Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 12 NO: 4 PAGE: 1028-1040 YEAR: 2023
DOI:10.17798/bitlisfen.1316881

Prediction of Lake Van Water Level Using an Artificial Neural Network Model with Meteorological Parameters and Multiple Linear Regression Analysis: A Comparative Study

Furkan SİDAL[1], Yener ALTUN[1*]

[1]*Department of Statistics, Institute of Science, Yüzüncü Yıl University, Van, Türkiye*
(ORCID: 0000-0002-9670-2618) (ORCID: 0000-0003-1073-5513)

Keywords: Artificial Neural Network, Comparative Study, Lake Water Level Prediction, Modeling, Multiple Linear Regression Analysis.

Abstract
The water level of Lake Van has shown changes over time. This study encompasses a statistical investigation conducted to understand the reasons behind the variation in the lake's water level. In this study, an attempt has been made to establish a predictive model by determining the effects of meteorological factors on the lake's water level. Artificial neural networks have been utilized to predict the water level of Lake Van using meteorological parameters such as precipitation, temperature, evaporation, wind speed, relative humidity, and atmospheric pressure. Furthermore, a model equation has been formulated by examining the relationship between independent variables and the changes in the water level of Lake Van through multiple linear regression analysis. The two models have been compared, and the results have been evaluated. The obtained results indicate that the artificial neural network model can provide more realistic predictions for the water level of Lake Van compared to the multiple regression analysis method, demonstrating that artificial neural networks serve as a tool for both temporal and spatial predictions.

1. Introduction

Van Lake, one of the largest lakes in Turkey, holds significant ecological and socioeconomic importance. In recent years, noticeable fluctuations have been observed in the water level of Van Lake. These fluctuations have had significant impacts on both the lake ecosystem and the lives of the people residing in its vicinity. Therefore, understanding the causes behind these changes in the water level of Van Lake and predicting future fluctuations are of great importance. Numerous studies have been conducted to identify the reasons behind the rise and fall of the water level in Van Lake. These studies not only focus on determining the physical factors responsible for these changes but also employ statistical approaches to analyze the lake's water level.

This study presents a statistical approach to investigate the changes in the water level of Van Lake. With the rapid advancement of technology, an innovative method utilizing artificial neural networks (ANNs) has been employed for predicting the lake's water level. Artificial neural networks (ANNs) are widely and effectively used in various fields today, and water level prediction is one of these areas.

Artificial neural networks (ANNs) are systems designed to analyze and process information similar to humans. They combine multiple neurons (cells) according to a specific rule to perform a task and provide solutions to statistical, mathematical, structural, and philosophical problems. ANNs have recently emerged and have gained wide acceptance across various disciplines for solving many real-world problems [1].

In this study, an attempt was made to establish a prediction model by determining the effects of meteorological factors on the water level of Van Lake. Artificial neural networks (ANNs) were used in this model to predict the Van Lake water level using meteorological parameters such as precipitation (P), temperature (T), evaporation (E), wind speed (WS), relative humidity (RH), and actual pressure (AP). Monthly average water level data of Van Lake from 2004 to 2022 obtained from the 17th Regional

*Corresponding author: *yeneraltun@yyu.edu.tr*
Received:19.06.2023 , Accepted:26.11.2023

Directorate of State Hydraulic Works (DSI) and monthly average meteorological parameter data (precipitation, temperature, evaporation, wind speed, relative humidity, and actual pressure) collected from various stations in the Van region from 2004 to 2022 obtained from the General Directorate of Meteorology were used to create a model and predict the impact of these meteorological variables on the lake level using ANNs. After completing the training of the model, it was tested using data from 2022. Additionally, another focus of the study is multiple linear regression analysis. With this analysis method, the relationship between the independent variables (precipitation, temperature, evaporation, wind speed, relative humidity, and actual pressure) and the dependent variable, Van Lake water level changes, was examined, and a model equation was formed. These two methods were compared, and the results obtained were evaluated.

In the literature, there are numerous studies focusing on the prediction of groundwater levels, where various models have been established.

In [2], Çobaner and others applied a novel genetic programming approach called multi-gene genetic programming (MGGP) to predict groundwater levels using meteorological data and historical groundwater levels. The construction of this model involved the use of four years of daily data from the Karacaviran observation well, covering the period from 2007 to 2010, as well as daily meteorological data from the Develi meteorology station. The accuracy of these models was evaluated and compared with conventional rating curve analysis (RCA). The results of the MGGP models were found to be superior to RCA based on four different criteria.

In [3], Yalova Gökçe Dam water levels were predicted for the year 2019 using artificial neural networks (ANNs) based on data from 2000 to 2019, including the flow rate of Selimandıra stream, evaporation and precipitation values, dam water discharge, seepage water quantity, and dam level. The developed models utilized the Levenberg-Marquardt algorithm with a multi-layered ANN function to evaluate monthly datasets. The prediction data showed a close approximation to the actual water level with a determination coefficient of 94.14%. In summary, the average predicted water level in the dam for 2019 was 73.77 m, while the actual average water level in the dam was 72.13 m. These results suggest that the ANN model yielded successful outcomes in predicting the water levels of Yalova Gökçe Dam.

