lake water area than the CART and RF methods.*

***Keywords:*** *GEE, Simple non-iterative clustering, Object based classification, Lake surface area, Sentinel 2, Machine learning*

## *Özet*

*Köyceğiz Gölü, Akdeniz Bölgesi'nin batı ucunda yer alan kükürt bakımından zengin, ülkemizin en kritik kıyı set göllerinden biridir. Dalyan Boğazı ile Akdeniz'e bağlanan Köyceğiz Gölü, bu özelliği ile de dünyadaki 7 gölden birisidir. Bu çalışmada, Nesne Tabanlı Görüntü Sınıflandırma yöntemi, makine öğrenimi algoritmalarından SRA (Sınıflandırma ve Regresyon Ağaçları), RO (Rasgele Orman) ve DVM (Destek Vektör Makineleri) algoritmaları ile bütünleştirerek Köyceğiz gölünün su değişim analizi gerçekleştirilmiştir. Görüntüyü süper piksellere bölerek nesne düzeyinde ayrıntılı bir analize olanak tanıyan Basit Yinelemesiz Kümeleme (BYK) segmentasyon yöntemi kullanılmıştır. Çalışma alanına ait Sentinel 2 Harmonized görüntüleri 2019, 2020, 2021 ve 2022 yılları için Google Earth Engine (GEE) platformundan elde edilmiş ve tüm hesaplamalar GEE'de yapılmıştır. Dört yılın sınıflandırma doğrulukları incelendiğinde BYK algoritması ile SRA, RO ve DVM makine öğrenme algoritmalarının kombinasyonu ile elde edilen nesne tabanlı sınıflandırma yöntemi kullanılarak göl su alanının bütün yöntemler için sınıflandırma doğruluklarının (ÜD, KD, GD ve Kappa) %92'nin üstünde, F-score'un 0.98'in üzerinde olduğu görülmüştür. SVM algoritmasının SRA ve RO yöntemlerine göre göl su alanının belirlenmesinde daha yüksek değerlendirme metriklerine sahip olduğu belirlenmiştir.*

***Anahtar kelimeler:*** *GEE, Basit yinelemeli kümeleme, Nesne tabanlı sınıflandırma, Göl yüzey alanı, Sentinel 2, Makine öğrenmesi*

## 1. Introduction

Different types of inland waters, including lakes, ponds, rivers, reservoirs, and wetlands, are extensively dispersed across the surface of the Earth. Water is a vital component of the planet and a necessary resource for human existence and productivity. The geographic and temporal distribution of water bodies has been shifting due to climate change and human activity, and the global water resource scarcity is worsening. In conclusion, for all these reasons, monitoring water bodies is important (Li et al., 2022).

Many disciplines, such as urban management, environmental planning, precision agriculture, and monitoring of natural resources, including water bodies, largely depend on the accurate classification of satellite images (Ouchra et al., 2023). In the accurate classification of satellite images, it has been observed that more accurate results are obtained with the object-based classification method, which performs the segmentation process by dividing the image into segments and eliminates the salt and pepper effect in the image (Tassi et al., 2021; Gao & Mas, 2008; Blaschke, 2010; Kaplan & Avdan, 2017). The most important stage of object based image classification is the image segmentation. The traditional segmentation algorithm is performed based on region growth, threshold, level set, and active contours. Because of its low computation amount, quicker processing speed, and superior anti-noise features, the super pixel segmentation (Wang et al., 2017) technique is extensively relevant to image segmentation and classification in numerous sectors in the state of technology today. Based on the idea of super pixel segmentation, Achanta and Süsstrunk (2017) presented SNIC segmentation algorithms in 2017. The SNIC technique efficiently addresses these problems by calculating the spatial proximity and color relationship. This allows the user to regulate the compactness of the super pixels, reducing an image to a small collection of connected super pixels. This makes it possible to segment data in grayscale and color. The final method yields better organized, compact super pixels with sharper border characteristics and a more lifelike detail display. Besides all these, Yang et al., (2020) study confirmed that SNIC is a robust preprocessing tool when directly mapping large-scale surface water to RS data. When the literature is examined, it has been revealed that the object-based classification method using the SNIC algorithm is superior to other methods in determining water areas (Mahdianpari et al., 2019; Mahdianpari et al., 2020; Liu et al., 2022; Li et al., 2021).

The newly deployed Sentinel-2 satellite can offer high spatial resolution multispectral images. This new dataset has the potential to be very helpful for mapping regional water bodies because it is freely accessible and allows for numerous revisits. The Sentinel-2 image has been widely used in mapping studies related to surface waters (Zhou et al., 2017b; Gašparović & Singh, 2022; Jiang et al., 2021; Jiang et al., 2023; Pan et al., 2023).

GEE is an open-source cloud platform with strong data processing, analysis, storage, and visualization capabilities that has been heavily utilized in the field of planetary-scale spatial analysis that helps address high-impact societal issues like deforestation, drought, disaster, sickness, food security, water management, climate monitoring, and environmental protection in recent years (Gorelick et al., 2017). Regarding this, the GEE-based Simple Non-Iterative Clustering (SNIC) method effectively recognized probable unique objects and grouped comparable pixels (Achanta & Süsstrunk, 2017). With this, machine learning has become widespread in recent years, with GEE coming to the fore for free analysis and storage of big data (Zhou et al.,2017a). However, different combinations of object-based classification and machine learning algorithms have emerged. Specifically, it has been observed that non-parametric machine learning classifiers, like Support Vector Machine (SVM), Random Forest (RF), and Classification and Regression Trees (CART), produce remarkably accurate classification results from remotely sensed images (Tassi & Vizzari, 2020; Bar et al., 2020). When examining machine learning algorithms, The Random Forest (RF) is an ensemble classifier composed of many tree classifiers. Each classifier is created by randomly sampling factors from the input vector data. Each tree is voted as a unit for a popular class to categorize an input vector (Ao et al., 2019; Pal, 2005). One benefit of the RF method is its ability to manage significant variations within land cover classes and effectively mitigate noise data (Slagter et al., 2020). Furthermore, this approach does not necessitate knowledge of the data distribution, unlike parametric classification algorithms like maximum likelihood, which require knowledge of the data distribution. The Support Vector Machine (SVM) uses an advanced kernel function to categorize datasets with intricate decision boundaries. SVM has the advantage of reducing uncertainty in model structure, and, like RF, it is not dependent on the data distribution (Oommen et al., 2008). CART is a tree-based classification technique that assesses the interdependence between one variable and other factors (Simioni et al., 2020). CART has an advantage in naturally modeling non-linear boundaries due to its hierarchical structure, as noted by Shelestov et al. (2017) and Simioni et al. (2020). For all these reasons, the study used these three machine learning methods. When the literature on mapping water bodies using machine learning techniques is examined, it is seen that there are studies on random forests (Jiang et al., 2022; Wangchuk & Bolch, 2020;), SVM (Ghamisi & Höfle, 2017; Schmitt, 2020; Liu et al., 2020), CART method (Gxokwe et al., 2022). When the studies are examined in detail, Sarp and Ozcelik (2017) utilized Landsat TM and ETM+ data to study the spatiotemporal variations in Lake Burdur, Turkey, from 1987 to 2011, focusing on applying machine learning methods for water area identification.

