Estimation of Incomplete Precipitation Data using the Adaptive Neuro-Fuzzy Inference System (ANFIS) Approach

Okan Mert Katipoglu[1,*]
[1]*Department of Civil Engineering, Erzincan Binali Yıldırım University, 24002, Erzincan, Turkey*

***Abstract*—The completeness and continuity of precipitation data, which is one of the basic components of the hydrological cycle, is of vital importance for the planning of water resources. In this study, the gaps in the missing precipitation data in the Erzincan precipitation observation station were filled by using the adaptive neuro-fuzzy inference system (ANFIS). While Erzincan precipitation station 17094 was used as output, Bayburt 17089, Tercan 17718 and Zara 17716 precipitation stations were selected as model inputs. In the ANFIS model, monthly total precipitation data (52 years) between 1966 and 2017 were used. In the model established, 80% of the data (1968) were used for training and 20% (492) for testing. In the ANFIS model, variables were tried by dividing them into sub-sets between 3 and 8. The most suitable ANFIS model was revealed by comparing various statistical indicators. As a result of the study, 3 sub-sets, hybrid learning algorithm, trimf membership function, and model with 600 epochs were selected as the most suitable model.**

***Index Terms*—Estimation, Hydrology, Machine learning.**

I. INTRODUCTION

Long, continuous, and consistent precipitation data play an important role in meteorological and hydrological studies and ensures reliable results.

Using a series of rainfall data with missing values critically affects the statistical strength and accuracy of a study. Various techniques have been proposed and adopted in filling incomplete data to obtain a continuous and long series of rainfall data. The highlights of these methods; deterministic, stochastic and artificial intelligence-based methods [1].

Adaptive neuro-fuzzy inference system (ANFIS) is an artificial intelligence method that combines the ability of artificial neural networks to compute and learn in parallel with the inference capability of fuzzy logic and uses Sugeno-type fuzzy logic system and hybrid learning together. Adaptive networks consist of directly connected nodes. Each node here represents a different operation, and the links between the nodes show an uncertain weight whose value is not exactly known. The learning rules show how the variable data should be changed in a way that minimizes the error with a different discourse between the output of the network and the target value [2].

Conventional methods such as regression analysis and autoregressive moving average models are widely used to predict hydrological processes. New methods such as fuzzy logic (FL) and ANFIS have recently been recognized as effective alternative tools for modelling complex hydrological systems and are widely used for predictive purposes. Fırat [3] was used the ANFIS method to estimate the daily streamflow and sediment amount for four-stream branches in the Büyük Menderes Basin. Piri, Kahkha [4] referred ANFIS model to forecast changes in level of the reservoir. Arslan, Üneş [5] estimated the water level change in the dam reservoir by ANFIS, support vector machines (SVM) and multiple linear regression (MLR) methods. Quej, Almorox [6] employed the ANFIS model to estimate daily global solar radiation. Tien Bui, Khosravi [7] used the ANFIS model to map flood susceptibility. Irwanto, Alam [8] used the ANFIS to estimate solar energy density. Adnan, Malik [9] applied the ANFIS model to estimate monthly pan evaporation. Suparta, Samah [10]; Çitakoğlu, Çoşkun [11] applied the ANFIS times series technique to predict rainfall. Faruq, Marto [12] developed a hybrid technique called ANFIS model to forecast flood river water level. Adnan, Petroselli [13]; Gerger, Gümüş [14] used the ANFIS in rainfall-runoff modeling. Nguyen, Li [15] used the ANFIS to forecast droughts [16, 17]. In [18] were used ANN to predict streamflow. Katipoglu [19] applied the ANFIS approach to fill in the gaps in the temperature data. The success of ANFIS and artificial neural networks (ANN) in daily evaporation prediction was evaluated by using various climate parameters as inputs [20, 21]. Acar, Saplıoğlu [22] investigated stream sediment transport in the Euphrates basin with artificial neural networks (ANN) and ANFIS methods.

In this study, ANFIS model was applied to complete the missing precipitation data gaps in the Erzincan precipitation observation station. Neighboring and similar stations were used to complete the data. Precipitation data are divided into 80% training and 20% testing, respectively.

II. MATERIAL AND METHODS

A. Study area and data

Erzincan number 17094, Bayburt number 17089, Tercan number 17718 and Zara numbered 17716 were selected as the study area. The data used in this study covers the years

Manuscript received July 12, 2021; accepted November 25, 2021.
*Corresponding author: okatipoglu@erzincan.edu.tr

between 1966 and 2017 and was obtained from the General Directorate of Meteorology. Detailed information about the stations used in the study is given in Table I.