In [4], Ispir conducted a study in the sandy region of Hatay Amik Plain. In this study, monthly measurements of groundwater levels were obtained from an observation well belonging to the General Directorate of State Hydraulic Works (DSI) between 2000 and 2015. Additionally, monthly measurements of total precipitation and average temperature were included to predict groundwater levels. Autoregressive (AR) and Artificial Neural Network (ANN) models were developed in the study, and the performance of groundwater level prediction was evaluated. As a result of the study, it was observed that the ANN model yielded better results in predicting groundwater levels in the region compared to the AR and AR models.

Artificial Neural Networks (ANNs) were used to model the temporal changes in water levels of Lake Van. The backpropagation algorithm was employed to train the developed model. This study aimed to evaluate the dynamic variations in the lake's water level. As a result, the study demonstrated that ANNs can successfully model the complex relationship between precipitation and consecutive water levels [5].

In [6], Panyadee and others modeled the water level fluctuations of Lake Drwęckie using Artificial Neural Network (ANN) techniques. Meteorological data from the Olsztyn meteorological station for the years 1980-2012 were used for modeling purposes. The study concluded that predicting lake water level changes using ANN yielded accurate results.

This study contributes to the literature by developing a statistical model that examines the relationship between meteorological factors and the water level changes in Lake Van. While some studies in the existing literature focus on the physical causes of water level variations in Lake Van, others analyze the lake level using statistical approaches. This study extensively investigates the effects of meteorological factors on the lake water level and provides a model that can predict future changes.

2. Material and Method

In this section of the study, the structure of the Artificial Neural Network (ANN), the functioning of neurons, its components, and the training of the network, as well as the significance of Regression Analysis (RA), Multiple Linear Regression (MLR), the study area, and the dataset are explained to describe the methods used. The modeling techniques employed in the study are also outlined.

2.1. Artificial Neural Networks (ANNs)

Artificial Neural Networks (ANNs) are systems that enable computers to learn from data and generate new information by mimicking the learning process of the human brain. They are based on a mathematical modeling of the biological and cognitive properties of the human brain. ANNs are models created by imitating the biological neural structure of living organisms [7].

The recognition that the computations performed by the human brain are significantly different from those of digital machines has been a driving force behind research on ANNs since their inception [8].

In humans, the process of learning occurs through synaptic connections between neurons. Neurons receive input signals through numerous dendrites. The input received by dendrites can be excitatory or inhibitory. Inputs are integrated into the neuron's cell body. When the combined input exceeds a certain threshold, the neuron transmits an output signal to other cells through its axon. This basic description forms a model for artificial neurons [9].

Table 1. Comparison of biological and artificial neurons

Biological Neuron Structure	Artificial Neuron Structure
Neuron	Input Value
Dendrite	Summation Function
Cell Body	Activation Function
Akson	Output Value
Synapse	Weights

The basic elements of ANNs are artificial neurons inspired by the functioning of biological neurons. These neurons form connections among themselves, which are grouped into layers to construct ANNs. Figure 1. depicts a graphical representation of how the cellular structure of a brain is mathematically modeled [8].

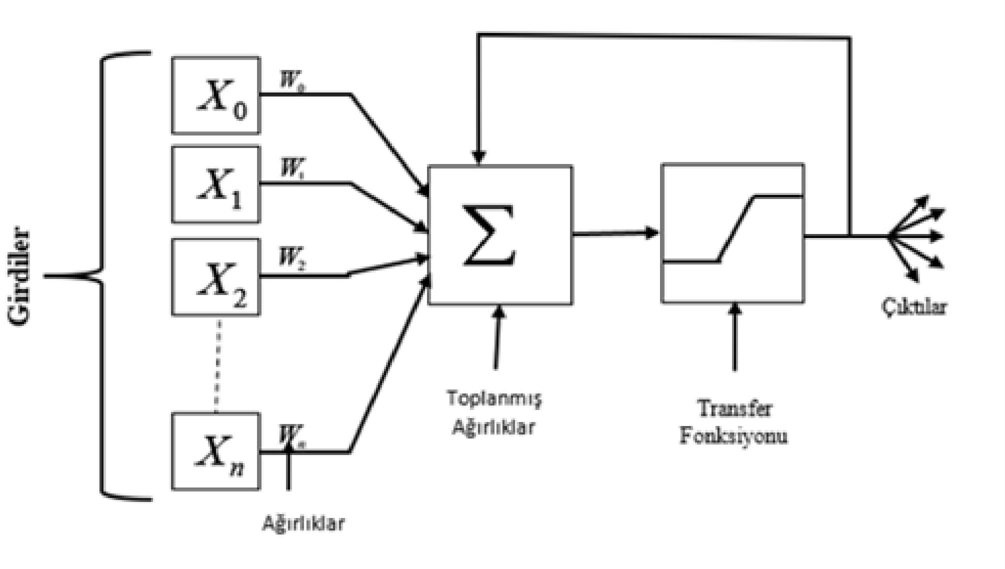

Figure 1. For captions, please use the figure description style.