The researchers compared Support Vector Machine (SVM), a machine learning methodology, with thresholding methods such as the Normalized Difference Water Index (NDWI), modified Normalized Difference Water Index (mNDWI), and Automated Water Extraction Index (AWEI) to detect surface water. SVM yielded the best overall results among the techniques examined, whereas mNDWI showed the highest accuracy for surface water detection among all water indices. A study conducted in Nepal utilized Landsat 8 OLI pictures to create a surface water map by employing six machine learning techniques, such as RF, SVM, and NN (Acharya et al., 2019). The surface water in Nepal displays several characteristics, such as depth, turbidity, temperature variations, and vegetation coverage. The RF method had the highest overall accuracy in classifying water extraction procedures.

This study aims to prove that the GEE platform and advanced machine learning algorithms have the potential to detect and monitor the Köyceğiz Lake water surface area using the Sentinel-2 multi-year composite image. In the study, an object-based classification algorithm combining the SNIC algorithm was developed to identify spatial clusters on the GEE platform using Sentinel 2 images. Three machine learning algorithms (CART, RF, and SVM) were used to determine the water surface area of Köyceğiz Lake in 2019, 2020, 2021, and 2022 to classify water and non-water areas. The lake surface water area was observed for 4 years with annual composite images from January 1 to December 31 to capture different land cover conditions and changes over time. Accuracy values for UA, PA, OA, and Kappa and F-score were determined for the lake surface area.

## 2. Material and Methods

### 2.1 Study Area

There are two important factors in the study of Köyceğiz Lake. First, there is no study in the literature on the change in the lake's water surface area. Another important factor is the features it has. These features are as follows: Köyceğiz Lake is a sulfur-rich lake within the district's borders at the western end of the Mediterranean Region. It is in the southeast of Muğla city center, 71 km from the center, and has a surface area of 54 $km^2$. It is 8 meters above sea level. It ranges in depth from 15 to 150 meters. The Namnam, Kargicak, and Yuvarlak Çay feed it, as seen in Figure 1 (T.C. Köyceğiz Kaymakamlığı, 2023).

Köyceğiz Lake, formed due to the closure of the mouth of a coastal bay by the alluviums brought by the Dalaman Stream, is among our country's most critical coastal barrier lakes. The canal system connecting the lake and the Mediterranean and the surrounding wetlands has a rich diversity of species. The areas outside the settlements are natural log forests, dunes, and reed areas on the lake shore, under protection. The lake environment, canals, and forests have a rich potential as a breeding and sheltering place for various animals (Ministry of Environment, Urbanisation and Climate Change, 2019).

The destination of Köyceğiz Lake is 'Dalyan Strait'. This channel, approximately 1.5 meters deep, connects Köyceğiz Lake to the Mediterranean. Such lakes are called standing lakes. There are only 7 lakes of this type in the world. In this state, the lake is protected by the state under the name "Köyceğiz-Dalyan Basin Special Environmental Protection Area". Its unique ecological structure makes the region worth protecting (Avşar & Kurtuluş, 2017; Goller.gen.tr, 2024). A typical Mediterranean climate prevails in the region. Summers are hot and dry; winters are warm and rainy (Türedi, 2006). The average annual temperature is 18.3°C, the average annual total precipitation is 1082.6 mm, and the average annual relative humidity is 61% (Selim et al., 2016).

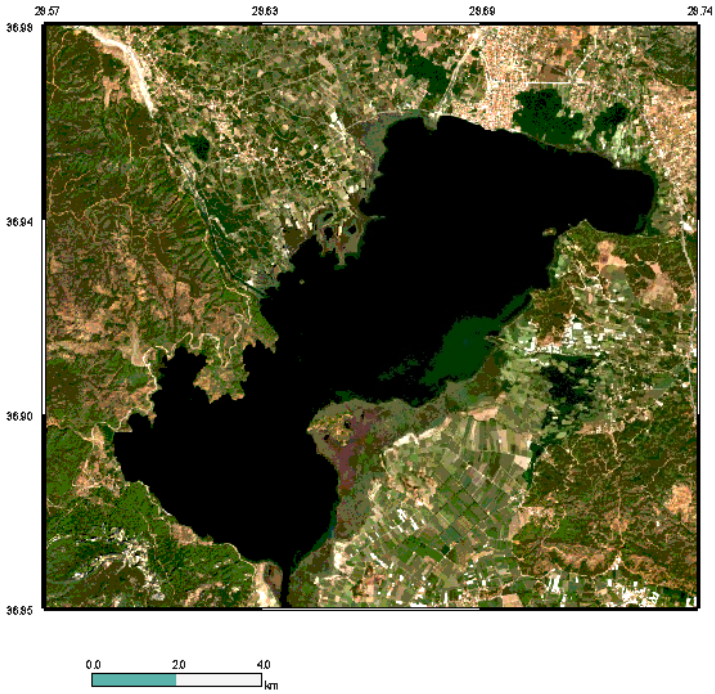



**Figure 1.** Study area

## 2.2 Data

In 2010, Google introduced GEE, a cloud computing platform that takes advantage of open-access remote sensing datasets and Google's computing resources. GEE hosts satellite imagery and archives it in a public data repository spanning four decades. The daily ingested images are then made available for data mining worldwide.

GEE integrates planetary-scale analysis capabilities with multi-petabyte geospatial datasets and a satellite image database. Earth Engine is a platform that allows academics, non-profits, businesses, and governments to analyze and visualize geographical datasets scientifically. Additionally, GEE offers tools and APIs for analyzing big datasets (Google Earth Engine, 2023).

A constellation of two similar satellites in the same orbit is the foundation for the Copernicus Sentinel-2 program. Each satellite is equipped with a cutting-edge, wide-swath, high-resolution multispectral imager with 13 spectral bands for a fresh look at our landscape and plants. Unprecedented views of Earth are made possible by the combination of high resolution, unique spectral capabilities, a 290 km swath width, and frequent revisit durations. Every five days at the equator, they collectively cover all of Earth's land surfaces, sizable islands, and inland and coastal waterways. On June 23, 2015, Sentinel-2A was launched, and on March 7, 2017, Sentinel-2B (European Space Agency, 2023).

The GEE database provided the Sentinel-2 Multispectral Instrument (MSI), level 2A (COPERNICUS/S2_SR) surface reflectance images used in this manuscript. These products include three QA bands, one of which (QA60) is a bitmask band containing cloud mask information, and the twelve spectral bands seen in Table 1 scaled to 10,000. They have already been atmospherically corrected using the Sen2cor (European Space Agency, 2023). The 2019-01-01 to 2023-01-01 was represented by the filter applied to the image stack and utilizing the codes ee.Filter.Date () and Image.filterBounds () to determine the region inside the chosen lake bounds.