TABLE I. PRECIPITATION STATIONS USED IN THE STUDY.

Station	Number	Latitude	Longitude	Altitude
Erzincan	17094	39.75	39.48	1216
Bayburt	17089	40.25	40.22	1584
Tercan	17718	39.78	40.39	1425
Zara	17716	39.89	37.75	1338

B. Adaptive neuro-fuzzy inference system (ANFIS)

The ANFIS model was proposed by Jang [23] and was created by using ANN and fuzzy logic together. With this method, the ability to learn and compute, which is characteristic of artificial neural networks, is provided to fuzzy logic inference systems. Likewise, the ability of fuzzy logic inference to make decisions and provide expert knowledge is given to artificial neural networks. Thanks to this idea, the superior features of the two models can be used together. ANFIS only operates Sugeno type models. This modeling type consists of output variables that have membership functions as a function of inputs. ANFIS network structure consists of 6 layers. These layers are; the input, fuzzification, rules, normalization, defuzzification and aggregation layer [24].

The basis of the ANFIS program is the artificial neural networks used in the prediction of the desired subjects. During the training phase of the program, the model is created by making trials until the given values that will make the least faulty and closest estimate are obtained. The fuzzy logic inference system evaluated in ANFIS is transformed into adaptive networks and the most suitable condition is created with a learning algorithm [2]. As an example, an ANFIS model with two inputs (x_1 and x_2), single output (y) and containing four rules is shown in Fig. 1.

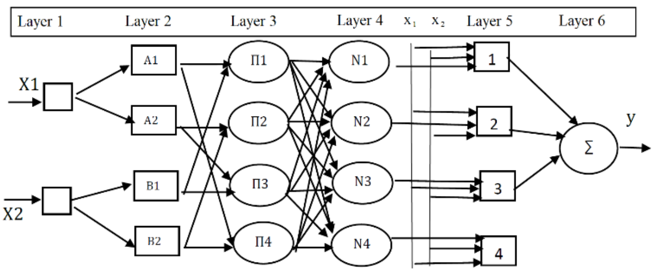

Fig. 1. Adaptive network-based fuzzy logic inference system [2].

Layer 1: It is called the input layer. Signals coming to each node input in this layer are transmitted to other parts and layers.

Layer 2: This layer is called the fuzzification layer. Each node in Layer 2 describes fuzzy clusters such as A_j and B_j (j = 1, 2). Certain membership functions are used, which are generalized as membership functions of input values.

Layer 3: This layer is regarded as the rule layer. Each node that makes up this layer specifies the number of rules based on Sugeno's fuzzy logic inference system.

Layer 4: This layer is expressed as the normalization layer. The normalized firing level of each rule is calculated by considering the nodes in this layer as input data from the other nodes from the rule layer.

Layer 5: This layer is considered the defuzzification layer. It is certain that there is a relationship between the result values of each node of the normalization layer in this layer and the input values x_1 and x_2 in this layer. In this layer, the weighted results of a rule processed on each node are calculated. Calculations are made based on the weighted values of the expressions after the parts of the rules.

Layer 6: This layer is regarded as the collection layer. There is only one node in the aggregation layer and this node is indicated by the symbol. In this layer, the output values of each node from the previous layer are summed up and as a result, the real value of the ANFIS system is found [23].

C. Evaluation of models

After the models were prepared, their performances were compared with the help of different statistical criteria. These; root mean square error (RMSE) and coefficient of determination (R^2). The fact that the error values are close to 0 and the R^2 value to 1 indicates that the predicted value is strongly converged to the line. The mentioned statistical calculations can be made with the help of (1) and (2) respectively.

$$\text{RMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(b_i-y_i)^2} \tag{1}$$

$$R^2=\frac{\sum_{i=1}^{N}(b_i-b_{ort})^2-\sum_{i=1}^{N}(b_i-y_i)^2}{\sum_{i=1}^{N}(b_i-b_{ort})^2} \tag{2}$$

b_i : expected values of models,
y_i : the outputs produced by models,
$b_i - y_i$: error (residual) value,
N: number of data.

III. Research Findings

In this study, the ANFIS model was applied to fill the 1.4% data gaps in the Erzincan precipitation observation station. Correlation analysis has been performed to select the inputs in the established model (Table 2). While neighboring stations with the highest correlation and similar climatic characteristics were presented as input to the model, the Erzincan station precipitation data were selected as outputs. ANFIS model has been obtained by testing various subsets, the number of iterations and membership functions.

