NÖHÜ Müh. Bilim. Derg. / NOHU J. Eng. Sci., 2025; 14(2), 615-629

Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi
Niğde Ömer Halisdemir University Journal of Engineering Sciences

Araştırma makalesi / Research article

www.dergipark.org.tr/tr/pub/ngumuh / www.dergipark.org.tr/en/pub/ngumuh

Spatial modeling of chlorophyll-a parameter by Landsat-8 satellite data and deep learning techniques: The case of Lake Mogan

Landsat-8 uydu verileri ve derin öğrenme teknikleri ile klorofil-a parametresinin mekansal modellenmesi: Mogan Gölü örneği

Osman Karakoç[1,*], İlkay Buğdaycı[2]

[1,2]*Necmettin Erbakan University, Department of Geomatic Engineering, 42090, Konya, Türkiye*

Abstract

Water is essential for the sustainability of life and the healthy functioning of ecosystems. Increasing pollution poses a serious threat to the world's waters, making the monitoring and protection of water quality a strategic imperative. Chlorophyll-a is one of the most important indicators of water quality and ecosystem health, as it is a measure of photosynthetic activity and phytoplankton density, the lifeblood of aquatic ecosystems. Remote sensed data provide a unique opportunity to analyse chlorophyll-a changes in lake ecosystems. In this study, chlorophyll-a concentration was modelled by machine and deep learning techniques using chlorophyll-a measurements, Landsat-8 surface reflectance values and spectral indices of Lake Mogan between 2018 and 2024. The RF, ANN, and CNN models achieved R^2 values of 0.84, 0.85, and 0.92, respectively. With its ability to learn spectral relationships, identify patterns in complex datasets, and its superior ability to process remote sensing imagery, thematic maps were generated using the CNN model, which performed best in the study. The results of the study demonstrate the potential of remote sensing-based deep learning approaches for monitoring chlorophyll-a. With its ability to produce highly accurate results, this study provides the literature with an effective tool for future strategic monitoring studies.

Keywords: Chlorophyll-a, Deep learning, Landsat-8, Remote sensing, Spectral indices

Öz

Su, yaşamın sürdürülebilirliği ve ekosistemlerin sağlıklı işleyişi için kritik öneme sahiptir. Artan çevresel kirlilik, dünyadaki su kütlelerine yönelik ciddi tehditler oluşturmakta olup, su kalitesinin izlenmesi ve korunmasını stratejik bir zorunluluk haline getirmiştir. Klorofil-a, su ekosistemlerinin yaşam kaynağı olan fotosentetik aktivenin ve fitoplankton yoğunluğunun bir göstergesi olarak, su kalitesini ve ekosistem sağlığını şekillendiren en kritik göstergelerden biridir. Uzaktan algılama tabanlı veri setleri, göl ekosistemlerindeki klorofil-a değişimlerini analiz etmek için eşsiz bir fırsat sunmaktadır. Bu çalışmada, Mogan Gölü'nün 2018-2024 yılları arasındaki klorofil-a ölçümleri, Landsat-8 yüzey yansıma değerleri ve spektral indeksleri kullanılarak klorofil-a konsantrasyonu makine ve derin öğrenme teknikleri ile modellenmiştir. RF, ANN ve CNN modelleri sırasıyla 0.84, 0.85 ve 0.92 R^2 değerlerine ulaşmıştır. Spektral ilişkileri öğrenme kapasitesi, karmaşık veri setlerindeki desenleri tanımlama becerisi ve uzaktan algılama görüntülerini işleme konusundaki üstün yetenekleriyle, araştırmada en iyi performansı gösteren CNN modeli kullanılarak tematik haritalar üretilmiştir. Çalışma sonuçları, klorofil-a parametresinin izlenmesinde uzaktan algılama tabanlı derin öğrenme yaklaşımlarının potansiyelini ortaya koymaktadır. Bu çalışma, yüksek doğruluklu sonuçlar üretme yeteneği ile gelecekteki stratejik izleme çalışmaları için literatüre etkili bir araç sunmaktadır.

Anahtar kelimeler: Derin öğrenme, Klorofil-a, Landsat-8, Spektral indeks, Uzaktan algılama

1 Introduction

Monitoring water quality is crucial for safeguarding ecosystems and supporting human well-being [1]. It is very important to minimize water pollution and develop controllable pollution mechanisms in water bodies [2]. Water quality analyses are carried out with fixed monitoring stations as well as traditional methods that require field observations. Traditional methods are labor intensive as they require field observations. Additionally, it is costly and lacks a regular sampling interval. Although fixed monitoring stations have regular temporal sampling, the limited number of samples causes both approaches to be spatially and temporally inadequate in analyzing water quality [3]. Satellite imagery has great potential for monitoring and assessing lake water quality on a regional scale. Indeed, the use of satellite imagery in the estimation of water quality parameters allows synoptic estimates for large regions with its continuous spatial coverage. In addition, it allows water quality estimation in inaccessible areas. Archived satellite imagery, such as Landsat, enables retrospective water quality determination in cases where field observations cannot be made [4].

The concentration of chlorophyll-a plays a significant role in assessing water quality dynamics [5]. Chlorophyll-a

* Sorumlu yazar / Corresponding author, e-posta / e-mail: osmankarakoc18@gmail.com (O. Karakoç)
Geliş / Received: 18.12.2024 Kabul / Accepted: 07.03.2025 Yayımlanma / Published: 15.04.2025
doi: 10.28948/ngumuh.1603421

is a critical pigment present in all phytoplankton and plays a pivotal role in the process of photosynthesis. Quantifying the concentration of chlorophyll-a in water serves as a pivotal approach for evaluating phytoplankton biomass, as well as for assessing the degree of eutrophication in water bodies and for maintaining lake water quality [6]. Although several methods are available to determine phytoplankton biomass, the most practical approach is to measure chlorophyll a concentration. Phytoplankton biomass is produced by photosynthetic reactions; therefore, determination of pigment concentration alone is sufficient to assess biomass. This method differs from other parameters related to phytoplankton metabolism. In this context, chlorophyll-a analysis is an effective tool for estimating the trophic state of a lake. In addition, chlorophyll-a measurement is faster, easier and more economical than estimating phytoplankton biomass by microscopic analysis. Therefore, chlorophyll-a concentration can be used as an important indicator to determine the eutrophication level of the lake [7]. High levels of chlorophyll-a indicate overproduction of algae, leading to algal blooms [8].

Satellite sensors provide an affordable and efficient method for monitoring and mapping chlorophyll-a concentrations, which serve as an indicator of phytoplankton biomass in lakes [9-11]. Chlorophyll-a estimation with satellite spectral data is mainly based on the absorption and scattering of certain wavelengths of radiation through phytoplankton [12]. Tarı and Kıyak [13] modeled chlorophyll-a concentrations in Lake Van, the largest alkaline lake in the world, using Landsat-8 satellite imagery and geographic information system technologies in an integrated manner. The results of the study show that chlorophyll-a concentrations were obtained with sufficient accuracy and high resolution despite the limited data set. Nas et al. [14] spatially modeled the chlorophyll-a parameter in Lake Beyşehir, the largest freshwater lake in Turkey and also an important source of drinking water, using Terra ASTER satellite imagery and simultaneous measurement results using regression technique. The coefficient $R^2 > 0.86$ calculated as a result of the research shows the relationship between remote sensing data and chlorophyll-a measurements. In addition to the research on monitoring chlorophyll-a concentration, more comprehensive and large amounts of data on water quality can be obtained due to the developments in remote sensing technology. In recent years, the application of machine learning techniques for predicting water quality parameters has become increasingly popular. Machine learning techniques can identify potential relationships between variables by analyzing the features of input data and minimize the gap between predictions and actual observations through model parameter optimization [15-17].

In the literature, studies combining remote sensing technology with machine learning and deep learning techniques for estimating chlorophyll-a parameters have achieved highly successful results. Freddy et al. [18] Measurement data from the country's monitoring network were used to estimate water quality parameters in two lakes in Mexico. In the study using Landsat-8 data, six different machine learning algorithms were developed for chlorophyll-a estimation. In two lakes, R^2 = 0.60, RMSE = 48.06 mg/m^3, MAE = 37.98 mg/m^3 and R^2 = 0.71, RMSE = 6.16 mg/m^3, MAE = 4.97 mg/m^3, respectively. Lien et al. [19] Four advanced machine learning models were used to predict chlorophyll-a concentration in Lake Ranco, Chile. Data from three sampling stations from 1987 to 2020 were used as the dataset. Three different cases were considered in the study: using only in situ data (Case-1), using integrated in situ data and meteorological data (Case-2), and using satellite images, in situ data and meteorological data together (Case-3). In all three cases, machine learning methods gave successful results in the estimation of the chlorophyll-a parameter. In Case-3, the Temporal Convolutional Network (TCN) model was very successful in predicting chlorophyll-a with R^2=0.96, RSME=0.13, MSE= 0.33 and MAE= 0.06. The findings of the study indicate that incorporating a broader range of variables related to chlorophyll-a, either directly or indirectly, enhances the predictive performance of the algorithm. Fangling et al. [20] A hierarchical CNN was designed to identify the relationship between Landsat-8 imagery and water quality parameters in Erhai and Chaohu Lake. A transfer learning strategy was developed in the CNN model to complement the shortcomings of the in situ data. After training the CNN model for Lake Erhai with Landsat-8 imagery and in situ water quality data, the water quality was classified. This model trained on Lake Erhai can be used to classify the water quality of Lake Chaohu. The results show that the CNN model outperforms traditional machine learning methods. Haibo et al. [21] Using Landsat, Sentinel-2 and GaoFen-2 data from Baiyangdian Lake with Spatial Temporal Fusion (STF) method, an inversion model for chlorophyll-a parameter was designed with CNN. A correlation of R^2=0.80 was calculated between measured and predicted chlorophyll-a. Karul et al. [22] Artificial neural networks were used for eutrophication modeling in Mogan and Eymir Lakes. The eutrophication status in Mogan and Eymir lakes was successfully determined due to their small area and more homogeneous structures. The results of the research revealed that the parameter best calculated by artificial neural networks was chlorophyll-a. All of these studies show that the integration of remote sensing techniques with machine and deep learning techniques gives successful results in chlorophyll-a estimation.