**Table 1.** Band information of the Sentinel-2 Level 2A data

| Band ID | Spatial resolution (m) | Band Description | Band ID | Spatial resolution (m) | Band Description |
|---|---|---|---|---|---|
| B1 | 60 | Aerosols | B8 | 10 | NIR |
| B2 | 10 | Blue | B8A | 20 | Red Edge 4 |
| B3 | 10 | Green | B9 | 60 | Water vapor |
| B4 | 10 | Red | B11 | 20 | SWIR 1 |
| B5 | 20 | Red Edge 1 | B12 | 20 | SWIR 2 |
| B6 | 20 | Red Edge 2 | QA60 | 60 | Cloud mask |
| B7 | 20 | Red Edge 3 | | | |

### 2.3 SNIC (Simple Non-Iterative Clustering) Algorithm

Image segmentation is the initial phase in object based classification. According to a predetermined set of criteria, the image is divided into several distinct, non-overlapping segments as part of the procedure (Dlamini et al., 2021). Individual pixels are combined in this process to create more significant things. This reduces "salt and pepper" noise in the final classification map and improves the distinction of spectrally similar objects using texture, shape, and contextual data (Dlamini et al., 2021; Mahdianpari et al., 2020). The composite in this study was segmented using the SNIC technique. According to Achata and Süsstrunk (2017), the SNIC algorithm was selected due to its simplicity, memory efficiency, processing speed, and capacity to integrate communication between pixels once the method has been started. The SNIC technique first initializes the centroid pixels on a regular grid picture to begin image segmentation process. Next, it uses the distance in a five-dimensional space of color and spatial coordinates to establish the dependency of each pixel relative to the centroids. Specifically, the distance creates homogeneous super pixels by integrating normalized spatial and color distances (Achanta & Süsstrunk, 2017). The smallest distance from the centroid determines which candidate pixel gets chosen (Achanta & Süsstrunk, 2017).

Some main parameters must be determined in the SNIC image segmentation process in GEE. These parameters are as follows. Superpixel seeds have compactness, connectivity, and neighborhood size values. Among these parameters, the value of the seed was determined as 36 as a result of the experiments carried out to find the optimal superpixel size to obtain the best segmentation results of SNIC (Wang et al., 2024). The code ee.Algorithms were used to run the SNICImage.Segmentation.SNIC() on GEE produces an image with super pixels and computed perimeters, sizes, textures, and areas for each super pixel.

### 2.4 CART

Breiman et al. (1984) created CART, a binary decision regression tree, allowing simple decision-making in logical if-then scenarios. Recursively, CART divides nodes until they cross a threshold and arrive at a terminal node. This method grows a large tree and then prunes it to size with the lowest cross-validation error estimate rather than using stopping rules. The clever pruning process uses a cost-complexity parameter to index the links and is based on the principle of weakest-link cutting. CART employs a sequence of "surrogate" splits, or splits on other variables that take the place of the chosen split in cases where the latter cannot be applied due to missing values, to deal with missing data values at a node. An importance score is also provided for each X variable by surrogate splits. These scores can aid in the detection of masking because they gauge how well the surrogate splits predict the preferred divides (Breiman et al. 2017; Loh, 2014).

### 2.5 RF

RF is a commonly used classifier to combine multiple CART trees into an ensemble classification. RF randomly generates several decision trees from training data sets and variables. The square of the variable set is the ideal number of variables to calculate, and the ideal number of trees to calculate falls between 100 and 500 (Brieman, 2001). The classifier is trained using the training sample set, and the sample set's quality directly influences the final maps' quality. Three things are usually needed to create a high-quality training set. Firstly, all samples have good representativeness; secondly, the categorization category is uniform; and thirdly, all samples must be dispersed uniformly throughout the study region. The sample training set is constructed initially based on the previous goals. It is essential to ensure that each class of samples is spread as evenly as possible throughout the study area while simultaneously ensuring sample diversity and completeness. All of the samples from the root node can then be entered into a tree to train the classifier. A splitting criterion is inserted at each intermediate node and the root node, which divides the data into two child nodes. When all elements in a subset at a node have the same value, the attribute values are exhausted, or the decision tree reaches other predetermined stopping conditions, the training process of the decision tree is based on attribute value testing and splitting the input training set into subsets. This process is repeated recursively in each split into a subset (Xue et al., 2023).

### 2.6 SVM

A machine learning technique called support vector machine (SVM) was created in the middle of the 1990s and is based on statistical theory. It was first put forth by Vapnik in 1995. Using a sum function that maximizes the distance between the nearest point and this surface, it locates a hyperplane (also known as a hypersurface) of the data in a high-dimensional space. Then, building on this, the support vector machine technique to classification maximizes the distance between the categories by splitting the dataset into many discrete categories compatible with the shape of the training

samples. SVM was first developed to address binary classification issues; however, an appropriate multi classifier must be constructed to address multiclass issues. There are two approaches available at the moment: "one-versus-rest" and "one-versus-one." To generate k SVMs for k classes of data, the one-versus-rest approach (OVR SVMs) classifies samples of one class into one class and the remaining samples into another class. The unknown samples are then classified into the class with the most significant classification function value. n*(n - 1)/2 SVMs are constructed for n classes of samples using the one-versus-one approach (OVO SVMs, or pairwise) to design an SVM between any two classes of samples. The outcome of this categorization is the one that obtained the most votes. The one-versus-one libsvm technique is employed in the GEE (Xue et al., 2023).

### 2.7 Evaluation Metrics

The error matrix was used for the accuracy assessment. The error matrix's rows show the classes to which the image's pixels have been assigned, and its columns show the classes to which the pixels belong in the ground truth. The correctly identified pixels are shown on the diagonal of the error matrix. Overall Accuracy (OA) was applied to ascertain the percentage of accurately mapped reference sites. The calculation divided the total samples by the number of correctly categorized samples (Jayaswal, 2020; Heydarian et al., 2022). Computed by dividing the total correct, or the sum of the major diagonal, by the total number of pixels in the error matrix, overall accuracy is the most basic and widely used accuracy metric (Congalton, 1991). Story and Congalton (1986) developed the terms user's accuracy (UA) to denote the likelihood that a pixel identified on the map accurately represents that category on the ground and producer's accuracy (PA) to indicate the probability of a reference pixel being correctly classified (Congalton, 1991). Another evaluation metric is the F-score. The F-score is the harmonic mean of recall and precision, equivalent to PA and UA, respectively (Solano et al., 2019).

## 3. Results and Discussion

The water surface areas of Köyceğiz Lake were acquired between 2019-2020-2021 and 2022 years from Sentinel-2 Level-2A images for this study. Water and non-water training points are determined. The data taken from the study region-based Sentinel image segment were point-by-point compared with the composite images for water surface area maps to show the accuracy of the results. Composite images seen in Figure 3a, Figure 4a, Figure 5a, and Figure 6a were obtained using date, cloud, study area filters to create composite images. Sentinel-2 images from 1 January 2019 to 31 December 2022 were used. This date range was chosen because there were no Sentinel 2 Level 2A images before March 2018 (European Space Agency, 2023).

Since spectral indices are known to improve classification accuracy, the NDVI spectral index was used in this study (Kobayashi et al., 2020). NDVI (Normalized Difference Vegetation Index) was calculated from composite images before the SNIC segmentation algorithm. Then, the SNIC segmentation algorithm was applied to the created composite images. Train polygons were selected from the points by adding the calculated NDVI values as a band. The composite images are divided into segments, as seen in Figure 3b, Figure 4b, Figure 6b, and Figure 6b. When using the SNIC method in GEE, attention should be paid to parameter settings. “Size” refers to the segmentation size, which is the pixel-wise range of superpixel kernel positions determined by the study area. It was chosen as 50 in the study. The term "compactness" describes the level of compactness. The closer the segmentation result is to a square, the higher the value. Nevertheless, the "compactness" is set to 0,1. "Connectivity" refers to the connection set at 8 in this study. The amount of image elements at the center of the segmentation process is called the "seed." The "seed" option is not set again because the spacing has already been established. The "size" parameter has the most significant impact on the categorization effect. It should be mentioned that the categorization effect improves with a greater or smaller value. The SNIC algorithm segmented the images based on the correlation between the pixels (Luo et al., 2021). Clusters were calculated as seen in the segmented images Figure 3c, Figure 4c, Figure 5c, and Figure 6c.