TABLE II. PEARSON CORRELATION COEFFICIENTS (R).

	Bayburt	Zara	Tercan
Erzincan	0.80	0.74	0.84

In Table II, correlation coefficients of selected stations are shown to complete the missing data of the Erzincan precipitation observation station.

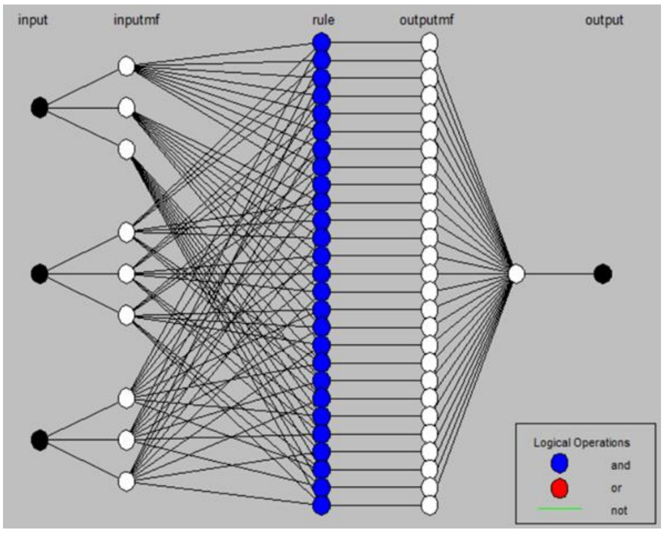

Fig. 2. Structure of the ANFIS model.

The ANFIS model consists of 3 inputs and 1 output. Trimf membership function is used in the input and output layer and 27 rules have been applied.

TABLE III. SELECTING THE MOST SUITABLE ANFIS MODEL.

Parameters	m333*	m444	m555	m666	m777	m888
Optim. method	hibrid	hibrid	hibrid	hibrid	hibrid	hibrid
RMSE of training	11.60	11.04	10.94	9.58	9.84	8.34
RMSE of testing	15.85	29.60	30.30	80.72	81.13	40.34
Training R^2	0.78	0.80	0.80	0.85	0.84	0.88
Testing R^2	0.66	0.53	0.43	0.08	0.09	0.04
Iteration	600	600	600	100	100	100
Membership function	trimf	trimf	trapmf	trimf	trapmf	trimf

Note: * sign indicates the most suitable ANFIS model.

In this study, precipitation data were divided into 3 to 8 subsets, the backpropagation and hybrid learning algorithm were tried and the missing data were estimated by using various membership functions. As a result of the study, the model divided into 3 sub-sets, the hybrid learning algorithm, the prediction model obtained with the trimf membership function were selected as the most suitable. Statistical coefficients of training and test data were tested in selecting the most suitable model. The model with the lowest RMSE (11.60 and 15.85) and the highest R^2 value (0.78 and 0.66) was chosen as the most appropriate prediction model (Table III).

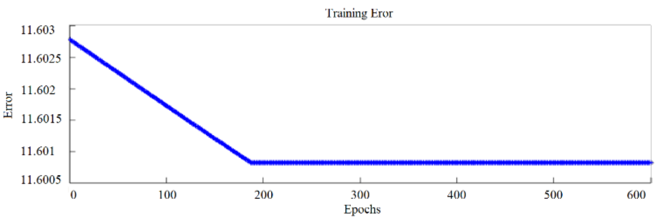

Fig. 3. Error propagation of the selected ANFIS model.

Fig. 3 shows the variation of the training error and the number of epochs. Fixing the training error after approximately 200 epochs indicates that the training was successful.

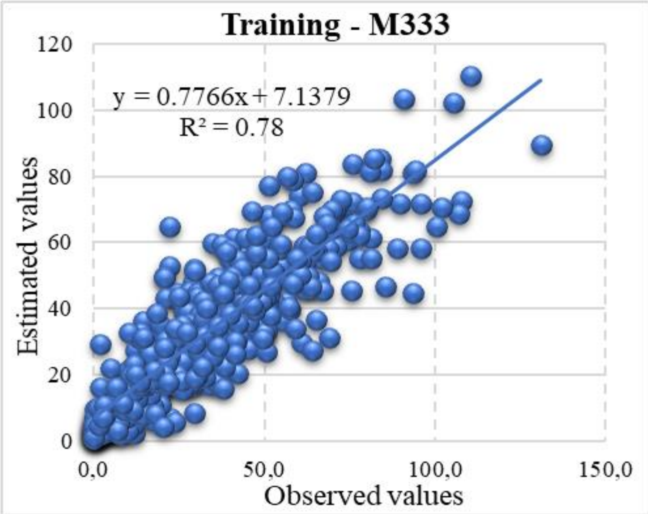

Fig. 4. Scatter plot of training data.

