Dicle University
Journal of Engineering
https://dergipark.org.tr/tr/pub/dumf
duje.dicle.edu.tr

Research Article

Estimation of missing temperature data by Artificial Neural Network (ANN)

Okan Mert KATİPOĞLU [1,*], Reşat ACAR [2]

[1]Erzincan Binali Yıldırım University, Department of Civil Engineering, Erzincan 24100, Turkey, https://orcid.org/0000-0001-6421-6087
[2]Atatürk University, Department of Civil Engineering, Erzurum 25100, Turkey, https://orcid.org/0000-0002-0653-1991

ARTICLE INFO

Article history:

Received 3 Januray 2021
Received in revised form 3 March 2021
Accepted 5 March 2021
Available online 30 March 2021

Keywords:

Missing data, Temperature, Artificial Neural Network (ANN), Network training, Network testing, and validation.

ABSTRACT

Ensuring more reliable and quality meteorological and climatological studies by providing data continuity and widening the data range. For this reason, missing values in meteorological data such as temperature, precipitation, evaporation must be completed. In this study, an artificial neural network (ANN) model was used to complete missing temperature data in the Horasan meteorology station. To establish the ANN model, monthly average temperature values of neighboring stations having similar climatic characteristics and altitude with Horasan were used as input. The monthly average temperature values of the Horasan station were used as output. Approximately 70% of the data was used for training, about 15% for testing, and about 15% for verification in the ANN model. Various statistical parameters were compared to determine the best network architecture and best model. As a result, the model's high determination coefficient (R2 = 0.99) and low mean absolute error (MAE = 0.61) showed that the ANN model can be used effectively in estimating missing temperature data.

Doi: 10.24012/dumf.852821

MAKALE BİLGİSİ

Makale geçmişi:

Geliş 3 Ocak 2021
Revizyon 3 Mart 2021
Kabul 5 Mart 2021
Online 30 Mart 2021

Anahtar Kelimeler

Eksik veri, Sıcaklık, Yapay Sinir Ağı (YSA), Sinir ağının eğitimi, Sinir ağının test ve geçerliliği.

ÖZ

Veri sürekliliğinin sağlanması ve aralığın genişletilmesi ile meteorolojik ve klimatolojik çalışmaların daha güvenilir ve kaliteli olmasını sağlamaktadır. Bu nedenle sıcaklık, yağış, buharlaşma gibi meteorolojik verilerde eksik olan değerlerin tamamlanması gerekmektedir. Bu çalışmada, Horasan meteoroloji istasyonundaki eksik sıcaklık verilerini tamamlamak için Yapay sinir ağı (YSA) modeli kullanılmıştır. YSA modelinin kurulması için Horasan ile benzer iklim özelliklerine ve rakıma sahip komşu istasyonların aylık ortalama sıcaklık değerleri girdi olarak kullanılmıştır. Horasan istasyonunun aylık ortalama sıcaklık değerleri ise çıkış olarak kullanılmıştır. YSA modelinde verilerin yaklaşık% 70'i eğitim için, yaklaşık% 15'i test için ve yaklaşık% 15'i doğrulama için kullanılmıştır. En iyi ağ mimarisini ve en iyi modeli belirlemek için çeşitli istatistiksel parametreler karşılaştırılmıştır. Sonuç olarak, modelin yüksek belirlilik katsayısı (R^2 = 0.99) ve düşük ortalama mutlak hataya (OMH = 0.61) sahip olması YSA modelinin eksik sıcaklık verilerini tahmin etmede etkin bir şekilde kullanılabileceğini göstermiştir.

* Corresponding author
Okan Mert KATİPOĞLU
✉ okatipoglu@erzincan.edu.tr

Please cite this article in press as O. M. Katipoğu, R. Acar, “Estimation of missing temperature data by Artificial Neural Network (ANN)”, DUJE, vol. 12, no.2, pp. 431-438, March 2021.

Introduction

Artificial Neural Networks (ANNs) are developed as a parallel processing modeling system, inspired by the brain work system, and have been recently applied to many fields of science. These models consist of input, output, and hidden layers and can be used to achieve high-performance models in the hydrology discipline as in other disciplines. Water resources systems and weather forecasts are composed of complex relationships that are non-linear and have many parameters. Such problems can be solved effectively thanks to ANN's ability to easily adapt to the problem [1,2]. Therefore, the ANN model was used in the estimation of missing air temperatures in this study.