After the images were segmented, the optimal parameters in this study are unknown in the SVM classifier as there is no prior knowledge about the physical nature of the prediction problem. The kernel type parameter that needs to be determined in the SVM method is linear. Among the other parameters, the gamma value, which is the gamma value in the kernel function, and the cost value, which is the cost parameter, are selected by default. The number of tree parameters that must be determined in the RF method is the number of decision trees to be created. This parameter was determined to be 70 in the study. The maxNodes parameter, which must also be determined in the CART method, specifies the maximum number of leaf nodes in each tree. This parameter was determined to be 70 (GEE) in the study.

After the parameters were determined, the lake water areas shown in Figure 3 were determined by applying CART as seen in Figure 3d, Figure 4d, Figure 5d and Figure 6d, RF as seen in Figure 3e, Figure 4e, Figure 5e and Figure 6e, and SVM classification methods as seen in Figure 3f, Figure 4f, Figure 5f and Figure 6f respectively. With the CART method, the lake area was found to be 51.31 $km^2$ in 2019, 51.41 $km^2$ in 2020, 49.38 $km^2$ in 2021 and 50.26 $km^2$ in 2022.

With the RF method, the lake area was found to be 51.26 km$^2$ in 2019, 51.91 km$^2$ in 2020, 50.56 km$^2$ in 2021 and 50.62 km$^2$ in 2022. With the SVM method, the lake area was found to be 51.31 km$^2$ in 2019, 51.36 km$^2$ in 2020, 50.96 km$^2$ in 2021, and 51.11 km$^2$ in 2022, as seen in Figure 2.

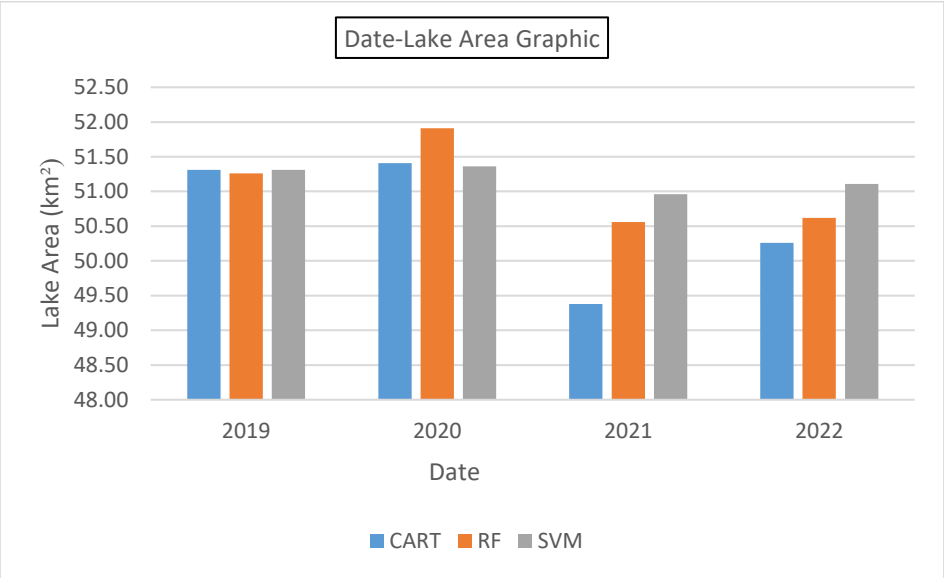
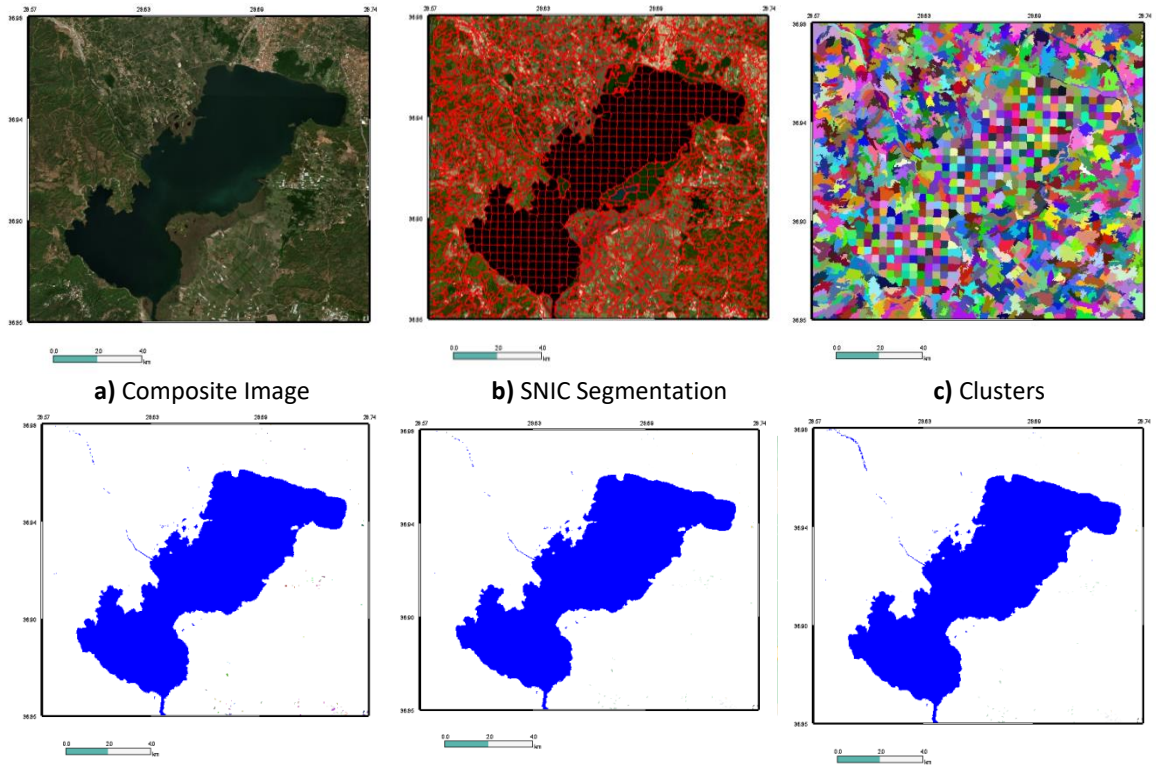



