Journal of AI

Volume: 7, **Issue No:** 1, January-December 2023, **Pages:** 15-31

Precipitation Forecast with Artificial Neural Networks Method

Serkan ANSAY [1], Bayram KÖSE [2]

[1] İzmir Bakırçay University, Faculty of Engineering-Architecture, İzmir, Türkiye.
Email: bayram.kose@bakircay.edu.tr, Orcid: 0000-0002-3368-3886
[2] İzmir Bakırçay University, Faculty of Engineering-Architecture, İzmir, Türkiye.
Email: serkanansay@gmail.com, Orcid: 0000-0003-0256-5921

Recieved: 07.06.2023 *Accepted: 05.08.2023* *Online: 08.08.2023* *Published: 31.12.2023*

Research Article

Abstract

Events in the atmosphere from past to present – wind, precipitation, humidity, temperature – have almost always been the subject of research to create a forecast in regions. The rapid development of the technological field in terms of software and hardware brings methods and techniques to be used in research. One of them is Artificial Neural Networks. In this study, precipitation data were estimated using the Feed Forward Backpropagation method of Artificial Neural Networks method using past data of meteorological parameters, and they were compared with the data of multiple linear regression analysis. Based on these models, six different models were studied, and regression and performance evaluations were made. While the error average of multiple linear regression is 0.2413, this value is 0.076 in artificial neural networks, and the correlation average for both is 0.90. As a result of this study, the best model has a coefficient of determination of 0.95 and an error value of 0.18 in multiple linear regression, as well as a coefficient of certainty of 0.99 and an error value of 0.0438 in artificial neural networks; It has been understood that the 1st model, which has 6 data sets as the input layer, exhibits the best performance.

Keywords: *Regression, artificial neural networks, precipitation forecast*

Cite this paper (APA)

Ansay, S., Köse, B. (2023). Precipitation Forecast with Artificial Neural Networks Method. Journal of AI. 7(1), 15-31.

Precipitation Forecast with Artificial Neural Networks Method

Highlights

- Multiple linear regression comparison with artificial neural network
- Precipitation forecasting models have been created
- No similar study was found in the research field.
- The best precipitation prediction model obtained in the study will contribute to the studies in the similar field.

Graphical Abstract

Among the parameters of the data obtained from the meteorology directorate, models were created with the best correlation, and the models were analyzed with ANN and MLR. All models were compared among themselves, and ANN and MLR were compared with each other. The flow of the study is given in the figure, where h represents humidity, t represents temperature, w represents wind speed and p represents precipitation.

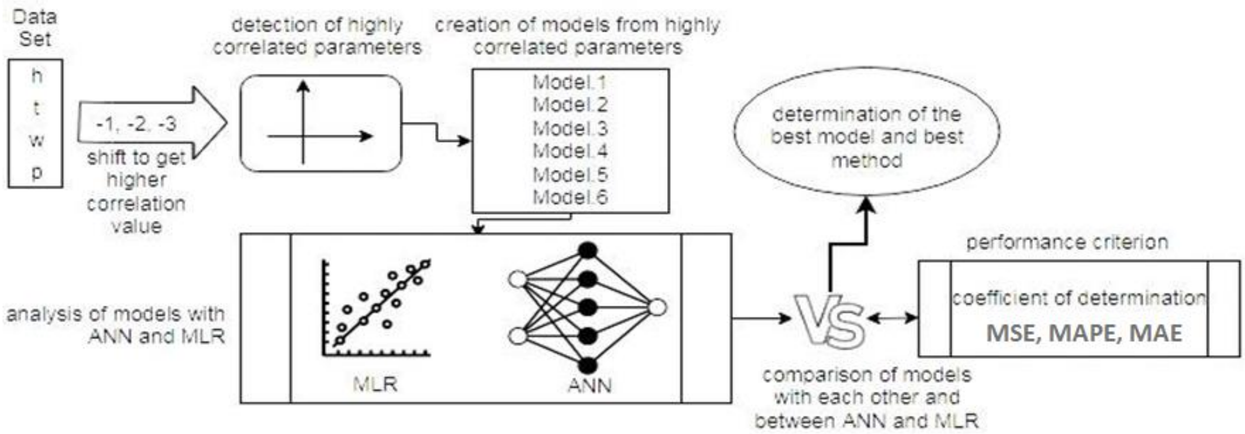

Figure. Flow of work

Aim

The aim of the study is to contribute to those who work in this field by creating a high-accuracy precipitation forecast model.

Design & Methodology

Models were created with high correlation of parameters and models were analyzed with ANN and MLR methods. The performance of the models was measured by the coefficient of determination, MSE, MAPE and MAE.

Originality

It is the first of its kind in terms of the region where the study was conducted in the literature review.

Findings

In this study, it was found that ANN is better than MLR in precipitation forecasting.

Conclusion

The results of artificial neural networks working with feedforward backpropagation algorithm gave more successful results than multiple linear regression analyzes in all models.

It is seen that with the best model obtained, it can be beneficial in many areas that require precipitation forecasting.

Declaration of Ethical Standards

The author(s) of this article declare that the materials and methods used in this study do not require ethical committee permission and/or legal-special permission.

1. INTRODUCTION

Water is undoubtedly one of the basic needs of life for the continuation of life. Sufficient water reserves and keeping them under control are essential for the sustainability of the ecological balance. It is inevitable that some parameters must consist of predictions with high accuracy in strategic decisions to be taken to provide and maintain this competence within the scope of hydrological modeling. The way to reach a faster and more accurate result with fewer data will be opened by reaching the prediction model that will be formed by mathematical processing of parameters such as wind, precipitation, humidity, and temperature for several years [1].