The information from the external world (x_i) is gathered and fed into the input layer. In other words, the desired learning information for the network is collected at the input layer. For example, data such as precipitation, temperature, wind speed, relative humidity, cloud cover, and atmospheric pressure that affect the water level of Lake Van are gathered as inputs in this layer, and there is a corresponding neuron for each input.

The weight values (w_i) represent the importance and impact of the incoming inputs on the neuron. This effect can be positive or negative [10].

The weights of the presented inputs in the network are not fixed. As the ANN is presented with new examples, it adjusts the weight values to achieve the most suitable result for itself. At the beginning of the training, the weights are assigned randomly. If the output value reaches an acceptable error level at the end of the process, the training of the network is considered complete. Otherwise, the process is repeated and the weights are adjusted again. Minimizing the error between the actual values and the predicted output values of the model is crucial [11].

The Net Input (s) value of the cells in the hidden layer is obtained by combining (summing) the inputs with their corresponding weights. Several methods can be used to calculate this value.

$$s = \sum_{i=1}^{n} x_i * w_i \tag{1}$$

In Equation (1), x_i represents the values of the inputs, and w_i represents their corresponding weights. Each input is multiplied by its randomly assigned weight, from the first input to the last input, and the results are summed to obtain the value of *s*.

After the computation of the activation function, the resulting *s* value is passed through an activation function. The output of the pooling function is transformed into the final output through the activation function. The activation function determines the output value produced by the cell based on the net input value received by the cell. If we consider an artificial neural network without an activation function, the network would resemble a simple linear regression. Different activation functions can be used depending on the problem type and network structure.

The resulting value after the computation of the activation function becomes the output of the cell. This value can serve as input to another cell or be directly used as information. A single cell obtains only one output value. The information that passes to other process elements remains consistent throughout the process. When the process is completed, the artificial neural network has fulfilled its task and generated the required information [11].

To train an ANN and enable it to produce results, several steps need to be taken. These steps are as follows:

- Presenting Data to the Network and Determining the Network Structure
- Determining the Number of Hidden Neurons
- Selecting Activation Functions for Combination
- Determining Stopping Criteria
- Performance (Error) Measures for Models
- Error Reduction

By following these steps, the ANN can be trained and utilized to produce the desired results.

2.2. Regression Analysis (RA)

Regression is the process of expressing the relationship between at least two variables through an equation [12]. Similarly, in Regression Analysis (RA), when variables are considered dependent and independent, their relationship is expressed through an equation.

The term 'regression' was first introduced by Francis Galton. Regression is the first technique that comes to mind when examining the relationship between variables. From a statistical perspective, the relationship between two variables is interpreted as a dependence that exists between the varying values of these variables. If the value of variable '*x*' changes and this change is accompanied by a change in the value of variable '*y*' that is dependent on '*x*', it can be said that there is a relationship between these variables [13]. There are various types of regression analysis found in the literature.

Linear regression is an approach aimed at modeling the relationship between a numerical dependent variable (*y*) and one or more independent variables (*x*) [14].

In regression analysis, instead of analyzing the entire set of data (population), the analysis is performed with samples selected from this data. Due to the large number of data points in the population, statistical analyses are conducted using a specific number of randomly selected data points from the population, taking into account time and research costs. The results obtained with the sample data are then used to predict the relationship in the population [15].

Simple linear regression analysis is used when there is a linear relationship between a single explanatory (independent) variable, denoted as *x*, and a response (dependent) variable, denoted as *y*. It is employed to predict the value of the *y* variable based on the *x* variable.

$$y = \beta_0 + \beta_1 x + \varepsilon \tag{2}$$

The equation (2) represents the equation for building a linear model. In this model, β_0 is the intercept term of the linear function, representing the point where the regression line intersects the vertical axis when $X = 0$. β_1, on the other hand, describes the slope of this function and is also known as the regression coefficient, indicating how much change in the dependent variable is associated with a one-unit change in the independent variable *x* in regression analysis. The error term ε is represented by ε [16].

When the dependent variable consists of two-choice categorical data, logistic regression analysis can be used to reveal the cause and effect relationship between the independent variables and the dependent variable [17]. The relationship between a dependent variable and independent variables is analyzed using simple regression analysis. However, in some analyses, multiple independent variables can influence a dependent variable. In cases where multiple independent variables affect a dependent variable, a multiple regression model should be constructed [18]. Multinomial models in which one independent variable takes more than two discrete variables often deal with more complex models, such as multivariate models with many independent variables [19]. Multiple linear regression has two general purposes. One is to predict the value of the dependent variable through variables identified to affect the dependent variable. The other is to determine which independent variable or variables, believed to influence the dependent variable, have a greater impact on it and describe the relationship between them.

$$Y = b_0 + b_1 X_{i1} + b_2 x_{i2} + \ldots\ldots + b_n x_{in} + e_i \tag{3}$$

$$Y = b_0 + \sum_{k=1}^{p} b_k X_{ik} + e_i \tag{4}$$

In equations (3) and (4), the estimation of the dependent variable is calculated in multiple linear regression analysis. Here, $b_0, b_1, \ldots, b_n$ represent the regression coefficients [20].