**Figure 2.** Lake water areas

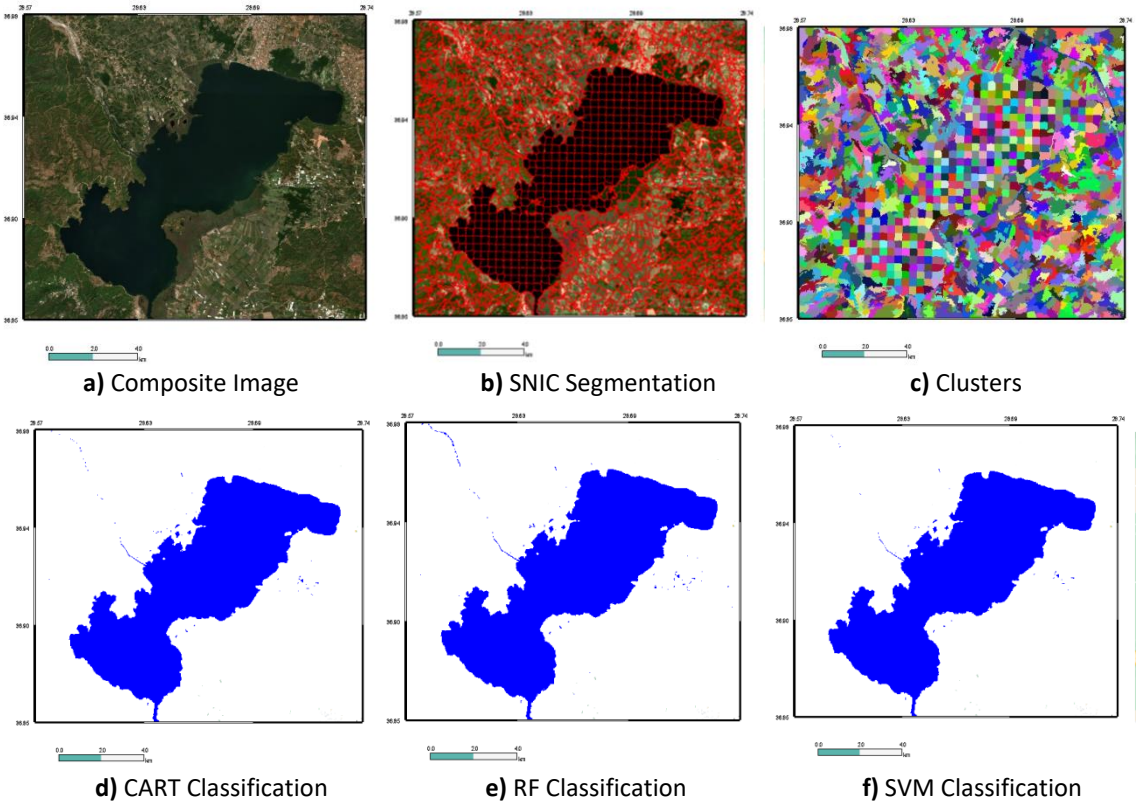
When the result maps of the classification algorithms (CART, RF and SVM) shown in Figure 3-d-e-f, Figure 4-d-e-f, Figure 5-d-e-f and Figure 6-d-e-f are visually compared with each other; The Dalyan Strait, where the lake flows into the Mediterranean, is clearly identified on all maps. However, Namnam Stream, which feeds the lake, has not been generally determined.