Lake Mogan, is located in the borders of Gölbaşı District in the south of Ankara Province. Mogan is within the "Gölbaşı Special Environmental Protection Zone" declared by the Council of Ministers Decree No. 90/1117 [23]. Gölbaşı Special Environmental Protection Area is an environmentally, biologically and culturally rich area with a registered protection status. Mogan Lake and its surroundings are used as a shelter, breeding and resting area for birds and is one of the most important bird habitats designated as a Ramsar candidate in Turkey. In addition, Centaurea tchihatcheffii, one of the endemic plant species, grows naturally in and around Lake Mogan. As Lake Mogan is an important part of the wetland ecosystem and provides various ecosystem services to the people of the region, its protection and regular monitoring is very important.

However, unplanned construction prior to the protection date and human activities that continue today are increasing pollution pressures in and around the lake. For all these reasons, Lake Mogan was selected as the study area in this study.

The aim of this study is to model chlorophyll-a concentration in Lake Mogan using chlorophyll-a measurements, Landsat-8 satellite data and deep learning techniques. Within the scope of the research, chlorophyll-a measurement results measured between 2018 and 2024 were obtained from the Ministry of Environment, Urbanization and Climate Change (MEUCC) General Directorate for the Protection of Natural Assets (GDPA). Using Landsat-8 satellite images, the relationship between the chlorophyll-a parameter and satellite image bands and spectral indices was tested with RF machine learning method and the dataset was expanded. Deeper models were built with ANN model and CNN to estimate the chlorophyll-a parameter affecting the water quality in Lake Mogan. As a result of the research, chlorophyll-a related thematic maps of Lake Mogan were produced with the CNN model, which provides the highest accuracy with deep learning success.

Remote sensing research on chlorophyll-a analysis typically uses empirical methods and traditional regression-based methods. However, these methods may not be sufficient to analyze complex and multidimensional relationships and increase model accuracy. The integrated application of machine learning and deep learning techniques to chlorophyll-a estimation is rare in the literature. In particular, there is no comprehensive study of advanced modeling techniques in Lake Mogan. In this study, chlorophyll-a analysis was performed with high accuracy using advanced modeling techniques such as RF, ANN and CNN. These approaches provide fast and reliable results in monitoring chlorophyll-a concentration and have significant potential for developing autonomous systems for future monitoring studies.

2 Materials and methods

2.1 Study area

Lake Mogan, is located at 39°47′ North Parallel, 32°47′ East Meridian in the borders of Gölbaşı District in the south of Ankara Province, is a shallow freshwater lake with an average surface area of 6 km^2 (Figure 1). The maximum depth of the lake is 4 meters and the average depth is 3 meters. The lake, which has an average height of 973 meters above sea level, has taken its current shape by undergoing changes over time [23]. The streams feeding Lake Mogan come mostly from the south and west. Gölcük and Tatlım in the west, Bağırsak in the northwest, Çolakpınar and Çayır in the southwest, Gölova in the south and Sukesen, Kepir, İğdeli, Başpınar and Yağlıpınar in the east. These streams and the valley floors that harbor them are vital for the survival of the lake. However, negative interventions on these streams and their beds may further reduce the already limited water volume of the lake, leading to deterioration of water quality and increased pollution [23].

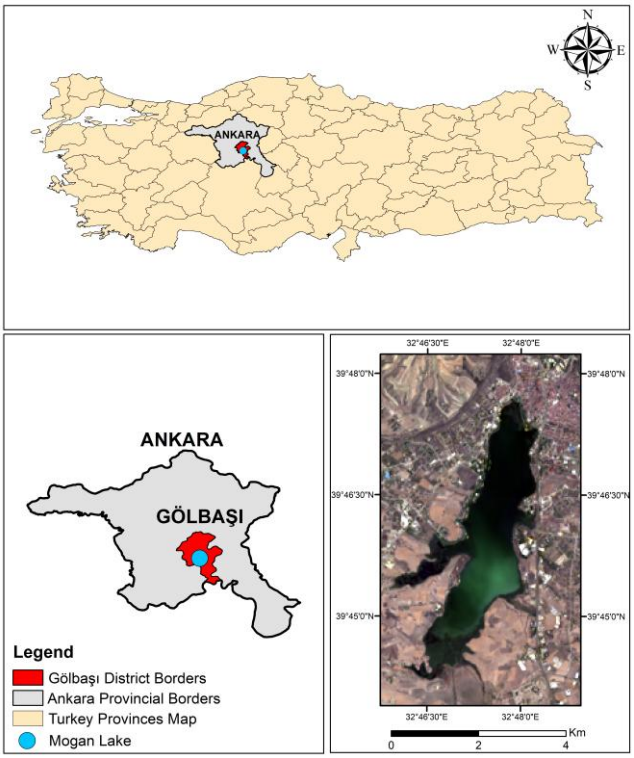

Figure 1. Study area.

Lake Mogan is fed by small streams in its northern and western parts, which often dry up in summer. The water of the lake reaches Eymir Lake through a regulator and a concrete canal and then reaches Ankara Stream through Imrahor Stream [24]. Mogan Lake is located within the "Gölbaşı Special Environmental Protection Area" declared by the Council of Ministers Decree No. 90/1117. The Presidency of the Special Environmental Protection Authority defines Special Environmental Protection Zones as "areas that have integrity in terms of historical, natural, cultural, etc. values and have ecological importance both at the national and world level" [25]. It is very important to protect and regularly monitor Lake Mogan and its surroundings, which are of strategic importance to Turkey as they are an important part of the ecosystem. A Ramsar candidate and protected wetland, Lake Mogan attracts attention for its ecological diversity, endemic species and ecosystem services. Strategically important for tourism and the local economy, the lake and its surroundings face environmental threats due to unplanned development. The conservation, regular monitoring and management of the lake, located in the capital of Turkey, has the potential to serve as a model for all other wetlands. For these reasons, Lake Mogan was selected as the study area.

2.2 Water quality parameters

The results of chlorophyll-a measurements between 2018 and 2024 were obtained from the MEUCC to be used in the study. The water samples used in this study were collected, stored and analyzed under the coordination of the Ministry in accordance with the provisions of the Water Pollution Control Regulation and the Communiqué on Sampling and

Analysis Methods in force in Turkey. The locations of the water samples in Mogan Lake are shown in Figure 2. Water samples taken from three points: the northern end, the southern end and the middle.

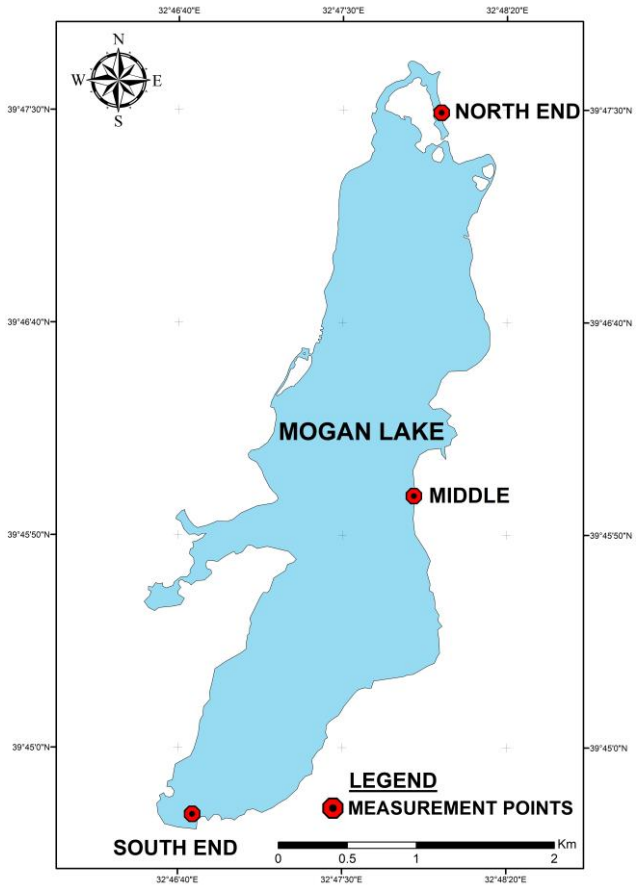

Figure 2. Sampling measurement locations.

Sampling procedures were carried out to represent the physical, chemical and biological characteristics of the lake water. The preservation and transportation processes of the samples collected from the lake were based on TS EN ISO 5667-3 standard. According to this standard, the samples were transported under 4°C temperature and in a dark environment to preserve their properties. They were stored in suitable containers to protect them from external contamination during transportation. Samples delivered to the laboratory were analyzed within the periods specified in the legislation to ensure reliable results. All processes from the sampling stage to the laboratory analysis stage were carried out in accordance with the Water Pollution Control Regulation and international standards and analysis results were obtained.

The current three sampling sites are sufficient to give an idea of the overall water quality status of the lake. However, in order to obtain a high accuracy of water quality and a more homogeneous representation of the lake, more sampling points should be sampled and the sampling frequency should be increased. Kırtıloğlu and Karabörk [26] collected a total of 26 water samples from 13 fixed stations on two different dates to estimate the chlorophyll a concentration in Lake Bafa. The large data set used in the study allowed the performance evaluation of different algorithms to be performed accurately and the remote sensing data to be calibrated with high accuracy. Kavurmacı et al. [27] modeled the chlorophyll-a parameter in Hirfanlı Dam Lake using water quality measurements collected from 54 different points in the lake. Due to the high number of sampling points and the homogeneous distribution over the lake surface, the calculated correlation coefficient of 0.97 for chlorophyll-a was successfully obtained. These studies show that the number of sampling points increases the accuracy of results in remotely sensed water quality monitoring studies.