Researches and forecasts in the field of meteorology provide benefits in various subjects. With the instant meteorological data of regions, cities, and villages, forecasts of meteorological parameters are made. The precipitation parameter has a vital role in these estimates. The aviation industry benefits from precipitation forecasting in many areas such as the economy, early warning and preparation systems of municipalities, the realization of sports competitions, the production and planning of energy companies, the protection of life and property, the production capacity of hydroelectric power plants, the occupancy rate of dams that provide mains water [2]. In this study, models were created from parameters with high correlation. Prediction models were created by passing these models through Artificial Neural Network and Multiple Linear Regression processes, the coefficient of determination and error rates of the models obtained by both methods were compared, and the best model was tried to be found.

There are various studies on estimation with regression. In a 2016 study, it was concluded that both linear and functional regression outperform professional weather forecasting services, and the performance will increase with the evaluated time [9]. In a study in 2019, precipitation forecasting was discussed with a regression model. It has been tried to prove that the forecast to be made is highly reliable by making a precipitation forecast based on previous records of a geographical area. It has been seen that the model's performance has higher accuracy compared to traditional precipitation forecasting systems [10]. In a study from 2017, linear regression was applied to the data set, and the coefficients were used to predict precipitation based on the corresponding values of the parameters. The advantage of the model is that it predicts precipitation based on the previous correlation between different atmospheric parameters [11]. In a study of 2020, Multiple Linear Regression was applied with various parameters to predict precipitation, and it was seen that this method gave successful results [12].

Various studies have been made with Artificial Neural Networks (ANN) and are continuing. In a study conducted in 2020, it was understood that landslides triggered by precipitation would shed light on future landslides due to the application of ANNs of past data. It has been seen that it provides benefits in this area [13]. A study in 2017 was carried out in an area with heavy flooding, and the ANN method was used. Due to the nonlinear relationships in the precipitation data, it has been revealed that ANN is a preferable approach compared to all existing approaches [14]. In a study of 2015, based on India's dependence on agriculture, research was conducted on various ANN architectures by referring to the importance of precipitation. The study shows that overall, ANN performs well for annual precipitation forecasting [15]. In a study of 2020, the feasibility and success of the ANN approach are demonstrated by developing efficient and reliable nonlinear forecasting models for meteorological analysis [16].

When the previous studies on the subject and the method are examined, studies comparing Multiple Linear Regression (MLR) and ANN are encountered. In a study, models were created with a feedforward backpropagation neural network to find missing current data from previous years. It was understood that the models prepared by using the monthly flow data of the same years gave successful results, and it was seen that the feedforward backpropagation artificial neural network could play an active role in the detection of missing flow data [3]. In a different study, artificial neural networks and multiple linear regression analysis methods were used to estimate the missing data in 4018-day data obtained from a meteorology station with temperature, relative humidity, and temperature parameters. As a result of the study, it was seen that artificial neural networks gave more successful results [1]. In the studies conducted for the province of Şanlıurfa, 16 models were created using the artificial neural network method to estimate the previous index values of the Standardized Precipitation Index values. In this study, which consists of monthly data of 78 years, the 16th model gave more successful results than other models in terms of Mean Squared Error (0.12) and Mean Absolute Error (0.23) values formed in terms of post-test with Artificial Neural Networks method after formulating the Standardized Precipitation Index [4]. In a study in 2018, the daily evaporation amount in the Cambridge basin was estimated using Hargreaves-Samani, Ritchie, and Turc methods and Artificial Neural Networks methods. As a result of the studies with 1081 daily data taken from the USA Massachusetts station, it has been understood that artificial neural networks can succeed in similar models [5]. In the studies carried out in 2019, the artificial neural network method was used for the monthly total precipitation data of 56 years and the estimation of the drought. It is stated that the neural network model is highly flexible and can be used as a very powerful tool for accurate simulation prediction [6]. In his thesis in 2019, Sezer performed the load estimation in Zonguldak province with the methods of Regression, Backpropagation Artificial Neural Networks, and Radial Based Artificial Neural Networks. A total of 8 different models were created and analyzed using the same independent variables in all developed models. The best model was found with the model's R^2 (determination) value [7]. In a study on air pollution in 2020, regression and artificial neural network models were created. According to the results of the study, it was understood that artificial neural networks are more efficient than regression analysis [8].

Looking at the studies carried out in this field by scanning the literature, no study was found for the province of Aydın. Aydın is a city with low annual average precipitation, intensive agricultural activities, and steam energy is used due to its geological location, and a need to inform the public, prevent possible flooding, and take precautions according to the occupancy rates of the dams, as in every province. For this reason, it is crucial to make accurate and fast precipitation forecasts. In this study, Artificial Neural Networks Method was used to achieve this. The best model was found among the models formed from the hourly data of various parameters obtained from Aydın Meteorology Directorate with high correlation.

2. MATERIAL and METHOD

2.1. Feedforward Backpropagation Neural Networks

The learning process takes place through neural networks in the human brain. The neural networks in the human brain consist of approximately 6 billion neurons and around 60 trillion connections called synapses responsible for interneuron communication. The action of each neuron is shaped by the signals from other neurons through the synapse. The learning process to understand the working principle of the human brain has been investigated, and studies have been made to convert it into a numerical model and are continuing. Various artificial neuron and network models have been developed in these studies to translate the way neurons work into a model. Artificial Neural Networks (ANNs), as a branch of science different from the computation method of computers, are expressed as complex information processing structures formed by the combination of parts that are related to each other by weight, the method of the human brain to perform a task[17]. Figure 1 shows the working logic of the artificial neural network. The input layer is the layer where the data that needs to be learned into a network is defined. The hidden layer is located between the input and output layers, and the number of layers and neurons varies according to the problem. Backward error propagation occurs at the same time as forward calculations are made in the hidden layer. The output layer is the layer where the sample data to be learned in the ANN is calculated as output [18].