Numerous studies have been conducted on the prediction of artificial neural network models and meteorological and hydrological variables. Some of those; Some of those; Güç [3] used for estimation of air temperature, Sanikhani et al. [4] used to reduce the biases of climate variables (temperature and precipitation), Vakili et al. [5] used for estimation of daily global solar radiation, Behmanesh, and Mehdizadeh [6] used for estimation of soil temperatures, Zhu et al. [7] used for river water temperature simulation, Taşar, et al. [8] used for estimation of the evaporation amount, Rahman, and Chakrabarty [9] used for estimation of transport of sediment, Yıldıran and Kandemir [10] used for estimating the amount of precipitation, Afzaal et al. [11] used for estimation of groundwater, Dalkiliç, and Hashimi, [12]; Kızılaslan, et al. [13] used for streamflow estimation.

In this study, ANN model was used to complete the missing temperature data in Horasan meteorology station. To establish the model, the monthly average temperature values of neighboring stations with the nearest, climatic characteristics and least elevation difference of the Horasan station were used as input, and missing temperature data were estimated as output.

Materials and Methods

Study Area and Data

17690 no Horasan, 17688 no Tortum, 17666 no İspir, 17099 no Ağrı, 17204 no Muş and 17718 no Tercan meteorological observation station temperature data were selected in the study area since the stations are close to each other and have similar hydrological and climate characteristics and high correlations. The data used in this study were between 1966 and 2017 and were obtained from the General Directorate of Meteorology. Detailed information on the stations used in the study is given in Table1.

Table 1. Stations used in the study

Station Name	Station Number	Latitude	Longitude	Altitude
Horasan	17690	40,04	42,17	1540
Tortum	17688	40,30	41,54	1576
İspir	17666	40,49	40,99	1223
Ağrı	17099	39,73	43,05	1646
Muş	17204	38,75	41,50	1322
Tercan	17718	39,78	40,39	1425

The distance between the meteorology observation stations and Khorasan, which are used as inputs in the artificial neural network model, is shown in Table 2.

Table 2. Distance between selected stations and Horasan

Station Name	Distance (KM)
Tortum	59,55
İspir	113,00
Ağrı	83,50
Muş	157,09
Tercan	154,82

The stations used in the study are expressed according to the Euphrates Basin boundaries (Figure 1).

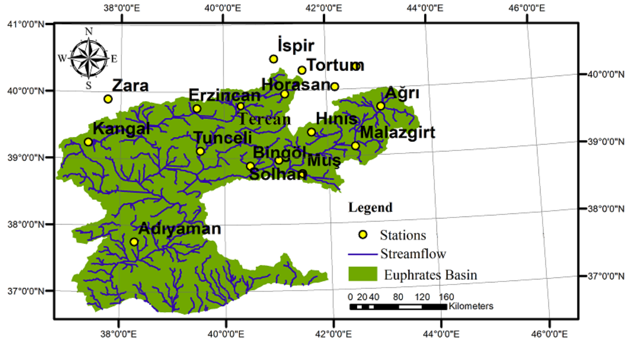

Figure 1. Euphrates basin location map

Artificial Neural Networks (ANNs)

ANNs are the computer systems that make the learning function which is the most basic feature of the human brain. They make the learning process with the help of examples. These networks consist of interconnected processing elements (Artificial nerve cells). Each connection has a weight value. The knowledge of the ANN is hidden in the weight value and spread into the network.

ANNs suggest a different calculation method than the known methods of calculation. It is possible to see the successful applications of this calculation method, which is adapted to the environment, can work with incomplete information, can make decisions about uncertainties, and tolerate errors. The interest in these networks increases day by day, although there is not a certain standard in the structure of the network to be formed and selection of the network parameters, the problems are shown only with numeric information, it is not known how education is finished and the behavior of the network cannot be explained. Especially in classification, pattern recognition, signal filtering, data compression, and optimization studies, ANNs are considered the most powerful techniques [14].