**a)** Composite Image **b)** SNIC Segmentation **c)** Clusters

**d)** CART Classification **e)** RF Classification **f)** SVM Classification

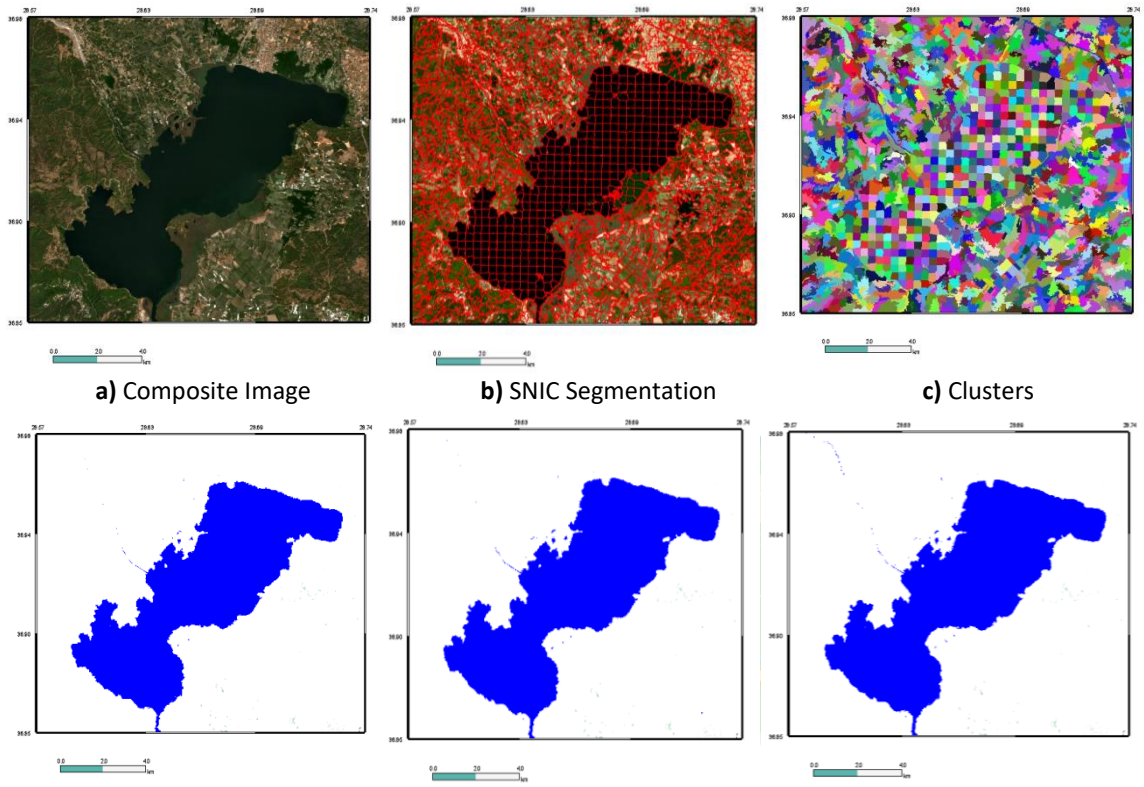
**Figure 3.** 2019



**a)** Composite Image **b)** SNIC Segmentation **c)** Clusters

**d)** CART Classification **e)** RF Classification **f)** SVM Classification

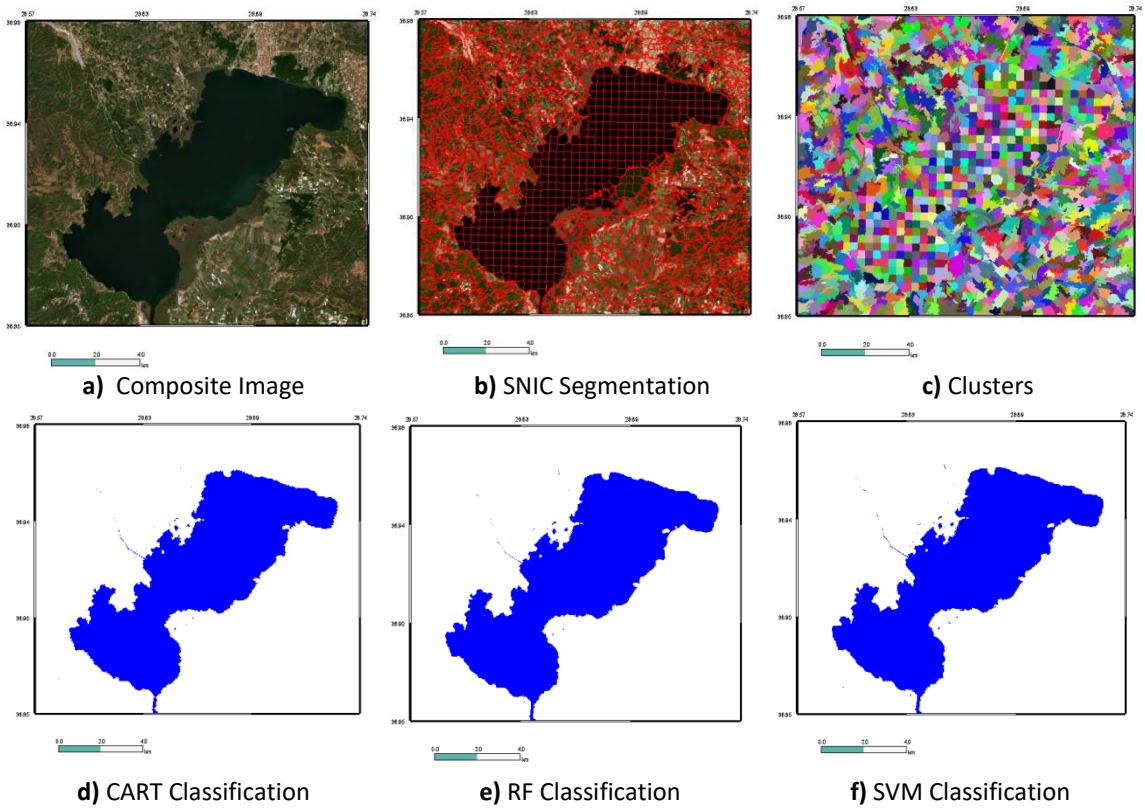
**Figure 4.** 2020



**a)** Composite Image **b)** SNIC Segmentation **c)** Clusters

**d)** CART Classification **e)** RF Classification **f)** SVM Classification

**Figure 5.** 2021



**a)** Composite Image **b)** SNIC Segmentation **c)** Clusters

**d)** CART Classification **e)** RF Classification **f)** SVM Classification

**Figure 6.** 2022

The accuracy values for UA, PA, OA, Kappa and F-score for the lake area are displayed in Table 2. When the accuracy of CART, RF, and SVM methods is evaluated over the years, In the CART method, the lowest OA is 97%, the lowest UA is 92%, the lowest PA is 94%, and the lowest kappa is 0.9441, and the lowest F-score is %99. In the CART method, the highest OA and UA is 98%, the highest PA is 96%, the highest kappa is 0.9658, and the highest F-score is %99. In the RF method, the lowest OA is 97%, the lowest UA is 92%, the lowest PA is 95%and the lowest F-score is %98. In the RF method, the highest OA, UA, PA is 98%, the highest kappa is 0.9738, and the highest F-score is %99. In the SVM method, the lowest OA, UA, is 98%, the lowest PA is 96%, and the lowest F-score is % 99, and the lowest kappa is 0.9712. In the SVM method, the highest OA, UA, PA are1%, the highest kappa and F-score are1.

**Table 2.** Four-year classification accuracies

| **CART** | **2019** | **2020** | **2021** | **2022** |
|---|---|---|---|---|
| OA (%) | 97.5 | 98.72 | 98.75 | 98,63 |
| UA (%) | 92,86 | 98,30 | 98.39 | 98,15 |
| PA (%) | 96,29 | 95.0 | 94.74 | 95 |
| Kappa | 0.9441 | 0.9658 | 0.9648 | 0.9650 |
| F-score | 0,9910 | 0,9915 | 0,9919 | 0,9906 |
| **RF** | | | | |
| OA (%) | 98,53 | 97.43 | 97,56 | 98,84 |
| UA (%) | 92,31 | 98.24 | 95,24 | 98,27 |
| PA (%) | 98,21 | 95.23 | 95,24 | 96,55 |
| Kappa | 95,10 | 0.9348 | 0,9360 | 0.9738 |
| F-score | 0,9910 | 0,9824 | 0,9836 | 0,9913 |
| **SVM** | | | | |
| OA (%) | 1 | 1 | 98,85 | 98,70 |
| UA (%) | 1 | 1 | 98,36 | 98,04 |
| PA (%) | 1 | 1 | 96,30 | 96,30 |
| Kappa | 1 | 1 | 0,9729 | 0.9712 |
| F-score | 1 | 1 | 0,9917 | 0,9900 |

The findings of the annual evaluation for the water areas showed an OA of more than 97%, UA of more than 92%, and PA of more than 94% were obtained for the lake areas after four years of computing a confusion matrix. The Kappa coefficient values are higher than 93%. The CART, RF, and SVM approaches successfully extracted the water surfaces, as evidenced by the accuracy of above 92% calculated for all the metrics utilized in the accuracy assessment. Compared with CART, SVM and RF classifiers shows the highest producer accuracy, overall accuracy, and user accuracy, as proven by other studies (Gxokwe et al., 2022; Aldiansyah & Saputra, 2023).

The findings show how well the Google Earth Engine cloud computing platform can characterize and map lake surfaces with respectable overall accuracy. Technological developments in data analytics offer rare chances to enhance the identification and tracking of lake surface areas with varying dimensions, which is impossible with conventional remote sensing methods. The advent of cloud computing platforms, like Google Earth Engine (GEE), brings benefits, including memory efficiency, quick image processing speed, sophisticated machine learning techniques, and parallel processing. In addition to evaluating the performance of GEE machine learning algorithms suitable for identifying and mapping such systems, the study aimed to assess the capabilities of the GEE cloud computing platform to characterize and map annual lake surface water fields at a site-specific scale.

Overall, the findings show that the Google Earth Engine cloud computing platform can accurately characterize and map the lake surface area with a generally acceptable degree of accuracy. All parameters used in accuracy evaluation resulted in values above 92%, and the lowest F-score is %98. The accuracy of the classification process often decreases with fewer training and validation points (Corcoran et al., 2015; Mahdianpari et al., 2020). The multi-year images utilized in the median composite computation did not account for seasonality and yearly changes which could have decreased the accuracy of the producer and user. It is important to account for seasonality and yearly changes to avoid decreasing the accuracy due to a small number of training and validation data points.

The precision with which different land cover classes are classified in GEE varies depending on the composition techniques and input image. Classification outputs have excellent accuracy values when considering temporal aggregation during median merging. According to Phan et al. (2020), this highlights the significance of temporal aggregations during the median aggregation phase. There are several restrictions related to the usage of the GEE cloud computing platform, even if the results highlight its applicability in characterizing and mapping lake surface area. These include space-related computing limits that prevent the huge training needed by sophisticated machine learning algorithms from being completed, the lack of sophisticated and precise image segmentation techniques, and the limitation of image processing tools. Additionally, the algorithms have constraints such as sluggish training and bias when handling categorical data in Random Forests, reduced performance with large datasets in Support Vector Machines, and longer training time in Classification and Regression Trees.

The findings of this study prove that the GEE platform and advanced machine learning algorithms have the potential to detect and monitor the Köyceğiz Lake water surface area using the Sentinel-2 multi-year composite image.

## 4. Conclusion

The findings show that the Google Earth Engine cloud computing platform, the SNIC segmentation method, and the subsequently used CART, RF, and SVM algorithms can effectively define and map lake water areas with generally acceptable levels of accuracy. When the classification accuracies of four years were examined, it was seen that the lake water area could be determined quite accurately using the object-based classification method obtained by combining the SNIC algorithm with CART, RF, and SVM machine learning algorithms. All parameters used in accuracy evaluation resulted in values above 92%. The study showed that, similar to the literature, the SVM algorithm had higher evaluation metrics for determining the lake water area (Wei et al., 2020). When the change of the lake area over the years was examined, it was seen that the largest lake area was in 2020, then in 2019, 2022, and at least in 2021. It is thought that the lake area did not pose a risk of quaking due to the streams feeding the lake during the examined period.