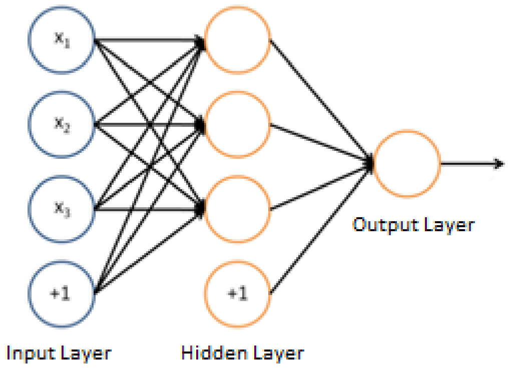

Figure 1. Artificial neural network

Studies on ANN started in the first half of the 20th century and have been used in many fields until today. When we look at the development process, it is compatible with many models and algorithms [19]. Single-layer models have left their place to multi-layer models over time. With the emergence of the backpropagation neural network, research has moved to the development phase of more active and fast learning algorithms. With the support of processes, new methods, and network types, ANN is a model in which applications are made in many areas such as modeling, learning, simulation, identification, and estimation [20, 21].

Each method of the ANN model aims to determine the network that can be best analyzed with the data at hand. In this study, for the nonlinear parameter estimation, Feed Forward Back Propagation neural network and functions called the activation function of the intercellular weight value inputs and the sigmoid activation function, which work efficiently at the point of transferring to the output layer, are used [22].

Feed Forward Backpropagation neural network consists of input, output, and at least one hidden layer. While the input layer has the workspace data count, the output layer has the same data count. The number of hidden layers; is determined by trial and error, considering various parameters (such as the number of data and correlation of data). Likewise, a trial and error process determines the number of hidden layer neurons. During the training process of the network, forward scanning is performed along the network, and the output of the consecutive node is calculated and transferred to the output layer. The desired values are compared with the values transferred to the output layer. In order to reduce the error rate, the network returns to the beginning, arranges the weight values between the layer elements, and repeats the process. The artificial neural network aims to minimize the error between the output layer weight values and the desired result values [1].

To minimize the errors, the weight values between the mesh cells should have the optimum value. As shown in Figure 2, the weighted input data obtained with the input data are summed with the transfer function. The network value, called net in ANN, is processed in the activation function, and in this way, the activation process is completed, and the data is transferred to the output layer.

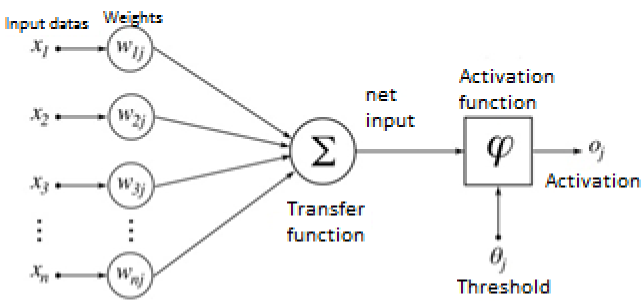

Figure 2. *Mathematical structure of artificial neural network [23]*

$$z_i = \sum_{i=1}^{n} (w_{ij} x_i + b_j) \tag{1}$$

Net inputs to the cell in the transfer function are calculated using Equation (1) formula. n is the number of entries; w weights; x represents the inputs, and b represents the threshold (bias) value. Since the activation function provides curvilinear matching between input and output, it significantly affects the network's performance. The most used activation function is the sigmoid function because it keeps the error at a minimum, and its equation is included in Equation (2).

$$f(z_i) = \frac{1}{1 + e^{-zi}} \tag{2}$$

The value obtained as a result of the sum function in Equation (1) is passed through a linear or nonlinear differentiable activation function, as seen in Figure 2, and a result value is obtained as shown in Equation (3) [24].

$$y = f(z_i) = f\left(\sum_{i=1}^{n} (w_{ij} x_i + b_j)\right) \tag{3}$$

2.2. Multiple Linear Regression

Regression is known to obtain a mathematical model with the correlation between two variables. Its general expression is given by Equation (4).

$$y = a + bx \tag{4}$$

In Equation (4), x represents the value of the selected independent variable, y: the y value for the chosen x value, a: the value of the point where the line intersects the y-axis, b: the slope of the line, a and b: the regression coefficients [22]. The regression graph containing this information is shown in Figure 3. This model is also known as the simple linear regression model.

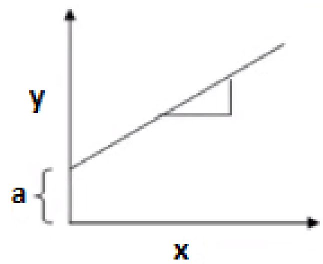

Figure 3. Linear regression graph

As shown in Figure 3, if the independent variable x is more than one, the y value can be calculated with MLR. The related regression equation is shown in Equation (5).

$$Y_i = a + a_1 X_1 + a_1 X_1 + \cdots + a_n X_n + e_i \tag{5}$$

In equation (5), a represents the constant coefficient, a_1,\ a_2,\ \ldots,a_n regression coefficients, a_i coefficients the effects of changing the value of the independent variable Y_i, and e_i the error term [25]. In Equation (5), p is the output data. Input data is obtained by shifting the data of t (temperature), humidity (h), p (precipitation) and w (wind) parameters up to 3 hours.

2.3. Performance Evaluation

In the study, performance is evaluated in 3 different criteria.

2.3.1. Correlation analysis

Correlation indicates the direction and strength of the linear relationship between two random variables in probability theory and statistics. In general, statistical use, correlation indicates how far away from independence has been achieved. The relationship of a variable with two or more variables is calculated through multiple correlations; One of these variables is fixed, and partial correlation techniques calculate its relationship with other variables. The correlation coefficient takes values between -1 and +1. If it is -1, there is a wholly negative linear relationship. If it is +1, there is a wholly positive linear relationship. If 0, there is no relationship between the two variables. It is expressed with R in the literature.

2.3.2. Mean squared error (MSE)

It is a type of statistics used primarily in cases where the magnitudes of error values are similar. Its formula is shown in Equation (6), and it is the process of summing the squares of the error between the actual output (y) and the expected output (d) and averaging [18, 26].

$$\boldsymbol{MSE = \frac{1}{N}\sum_{i=1}^{N}(d_i - y_i)^2} \qquad (6)$$

2.3.3. Mean absolute percentage error (MAPE)

MAPE is an error measure calculated as an absolute percentage error between predicted values and actual values. MAPE is often used to evaluate the accuracy of prediction models. The MAPE value is usually expressed as a percentage (%). It is calculated by Equation (7).

$$\boldsymbol{MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{(d_i - y_i)}{d_i}\right|} \qquad (7)$$

There are some general assumptions for the interpretation of the MAPE value.