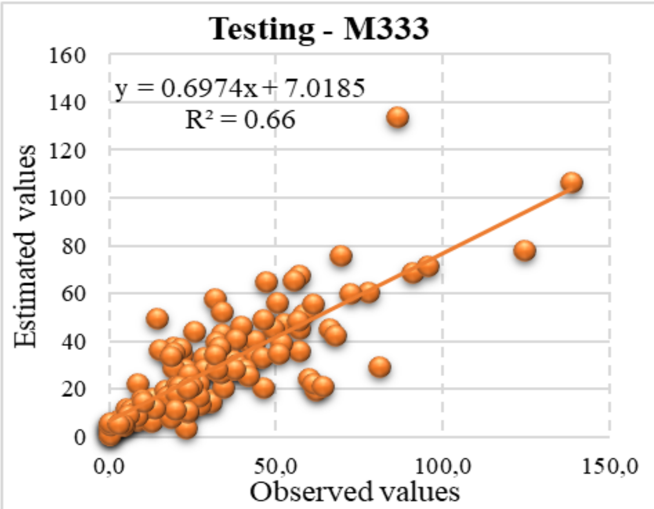

Fig. 5. Scatter plot of testing data.

Performance analysis of the ANFIS model is realized using MATLAB's ANFIS Toolbox. The regression results of training and testing procedures are shown in Fig. 4 and 5. All data points scatter around the regression line, indicating that the ANFIS model performance was statistically successful. In addition, the results show that the model can be applied to produce an accurate and reliable precipitation prediction.

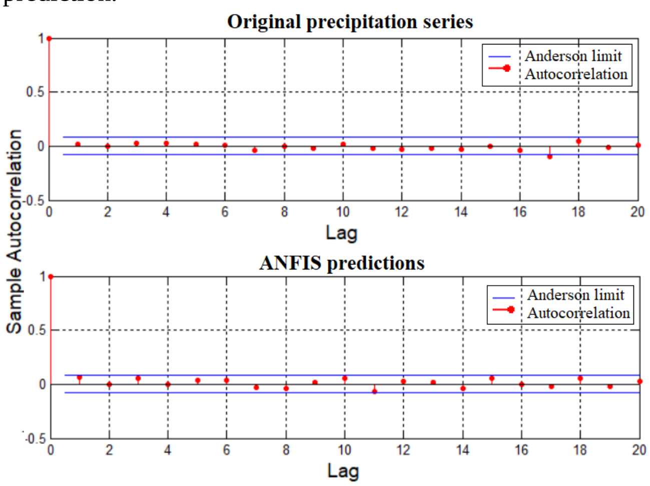

Fig. 6. Correlogram of original precipitation series and predicted precipitation series with ANFIS.

To test the prediction accuracy of the established ANFIS model, the autocorrelation graphs of the original precipitation series and the precipitation series produced by the ANFIS model were examined (Figure 6). In order to compare the series and determine their autocorrelation, the periodicity of the series was eliminated by the standardization process. It is seen that the original series and the ANFIS estimations are not autocorrelated since the Anderson's limits are not exceeded in the monthly time period. This situation shows that the ANFIS prediction does not disturb the structure of precipitation and that the accuracy of the prediction is high.

The comparison of precipitation predicted by the ANFIS model with the observed values is shown in Fig. 7. Since the distribution of training and test series overlap, it can be said that the model established gave very successful results.