Although water quality measurements are aimed to be made on a monthly basis, measurements could not be made in some months due to various reasons such as adverse weather conditions. For this reason, there are no measurement results on some dates in the data provided for 2018-2024. Missing measurement data has the potential to adversely affect seasonal analysis and model performance. Polatgil [28], in his research, examined the success of machine learning techniques in completing missing data. As a result of the research, it was concluded that the completion of missing data using appropriate methods positively affects the model performance. Döş and Uysal [29] focused on the classification of remote sensing data using deep learning techniques. The research results revealed the potential of deep learning algorithms to improve the classification performance of remote sensing data. These studies show the impact of the missing data problem on model performance and the potential of machine and deep learning methods to overcome these deficiencies. In this study, machine and deep learning techniques are integrated with remote sensing data to address the problem of limited and missing data. A dataset was created using chlorophyll a measurement data between 2018 and 2024.

2.3 Satellite data

It is very important to obtain satellite images closest to the chlorophyll-a measurement dates used in the study. On the other hand, another factor affecting the accuracy of the study is that the satellite images to be used should be cloud-free. Researchers worldwide have employed a variety of methods to identify water quality parameters using Landsat images across different regions [30-34]. Landsat-8 OLI sensors provide successful data for water quality monitoring with their radiometric and temporal resolution [35]. For this reason, Landsat-8 satellite images were used in this study. Table 1 shows the spectral bands of Landsat-8 satellite.

Satellite images for the study area between 2018 and 2024 were filtered through USGS (United States Geological Survey). Simultaneous and close-time satellite images were identified with the chlorophyll-a measurement dates obtained from the MEUCC. Since water quality parameters are directly affected by atmospheric conditions, it was ensured that the satellite images covering the study area were cloudless. As a result of detailed analysis, Landsat-8 satellite images from eight different dates were matched with the measurement data. The satellite images used in the study are given in Table 2.

Table 1. Landsat-8 satellite spectral bands.

Bands	Defined Spectral Position	Band Range (µm)	Spatial Resolution (m)	Sensor
Band-1	Coastal Aerosol	0.43 – 0.45	30	OLI
Band-2	Blue	0.45– 0.51	30	OLI
Band-3	Green	0.53 – 0.59	30	OLI
Band-4	Red	0.64 – 0.67	30	OLI
Band-5	Near Infrared	0.85 – 0.88	30	OLI
Band-6	SWIR-1	1.57– 1.65	30	OLI
Band-7	SWIR-2	2.11– 2.29	30	OLI
Band-8	Panchromatic	0.50 – 0.68	15	OLI
Band-9	Cirrus	1.36– 1.38	30	OLI
Band-10	TIRS-1	10.60 – 11.19	100	TIRS
Band-11	TIRS-2	11.50 – 12.51	100	TIRS

Table 2. Landsat-8 satellite images used.

DATE	PATH /ROW	CLOUD COVER (%)	ID
18/04/2018	177/32	0.13	LC08_L2SP_177032_ 20180418_20201015_02_T1
23/07/2018	177/32	8.8	LC08_L2SP_177032_ 20180723_20200831_02_T1
27/10/2018	177/32	0.34	LC08_L2SP_177032_ 20181027_20200830_02_T1
31/07/2021	177/32	1.08	LC08_L2SP_177032_ 20210731_20210804_02_T1
16/08/2021	177/32	0.3	LC08_L2SP_177032_ 20210816_20210826_02_T1
17/09/2021	177/32	11.75	LC08_L2SP_177032_ 20210917_20210925_02_T1
05/07/2023	177/32	0.18	LC08_L2SP_177032_ 20230705_20230717_02_T1
08/08/2024	177/32	1.71	LC08_L2SP_177032_ 20240808_20240814_02_T1

In this study, Collection-2 images from the Landsat-8 satellite were used. These images are geometrically and atmospherically corrected by the USGS through pre-processing steps. Atmospheric correction is performed using LaSRC (Land Surface Reflectance Code) to account for various atmospheric effects such as water vapor, aerosols, and ozone. Geometric correction is performed using a Level-1 L1TP (Land Precision) process with DEM (Digital Elevation Models). Since the satellite imagery used is atmospherically and geometrically corrected, no additional pre-processing steps were applied. Erdas Imagine software was used to clip Landsat-8 satellite images to cover the study area and organize the bands to be used for chlorophyll-a analysis. In addition, Digital Number (DN) values were converted into reflectance values and made available in the data set. Spectral reflectance values were extracted for Band 4 (Red) and Band 5 (NIR) bands of the satellite images used.

2.4 Spectral indices

The Normalized Difference Vegetation Index (NDVI), widely used in areas with extensive vegetation cover, is a frequently used method for analyzing vegetation cover [36]. There are significant relationships between chlorophyll-a concentrations in water bodies and NDVI values [37]. NDVI, commonly utilized for chlorophyll-a estimation, was created by analyzing the significant absorption of chlorophyll in the red wavelengths and its strong reflectance in the near-infrared region [38]. The NDVI index is derived from the red and near-infrared (NIR) spectral bands and its formula is given in Equation-1.

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (1)$$

Modified Normalized Difference Water Index (MNDWI) is a special adaptation of the Normalized Difference Water Index (NDWI) index, specifically designed for the detection of water bodies. It provides valuable information for analyzing water bodies using remote sensing techniques [39]. The MNDWI index was developed to overcome the shortcomings of the NDWI index in urban water bodies and to achieve more successful results. In the research conducted by Xu [40], it was found that the MNDWI index performed better than the NDWI index in detecting water bodies in urban areas. MNDWI allows for more accurate mapping of water bodies by preventing them from mixing with other areas. The MNDWI index is calculated using the green and Shortwave Infrared (SWIR) bands and its formula is given in Equation-2.

$$MNDWI = \frac{Green - SWIR}{Green + SWIR} \quad (2)$$

In this study, NDVI and MNDWI indices, which have been proven in the literature to provide successful solutions in water quality analysis, were selected due to their positive effects on model performance. NDVI is an index that captures the spectral characteristics of vegetation and is directly related to the chlorophyll-a parameter. Since phytoplankton in water bodies show strong absorption in the red wavelength range and high reflectance in the near infrared wavelength range, it is known that areas with high NDVI values also have high chlorophyll-a concentrations. Praeger et al. [41] show in their research that there is a strong relationship between NDVI and algal density and the success of the index in predicting algal density. Yeonwoo et al. [42] investigated the relationship between chlorophyll a concentrations and NDVI values. As a result of the research, it was found that the NDVI index can be used in water quality management and algal monitoring. The significant relationship between NDVI and chlorophyll-a has been confirmed by several studies in the literature. For this reason, the NDVI index was calculated to improve the model performance and improve the prediction performance. On the other hand, the MNDWI index is very important for identifying non-water features in water bodies. Xu [40] and

Feyisa et al. [43] used the MNDWI index to obtain a water mask. It was concluded that the MNDWI index outperforms the NDWI in identifying complex water bodies and improves model performance. Therefore, the MNDWI index was included in the study. NDVI and MNDWI spectral indices were calculated for all satellite images to be included in the dataset. ArcGIS software was used to calculate the spectral indices. The workflow diagram showing all steps of the study methodology is shown in Figure 3.

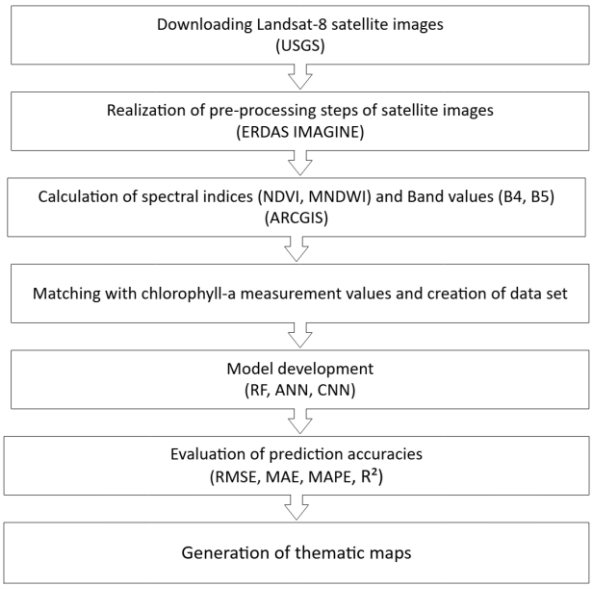

Figure 3. Research workflow diagram.

2.5 *Creating the data set*

Reflectance values (B4, B5), spectral indices (NDVI, MNDWI) obtained from Landsat-8 satellite images were matched with chlorophyll-a measurement data obtained from the MEUCC. A dataset containing a total of 24 measurement data was created.

2.6 *Machine and deep learning methods*

2.6.1 *Random forest method*

The random forest algorithm is a machine learning method based on decision trees. This algorithm randomly selects parameters from the input parameter set and creates multiple decision trees based on these parameters. The decision trees are combined to obtain the classification output of the algorithm. With this approach, independent or irrelevant parameters can be analyzed with different decision trees and with this feature, it provides high accuracy results [44].