• If the MAPE value is between 0% and 10%: This indicates that the prediction is quite good and has high accuracy.

• If the MAPE value is between 10% and 20%: This indicates that the estimate has a reasonable accuracy.

• If the MAPE value is greater than 20%: This indicates that the forecast has low accuracy and needs improvement. Higher MAPE values reflect larger errors and lower prediction accuracy.

2.3.4. Mean absolute error (MAE)

MAE, an error measure that measures the absolute amount of error between available forecasts and actual values, is a common measure used especially to evaluate the performance of forecasting models. Equation (8) is used to calculate the MAE.

$$\boldsymbol{MAE = \frac{1}{N}\sum_{i=1}^{N}|d_i - y_i|} \qquad (8)$$

Lower MAE values indicate that the predictions are closer to the true values and have better predictive performance. MAE is particularly useful for evaluating the performance of prediction models because it uses absolute values of errors. This means that negative and positive errors do not affect each other.

2.3.5. Coefficient of determination (R^2)

After calculating the regression coefficients, the coefficient of determination is calculated after the regression estimation model is established. Thus, the coefficients' significance and the model's suitability will become understandable. The coefficient of determination showing the suitability of the multiple regression model is calculated by Equation (9) [27].

$$R^2 = 1 - \frac{\sum_i^n (y_{i(real)} - y_{i(estimated)})^2}{\sum_i^n (y_{i(real)} - mean(y_{real}))^2} \tag{9}$$

2.3.4. t-static

The t test is a hypothesis test used in statistical analysis. It is used to compare means between two groups. The t test is used to determine whether the difference between groups is statistically significant. The t-test is based on a null hypothesis and an alternative hypothesis. The null hypothesis states that there is no difference between the groups or that the mean of the two groups is equal. The alternative hypothesis states that there is a difference between the groups. The main purpose of the t-test is to test the validity of the null hypothesis. This test is based on the means and variances of sample data between groups.

As a result of the t test, the t value is calculated. This t value expresses the size and statistical significance of the difference between groups. Also, the p-value is calculated. The p-value is a statistical measure that evaluates the validity of the null hypothesis. The p value is compared with the acceptable error level (alpha level) to determine whether the null hypothesis will be rejected.

In conclusion, the t-test is a hypothesis test used to compare means between groups and to evaluate the statistical significance of the difference. This test assists in the analysis of statistical results and decision making.

2.3.5. Introduction of input parameters

For the study, 45384 data from Aydın numbered 17234 meteorology stations recorded between May/2016 - May/2021, for which humidity, temperature, wind, and precipitation values were not recorded, were eliminated, and the remaining 9639 hours of data were used. The distribution of these data according to the humidity data is given in Figure 5, the distribution according to the temperature data in Figure 6, the distribution according to the wind speed data in Figure 7, and the distribution according to the precipitation data in Figure 8.

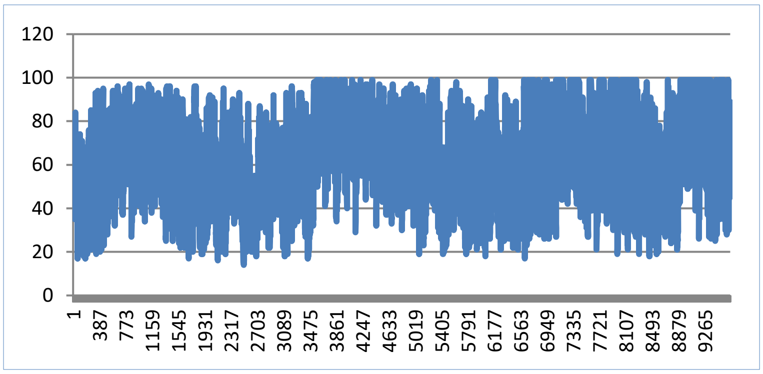

Figure 5. Humidity chart by data

In Figure 5, it is understood that within the scope of the available data, the humidity reaches 100%, which is the saturation ratio, often between 40% and 60% on average. The lowest humidity rate is between 15% and 20%.

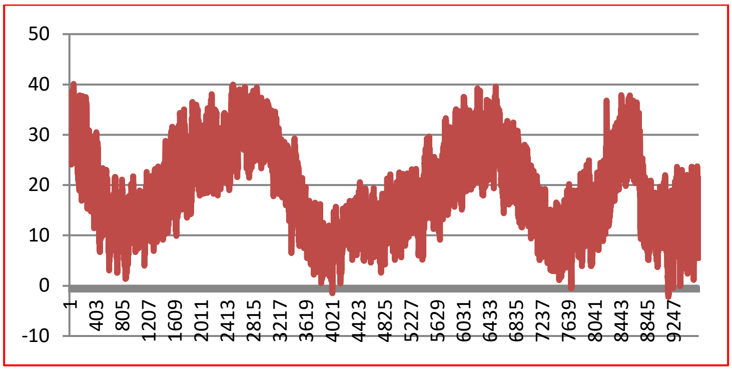

Figure 6. Temperature chart by data

In Figure 6, within the scope of the data, fluctuations between 0-40 °C are striking in general. The highest temperature occurs in the 40-45°C, while the lowest temperature occurs between -5–0 °C. It can be said that the average value is a temperature in the range of 10-30 °C.