2.3. Research Field

Van Lake, located in the lowest part of the closed basin, is a soda lake with a salinity rate of 2.6%. It has a surface area of 3,626 km², a volume of 607 km³, a drainage area of 12,470 km², a pH value of 9.7, and a maximum depth of 451 meters, making it the largest body of water in Turkey. Van Lake is situated within the boundaries of Van and Bitlis provinces. The borders of Van Lake are indicated in Figure 2.

Figure 2. Map of Lake Van

As Lake Van is a closed basin, the water level of the lake usually varies according to climate and seasons. Observable changes in the lake's water level are noticeable fluctuations that occur at different times, resulting in both rising and falling levels. It can be said that such changes are largely influenced by factors such as rainfall, temperature, and evaporation control [21].

2.4. Data Set

The data used in the study were obtained from various measurement stations belonging to the Meteorology 14th Regional Directorate and the General Directorate of State Hydraulic Works (DSI) 17th Regional Directorate, located within the boundaries of Van province.

Table 2. Comparison of biological and artificial neurons

Parameter	Unit	Type of Use	Institution
Rainfall	mm=kg÷m²	Independent variable	General Directorate of Meteorology
Temperature	ºC	Independent variable	General Directorate of Meteorology
Evaporation	mm	Independent variable	General Directorate of Meteorology
Wind Speed	m/sn	Independent variable	General Directorate of Meteorology
Relative Humidity	%	Independent variable	General Directorate of Meteorology

In Table 2, the parameters of rainfall, temperature, evaporation, wind speed, relative humidity, atmospheric pressure, and Lake Van water level, along with their respective units of measurement (mm for millimeters, °C for degrees Celsius, m/s for meters per second, % for percentage, hPa for hectopascals, m for meters), their purpose in the model, and their sources of acquisition are specified. These parameters were measured using stations belonging to the institutions obtained between the years 2004 and 2022, and they cover monthly data.

2.5. Data Preparation

Model development, training, comparison of obtained output values with the defined target values, and evaluation of results primarily started with the collection of input and output values from past records. The data consists of monthly measurements taken from various measurement stations belonging to the General Directorate of Meteorology in the 14th Region and the General Directorate of State Hydraulic Works in the 17th Region, covering the years 2004-2022 within the boundaries of Van province.

In Table 3, the results of the normal distribution tests for these data are presented. The Kolmogorov-Smirnov and Shapiro-Wilk tests have been applied. According to the results of these tests, since the p-value is less than 0.05, the water level variable does not follow a normal distribution.

Table 3. Tests of normality

	Kolmogorov-Smirnov	Shapiro-Wilk
Water Level	0.001	0.004

$$P_{(avg.parameter)} = \frac{1}{n}\sum_{i=1}^{n} P_{(parameter)} \tag{5}$$

The values measured by multiple stations have been transformed into a single value as shown in Equation (5).

Here, P(parameter) represents the monthly measured values of the respective parameter, n represents the total number of stations where this parameter is measured, and $P_{(avg.parameter)}$ represents the monthly average value of this parameter. After this process, a total of 225 data points covering the years 2004-2022 were obtained. Among them, 216 data points from 2004 to 2021 were used for model development, while 9 data points from 2022 were separated to test the established models, forming two distinct groups.

Lastly, before constructing the models, the data was normalized using the adjusted minimum-maximum method to enhance the efficiency of the models, as shown in Equation (6). This normalization process ensured that the data set takes on a dimensionless form within the range of 0.1 to 0.9, aiming to improve the effectiveness of the models.

$$X_i = 0.8\frac{(X - X_{min})}{(X_{max} - X_{\min}\ )} + 0.1 \tag{6}$$

Here, X_i represents the value to be normalized, X represents the measured value, X_{min} represents the minimum value for X, and X_{max} represents the maximum value for X. The equation is used to normalize X_i within the range of 0.1 to 0.9 using the minimum and maximum values of X.

Table 4. Descriptive statistics of the data

	Mean	Median	Std. Deviation	Minimum	Maximum
Rainfall	0.2977	0.2812	0.1416	0.1000	0.9000
Temperature	0.5355	0.5510	0.2381	0.1000	0.9000
Evaporation	0.3750	0.2938	0.2436	0.1000	0.9000
Wind Speed	0.4306	0.4114	0.1448	0.1000	0.9000
Relative Humidity	0.5050	0.5227	0.1903	0.1000	0.9000
Actual Pressure	0.4932	0.4856	0.1516	0.1000	0.9000
Water Level	0.4694	0.5339	0.2386	0.1000	0.9000

In Table 4, descriptive statistics such as mean, median, standard error, and other measures have been provided for the normalized data.