The Structure and Elements of ANN

Artificial neural networks are computer systems developed inspired by the properties of the nervous system (information generation, description, prediction, etc.). Artificial neural networks are formed by the combination of cells as in biological nervous systems and generally artificial neural network architecture is defined in 3 layers. These are;

- Input layer
- Hidden layer
- Output layer

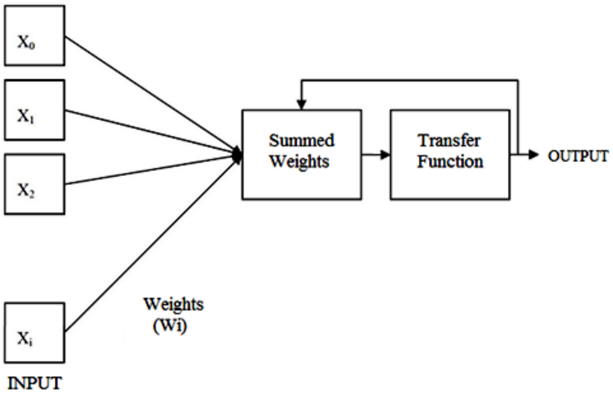

Figure 2. The model of Artificial neural network. X_i is input values, W_i, Connection weight

Information is transmitted from the input layer to the network. They are processed in interlayers and sent to the output layer. Information processing is the conversion of the incoming information to the network using the weight values of the network. The weights should be evaluated correctly for the network to produce the correct outputs for the inputs. The network needs to be trained to find the right weights. The process of determining the correct weights is called network training. Weights are initially assigned randomly. Then, when each sample is presented to the network during training, weights are changed according to the learning rule of the network. Then another sample is presented to the network and the weights are changed again. These operations are repeated until the correct outputs are produced for all the samples in the network training set [15, 16]

Determination of Model Performance

The performances of the established models have been tested with the help of different statistical criteria. These criteria are; The Determination Coefficient (R^2) and the Mean Absolute Error (MAE) values. The statistical calculations used can be calculated with the help of Equations 1 and 2, respectively. MAE is a statistical measure that measures the predictive accuracy by determining the differences between the predicted values and the observed values. The Determination Coefficient (R^2) is a statistical measure that shows how close the data is to the fitted regression line.

$$R^2 = \frac{\sum_{i=1}^{N}(x_i-\bar{x})^2-\sum_{i=1}^{N}(x_i-y_i)^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2} \qquad (1)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |(x_i - y_i)| \qquad (2)$$

In these equations, x_i shows expected (observed) values of models; y_i shows outputs of models, $x_i - y_i$: error (residue) and N: the number of data. The model with the largest R^2 (near 1) and the lowest error rate (near 0) is considered the best.

Results

The completion missing temperature data with Artificial Neural Network (ANN)

The quality, reliability, and completeness of the data used while working in the field of hydrology and meteorology is very important in terms of the correct result of the established model. Studies using incomplete data have inadequate and incorrect results. For this, the missing records must be completed with various methods before starting the study. In this study, missing temperature data were completed with an artificial neural network model by using neighboring station data.

The collection of data and analysis

In this study, the temperature data of the Tortum, İspir, Ağrı, Muş, and Tercan stations were presented as inputs to the artificial neural network model, and the missing temperature data of the Horasan station was completed. The data sets are divided into three sections: training set, test set, and validation set. About 70% of the data was used for training, about 15% for testing, and about 15% for validation.

Table 2. Altitude, climate, and correlation coefficients of the stations used in the study

Station	Correlation with Horasan	Climatic similarity with Horasan	Altitude difference with Horasan (m)
Tortum	0,98	Similar	36
İspir	0,99	Similar	317
Ağrı	0,99	Similar	106
Muş	0,99	Similar	218
Tercan	0,99	Similar	111

To complete the missing temperature data of Horasan, the stations closest to the Horasan station, with the highest correlation and with similar climatic characteristics and elevations were used.

Determining the network architecture

The information coming from the input layer is processed according to certain standards and transmitted to the output layer. The main function of the network is the hidden layer and the number of hidden layers varies from network to network in line with the purpose to be achieved [21]. In this study, since a single hidden layer is sufficient to solve the problem, the best network architecture has been chosen as a single layer architecture with the smallest training, testing and verification errors.