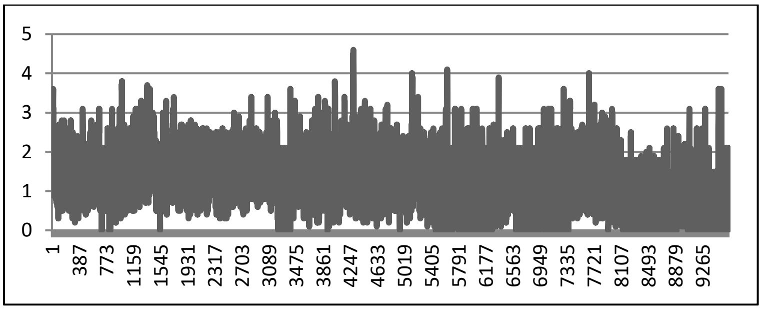

***Figure 7.** Wind chart by data*

In Figure 7, the wind speed is shown as m/s, and within the scope of the available data, it is understood that the maximum speed is 4-5 m/sec, often 0, and generally 1-2 m/sec.

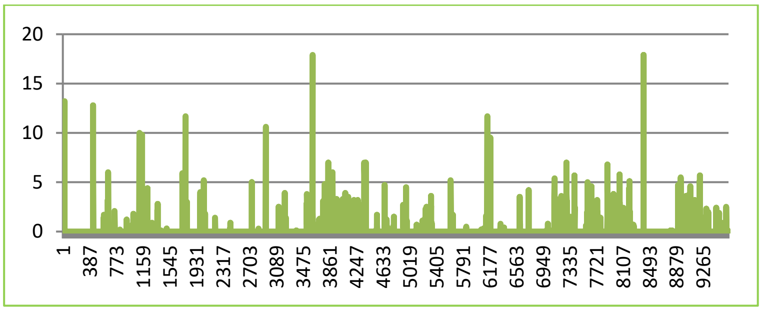

Figure 8. Precipitation chart by data

In Figure 8, the amount of precipitation is given in mm=kg/, and it is understood that it is 4-5 mm at most, 0 most of the time, and 1-2 mm generally.

Different input layers are derived by shifting the data of the parameters to form a model so that they do not lose their meaningful relationships. The correlation between data is given in Table 1.

h in Table 1; relative humidity (%), t; temperature (c°), w; wind speed (m/s) and p; represents precipitation (mm=kg/) data and -1, -2, -3 represents the number of shifted hours. There are linear, inverse correlations in the table, with the highest correlation being linear with a value Models created with highly correlated parameters are given in Table 3. of 0.956 and between y-1 and y (correlation of precipitation between the amount of precipitation one hour ago). On the other hand, the lowest correlation is between r-1 and s, with a value of -0.021.

Table 1. Correlation of data

	h	h-1	t	t-1	w	w-1	p	p-1	p-2	p-3
h	1									
h-1	0.762	1								
t	-0.695	-0.568	1							
t-1	-0.557	-0.695	0.898	1						
w	0.040	0.043	-0.023	-0.025	1					
w-1	0.040	0.040	-0.021	-0.022	0.667	1				
P	0.194	0.189	-0.101	-0.099	0.081	0.083	1			
p-1	0.199	0.194	-0.104	-0.101	0.085	0.080	0.956	1		
p-2	0.204	0.199	-0.108	-0.104	0.087	0.084	0.909	0.955	1	
p-3	0.207	0.203	-0.107	-0.108	0.089	0.086	0.877	0.908	0.955	1

Table 2. t-test on datas

Variables	Parameter	Values
X=h, Y=p	t Stat	285,4
	P(T<=t)	0
X=h-1, Y=p	t Stat	285,3
	P(T<=t)	0
X=t, Y=p	t Stat	217,0
	P(T<=t)	0
X=t-1, Y=p	t Stat	217,0
	P(T<=t)	0
X=w, Y=p	t Stat	135,2
	P(T<=t)	0
X=w-1, Y=p	t Stat	133,3
	P(T<=t)	0
X=p-1, Y=p	t Stat	0,85
	P(T<=t)	0,20
X=p-2, Y=p	t Stat	0,62
	P(T<=t)	0,27
X=p-3, Y=p	t Stat	0,38
	P(T<=t)	0,35

** X is independent and Y is dependent variable*

As can be seen in Table 1, there are 10 different parameters belonging to the data set. 9 of these parameters are independent variables and 1 (precipitation) is dependent variable. The t and P values obtained as a result of the t test are given in Table 2.

Accordingly, a high tStat value and a low P value (the acceptable level is usually 0.05) means that the difference between the tested parameters is large and statistically significant. According to the t test of precipitation data shifted with precipitation data, it is seen that the results are outside the acceptable level.

When Table 3 is examined, it is seen that the average precipitation between the years 2016-2021 is between 1.3 and 1.7 mm. On the other hand, it is seen that the highest precipitation occurred in 2017, and the lowest precipitation occurred in 2016.

$$\sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} \tag{10}$$

The formula for the standard deviation is shown in Equation (10). x represents each data in the data set, the mean, and n the sample size. When Table 3 is evaluated according to years, it is understood that the distribution of precipitation in 2017 was higher than in other years, and it showed a closer spread to the average in 2019 and 2020.

Table 3. Descriptive statistical parameters of datas

Years	2016	2017	2018	2019	2020	2021
Average precipitation*	1,689	1,537	1,469	1,234	1,367	1,315
Maximum precipitation *	12,100	44,400	17,900	20,700	17,900	19,000
Standart deviation	2,491	3,039	2,382	1,793	1,793	2,151
Skewness	2,660	7,789	3,638	4,354	3,873	3,922
Kurtosis	7,623	94,277	16,593	32,761	25,524	20,808
Coefficient of variation	147,483	197,722	162,151	145,299	131,163	163,574
Total number of data (hourly)	282	1921	1939	2174	1555	1768

* mm=kg/m^2

$$\frac{n}{(n-1)(n-2)}\sum\left(\frac{x_j-\overline{x}}{s}\right)^3 \quad (11)$$

Skewness indicates the degree of asymmetry around the mean of a distribution. Positive skewness indicates a skewness whose asymmetric end expands towards higher positive values, while negative skewness indicates a distribution whose asymmetric end expands towards lower negative values. It can be said that the skewness decreases when it approaches zero. Its formula is given in Equation (11). s is the standard deviation, x_j represents each data in the data set, \bar{x} the mean, and n is the sample size. In Table 3, it can be seen that 2016 was the year with the least skewness and the closest to normality in the data set, while 2017 was the year with the highest skewness, and there was a positive skewness in all the data.