3. Results and Discussion

3.1. Results of the ANN Model

In order to develop and train an ANN model in the MATLAB environment, the data needs to be classified into training, testing, and validation sets.

This classification process is generally done randomly. Out of the total 216 data points allocated for training the ANN model, 70% are assigned for learning, 15% for testing, and 15% for validation purposes.

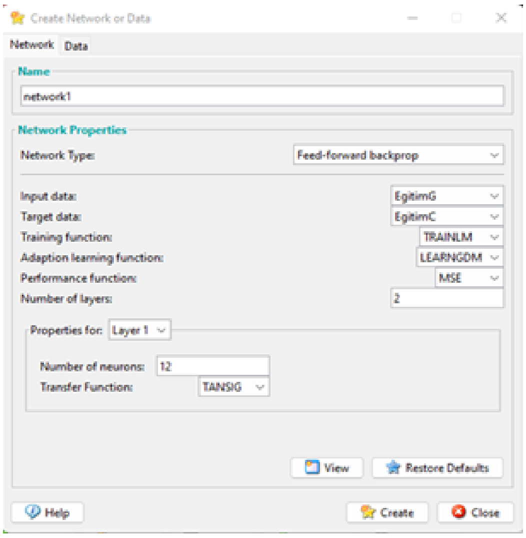

Figure 3. The characteristics of an Artificial Neural Network (ANN) structure

Figure 3 illustrates the characteristics of the network structure. Due to the nonlinear relationship between the data, a feedforward backpropagation multilayer network, which is widely used, has been preferred. In the weight determination for prediction, the "TANSIG" function, which provided the most successful results through trial and error, was selected as the activation function. Various training functions available in Matlab were tested, and it was observed that the network learned best using the "TRAINLM (Levenberg-Marquardt)" function, which was consequently chosen.

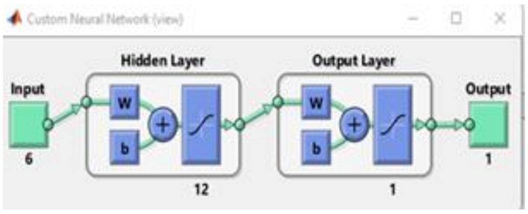

Figure 4. The structure of the ANN

The network architecture is depicted in Figure 4. It consists of 6 input data points and 1 output data point. The number of layers and hidden cells is determined as 2 and 12, respectively. These values were chosen through trial and error as they yielded the most successful results.

The network achieved a low error value of 0.0131, which indicates a high level of accuracy. The fact that this value is close to zero implies that the network has performed well. It took approximately 11 seconds for the network to reach its best result and complete the training process. The performance graph in Figure 5 illustrates how the training, validation, and testing data evolved during each iteration. According to this result, the completed model achieved its best performance at iteration 4.674.

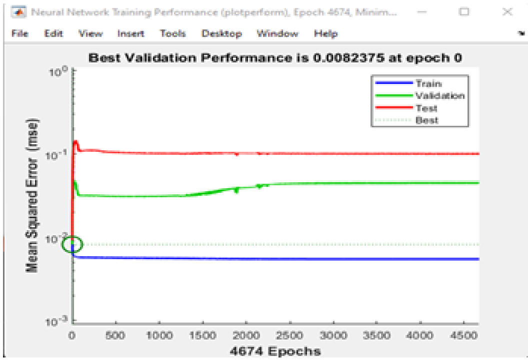

Figure 5. Performance Graph after YSA Training

Figure 6 illustrates the relationship between the input values entered into the model and the predictions made by the model. In this context, the regression (R) values for training, validation, and testing, as well as their overall value, are observed to be 0.88. Since these values are close to 1, the learning process has been successfully carried out.