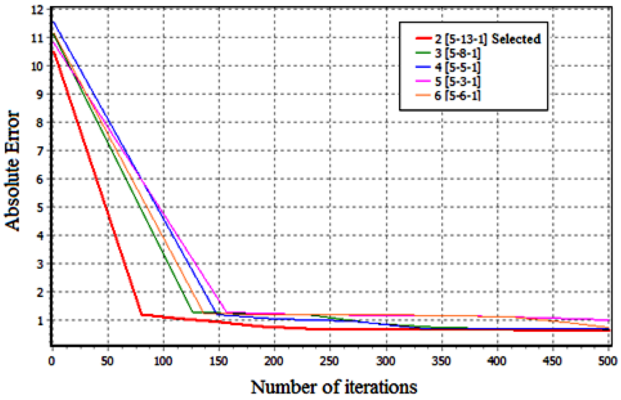

Figure 3. Comparison of the absolute errors of the top 5 network architectures

Figure 3 shows the graph of the absolute error values of the top 5 network architectures that fit the model. The smallest absolute error value of these architectures is seen in [5-13-1] architecture. Besides, when the other criteria are compared, the fitness criterion gives the best model, the highest correlation coefficient, and the determination coefficient, and the lowest of training, test, and validation errors. When the model parameters are compared, [5-13-1] architecture has been found to represent the best network structure (Table 3).

Table 3. Comparison of statistical parameters to determine the best network architecture

Network architecture	**5-13-1**	5-8-1	5-5-1	5-3-1	5-6-1
Fitness criterion	1,32	1,24	1,26	0,92	1,20
Training error	0,62	0,66	0,69	0,96	0,72
Test Error	0,76	0,80	0,79	1,08	0,83
Validation Error	0,70	0,73	0,71	1,07	0,77
Correlation	0,996	0,996	0,995	0,994	0,995
Determination	0,992	0,992	0,99	0,988	0,990

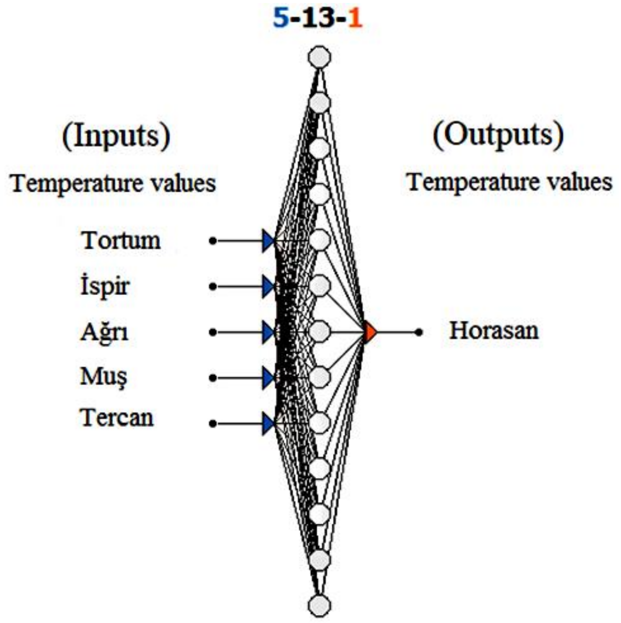

Figure 4. Selected network architecture

Figure 4 shows the most suitable network architecture (5-13-1) selected to complete missing temperature data. Here, 5 represents the number of inputs, 13 is the number of neurons in hidden layers, and 1 is the number of outputs.

Training the network

As a result of the training process, the error calculated in the artificial neural network is expected to decrease to an acceptable error rate. The training was completed with minimum error by selecting the network training process, various learning rates, the number of neurons in the hidden layer, the activation function, the number of iterations and the training algorithm.

Testing and verification

Testing is a process used to estimate the quality of a trained neural network. During this process, some of the data not used during training are presented to the trained network, as appropriate. Then the estimation error is measured in any case and used to estimate the network quality. This step determines the success of education by simply comparing the statistical data with the data predicted by ANN [17]. The correlation coefficient (R = 0,996) between the actual data and the predicted (modeled) data, while the Determination coefficient (R^2 = 0,992) and the errors are small. In addition, the correlation indicates that the model, which is estimated to be high and positive, is appropriate and correct (Table 4, Table 5).

Tablo 4 Artificial neural network model summary table

Artificial neural network parameters	
Activation function:	Logistics sigmoid
Model architecture:	(5-13-1)
Learning algorithm:	Quick propagation
Iteration number:	1000
R:	0,996
R^2:	0,992

Note. R: Correlation coefficient, R^2: Determination coefficient

Correlation and determination coefficients are used to compare predicted values and actual values and to determine the best estimation performance. The determination coefficient (R^2), which takes the values of the prepared model in the range of (0-1), indicates its fitness. This coefficient is the square of the correlation coefficient between the observed value of the dependent variable and the estimated value in the model. The value reflects how many percent of the fluctuations in the dependent variable are due to variations in the independent variable [18,19].