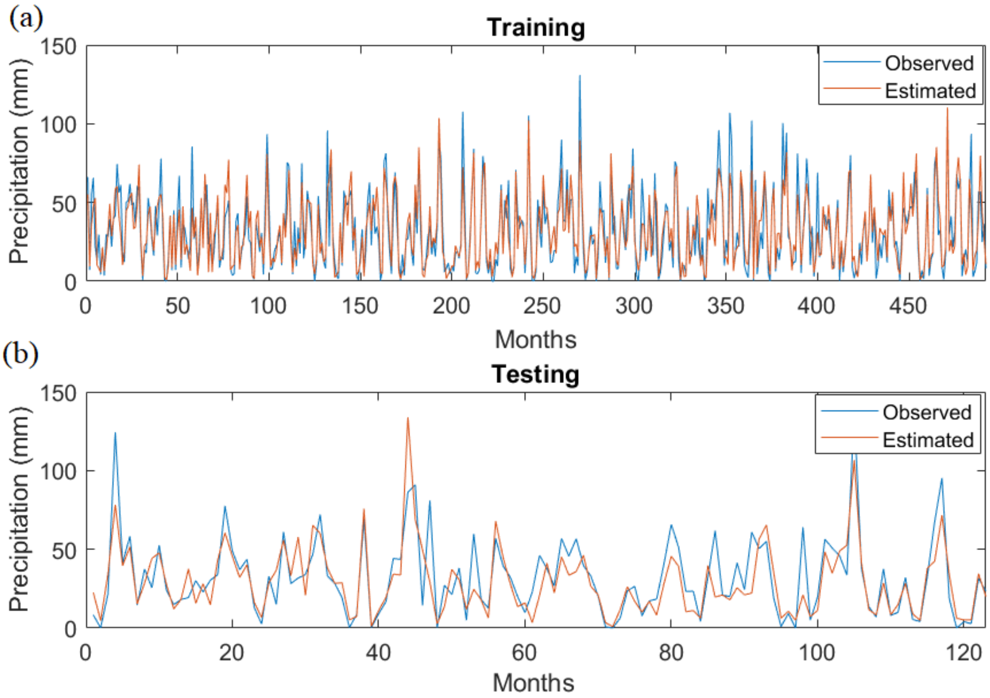

Fig. 7. ANFIS model precipitation forecast results.

IV. Conclusions

In this study, it was aimed to estimate the missing precipitation data at the Erzincan station by using the ANFIS method. For this, the closest and neighboring stations with similar characteristics were used. In model development, 80% of the training and 20% of them were used for testing. As a result of the study, it was determined that the ANFIS model can be used in missing data estimation. In addition, based on various statistical criteria, the model with 3 subsets, hybrid learning algorithm and trimf membership function has been shown to yield the most successful results.

As a result of the study, the statistical indicators of the most successful model were determined as RMSE=11.60, R^2=0.78 and RMSE=15.85 ve R^2=0.66 for the training and testing stages, respectively. It is thought that the predictive power can be increased in future studies by using precipitation, temperature and evapotranspiration data as inputs and separating the input values into subcomponents with wavelet transform.

Acknowledgment

The author thanks the General Directorate of Meteorology for the monthly total precipitation data provided. Part of this work was presented orally at the IV. International Conference on Data Science and Applications 2021 (ICONDATA'21).

References

[1] L. Campozano, E. Sánchez Á. Avilés, and E. Samaniego, "Evaluation of infilling methods for time series of daily temperature data: case study of Limpopo Province, South Africa". *Climate*, 7(7), pp. 86, 2019.

[2] M. Kaya, The completion with ANFIS of the missing currents data stream. Süleyman Demirel University Graduate School of Natural and Applied Sciences Department of Civil Engineering, MSc Thesis, Isparta, Turkey, 2018.

[3] M. Fırat, Modeling of watershed using adaptive neuro-fuzzy inference system approach. Pamukkale University Institute of Science and Technology, Doctoral dissertation, Denizli, Turkey, 2007.

[4] J. Piri, and M. R. R. Kahkha, "Prediction of water level fluctuations of Chahnimeh reservoirs in Zabol using ANN, ANFIS and cuckoo optimization algorithm." *Safety, and Environment*, 4(2), pp. 706-715, 2017.

[5] H. Arslan, F. Üneş, M. Demirci, B. Taşar, and A. Yilmaz, "Estimation of Keban Dam Lake Level Change Using ANFIS and Support Vector Machines." *Osmaniye Korkut Ata University Journal of Natural and Applied Sciences*, 3(2), pp. 71-77, 2020.

[6] V. H. Quej, J. Almorox, J. A. Arnaldo, and L. Saito, "ANFIS, SVM and ANN soft-computing techniques to estimate daily global solar radiation in a warm sub-humid environment." *Journal of Atmospheric and Solar-Terrestrial Physics*, 155, pp. 62-70, 2017.

[7] D. Tien Bui, K. Khosravi, S. Li, H. Shahabi, M. Panahi, V. P. Singh, *et al.* New hybrids of anfis with several optimization algorithms for flood susceptibility modeling." Water, 10(9), pp.1210, 2018.