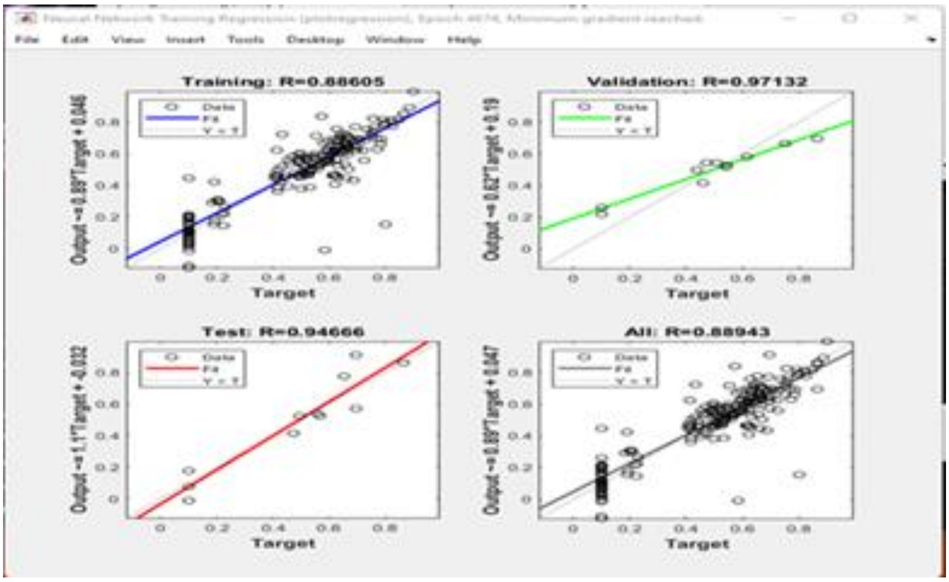

Figure 6. YSA training regression results

The water level prediction using YSA was achieved with an accuracy of 99%. The Mean Absolute Percentage Error (MAPE) was found to be very close to 0, specifically 0.34, indicating a 0.34% margin of error. The Mean Absolute Deviation (MAD) value was found to be 0.05. Having MAPE and MAD values close to 0 in predictions signifies that the network provides more realistic results.

Table 5 presents real data that has not been introduced to the model before. These data were used as input for testing the generated model. Table 6 shows the actual and predicted water levels for these data, along with the deviations between them. It is observed that the predicted values by the YSA model are very close to the actual values. For example, if the actual water level measured during a certain period is 1648.59, the YSA model predicts it as 1648.56 based on the meteorological parameters of that period. The difference between this predicted value and the actual water level is determined to be 0.0297.

Table 5. Test data

AB (hPa)	B (mm)	NN (%)	RH (m÷sn)	S (ºC)	D (mm=kg÷m²)
819.62	8.90	78.44	1.83	-6.51	61.68
821.83	14.60	79.65	1.57	-3.02	22.87
818.33	38.15	77.17	2.43	-1.65	61.29
821.54	24.80	58.82	2.39	8.39	24.80
821.22	115.30	66.83	2.37	10.50	47.77
820.56	189.30	45.02	2.33	18.51	12.17
819.34	255.20	35.76	2.30	22.11	1.34
820.90	298.70	30.34	2.21	23.26	0.26
822.86	230.20	35.15	2.11	18.46	0.64

Table 6. Actual and predicted water level values

Actual Water Level (m)	Predicted Water Level (m)	|Actual - Predicted| (m)
1648.49	1648.41	0.0713
1648.49	1648.42	0.0721
1648.52	1648.46	0.0650
1648.58	1648.57	0.0120
1648.67	1648.59	0.0733
1648.66	1648.53	0.1341
1648.59	1648.56	0.0276
1648.59	1648.56	0.0297
1648.33	1648.36	0.0284

3.2. Results of the MRA Model

The 216 data points used in the training of the ANN model were analyzed using Multiple Linear Regression techniques. The independent variables chosen were precipitation, temperature, evaporation, wind speed, relative humidity, and actual pressure, while the dependent variable was the lake water level.

Table 7. Regression analysis results

Regression Statistics	
R	0.50
R^2	0.25
R-squared (Coefficient of determination)	0.23
Standard deviation	0.21
Number of observations	216

The summary of the model created using regression analysis is presented in Table 7. The R-value is observed as 0.50. The proximity of this value to 1 indicates the strength of the relationship between the variables. The coefficient of determination (R-squared) is calculated as 0.23, meaning that the independent variables can explain 23% of the variation in the dependent variable. This value is relatively low, indicating that the model's fit to the data is considered inadequate.

Table 8 presents the information regarding the p-values. The variables of actual pressure, relative humidity, wind speed, temperature and precipitation have p-values smaller than 0.05, indicating that they significantly influence the water level.

Table 8. The analysis results of the regression analysis

	p-Values
Intercept	0.000
Actual Pressure (hPa)	0.000
Evaporation (mm)	0.246
Relative Humidity (%)	0.002
Wind Speed (m/s)	0.000
Temperature (°C)	0.021
Precipitation (mm=kg/m²)	0.028

Table 9. The weight coefficients of the regression analysis

	Coefficients
Intercept Coefficient (Q)	1.185
AB	-0.631
B	-0.205
NN	-0.575
RH	-0.438
S	0.104
Y	0.321

The coefficients of the variables obtained from the regression analysis are shown in Table 9. These calculated coefficients indicate how much the respective parameters affect the water level.

$$Y = Q + (-0.631 * AB) + (-0.205 * B) + (-0.575 * NN) + (-0.438 * RH) + (0.104 * S) + (0.321 * Y) \quad (7)$$

Table 10. The prediction results of the regression analysis

Actual Water Level (m)	Predicted Water Level with CRA (m) |	|AWL- PWL|(m)
1648.49	1648.07	0.4168
1648.49	1647.98	0.5093
1648.52	1648.00	0.5157
1648.58	1647.95	0.6343
1648.67	1647.95	0.7183
1648.66	1647.99	0.6721
1648.59	1648.03	0.5577
1648.59	1648.01	0.5712
1648.33	1647.97	0.3603

The CRA equation is formulated as Equation (7). In this equation, the effects of independent variables on the dependent variable are determined. For example, a unit change in the temperature variable will result in a 0.104 change in the water level. By substituting the parameter values into the equation, the water level value is calculated. Using the test data in Table 5, the water level is predicted using this equation. Table 10 shows the actual and predicted water levels for these data, along with the deviations. For instance, if the measured water level in a certain period is 1648.49, the CRA model predicts it as 1648.07 based on the meteorological parameters for that period. The difference between the predicted and actual water level values is determined as 0.4168.