In order to successfully expand the limited data set created in the study, the Random Forest method, which gives successful results with limited data, was used. In the training of the model, chlorophyll-a measurement data were determined as the dependent variable, while NDVI, MNDWI, B4 and B5 were determined as independent variables. The data set was randomly split into 80% training and 20% for testing. The number of trees (n_estimators) was set to 100 and the randomness control (random_state) was set to 42. In order to prevent the model from overlearning, the depth was limited and automatic. In the RF model, the choice of 100 trees balances model accuracy and computational efficiency, while the depth of the trees is unlimited to capture the complex relationships between input features. Breiman [45] states that 100 trees provides a balance in model accuracy. To better analyze high-dimensional and complex remotely sensed data, and to allow the model to best learn the relationships, the depth setting is left unrestricted. In the model, the minimum number of data points that each leaf should contain is set to 1, and the minimum number of samples required to split a node is set to 2. These settings ensure that all points are represented, and the model gains depth and best fits the data by preventing oversimplification of the model in a limited data set. After the training was completed, the model was validated. An extended data set was created by generating chlorophyll-a estimates covering the entire lake using the trained model. This data set was then used as the basic input data for ANN and CNN models.

2.6.2 *Artificial neural networks (ANN)*

ANN are a subset of machine learning, which is inspired by the human mind and forms the basis of deep learning methods [46] ANNs are complex, non-linear systems consisting of multiple processing units designed to emulate the behavior of biological neurons. An ANN generally consists of three primary layers: the input layer, hidden layers, and the output layer, with each layer containing a variable number of neurons or nodes. These neurons are interconnected through various mathematical functions, enabling the network to process and learn from data [47]. Figure 4 shows the basic architecture of the ANN.

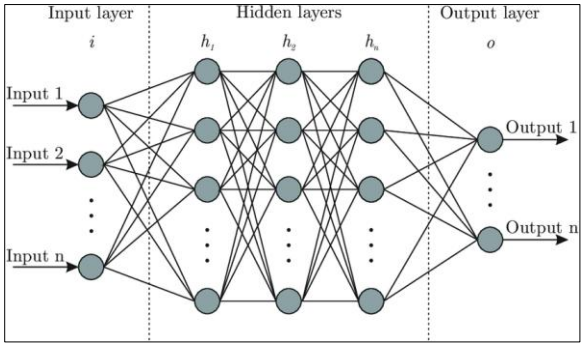

Figure 4. ANN basic architecture [48].

The ANN model was chosen to obtain more accurate predictions using the expanded dataset. The chlorophyll-a value estimated for each pixel from the RF model was used as the dependent variable in the ANN model. NDVI, MNDWI, B4 and B5 values are the independent variables of the model. Since there was a need to scale the features in the ANN model, the independent variables were brought to the range of 0-1 with Min-Max normalization. A total of 6344 data points were used for the ANN model. The data set was randomly split into 80% for training and 20% for testing. The training data consists of 5075 data points and the test data consists of 1269 data points. Adam optimization and 100

epochs were used in model training. The model is structured with 64 neurons in a hidden layer and a ReLU activation function. ReLU was chosen because of its success in modeling nonlinear relationships and the fast and stable optimization properties of the Adam algorithm. These hyperparameters are optimized to increase the generalization capacity and accuracy of the model. The ANN model provided high accuracy in chlorophyll-a estimation. The generalization ability of the model increased the accuracy of the predicted values across the lake.

2.6.3 Convolutional neural networks (CNN)

CNNs are artificial intelligence systems consisting of multilayer neural networks. They are capable of detecting, recognizing, classifying and parsing objects [49]. CNNs have been developed especially for processing and analyzing visual data. They are highly effective in tasks such as image recognition and classification. These networks are considered one of the basic building blocks of computer vision and image processing [50]. Figure 5 shows the basic architecture of the CNN model.

In order to better learn the hidden spatial relationships between the spectral features in the data set extended with the RF model and to increase the prediction accuracy, the CNN model was used. 2-D CNN is successful in learning hidden spatial relationships in spatially dependent data sets, especially in data obtained from satellite images.

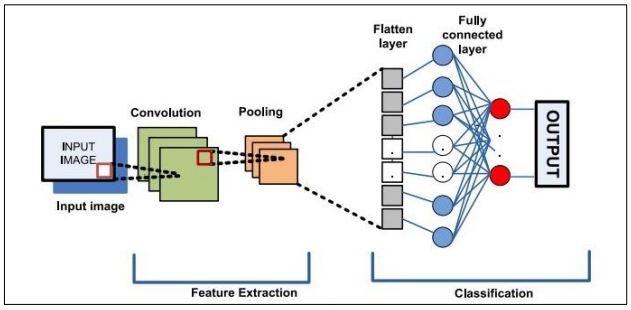

Figure 5. Basic architecture of the CNN model [51].

The dependent variable of the model is the chlorophyll-a measure and the independent variables are NDVI, MNDWI, B4 and B5. Since the CNN requires scaling of the features, all independent variables were set to the 0-1 range. The data set was divided into 80% training and 20% test data set. Out of a total of 6344 data points, 5075 were divided as training data set and 1269 as test data set. The data set used for the CNN model was transformed into a 2D matrix to be compatible with the model. Each data sample was organized into 5x5 spatial blocks to fit the CNN input format. In the first convolution layer, 32 filters and a 3x3 kernel were applied. In the second layer, more complex features were learned with 64 filters and 3x3 kernels. ReLU (Rectified Linear Unit) function was used as the activation function. The 3x3 and 5x5 kernel configurations were chosen to most effectively capture complex structures. The 3x3 kernels capture small-scale details and subtle structural features, while the 5x5 kernels allow for the detection of larger-scale patterns [52]. The combination of these two dimensions increases the learning capacity and improves the learning of relationships at different spatial scales [53]. The Adam algorithm is used to optimize the CNN model. This algorithm is an extension of the scotastic gradient descent method and provides fast and stable optimization during the learning phase. The Adam algorithm updates by taking into account the mean and variance of past gradients. Thanks to this feature, it produces more successful results in deep learning models [54]. The CNN model was optimized as the best performing structure after intensive testing of different hyperparameter combinations. During model training, the learning rate was set to 0.001 and the maximum epoch limit was set to 100. To prevent overlearning, an early stopping method was applied and model training was stopped when no improvement was observed. MSE was used as the loss function. To prevent overfitting, 50% dropout layers were added to different layers to prevent the model from overlearning. As part of the data augmentation method, random rotation, noise addition, and horizontal-vertical translation were applied to the data set to increase the diversity of the training data. These methods took into account the limited size of the dataset and allowed the model to better generalize to different data distributions. However, cross-validation methods were not applied due to the limited size of the dataset. The final model is the result of these extensive optimizations. The developed CNN model showed superior performance in predicting chlorophyll-a concentration. After the training process, the model produced chlorophyll-a estimates covering the entire lake. The obtained predictions were exported in raster format and used to create thematic maps. Python programming language and necessary libraries were used to manage the data processing and modeling processes of RF, ANN and CNN models.

2.7 Accuracy assessment

Coefficient of Determination (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) metrics were used to evaluate the accuracy of the models. These metrics are frequently used to evaluate the accuracy of estimation of water quality parameters with remote sensing techniques [55-57]. R^2 indicates the model's ability to make predictions. As the predicted values align more closely with the actual data, the R^2 value approaches 1, reflecting a stronger predictive performance of the model. RMSE reflects the difference between the predicted values and the actual values. A lower RMSE indicates higher prediction accuracy of the model. MAE represents the average of the absolute differences between the predicted values and the actual values, indicating the error magnitude [58]. MAPE is the average absolute percentage difference between model predicted values and actual values. The mathematical formulas of these parameters used in the accuracy assessment of the models in the study are given in Equation-3, Equation-4, Equation-5 and Equation-6 [59].

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(X_i - Y_i)^2}{\sum_{i=1}^{m}(\bar{Y} - Y_i)^2} \quad (3)$$

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(X_i - Y_i)^2} \quad (4)$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|X_i - Y_i| \quad (5)$$

$$MAPE = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{Y_i - X_i}{Y_i}\right| \quad (6)$$

It's here;

X_i = Estimated value
Y_i = Actual value
m = Number of samples
$\bar{Y}$ = It is the average of actual values.

3 Results and discussions

3.1 Evaluation of model performances

In this study, RF, ANN and CNN models were developed for chlorophyll-a estimation and their performance analysis was evaluated with R^2, RMSE, MAE and MAPE metrics.

The RF model provided reliable results with R^2=0.88 (training) and R^2=0.84 (test), RMSE=0.15 (training) and RMSE=0.18 (test), MAE=0.14 (training) and MAE=0.17 (test) and MAPE=38.50% (training) and MAPE=42.22% (test) for lake-wide chlorophyll-a estimation. This model was a strong starting point due to its ability to generalize with a limited number of measurement data.

The ANN model slightly improved the accuracy with R^2= 0.89 (training) and R^2=0.85 (test) using the data set extended with the RF model. However, RMSE=0.25 (training) and RMSE=0.29, MAE= 0.21 (training) and MAE=0.24 (test), and MAPE=39.80% (training) and MAPE=43.13% (test) which are higher than the RF model. The ANN model slightly improved the performance thanks to its capacity to learn non-linear relationships, but it was less effective compared to the RF model.

The CNN model showed to be the most effective method for chlorophyll-a estimation, providing the highest accuracy among all models. With the CNN model, R^2=0.96 (training) and R^2=0.92 (test), RMSE=0.16 (training) and RMSE=0.20 (test), MAE=0.10 (training) and MAE=0.12 (test) and MAPE=35.50% (training) and MAPE=41.22% (test). The CNN model produced very successful results compared to other methods thanks to its capacity to learn spatial relationships. Table 3 shows the model accuracies.

Partial dependence plots were created to visualize the effects of the RF model and features such as NDVI, MNDWI, B4 and B5 on chlorophyll-a detection. These plots isolate the effect of each independent variable on the predicted chlorophyll-a value with other traits held constant and are given in Figure 6.