$$\left\{\frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum\left(\frac{x_j-\overline{x}}{s}\right)^4\right\}-\frac{3(n-1)^2}{(n-2)(n-3)} \quad (12)$$

Kurtosis gives the relative steepness or flatness of a distribution compared to the normal distribution. Positive kurtosis indicates a comparatively steep distribution, while negative kurtosis indicates a comparatively flat distribution. Its formula is shown in Equation (12). s is the standard deviation, x_j represents each data in the data set, \bar{x} the mean, and n is the sample size. , it is seen that there is a steep distribution due to positive kurtosis, and the steepest distribution was in 2017.

$$\frac{\sqrt{\frac{\sum(x-\overline{x})^2}{n-1}}}{Average\ value}*100 \quad (13)$$

The coefficient of variation is a statistical measure in which the variation in a data set is expressed as a percentage. The coefficient of variation gives the ratio of the variability of a measurement to the mean. A high coefficient of variation indicates that the values in the data set deviate significantly from the mean, while a low coefficient of variation indicates that the values are closer to the mean. The coefficient of variation is calculated as the ratio of the standard deviation to the mean and is usually expressed as a percentage (%). It is calculated by Equation (13).

When the data in Table 3 are examined, it is seen that there are large deviations from the mean. The reason for this is that all data belonging to the hour when at least one of the parameters is empty among the collected 45384 data were removed, and the number of data decreased to approximately 1 in 5 for this reason.

3. STUDY AREA

3.1. Precipitation Station Data

Station; Cumhuriyet, 1999. Sk. The data were obtained from the precipitation observation station in Aydın province, numbered 17,234, which has the "Automatic Meteorological Observation Station AMOS - Synoptic - Daily Climate" observation type. No:3, 09020 Efeler Aydın, located next to the Aydın Meteorology Directorate building. The station is located at latitude 37° 50` 26.32`` and longitude 27° 50` 16.19`` with an altitude of 53m and observation type mm [28]. Satellite image is shown in Figure 9.

Figure 9. *Weather station location*

3.2. Information About the Region

Aydın province is located in the Aegean region and has a Mediterranean climate with warm and rainy winters and hot and dry summers. It covers the provinces of Uşak, Aydın and Denizli in the Büyük Menderes Basin and is located between Aydın Mountains and Menteşe Mountains. In the basin, which has an area of 24976 m^2, Aydın province has a surface area of 1582 m^2 [29].

4. RESULTS

In Table 4, using Table 1, models were created from parameters with high correlation to create an input layer to work in a flow, as seen in Figure 1. In Table 4, the coefficients obtained with the MLR are given as parameters. Table 5 shows the variable coefficients and related equations obtained as a result of applying the data to the MLR process according to the input/output parameters created in Table 4.

Table 4. Models prepared for multiple linear regression and artificial neural networks.

Model	Input Layer	Output Layer
1	h, t, w, p-1, p-2, p-3	P
2	h, t, w, p-1, p-2	P
3	h, h-1, t, w, y-1	P
4	h, t, t-1, w, p-1, p-2	P
5	h, t, w, w-1, p-1	P
6	h, t, w	P

Table 5. Equations obtained as a result of multiple linear regression.

Model	Equation
1	$0.0006x_1 + 0.0006x_2 + 0.0045x_3 + 0.1502x_4 - 0.195x_5 + 0.9943x_6 - 0.0475$
2	$0.0005x_1 + 0.0001x_2 + 0.0056x_3 - 0.0463x_4 + 0.9856x_5 - 0.0316$
3	$0.0004x_1 + 0.0002x_2 + 0.0004x_3 + 0.0061x_4 + 0.941x_5 - 0.0472$
4	$0.0006x_1 + 0.0012x_2 - 0.0009x_3 + 0.0061x_4 + 0.9411x_5 - 0.0451$
5	$0.0005x_1 + 0.0004x_2 + 0.0102x_3 - 0.0061x_4 + 0.9415x_5 - 0.0399$
6	$0.0006x_1 + 0.0005x_2 + 0.6665x_3 + 0.3998$

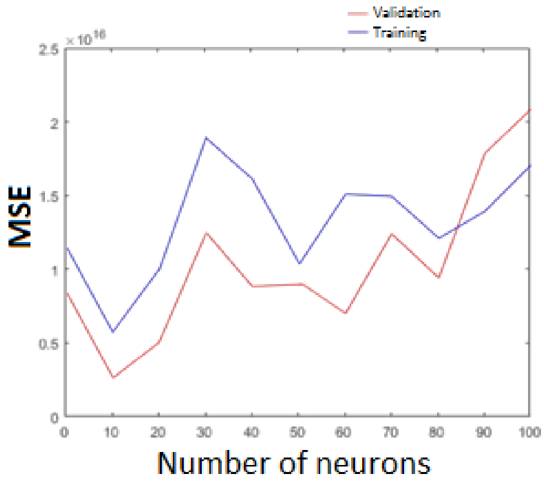

Figure 10. MSE values according to the number of neurons in neural network training.

In the ANN model created with the Neural Fitting application of the MATLAB program, the number of hidden layer neurons containing the sigmoid function was tried to be determined according to the lowest MSE value between 0-100 at the validation and education level with the MATLAB program, as shown in Figure 10. The optimal neuron value of the data set is 10, and 70% (6747) of the 9639 data were used for training, 15% (1446) for validation, and 15% for testing (1446). Additionally, MAPE and MAE are included in the calculations.

Levenberg-Marquard training algorithm was used for training. ANN's working style is shown in Figure 11 Levenberg-Marquard training algorithm was used for training. ANN working style is shown in Figure 11.