## References

Achanta, R., & Süsstrunk, S. (2017). Superpixels and polygons using simple non-iterative clustering. *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4895–4904). IEEE. https://doi.org/10.1109/CVPR.2017.520

Acharya, T. D., Subedi, A., & Lee, D. H. (2019). Evaluation of machine learning algorithms for surface water extraction in a Landsat 8 Scene of Nepal. *Sensors*, *19*(12), 2769. https://doi.org/10.3390/s19122769

Aldiansyah, S., & Saputra, R. A. (2023). Comparison of machine learning algorithms for land use and land cover analysis using Google Earth Engine (case study: Wanggu Watershed). *International Journal of Remote Sensing and Earth Sciences*, *19*(2), 197-210.

Ao, Y., Li, H., Zhu, L., Ali, S., & Yang, Z. (2019). The linear random forest algorithm and its advantages in machine learning assisted logging regression modeling. *Journal of Petroleum Science and Engineering*, 174, 776-789.

Avşar, Ö., & Kurtuluş, B. (2017). Köyceğiz Gölü su ve taban sedimanlarının sıcaklık dağılımı. *Jeoloji Mühendisliği Dergisi*, *41*(2), 117-136. https://doi.org/10.24232/jmd.334546

Bar, S., Parida, B. R., & Pandey, A. C. (2020). Landsat-8 And Sentinel-2 based forest fire burn area mapping using machine learning algorithms on GE cloud platform over Uttarakhand, Western Himalaya. *Remote Sensing Applications: Society and Environment*, *18*, 100324. https://doi.org/10.1016/j.rsase.2020.100324

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *65*, 2–16. https://doi.org/10.1016/j.isprsjprs.2009.06.004

Breiman, L., Friedman, J., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees* (1st edition). Chapman and Hall/CRC.

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

Breiman, L. (2017). *Classification and regression trees*. Routledge.

Corcoran, J., Knight, J., Pelletier, K., Rampi, L., & Wang, Y. (2015). The effects of point or polygon based training data on Random Forest classification accuracy of wetlands. *Remote Sensing*, *7*(4), 4002-4025. https://doi.org/10.3390/rs70404002

Dlamini, M., Adam, E., Chirima, G., & Hamandawana, H. (2021). A remote sensing-based approach to investigate changes in land use and land cover in the lower uMfolozi floodplain system, South Africa. *Transactions of the Royal Society of South Africa*, *76*(1), 13–25. https://doi.org/10.1080/0035919X.2020.1858365

European Space Agency. (2023, December 19). *Introducing Sentinel-2*. https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2/Introducing_Sentinel-2

Gao, Y., & Mas, J. F. (2008). A Comparison of the performance of pixel-based and object-based classifications over images with various spatial resolutions. *Online Journal of Earth Sciences*, *2*(1), 27-35.

Gašparović, M., & Singh, S. K. (2022). Urban surface water bodies mapping using the automatic k-means based approach and sentinel-2 imagery. *Geocarto International*, *38*(1), 2148757. https://doi.org/10.1080/10106049.2022.2148757

Ghamisi, P., & Hoefle, B. (2017). LiDAR data classification using extinction profiles and a composite kernel support vector machine. *IEEE Geoscience and Remote Sensing Letters*, *14*(5), 659-663.

Google Earth Engine. (2023, December 19). *FAQ*. https://earthengine.google.com/faq/

Goller.gen.tr. (2024, January 29). *Köyceğiz Gölü*. https://www.goller.gen.tr/koycegiz-golu.html

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.

Gxokwe, S., Dube, T., & Mazvimavi, D. (2022). Leveraging Google Earth Engine Platform to characterize and map small seasonal wetlands in the semi-arid environments of South Africa. *Science of The Total Environment*, *803*, 150139. https://doi.org/10.1016/j.scitotenv.2021.150139

Heydarian, M., Doyle, T. E., & Samavi, R. (2022). MLCM: Multi-Label confusion matrix. *IEEE Access*, *10*, 19083-19095. https://doi.org/10.1109/ACCESS.2022.3151048

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, *37*(1), 35-46. https://doi.org/10.1016/0034-4257(91)90048-B

Jayaswal, V., (2020, September 14). *Performance metrics: Confusion matrix, Precision, Recall, and F1 Score*. Towards Data Science. https://towardsdatascience.com/performance-metrics-confusion-matrix-precision-recall-and-f1-score-a8fe076a2262

Jiang, W., Ni, Y., Pang, Z., Li, X., Ju, H., He, G., Lv, J., Yang, K., Fu, J., & Qin, X. (2021). An effective water body extraction method with new water index for sentinel-2 imagery. *Water*, *13*(12), 1647. https://doi.org/10.3390/w13121647

Jiang, L., Zhou, C., & Li, X. (2023). Sub-Pixel Surface Water Mapping for Heterogeneous Areas from Sentinel-2 Images: A Case Study in the Jinshui Basin, China. *Water*, *15*(8), 1446. https://doi.org/10.3390/w15081446

Jiang, Z., Wen, Y., Zhang, G., & Wu, X. (2022). Water information extraction based on multi-model RF algorithm and Sentinel-2 image data. *Sustainability*, *14*(7), 3797. https://doi.org/10.3390/su14073797

Kaplan, G., & Avdan, U. (2017). Object-based water body extraction model using Sentinel-2 satellite imagery. *European Journal of Remote Sensing*, *50*(1), 137-143.

Kobayashi, N., Tani, H., Wang, X., & Sonobe, R. (2020). Crop classification using spectral indices derived from Sentinel-2A imagery. *Journal of Information and Telecommunication*, 4(1), 67-90.

Li, H., Zech, J., Ludwig, C., Fendrich, S., Shapiro, A., Schultz, M., & Zipf, A. (2021). Automatic mapping of national surface water with OpenStreetMap and Sentinel-2 MSI data using deep learning. *International Journal of Applied Earth Observation and Geoinformation*, *104*, 102571. https://doi.org/10.1016/j.jag.2021.102571

Li, J., Ma, R., Cao, Z., Xue, K., Xiong, J., Hu, M., & Feng, X. (2022). Satellite Detection of Surface Water Extent: A Review of Methodology. *Water*, *14*(7), 1148. https://doi.org/10.3390/w14071148

Liu, Q., Huang, C., Shi, Z., & Zhang, S. (2020). Probabilistic river water mapping from Landsat-8 using the support vector machine method. *Remote Sensing*, *12*(9), 1374. https://doi.org/10.3390/rs12091374

Liu, Q., Tian, Y., Zhang, L., & Chen, B. (2022). Urban Surface Water Mapping from VHR Images Based on Superpixel Segmentation and Target Detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *15*, 5339-5356.