Table 3. Model accuracy.

Metric	RF	ANN	CNN
R^2 (Training)	0.88	0.89	0.96
R^2 (Test)	0.84	0.85	0.92
RMSE (Training)	0.15	0.25	0.16
RMSE (Test)	0.18	0.29	0.20
MAE (Training)	0.14	0.21	0.10
MAE (Test)	0.17	0.24	0.12
MAPE (Training)	38.50	39.80	35.50
MAPE (Test)	42.22	43.13	41.22

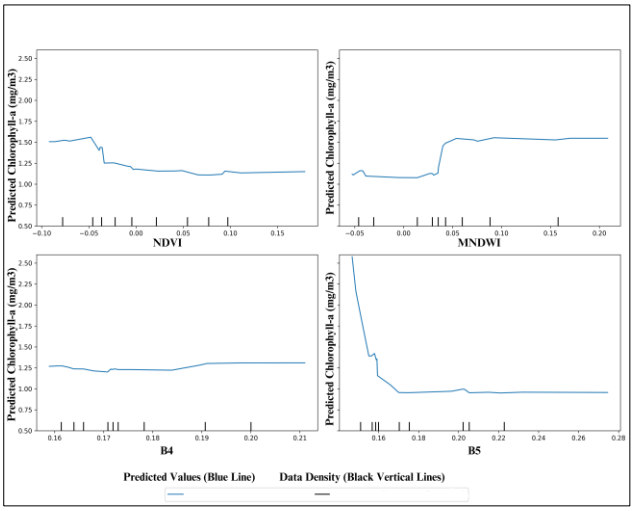

Figure 6. Partial dependency graphs.

A decrease in chlorophyll-a values is observed with the increase in NDVI. This indicates that chlorophyll-a concentration is suppressed with increasing vegetation density. With the increase in MNDWI, chlorophyll-a values first increase and then tend to reach a constant level. While there was no significant fluctuation in the B4 value, it was observed that the chlorophyll-a concentration decreased rapidly with the increase in the B5 value. In order to analyze the importance of each feature in the chlorophyll-a prediction of the RF model in more detail, the graph showing the importance of the features is shown in Figure 7.

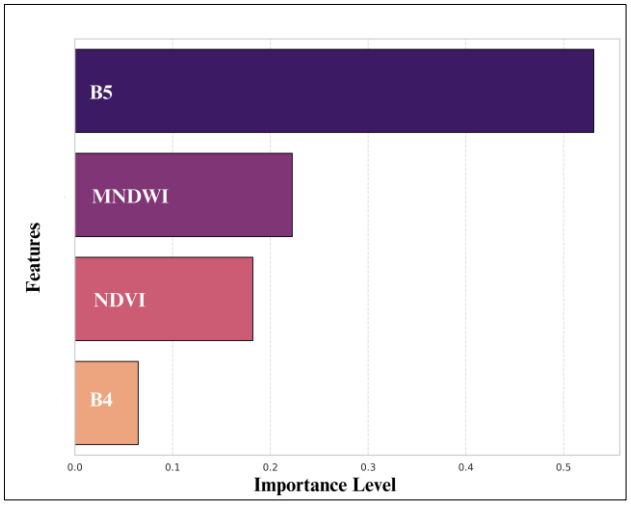

Figure 7. Importance of features.

Among the independent variables used in the RF model, B5 is the most effective feature in model predictions. This feature is followed by MNDWI, NDVI and B4, respectively. Feature B5, NIR, has a strong relationship with the reflectance characteristics of vegetation. Therefore, it plays a critical role in chlorophyll-a estimation. In the literature, the NIR band of Landsat-8 satellite is frequently used for chlorophyll-a estimation [60-61]. Although the Landsat-8 satellite imagery used in this study has great advantages for monitoring large areas, its 30-meter spatial resolution can impose various limitations on the detailed analysis of relatively small water bodies such as lakes. Therefore, the use of satellite imagery with higher spatial resolution can have a positive impact on model performance. Sentinel-2 satellite imagery, with its 10-meter spatial resolution and wide spectral bandwidth, is an alternative that allows for more precise results in the analysis of water quality parameters. Mandanici and Bitelli [62] analyzed the spatial and spectral characteristics of Landsat-8 and Sentinel-2 satellite imagery. The results show that Sentinel-2 satellite imagery performs better in monitoring small and complex areas. Seleem et al. [63] compared the performance of Landsat-8 and Sentinel-2 satellite imagery in monitoring lake water quality. Sentinel-2 satellite imagery, with its higher spatial resolution, provided greater accuracy in detecting water quality parameters. In this study, Landsat-8 satellite imagery was preferred because of its reliability in the literature and its extensive use in monitoring chlorophyll-a concentrations. In future studies, satellite imagery with higher spatial resolution, such as Sentinel-2, can be used to improve model accuracy. Figure 8 shows the error analysis of all models, Figure 9 shows the error analysis of all models scatter and line graphs are shown.

MNDWI, the second most important parameter of the model, is an important indicator for analyzing the optical properties and areas of water bodies. Vivek et al. [64] In their research conducted in Bangalore, the capital of India, they used the MNDWI index to monitor changes in water.

The NDVI index ranks third in importance. Akgün [37] In a study conducted in the Kura River, it was found that there were statistically significant relationships between chlorophyll-a values and NDVI index values. The data set was expanded by generating predictions for each pixel to cover the entire lake area with the random forest model. The expanded data set was checked with chlorophyll-a measurement values that were not included in the model. It was seen that the prediction values produced with RF were compatible with the actual measurement values. This expanded data set was used for training ANN and CNN models to increase the prediction accuracy.

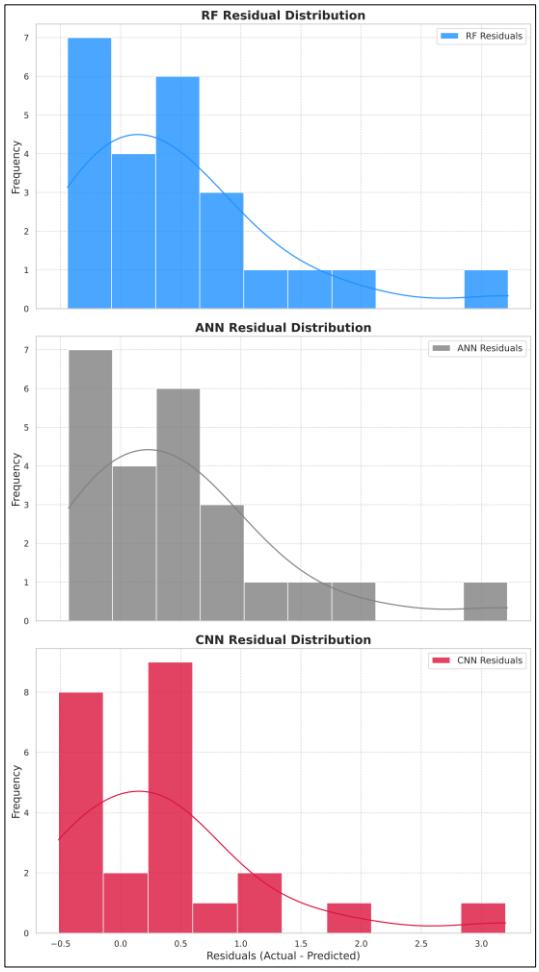

Figure 8. Error analysis of models for chlorophyll-a predictions.

The results show the advantages of machine and deep learning methods in chlorophyll-a estimation. The RF model is characterized by low error values as it was developed using a limited number of measurement data. The ANN model provided some improvement in accuracy by training with an expanded data set. RF and ANN models are frequently used in water quality monitoring studies. Dewi et al. [65] RF and ANN models were used to estimate the Water Quality Index (WQI). MAE= 121.40 and RMSE= 215.04 for the ANN model and MAE= 7.87 and RMSE= 28.99 for the RF model. As a result of the research, it was found that the RF model outperformed the ensemble of decision trees and was more reliable for modeling complex patterns.