3.3. Comparison of ANN and MRA Models

In this study, two different models, ANN and MRA, were developed using the same dataset of 216 samples. Table 11 presents the performance results of the two models. Upon completion of model development, the R-squared values for ANN and MRA were found to be 0.88 and 0.50, respectively. The MAPE values were 0.34 for ANN and 3.34 for MRA, while the MAD values were 0.57 for ANN and 0.10 for MRA. Based on these results, the error rate of the ANN model was calculated as 0.34%, while the error rate of the MRA model was 3.34%. The mean absolute deviation (MAD) was found to be 0.05 for ANN and 0.10 for MRA. According to the MAD results, it was observed that the ANN model provided predictions closer to the actual water levels compared to the MRA model. The R-squared results of both models were compared to measure the level of relationship between variables. The comparison results indicated that the ANN model had a significantly higher level of relationship between variables compared to the MRA model. This value is expected to be close to 1 at all times.

Table 11. The performance results of the YSA and MRA models

Performance	ANN	MRA
R	0.88	0.50
MAPE (%)	0.34	3.34
MAD	0.05	0.10

The developed models were tested by predicting the water level using 9 identical datasets. The predicted water level values obtained from the test results were compared with the actual water level values.

Table 12. YSA and MRA Comparison

Actual Water	Level ANN Water	Level MRA Water	|AW-ANN|	|AW-MRA|
1648.49	1648.41	1648.07	0.0713	0.4168
1648.49	1648.42	1647.98	0.0721	0.5093
1648.52	1648.46	1648.00	0.0650	0.5157
1648.58	1648.57	1647.95	0.0120	0.6343
1648.67	1648.59	1647.95	0.0733	0.7183
1648.66	1648.53	1647.99	0.1341	0.6721
1648.59	1648.56	1648.03	0.0276	0.5577
1648.59	1648.56	1648.01	0.0297	0.5712
1648.33	1648.36	1647.97	0.0284	0.3603

Table 12 displays the actual water level values, the predicted values from both models, and the absolute differences between the actual and predicted values. For example, for a specific period, if the actual water level is 1648.49, the YSA model predicts it as 1648.41, and the MRA model predicts it as 1648.07. The absolute difference between the YSA prediction and the actual value is 0.0713, while the absolute difference between the MRA prediction and the actual value is 0.4168. When comparing these two differences, it can be observed that the YSA model provides predictions that are closer to the actual values compared to the MRA model.

4. Conclusion and Suggestions

One of the sciences developed to study phenomena in nature and solve existing or potential problems is the science of data analysis. The science of data analysis aims to explain the subject with a certain probability and a limited number of observations using methods and theories appropriate to the data structure and to shed light on future research. In nature, there is a lot of data in every field of science, but it is important to know how to use these data and produce new information from them by analyzing them [22]. In this study, the relationship between changes in lake water level and some meteorological parameters was analyzed using ANN and MRA models, which are machine learning methods.

The results of the study indicate a close relationship between changes in the lake water level and meteorological parameters. The analyses conducted have shown that meteorological factors such as precipitation, temperature, evaporation, wind speed, relative humidity, and atmospheric pressure play a significant role in explaining the variations in the lake water level. The artificial neural network model used in this study has yielded highly accurate results in predicting the lake water level. With its ability to model complex relationships, the artificial neural network has achieved high levels of accuracy in predicting the lake water level. Additionally, it has been observed that artificial neural networks can be applied not only for temporal but also for spatial predictions of the lake water level. Furthermore, multiple linear regression analysis has been employed in the study to examine the relationship between independent variables and changes in the lake water level. This analysis has provided a more detailed understanding of the effects of meteorological parameters on the lake water level. However, it has been observed that the predictions obtained through the artificial neural network model offer more realistic results compared to multiple linear regression analysis.

The findings of this study demonstrate that changes in the water level of Lake Van are associated with meteorological factors and can be accurately predicted. These results will contribute to a better understanding of lake water level variations for decision-makers and researchers in areas such as sustainable water resource management, environmental planning, and climate change. They will also assist in forecasting future scenarios related to lake water level changes.

In conclusion, this study examined the relationship between the lake's water level and meteorological parameters, and this relationship was evaluated through specific statistical analyses. Statistical power can be assessed based on the results of these analyses. Specifically, it has been demonstrated how accurate the artificial neural network model's predictions are and how it outperforms multiple linear regression analysis. This can be a significant factor in demonstrating the statistical strength of the study.

Contributions of the authors

The authors' contributions to the study are equal.