Loh, W. Y. (2014). Fifty years of classification and regression trees. *International Statistical Review*, *82*(3), 329-348. https://doi.org/10.1111/insr.12016

Luo, C., Qi, B., Liu, H., Guo, D., Lu, L., Fu, Q., & Shao, Y. (2021). Using time series Sentinel-1 images for object-oriented crop classification in Google Earth Engine. *Remote Sensing*, *13*(4), 561. https://doi.org/10.3390/rs13040561

Mahdianpari, M., Salehi, B., Mohammadimanesh, F., Homayouni, S., & Gill, E. (2019). The first wetland inventory map of newfoundland at a spatial resolution of 10 m using sentinel-1 and sentinel-2 data on the Google Earth Engine cloud computing platform. *Remote Sensing*, *11*(1), 43. https://doi.org/10.3390/rs11010043

Mahdianpari, M., Salehi, B., Mohammadimanesh, F., Brisco, B., Homayouni, S., Gill, E., Delancey, E.R., Bourgeau-Chavez, L., (2020). Big Data for a Big Country: The First Generation of Canadian Wetland Inventory Map at a Spatial Resolution of 10-m Using Sentinel-1 and Sentinel-2 Data on the Google Earth Engine Cloud Computing Platform. *Canadian Journal of Remote Sensing*, *46*(1), 15–33. https://doi.org/10.1080/07038992.2019.1711366

Ministry of Environment, Urbanisation and Climate Change. (2019, October 12). *Köyceğiz-Dalyan Özel Çevre Koruma Bölgesi*. Retrieved October 12, 2019, from https://ockb.csb.gov.tr/koycegiz-dalyan-ozel-cevre-koruma-bolgesi-i-2753

Ouchra B, H., Belangour, A., & Erraissi, A. (2023). Comparison of Machine Learning Methods for Satellite Image Classification: A Case Study of Casablanca Using Landsat Imagery and Google Earth Engine. *Journal of Environmental & Earth Sciences*, *5*(2), 118-134.

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

Pan, H., Chen, H., Hong, Z., Liu, X., Wang, R., Zhou, R., ... & Ma, Z. (2023). A Novel Boundary Enhancement Network for Surface Water Mapping Based on Sentinel-2 MSI Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *16*, 9207-9222. https://doi.org/10.1109/JSTARS.2023.3308046

Phan, T. N., Kuch, V., & Lehnert, L. W. (2020). Land cover classification using Google Earth Engine and random forest classifier—The role of image composition. *Remote Sensing*, *12*(15), 2411. https://doi.org/10.3390/rs12152411

Sarp, G., & Ozcelik, M. (2017). Water body extraction and change detection using time series: A case study of Lake Burdur, Turkey. *Journal of Taibah University for Science*, *11*(3), 381-391. https://doi.org/10.1016/j.jtusci.2016.04.005

Schmitt, M. (2020). Potential of large-scale inland water body mapping from sentinel-1/2 data on the example of Bavaria's lakes and rivers. PFG – Journal of Photogrammetry, *Remote Sensing and Geoinformation Science*, 88(3-4), 271-289. https://doi.org/10.1007/s41064-020-00111-2

Selim, S., Çoşlu, M., Sönmez, N., & Karakuş, N. (2016). Köyceğiz Gölü ve Dalyan Kanallarında Kıyı Kenar Çizgisinin UA ve CBS Teknikleri ile Belirlenmesi, Alanda Karşılaşılan Sorunlar. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, *20*(2), 254-260. https://doi.org/10.19113/sdufbed.78402

Simioni, J. P., Guasselli, L. A., de Oliveira, G. G., Ruiz, L. F., & de Oliveira, G. (2020). A comparison of data mining techniques and multi-sensor analysis for inland marshes delineation. *Wetlands Ecology and Management*, *28*(4), 577-594.

Slagter, B., Tsendbazar, N. E., Vollrath, A., & Reiche, J. (2020). Mapping wetland characteristics using temporally dense Sentinel-1 and Sentinel-2 data: A case study in the St. Lucia wetlands, South Africa. *International Journal of Applied Earth Observation and Geoinformation*, *86*, 102009. https://doi.org/10.1016/j.jag.2019.102009

Solano, F., Di Fazio, S., & Modica, G. (2019). A methodology based on GEOBIA and WorldView-3 imagery to derive vegetation indices at tree crown detail in olive orchards. *International Journal of Applied Earth Observation and Geoinformation*, *83*, 101912. https://doi.org/10.1016/j.jag.2019.101912

Tassi, A., & Vizzari, M. (2020). Object-Oriented LULC Classification in Google Earth Engine Combining SNIC, GLCM, and Machine Learning Algorithms. *Remote Sensing*, *12*(22), 3776. https://doi.org/10.3390/rs12223776

Tassi, A., Gigante, D., Modica, G., Di Martino, L., & Vizzari, M. (2021). Pixel-vs. Object-Based Landsat 8 Data Classification in Google Earth Engine Using Random Forest: The Case Study of Maiella National Park. *Remote Sensing*, *13*(12), 2299. https://doi.org/10.3390/rs13122299

T.C. Köyceğiz Kaymakamlığı. (2023, October 12). *Köyceğiz Gölü*. http://www.koycegiz.gov.tr/koycegiz-golu

Türedi, M. (2006). *Köyceğiz Gölü (Limnolojik Etüt)* [Yüksek lisans tezi, Marmara Üniversitesi]. YÖK Ulusal Tez Merkezi. https://tez.yok.gov.tr/UlusalTezMerkezi

Vapnik, V. (1995). *The nature of statistical learning theory*. Springer.

Wang, M., Liu, X., Gao, Y., Ma, X., & Soomro, N. Q. (2017). Superpixel segmentation: A benchmark. *Signal Processing: Image Communication*, *56*, 28–39. https://doi.org/10.1016/j.image.2017.04.007

Wang, G., Meng, D., Chen, R., Yang, G., Wang, L., Jin, H., ... & Feng, H. (2024). Automatic Rice Early-Season Mapping Based on Simple Non-Iterative Clustering and Multi-Source Remote Sensing Images. *Remote Sensing*, *16*(2), 277. https://doi.org/10.3390/rs16020277

Wangchuk, S., & Bolch, T. (2020). Mapping of glacial lakes using Sentinel-1 and Sentinel-2 data and a random forest classifier: strengths and challenges. *Science of Remote Sensing*, *2*, 100008. https://doi.org/10.1016/j.srs.2020.100008

Wei, X., Xu, W., Bao, K., Hou, W., Su, J., Li, H., & Miao, Z. (2020). A water body extraction methods comparison based on FengYun Satellite data: a case study of Poyang Lake Region, China. *Remote Sensing*, 12(23), 3875. https://doi.org/10.3390/rs12233875

Yang, L., Wang, L., Abubakar, G. A., & Huang, J. (2021). High-resolution rice mapping based on SNIC segmentation and multi-source remote sensing images. *Remote Sensing*, 13(6), 1148. https://doi.org/10.3390/rs13061148

Zhou, L., Pan, S., Wang, J., & Vasilakos, A. V. (2017a). Machine learning on big data: opportunities and Challenges. *Neurocomputing*, *237*, 350-361. https://doi.org/10.1016/j.neucom.2017.01.026

Zhou, Y., Dong, J., Xiao, X., Xiao, T., Yang, Z., Zhao, G., ... & Qin, Y. (2017b). Open surface water mapping algorithms: A comparison of water-related spectral indices and sensors. *Water*, *9*(4), 256. https://doi.org/10.3390/w9040256

Xue, H., Xu, X., Zhu, Q., Yang, G., Long, H., Li, H., ... & Li, Y. (2023). Object-Oriented Crop Classification Using Time Series Sentinel Images from Google Earth Engine. *Remote Sensing*, 15(5), 1353. https://doi.org/10.3390/rs15051353