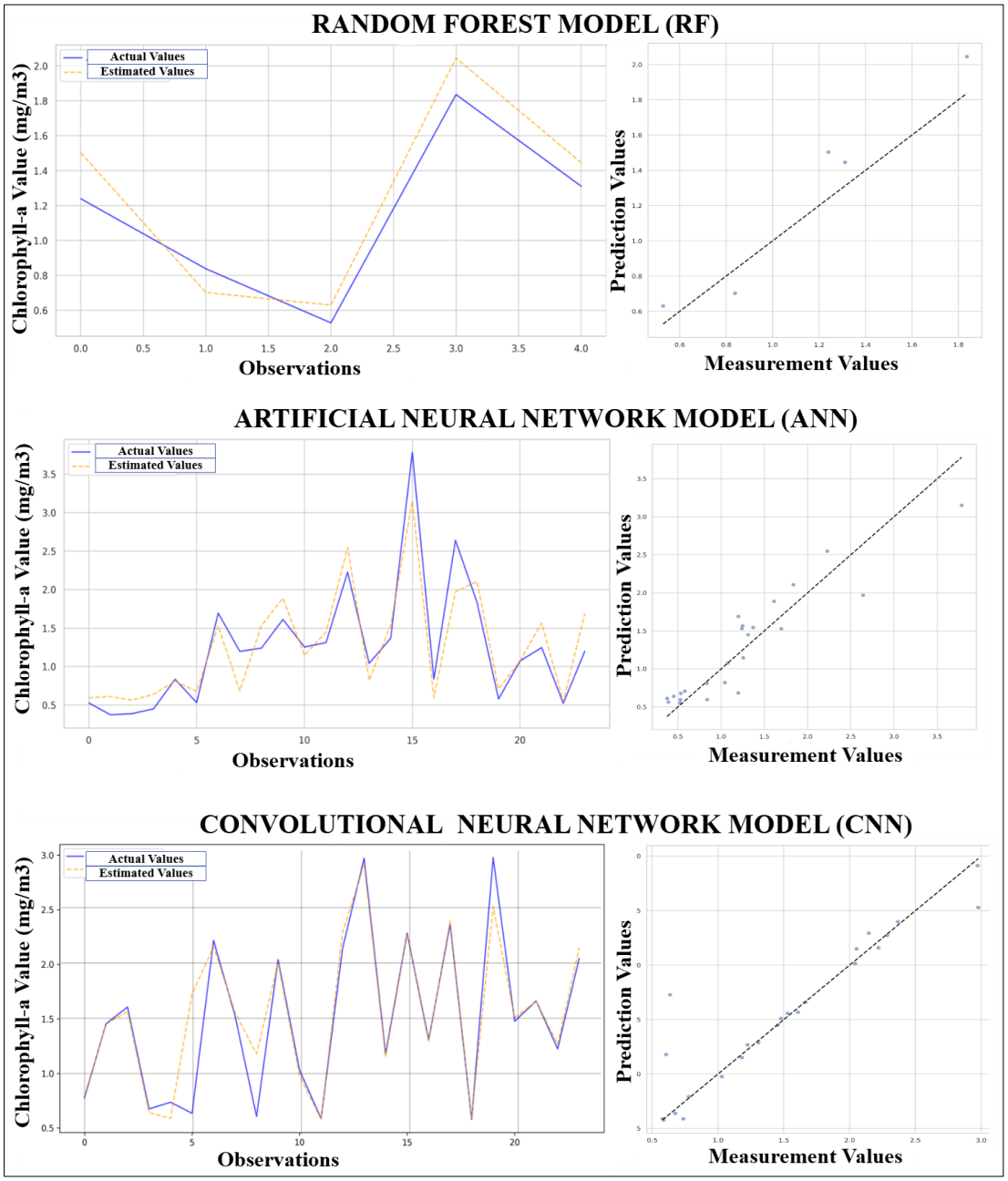

Figure 9. Scatter and line graphs of models.

The research results support the results obtained in this study. Ming et al. [66] In their research, they used four popular machine learning methods to obtain an optimal chlorophyll-a algorithm. RF was the machine learning method that provided the highest accuracy for chlorophyll-a. He mapped the changes in chlorophyll-a concentration between 2016 and 2020 for 163 large lakes in eastern China using the RF model. The CNN model showed the best performance in chlorophyll-a estimation with the highest R^2 and the lowest MAE value. In addition, a lower RMSE value was obtained with CNN model than ANN. CNN models have been popularly used in remote sensing and water quality monitoring studies in recent years. To cope with the limited prediction ability of traditional machine learning methods, deeper neural networks are used to achieve high accuracy results. Syariz et al. [67] Based on the fact that the ANNs proposed in previous studies do not represent the spatial features of satellite images in predicting chlorophyll-a concentrations, a new CNN model called WaterNet is proposed.

In the study, it was concluded that chlorophyll-a concentrations decreased to the range of RMSE= 0.509-0.975 µg/L with the activation of the CNN model.

As a result of the research, it was found that the CNN model gave better results and superior performance than all other models in chlorophyll-a estimation. Zhang et al. [68] several CNN models were developed to reconstruct chlorophyll-a concentrations. The RMSE values ranging

from 0.2916-0.3744 calculated for different inputs show the superior performance of the model.

3.2 Production of thematic maps

In the study, the CNN model, which provides the highest pryaediction accuracy, was used to produce chlorophyll-a forecasts for simultaneous dates with satellite images. The forecast data was exported in raster format using Python software language. Thematic maps showing the chlorophyll-a concentration in Lake Mogan for 8 different dates were produced by classifying the raster images with ArcGIS software. Thematic maps showing the distribution of chlorophyll-a in Lake Mogan are given in Figure 10.

The thematic maps produced within the scope of the study are very important in terms of monitoring the chlorophyll-a change in Lake Mogan. Thematic maps reveal the distribution of chlorophyll-a parameter on the lake surface area. In addition, thematic maps are extremely important in explaining the effect of natural events or human activities on chlorophyll-a change.

The forecast data obtained with the CNN model are based on the data obtained from the MEUCC. Chlorophyll-a measured values were compared. The thematic maps were found to be in agreement with the actual measured values. The results of the study reveal the superior success of CNN in predicting water quality parameters.

When the change in chlorophyll-a concentration in Lake Mogan between 2018 and 2024 is analyzed, it is seen that the chlorophyll-a concentration in the lake has increased from past to present. Although Lake Mogan and its surroundings are located in a special environmental protection zone, the intense urbanization pressure in the region has negatively affected the water quality from past to present. Since traditional monitoring methods are both costly and time consuming, monitoring water quality with remote sensing techniques offers very successful solutions. This study provides results that will shed light on future studies in Lake Mogan, which is located in a special environmental protection zone and has a critical importance in terms of continuous monitoring of water quality.

4 Conclusions

In this study, the chlorophyll-a concentration in Lake Mogan was estimated using Landsat-8 satellite imagery, machine learning and deep learning techniques. R^2 values of 0.84, 0.85 and 0.92 were calculated from RF, ANN and CNN models, respectively. The CNN model showed higher accuracy compared to the other models, and thematic maps of chlorophyll-a distribution were generated using the model predictions. The model performances show that there are significant relationships between remote sensing data and chlorophyll-a values. The research provides a successful solution for chlorophyll-a analysis and gives promising results for the autonomization of future monitoring studies.

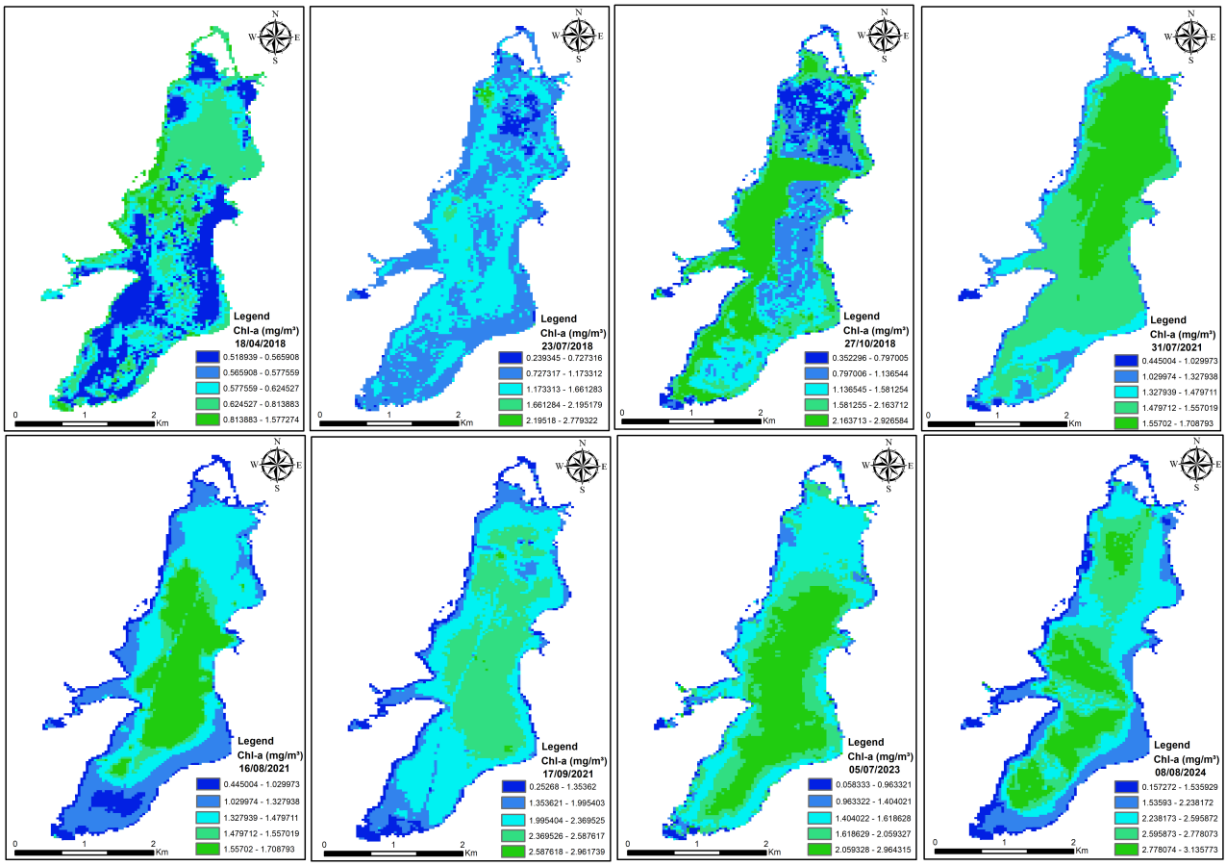

Figure 10. Chlorophyll-a thematic maps.

In light of the results obtained, the use of Landsat-8 satellite imagery with a spatial resolution of 30 meters in the study may limit the detection of rapid changes in water quality or fine details. In future studies, higher resolution satellite imagery, such as Sentinel-2, can be used for detailed analysis. In addition, gaps in the data set have the potential to negatively affect model performance. Model performance can be improved by filling the gaps in the data set with regular monitoring studies or measurements scheduled to coincide with the satellite imagery data.

Acknowledgment

This article is derived from PhD thesis titled Modeling of Water Quality Parameters with Remote Sensing and Deep Learning Techniques. We would like to thank the Ministry of Environment, Urbanization and Climate Change (MEUCC) for providing the necessary permission for the research and supporting us in obtaining the data used.