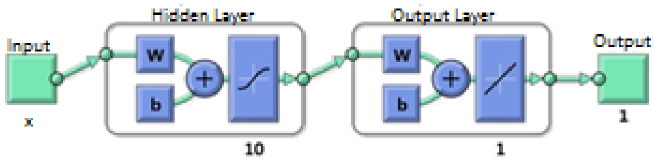

Figure 11. Artificial neural networks working steps for models.

Figure 11, x value in the input layer; is the number of coefficients of each model, that is, the number of parameters in the Input Layer column in Table 4, Introduction section; from the input of the data to be trained, the Hidden Layer; From training the data using the sigmoid function, the Output Layer; It is responsible for verifying the data and transferring the trained data to the output section. MSE, MAPE and MAE values are given in Table 6. As these values approach zero, the error decreases. The R value represents correlation; Approaching +1 or -1 means that the relationship is increasing (the coefficient of determination), the square of the correlation value between the model result and the actual result (included in Table 6) and measures the correlation between outputs and targets. The closer the value is to 1, the higher the correlation and reliability.

Table 6. R^2 , MSE, MAPE and MAE values of the data obtained with multiple linear regression and artificial neural networks.

Model	MLR				ANN			
	R^2	MSE	MAPE	MAE	R^2	MSE	MAPE	MAE
1	0.91	0.1838	0.2966	0.0186	0.98	0.0438	0.0766	0.0036
2	0.91	0.1870	0.2945	0.0271	0.88	0.0433	0.0745	0.0031
3	0.91	0.1872	0.2960	0.0278	0.90	0.0384	0.0650	0.0038
4	0.91	0.1872	0.2915	0.0274	0.90	0.0375	0.0635	0.0034
5	0.91	0.1872	0.3213	0.0325	0.90	0.0382	0.0643	0.0035
6	0.44	0.5153	0.6824	0.0925	0.45	0.2552	0.3724	0.0185

When these data in Table 6 are examined, it is seen that the 6th model has a bad value compared to other models in terms of error and regression values. On the other hand, the regression values of the first five models are above 90% and have a good deal. When we include the errors in the evaluation, it is seen that the model with both the highest regression value and the lowest errors is the 1st model.

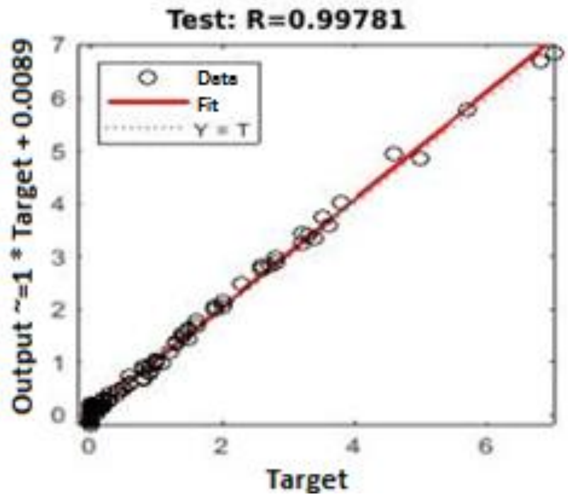

Figure 12. Neural networks regression graph.

As seen in Figure 12, if the R value is very close to 1, then the model is the most successful. This represents the success rate in training the data, so when the graph is interpreted, it can be stated that the data has undergone successful training.

In error evaluation, as shown in Figure 13, the error value in the training, validation, and testing areas is very close to the orange line representing zero error. It can be stated that it is in the range of -1 to 1 for training, validation, and testing, and a situation far from high error occurs.

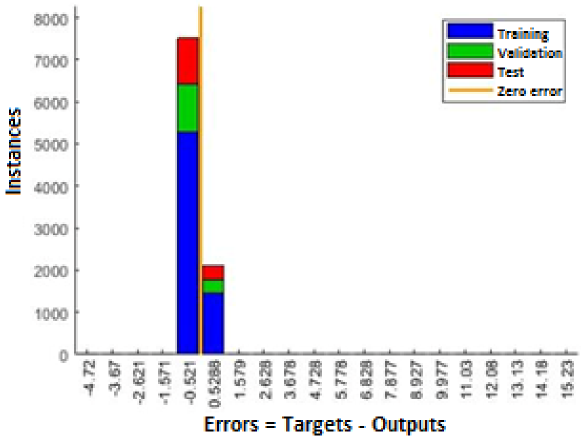

Figure 13. Neural networks error histogram.

7. CONCLUSION

Multiple Linear Regression and Artificial Neural Networks methods were used to forecast precipitation with temperature, humidity, wind speed, and precipitation data. To get meaningful results, six different models were created by shifting the data up to 3 hours, paying attention to the correlation coefficient between them. Models were calculated with MLR and ANN methods, performance criterias (\left(R^2\right), MSE, MAPE and MAE) were obtained as a result of the processes, and the 1st model, the most efficient model, was selected according to these criteria. The first model consists of humidity (h), temperature (t), wind (w), and precipitation (p-1, p-2, p-3) parameters shifted up to 3 hours. It is understood that the more precipitation parameters in the input parameters compared to the other models contribute to better results in terms of performance criteria.

The results of artificial neural networks working with a feedforward backpropagation algorithm gave more successful results than multiple linear regression analyzes in all models.

With the application of the obtained model, it is seen that benefits can be obtained in many areas that require precipitation forecasting.

There are some limitations in the study in terms of Multiple Linear Regression and Artificial Neural Networks.