Conflict of Interest Statement

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics

This study was conducted with the aim of examining and predicting the changes in the water level of Lake Van. The main objective of the study was to determine the effects of meteorological factors on the lake water level and to create a prediction model using these effects.

References

[1] M. F. Keskenler and E. F. Keskenler, "Geçmişten Günümüze Yapay Sinir Ağları," *Takvimi Vakiye.*, vol. 5, no. 2, pp. 8–18, 2017.

[2] M. Çobaner, E. Babayiğit, and B. Babayiğit, "Meteorolojik Veriler Kullanılarak Yeraltı Su Seviyesinin Genetik Programlama ile Tahmini," *Ömer Halisdemir Üniv. Mühendis. Bilim. Derg.*, vol. 5, no. 2, pp. 177–187, 2016.

[3] Y. Damla, T. Temiz, and E. Keskin, "Yapay Sinir Ağı Kullanılarak Su Seviyesinin Tahmin Edilmesi: Yalova Gökçe Barajı Örneği," *Kırklareli Üniversitesi Mühendislik ve Fen Bilimleri Dergisi*, vol. 6, no. 1, pp. 32–49, 2020.

[4] E. İspir, "Hatay Amik Ovası Kumlu Bölgesindeki Yeraltı Suyu Seviyesinin Yapay Sinir Ağları Yöntemi ile Tahmini," Iskenderun Technical University, Turkey, 2017.

[5] A. Altunkaynak and Z. Şen, "Fuzzy logic model of lake water level fluctuations in Lake Van, Turkey," *Theor. Appl. Climatol.*, vol. 90, no. 3–4, pp. 227–233, 2007.

[6] P. Panyadee, P. Champrasert, and C. Aryupong, "Water level prediction using artificial neural network with particle swarm optimization model," in *2017 5th International Conference on Information and Communication Technology (IcoIC7)*, 2017.

[7] V. Nabiyev and A. K. Erümit, Eds., *Eğitimde Yapay Zeka Kuramdan Uygulamaya*. Ankara Pegem Akademi Yayıncılık, 2020.

[8] S. Haykin, *Neural networks: Comprehensive foundation*, 2nd ed. Piscataway, NJ: I.E.E.E. Press, 1999.

[9] E. Ersoy and Ö. Kartal, "Yapay Sinir Ağları ve İnsan Beyni," *İnsan ve Toplum Bilimleri Araştırmaları Dergisi II*, vol. 1, no. 2, pp. 188–205, 2012.

[10] M. Caner and E. Akarslan, "Mermer Kesme İşleminde Spesifik Enerji Faktörünün ANFIS ve YSA Yöntemleri ile Tahmini," *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, vol. 15, no. 2, pp. 221–226, 2009.

[11] F. Sönmez Çakır, *Yapay Sinir Ağları Matlab Kodları ve Matlab Toolbox Çözümleri*. Turkey: Nobel Akademik Yayıncılık, 2020.

[12] B. Çil, *İstatistik*. Ankara: Detay Yayıncılık, 2014.

[13] B. A. Köksal, *İstatistik Analiz Metotları*. Istanbul: Çağlayan Kitapevi, 2013.

[14] S. Kılıç, "Doğrusal Regresyon Analizi," *İstatistiki İfadeyle*, vol. 3, no. 2, pp. 90–92, 2013.

[15] F. Yüzük, "Çoklu Regresyon Analizi ve Yapay Sinir Ağları ile Türkiye Enerji Talep Tahmini," Cumhuriyet University Institute of Social Sciences, Turkey, 2019.

[16] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning: With applications in R*, 1st ed. New York, NY: Springer, 2013.

[17] H. Eygü, "Covid-19 Sürecinde Üniversite Öğrencilerinin Uzaktan Eğitim Eğiliminin Araştırılması," *Mehmet Akif Ersoy Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, vol. 10, no. 2, pp. 882-903, 2023.

[18] C. Okutkan, "Borsa İstanbul Şirketlerinin Hisse Senedi Getirilerinin Yapay Sinir Ağları ve Çoklu Regresyon Yöntemleri Kullanarak Analizi," Kocaeli University Graduate School of Natural and Applied Sciences, Turkey, 2014.

[19] H. Eygü, *Çözümlü güncel örneklerle olasılık ve istatistik*. Ankara: Nobel Yayınları, 2020.

[20] M. Öztürkcan, *Regresyon analizi*. 2009.

[21] M. Z. Yıldız, "Kapalı Havza Göllerinde Seviye Değişimlerinin Kıyı Yerleşmelerine Etkisi: Van Gölü Örneği," *Fırat Üniversitesi Sosyal Bilimler Dergisi*, vol. 15, no. 1, pp. 15–31, 2005.

[22] G. A. A. Mulla, Y. Demir, and M. M. Hassan, "Combination of PCA with SMOTE Oversampling for Classification of High-Dimensional Imbalanced Data," *BEU Journal of Science*, vol. 10, no. 3, pp. 858-869, 2021.