Conflict of interest

The authors declare that there is no conflict of interest.

Similarity rate (iThenticate): 16%

References

[1] S. Prashant, M. Ramandeep, P. Prem, A. Akash, S. Prachi, P. Manish and G. Ayushi, Revisiting hyperspectral remote sensing: Origin, processing, applications and way forward. Hyperspectral Remote Sensing, Elsevier, 3-21, 2020. https://doi.org/10.1016/B978-0-08-102894-0.00001-2.

[2] M. I. H. Zaidi Farouk, Z. Jamil, M.F. Abdul Latip, Towards online surface water quality monitoring technology: A review. Environmental research, 238, 117147, 2023. https://doi.org/10.1016/j.envres.2023.117147.

[3] S. G. Virdis, W. Xue, E. Winijkul, V. Nitivattananon and P. Punpukdee, Remote sensing of tropical riverine water quality using sentinel-2 MSI and field observations. Ecological Indicators, 144, 109472, 2022. https://doi.org/10.1016/j.ecolind.2022.109472.

[4] A. Kulkarni, Water quality retrieval from Landsat TM imagery. Procedia Computer Science, 6, 475-480, 2011. https://doi.org/10.1016/j.procs.2011.08.088.

[5] J. Kravitz, M. Matthews, S. Bernard and D. Griffith, Application of Sentinel 3 OLCI for chl-a retrieval over small inland water targets: Successes and challenges. Remote Sensing of Environment, 237, 111562, 2020. https://doi.org/10.1016/j.rse.2019.111562.

[6] H. Huang, W. Wang, J. Lv, Q. Liu, X. Liu, S. Xie, F. Wang and J. Feng, Relationship between chlorophyll a and environmental factors in lakes based on the random forest algorithm. Water, 14, 19, 3128, 2022. https://doi.org/10.3390/w14193128.

[7] S.T.P. Phu, Research on the correlation between chlorophyll-a and organic matter BOD, COD, phosphorus, and total nitrogen in Stagnant Lake Basins. In: Kaneko, Sustainable Living with Environmental Risks, Springer, Tokyo, 2014. https://doi.org/10.1007/978-4-431-54804-1_15.

[8] A. Sudradjat, B. S. Muntalif, N. Marasabessy, F. Mulyadi and M. I. Firdaus, Relationship between chlorophyll-a, rainfall, and climate phenomena in tropical archipelagic estuarine waters, Heliyon, 10, 4, 2024. https://doi.org/10.1016/j.heliyon.2024.e25812.

[9] C. Kislik, I. Dronova, T. E. Grantham and M. Kelly, Mapping algal bloom dynamics in small reservoirs using Sentinel-2 imagery in Google Earth Engine, Ecological Indicators, 140, 109041, 2022. https://doi.org/10.1016/j.ecolind.2022.109041.

[10] J. Wang and X. Chen, A new approach to quantify chlorophyll-a over inland water targets based on multi-source remote sensing data. Science of The Total Environment, 906, 167631, 0048-9697, 2024. https://doi.org/10.1016/j.scitotenv.2023.167631.

[11] K. Dörnhöfer, P.Klinger, T. Heege and N. Oppelt, Multi-sensor satellite and in situ monitoring of phytoplankton development in a eutrophic-mesotrophic lake. Science of The Total Environment, 612, 1200-1214, 2018. https://doi.org/10.1016/j.scitotenv.2017.08.219.

[12] M. W. Matthews, A current review of empirical procedures of remote sensing in inland and near-coastal transitional waters. International Journal of Remote Sensing, 32, 21, 6855-6899, 2011. https://doi.org/10.1080/01431161.2010.512947

[13] U. Tarı and N. Olğun Kıyak, Uydu verisi ve CBS ile Van Gölü klorofil-a dinamiklerinin izlenmesi [English Translation: Monitoring the chlorophyll-a dynamics of Van Lake with satellite data and GIS]. Journal of Advanced Research in Natural and Applied Sciences, 10, 1, 60-79, 2024. https://doi.org/10.28979/jarnas.1317247.

[14] B. Nas, H. Karabork and S. Ekercin, Mapping chlorophyll-a through in-situ measurements and Terra ASTER satellite data. Environ Monit Assess, 157, 375–382, 2009. https://doi.org/10.1007/s10661-008-0542-9.

[15] H. Cen, J. Jiang, G. Han, X. Lin, Y. Liu, X. Jia, Q. Ji and B. Li, Applying deep learning in the prediction of chlorophyll-a in the east china sea. Remote Sensing. 14, 21, 5461, 2022. https://doi.org/10.3390/rs14215461

[16] J. G. N. Paulino, G. G. Esperanza, R. A. F. José and D. M. Cristina, Forecast of chlorophyll-a concentration as an indicator of phytoplankton biomass in El Val reservoir by utilizing various machine learning techniques: A case study in Ebro river basin, Spain. Journal of Hydrology, 639, 131639,0022-1694, https://doi.org/10.1016/j.jhydrol.2024.131639.

[17] K. Doyun, L. KyoungJin, J. SeungMyeong, S. MinSeok, K. ByeoungJun, P. Jungsu and H. Tae-Young, Real-time chlorophyll-a forecasting using machine learning framework with dimension reduction and hyperspectral data. Environmental Research, 262, 119823, 0013-9351, 2024. https://doi.org/10.1016/j.envres.2024.119823.

[18] F. H. Villota-González, B. Sulbarán-Rangel, F. Zurita-Martínez, K. J. Gurubel-Tun and V. Zúñiga-Grajeda, Assessment of machine learning models for remote sensing of water quality in Lakes Cajititlán and Zapotlán. Jalisco-Mexico, Remote Sensing. 15, 23, 5505, 2023. https://doi.org/10.3390/rs15235505

[19] L. Rodríguez-López, L. Bravo Alvarez, I. Duran-Llacer, D. E. Ruíz-Guirola, S. Montejo-Sánchez, R. Martínez-Retureta, E. López-Morales, L. Bourrel, F. Frappart and R. Urrutia, Leveraging machine learning and remote sensing for water quality analysis in lake ranco, Southern Chile. Remote Sensing, 16, 18, 3401, 2024. https://doi.org/10.3390/rs16183401

[20] F. Pu, C. Ding, Z. Chao, Y. Yu and X. Xu, Water-quality classification of inland lakes using Landsat8 images by convolutional neural networks. Remote Sensing, 11, 14, 1674, 2019. https://doi.org/10.3390/rs11141674

[21] H. Yang, Y. Du, H. Zhao and F. Chen, Water quality chl-a inversion based on spatio-temporal fusion and convolutional neural network. Remote Sensing. 14, 5, 1267, 2022. https://doi.org/10.3390/rs14051267

[22] C. Karul, S. Soyupak, A. F. Çilesiz, N. Akbay and E. Germen, Case studies on the use of neural networks in eutrophication modeling. Ecological Modelling, 134, 145-152, 2000. https://doi.org/10.1016/S0304-3800(00)00360-4.

[23] Anonim, Gölbaşı özel çevre koruma bölgesi yönetim planı, Çevre, Şehircilik ve İklim Değişikliği Bakanlığı Tabiat Varlıklarını Koruma Genel Müdürlüğü, 213, Ankara, 2015.

[24] A. Velioğlu and M. Kırkağaç, Seasonal variation of zooplankton in Lake Mogan. Aquatic Sciences and Engineering, 32, 3, 146-153, 2017.

[25] Özel Çevre Koruma Kurumu Başkanlığı, Ö.Ç.K.K.B. Yayınları, Ankara, 1994.

[26] E. Kırtıloğlu and H. Karabörk, Evaluating the performance of algorithms in estimating the Chl-a concentration of Lake Bafa. Turkish Journal of Geosciences, 3, 1, 30-38, 2022. https://doi.org/10.48053/turkgeo.1118373.

[27] M. Kavurmacı, S. Ekercin, L. Altaş and Y. Kurmaç, Hirfanlı baraj gölü su kalitesinin cbs ve uzaktan algılama teknikleri kullanılarak değerlendirilmesi [English Translation: Evaluation of water quality of Hirfanlı Dam Lake using gis and remote sensing techniques]. 65. Türkiye Jeoloji Kurultayı, 2-6 Nisan, 2012.

[28] M. Polatgil, Veri ölçekleme ve eksik veri tamamlama yöntemlerinin makine öğrenmesi yöntemlerinin başarısına etkisinin incelenmesi. Duzce University Journal of Science and Technology, 11, 1, 78-88, 2023 https://doi.org/10.29130/dubited.948564.

[29] M. E. Döş and M. Uysal, Uzaktan algılama verilerinin derin öğrenme algoritmaları ile sınıflandırılması. Türkiye Uzaktan Algılama Dergisi, 1, 28-34, 2019.

[30] T. Perivolioti, A. Mouratidis, D. Bobori, G. Doxani and D. Terzopoulos, Monitoring water quality parameters of Lake Koronia by means of long time-series multispectral satellite images. Acta Universitatis Carolinae, Geographica, Univerzita Karlova, 52, 2017. https://doi.org/10.14712/23361980.2017.14.

[31] C. Qi, S. Huang and X. Wang, Monitoring water quality parameters of Taihu Lake based on remote sensing images and LSTM-RNN. IEEE Access, 8, 188068-188081, 2020. https://doi.org/10.1109/ACCESS.2020.3030878

[32] B. Nas, S. Ekercin, H. Karabörk, A. Berktay and D. Mulla, An application of Landsat-5TM image data for water quality mapping in Lake Beysehir, Turkey. Water Air Soil Pollution 212, 183-197, 2010. https://doi.org/10.1007/s11270-010-0331-2.