[8] M. Irwanto, H. Alam, M. Masri B. Ismail, W. Z. Leow, and Y. M. Irwan, "Solar energy density estimation using ANFIS based on daily maximum and minimum temperature." *International Journal of Power Electronics and Drive Systems*, 10(4), pp. 2206, 2019.

[9] R. M. Adnan, A. Malik A. Kumar, K. S. Parmar, and O. Kisi, "Pan evaporation modeling by three different neuro-fuzzy intelligent systems using climatic inputs. " *Arabian Journal of Geosciences*, 12(19), pp. 1-14, 2019.

[10] W. Suparta, and A. A. Samah, "*Rainfall prediction by using ANFIS times series technique in South Tangerang, Indonesia.*" *Geodesy and Geodynamics,* 11(6), pp. 411-417, 2020.

[11] H. Çıtakoğlu, and Ö. Coşkun, "Precipitation Prediction of Central Anatolia Regional Stations Using Artificial Intelligence Techniques with Wavelet Transform Model." *Harran University Journal of Engineering,* 6(1), pp. 39-54, 2021.

[12] A. Faruq A. Marto, N. K. Izzaty, A. T. Kuye, S. F. M. Hussein, and S. S. Abdullah, "Flood disaster and early warning: application of ANFIS for river water level forecasting." *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control,* pp. 1-10, 2021.

[13] R. M. Adnan, A. Petroselli, S. Heddam, C. A. G. Santos, and O. Kisi, "Short term rainfall-runoff modelling using several machine learning methods and a conceptual event-based model." *Stochastic Environmental Research and Risk Assessment*, 35(3), pp. 597-616, 2021.

[14] R. V. Gerger, Gümüş, and D. Selmin, "The Evaluation of Different Artificial Intelligence Methods in Determination of Tigris Basin's Rainfall Runoff Relationship." *BSEU Journal of Science,* 8(1), pp. 300-311. 2021.

[15] V. Nguyen, Q. Li, and L. Nguyen, "Drought forecasting using ANFIS-a case study in drought prone area of Vietnam." *Paddy and water environment*, 15(3), pp. 605-616, 2017.

[16] E. D. Taylan, Ö. Terzi, and T. Baykal, "Hybrid wavelet–artificial intelligence models in meteorological drought estimation." *Journal of Earth System Science,* 130(1), pp. 1-13, 2021.

[17] Ö. Terzi, and A. Terzi, "Drought Estimation by using Adaptive Network Based Fuzzy Inference System in Çukurova Basin." *Duzce University Journal of Science & Technology*, 8(1), pp. 578-588. 2020.

[18] C. Belvederesi, J. A. Dominic Q. K. Hassan, A. Gupta, and G. Achari, "Predicting river flow using an AI-based sequential adaptive neuro-fuzzy inference system." *Water,* 12(6), pp. 1622, 2020.

[19] O. M. Katipoglu, "Completion of missing temperature data using Adaptive NetworkBased Fuzzy Inference System (ANFIS)." *4th International Conference on Data Science and Applications (ICONDATA'21),* June 4-6, Turkey, 2021, pp. 6.

[20] D. Yıldırım, B. Cemek, and E. Küçüktopcu, "Estimation of Daily Evaporation Using Fuzzy Artificial Neural Network (ANFIS) and Multilayer Artificial Neural Network System (ANN)." *Toprak Su Dergisi*, pp. 24-31, 2019.

[21] V. Gümüş, O. Şimşek, N. G. Soydan, M. S. Aköz, and K. Yenigün "Adana istasyonunda buharlaşmanın farklı yapay zeka yöntemleri ile tahmini." *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, 7(2), pp. 309-318, 2016.

[22] A. Ramazan, and K. Saplıoğlu, Detection of sediment transport in streams by using artificial neural networks and ANFIS methods." *Nigde Omer Halisdemir University Journal of Engineering Sciences*, 9(1), pp. 437-450, 2020.

[23] J. S. Jang, "ANFIS: adaptive-network-based fuzzy inference system." *IEEE transactions on systems, man, and cybernetics*, *23*(3), pp. 665-685.

[24] A. Özel, and M. Büyükyıldız, "Usability of artificial intelligence methods for estimation of monthly evaporation." Niğde Ömer Halisdemir University Journal of Engineering Sciences, 8(1), pp. 244-254, 2019.