- ANNs often require large amounts of data. There was no problem with this limitation with 9639 pieces of data.
- High-capacity ANNs can overfit the training dataset. This can negatively impact generalizer performance and may fail with new data. In the study, which reached a rate of 99%, no action was taken to prevent this situation in education. To prevent this, data splitting, data generalization, dropout, early stopping methods can be tried.
- Training large and complex ANNs can be time consuming. Such a situation was not encountered in this study.
- ANNs are generally less interpretable models due to their complex nature. Since the obtained data were interpreted on the basis of the performance evaluation criteria, this limitation did not constitute an obstacle to a large extent.
- MLR is based on the assumption of a linear relationship between dependent and independent variables. If the relationship is not linear, the MLR model may be inadequate. To reduce this limitation, highly correlated data sets were studied.
- MLR is based on the assumption of complete independence between independent variables. If there are multiple correlations or other relationships between the independent variables, this assumption is violated and the results may be misleading. In order to obtain a high correlation in the study, new parameters were derived by clock shifting the parameters. For this reason, there is a possibility that the models obtained with MLR may be misleading.

THANKS

The authors thank the Aydın Meteorology Directorate for their support.

AUTHOR CONTRIBUTION STATEMENT

In the work carried out, Serkan ANSAY contributed to data collection, design, technical implementation, literature review and analysis; Bayram KÖSE consisted of creating the idea, checking the spelling, controlling the content and evaluating the results

CONFLICT OF INTEREST

There is no conflict of interest in this study.

REFERENCES

Yıldıran A. and Kandemir S. Y., "Estimation of rainfall amount with artificial neural networks", BSEU Journal of Science, 5(2): 97–104, (2018).

https://www.mgm.gov.tr/genel/meteorolojinedir.aspx

Turhan E. and Çağatay H. Ö., "Using of Artificial Neural Network (Ann) for setting estimation model of missing flow data: Asi river-Demirköprü flow observation station (fos)", Çukurova University Journal of the Faculty of Engineering and Architecture, 31(1): 93–106, (2016).

Gümüş V., Başak A. and Yenigün K., "Drought Estimation of Şanlıurfa Station with Artificial Neural Network", Gazi University Journal of Science Part C: Design and Technology, 6(3): 621–633, (2018).

Ünes F., Taşar B., Demirci M. and Kaya Y. Z., "Forecasting of daily evaporation amounts using Artificial Neural Networks technique", Dicle University Journal of Engineering, 9(1): 543–551, (2018).

Tufaner F., Dabanlı İ. and Özbeyaz A., "Analysis of Drought with Artificial Neural Networks: Adıyaman Example", 4th International Water and Environment Congress, 4(1): 25–32, (2019).

Sezer M. S., "Long term load forecast thorugh Artificial Neural Network and different forecasting methods: Zonguldak case", Master's Thesis, Institute of Science Kütahya Dumlupınar University, (2019).

Akbulut İ. and Özcan B., "Air pollution forecast: A comparison with Artificial Neural Networks and Regression methods", Kocaeli Üniversitesi Science Journal. 3(1): 12–22, (2020).

Holmstrom C., Liu D. and Vo C., "Machine learning applied to weather forecasting" Stanford University, (2016).

Refonaa M., Lakshmi J., Abbas M., Raziullha R., "Rainfall Prediction using Regression Model" International Journal of Recent Technology and Engineering, 8(2S3): 543-546, (2019).

Prabakara P. S. M., Kumar S. and Tarun P. N., "Rainfall prediction using modified linear regression", ARPN Journal of Engineering and Applied Sciences, 12(12), (2017).

Ahmed S. W. A., and Mohamed H. A. Y., "Rainfall prediction using multiple linear regressions model", International Conference on Computer, Control, Electrical, and Electronics Engineering, 1-5, (2021).

Srivastava L. K., Anand S., Sharma N., Dhar S. and Sinha S., "Monthly rainfall prediction using various machine learning algorithms for early warning of landslide occurrence" International Conference for Emerging Technology, 1–7, (2020).

Parmar M., Mistree A. and Sompura K., "Machine learning techniques for rainfall prediction: A review", International Conference on Innovations in information Embedded and Communication Systems, 3, (2017).

Darji H. B., Dabhi M. P. and Prajapati V. K., "Rainfall forecasting using neural network: A survey", International Conference on Advances in Computer Engineering and Applications, 706–713, (2015).

Hatim R., Siddiqui M. and Kumar F., “Addressing Challenges and Demands of Intelligent Seasonal Rainfall Forecasting using Artificial Intelligence Approach”, International Conference on Computation, Automation and Knowledge Management, 263–267, (2020).

Ataseven B., “Forecasting by using artificial neural networks”, Marmara University Öneri Journal, 10(39): 101-115, (2013).

Köse B., “A new analytical approach for predicting hourly and daily wind speed and comparison with Artificial Neural Networks", 10th International Clean Energy Symposium, 928-938, (2016).

Kaya S., “The buckling analysis of axially loaded columns with artificial neural networks”, Electronic Letters on Science & Engineering, 2(2): 36–45, (2006).

Jacobs R. A., “Increased rates of convergence through learning rate adaptation” Neural Networks, 1(4): 295–307, (1998).

Riedmiller M. and Braun H., “A direct adaptive method for faster backpropagation learning: the RPROP algorithm”, IEEE International Conference on Neural Networks, 1: 586-591, (1993).

Doğru F., “Parameter estimation from residual gravity anomalies using actual optimization methods”, Bulletin of the Earth Sciences Application and Research Centre of Hacettepe University, 36(1): 31-43, (2015).

http://kod5.org/yapay-sinir-aglari-ysa-nedir/

Yavuz S. and Deveci M., “The effect of statistical normalization techniques on the performance of artificial neural network”, Erciyes University Journal of Faculty of Economics and Administrative Sciences, (40): 167–187, (2015).

Özdamar K., "SPSS ile biyoistatistik", Nobel Kitabevi, 11, Ankara, (2019).

https://avys.omu.edu.tr/storage/app/public/vceyhan

https://acikders.ankara.edu.tr/pluginfile.php/117326/mod_resource/content/1/11-Coklu Regresyon 1.pdf

http://www1.mgm.gov.tr/kurumsal/istasyonlarimiz.aspx?sSirala=AL&m=AYDIN

Yeşilırmak E., “Analysis of daily precipitation concentration in büyük menderes basin”, Journal of Adnan Menderes University Agricultural Faculty, 12(2): 55–71, (2015).