[33] W. Ahmed, S. Mohammed, A. El-Shazly and S. Morsy, Tigris River water surface quality monitoring using remote sensing data and GIS techniques. The Egyptian Journal of Remote Sensing and Space Science. 26, 816-825, 2023. https://doi.org/10.1016/j.ejrs.2023.09.001.

[34] L. Rodríguez-López, I. Duran-Llacer, L. Bravo Alvarez, A. Lami, R. Urrutia, Recovery of water quality and detection of algal blooms in Lake Villarrica through Landsat satellite images and monitoring data. Remote Sensing, 15, 7, 1929, 2023. https://doi.org/10.3390/rs15071929

[35] C. Martin, J. Junchang, G. M. Jeffrey, L. D. Jennifer, F. V. Eric, R. Jean-Claude, V. S. Sergii and J. Christopher, The harmonized Landsat and Sentinel-2 surface reflectance data set. Remote Sensing of Environment, 219, 145-161, 2018. https://doi.org/10.1016/j.rse.2018.09.002.

[36] H. Yıldız, A. Mermer, E. Ünal and F. Akbaş, Türkiye bitki örtüsünün NDVI verileri ile zamansal ve mekansal analizi [English Translation: Spatial and Temporal analysis of Turkey vegetation with NDVI images]. Tarla Bitkileri Merkez Araştırma Enstitüsü Dergisi, 21, 2, 50-56, 2012.

[37] M. Akgün, Uzaktan algılama ile klorofil-a izlenmesi üzerine bir çalışma [English Translation: A study on monitoring of chlorophyll-a level by remote sensing]. Jsat, 1,41–47, 2023. https://doi.org/10.5281/zenodo.8074879.

[38] B. Datt, Remote sensing of chlorophyll a, chlorophyll b, chlorophyll a+ b, and total carotenoid content in eucalyptus leaves. Remote sensing of environment, 66, 2, 111-121, 1998. https://doi.org/10.1016/S0034-4257(98)00046-7.

[39] Ü. H. Atasever and H. H. Abbas, Drought monitoring in Burdur Lake using Sentinel-2 images. Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi, 13, 3, 882-891, 2024. https://doi.org/10.28948/ngumuh.1411803

[40] H. Xu, Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. International Journal of Remote Sensing, 27, 14, 3025-3033, 2006. https://doi.org/10.1080/01431160600589179

[41] C. Praeger, M. J. Vucko, L. I. McKinna, R. D. Nys and A. J. Cole, Estimating the biomass density of macroalgae in land-based cultivation systems using

spectral reflectance imagery. Algal Research-Biomass Biofuels and Bioproducts, 50, 102009, 2020. https://doi.org/10.1016/j.algal.2020.102009.

[42] Y. Choo, G. Kang, D. Kim and S. Lee, A study on the evaluation of water-bloom using image processing. Environmental science and pollution research international, 25, 36, 36775–36780, 2018. https://doi.org/10.1007/s11356-018-3578-6

[43] G. L. Feyisa, H. Meilby, R. Fensholt and S. R. Proud, Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. Remote Sensing of Environment, 140, 23-35, 2014. https://doi.org/10.1016/j.rse.2013.08.029.

[44] Y. Dokuz, A Bozdağ and B. Gökçek, Hava kalitesi parametrelerinin tahmini ve mekansal dağılımı için makine öğrenmesi yöntemlerinin kullanılması [English Translation:Use of machine learning methods for estimation and spatial distribution of air quality parameters]. Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi, 9, 1, 37-47. https://doi.org/10.28948/ngumuh.654092

[45] L. Breiman, Random forests. Machine Learning, 45, 1, 5-32, 2001. https://doi.org/10.1023/A:1010933404324

[46] R. Qamar and B. Zardari, Artificial neural networks: an overview. Mesopotamian Journal of Computer Science. 130-139, 2023 https://doi.org/10.58496/MJCSC/2023/015.

[47] P. P. Biswas, W. H. Chen, S. S. Lam, Y. K. Park, J. S. Chang and A. T. Hoang, A comprehensive study of artificial neural network for sensitivity analysis and hazardous elements sorption predictions via bone char for wastewater treatment. Journal of hazardous materials, 465, 133154, 2024. https://doi.org/10.1016/j.jhazmat.2023.133154

[48] F. Bre, J. Gimenez and V. Fachinotti, Prediction of wind pressure coefficients on building surfaces using artificial neural networks. Energy and Buildings, 158, 2017. https://doi.org/10.1016/j.enbuild.2017.11.045.

[49] M. M. Taye, Theoretical understanding of convolutional neural network: concepts, architectures, applications, Future Directions. Computation, 11, 3, 52, 2023. https://doi.org/10.3390/computation11030052

[50] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci and M. Parmar, A review of convolutional neural networks in computer vision. Artificial Intelligence Review, 57, 99, https://doi.org/10.1007/s10462-024-10721-6

[51] N. Pravati, R. D. Shitya, K. M. Ranjan, M. Sairam, A. Ahmed, M. Alsharef and A. Flah, 2D-convolutional neural network based fault detection and classification of transmission lines using scalogram images. Heliyon, 10, 19, 2405-8440, 2024. https://doi.org/10.1016/j.heliyon.2024.e38947.

[52] K. Simonyan, A. and Zisserman, Very deep convolutional networks for large-scale image recognition. The 3rd International Conference on Learning Representations (ICLR2015), 2015. https://doi.org/10.48550/arXiv.1409.1556.

[53] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, Going deeper with convolutions. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1-9, 2015. https://doi.org/10.1109/CVPR.2015.7298594.

[54] D. P. Kingma and J. Ba, Adam: A Method for Stochastic Optimization. CoRR, 2014. https://doi.org/10.48550/arXiv.1412.6980

[55] P. D. Devi and G. Mamatha, Machine learning approach to predict the turbidity of Saki Lake, Telangana, India, using remote sensing data, Measurement: Sensors, 33, 101139, 2665-9174, 2024. https://doi.org/10.1016/j.measen.2024.101139.

[56] H. Zhang, B. Xue, G. Wang, X. Zhang and Q. Zhang. Deep learning-based water quality retrieval in an impounded lake using Landsat 8 imagery: an application in Dongping Lake. Remote Sensing, 14, 18, 4505, 2022. https://doi.org/10.3390/rs14184505

[57] A. Ali, G. Zhou, F. Pablo, L. F. Antezana, C. Xu, G. Jing and Y. Tan, Deep learning for water quality multivariate assessment in inland water across China. International Journal of Applied Earth Observation and Geoinformation. 133, 104078, 2024. https://doi.org/10.1016/j.jag.2024.104078.

[58] C. Peng, W. Biao, W. Yanlan, W. Qijun, H. Zuoji and W. Chunlin, Urban river water quality monitoring based on self-optimizing machine learning method using multi-source remote sensing data. Ecological Indicators, 146, 109750, 2023. https://doi.org/10.1016/j.ecolind.2022.109750.

[59] D. Chicco, M. J. Warrens and G. Jurman, The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science 7:e623 2021. https://doi.org/10.7717/peerj-cs.623

[60] U. Tarı, N. Olğun Kıyak, Uydu verisi ve CBS ile Van Gölü klorofil-a dinamiklerinin izlenmesi [English Translation: Monitoring the chlorophyll-a dynamics of Van Lake with satellite data and GIS]. Journal of Advanced Research in Natural and Applied Sciences, 10, 1, 60-79. 2024. https://doi.org/10.28979/jarnas.1317247

[61] W. G. Buma, S. I. Lee, Evaluation of Sentinel-2 and Landsat 8 images for estimating chlorophyll-a concentrations in Lake Chad, Africa. Remote Sensing, 12, 15, 2437, 2020. https://doi.org/10.3390/rs12152437

[62] Mandanici, E., & Bitelli, G. (2016). Preliminary Comparison of Sentinel-2 and Landsat 8 Imagery for a Combined Use. Remote Sensing, 8(12), 1014. https://doi.org/10.3390/rs8121014.

[63] Seleem, T., Bafi, D., Karantzia, M. et al. Water Quality Monitoring Using Landsat 8 and Sentinel-2 Satellite Data (2014–2020) in Timsah Lake, Ismailia, Suez Canal Region (Egypt). J Indian Soc Remote Sens 50, 2411–2428 (2022). https://doi.org/10.1007/s12524-022-01613-9

[64] K. G Vivek, K. G. Piyush, P. Murugan, M. Annadurai, Assessment of surface water dynamicsin bangalore Using WRI, NDWI, MNDWI, supervised classification and K-T transformation,Aquatic Procedia, 4, 739-746, 2214-241X, 2015. https://doi.org/10.1016/j.aqpro.2015.02.095

[65] D. Dewi, A. Wei, L. Lin and C. Heng, Water quality prediction using random forest algorithm and optimization. Journal of Applied Data Sciences, 5, 3, 1354-1362, 2024. https://doi.org/10.47738/jads.v5i3.348

[66] S. Ming, L. Juhua, C. Zhigang, X. Kun, Q. Tianci, M. Jinge, L. Dong, S. Kaishan, F. Lian, D. Hongtao, Random forest: An optimal chlorophyll-a algorithm for optically complex inland water suffering atmospheric correction uncertainties. Journal of Hydrology; 615, Part A, 2022. https://doi.org/10.1016/j.jhydrol.2022.128685

[67] M. A Syariz and C. H. Lin, M.V. Nguyen, L. M. Jaelani and A. C. Blanco, WaterNet: A convolutional neural network for chlorophyll-a concentration retrieval. Remote Sensing, 12, 12, 1966, 2020. https://doi.org/10.3390/rs12121966

[68] X. Zhang and M. Zhou, A general convolutional neural network to reconstruct remotely sensed chlorophyll-a concentration. Journal of Marine Science and Engineering, 11, 4, 810, 2023. https://doi.org/10.3390/jmse11040810