In Figure 5, the scatter plot of the actual temperature values and artificial neural network of Horasan and the estimated temperature values are shown. Scattering around the dots indicates high fit.

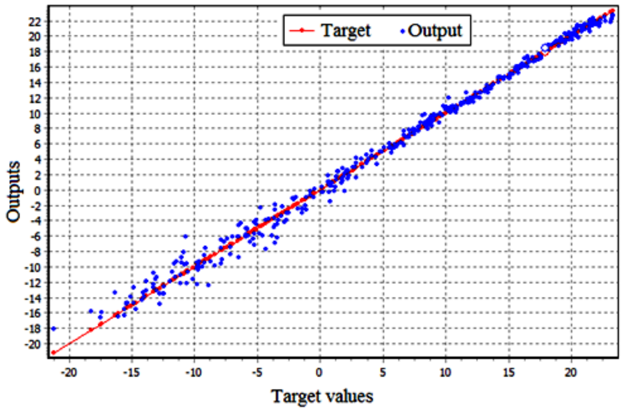

Figure 5. Comparison of targeted (actual) values and outputs (estimated values)

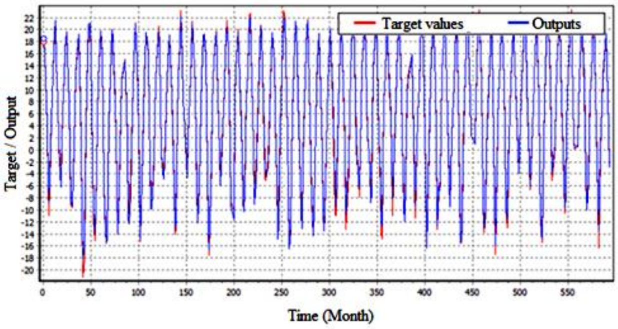

Figure 6. The relationship between targeted (real) values and artificial neural network and predicted outputs

Figure 6 shows the real temperature values of Horasan and the estimated temperature values with artificial neural network. The complete overlap of the target and output values show the success of the model.

Table 5. Test results of artificial neural network

	Target	Output	MAE
Mean	5,96	5,99	0,03
Standard deviation	11,34	11,28	0,06
Minimum	-21,20	-18,10	3,10
Maximum	23,30	22,78	0,52

Testing is a process used to estimate the quality of a trained neural network. In this process, some data that is not used during the training is presented to the trained network depending on the situation. In this study, estimation error was used to estimate network quality. To estimate the success of the analysis, targeted (real) values and output values were compared. The results of this comparison are shown in Table 5 and Figure 5. The fact that the errors are low enough shows that the model established when estimating the temperatures of the Horasan station is correct. Here the absolute error expresses the difference between the targeted values and the output values of the network.

Discussion

Dombaycı and Gölcü [20] presented the average temperature values of the previous days as input to the ANN model to estimate the daily average temperatures. In their study, Levenberg-Marquardt (LM) feed-forward backpropagation algorithm, which has 6 neurons in the hidden layer, was chosen as the most suitable model because it gives the biggest R^2 (0,99) and the smallest error rate (1,85). Besides, as a result of the study, it was determined that the estimated temperature values expressed as the output of the networks are very close to the real values. This shows that the ANN model is effective in temperature estimation. When the statistical parameters obtained as a result of the study are compared with our study, it is seen that our study is quite superior. This situation is thought to be due to the daily temperature estimation of the stated study, while monthly temperatures were used in our study. Besides, it is suggested that more effective estimation can be made by including parameters such as altitude, humidity, precipitation, and evaporation in the model for the estimation of temperatures in future studies.

Akyüz et al. [21] the average air temperature of Antalya was estimated by the artificial neural network method. In the study, real monthly average vapor pressure, monthly average relative humidity, related month, and year data were used as an input. As a result, it has been observed that the predicted values in the artificial neural network model are compatible with the actual average air temperature values. The error rate of the most suitable model obtained in their study was determined as 0.029 and the value of R^2 as 0.99. Compared with the statistical parameters obtained with our study, it is seen that it is almost the same in temperature estimation.

Conclusions

This study, it is aimed to estimate missing air temperature data with the ANN model. Accordingly, stations with similar features were used near the meteorological station with missing data. The results obtained show that a satisfactory estimate is obtained by the artificial neural network method. The quick propagation learning algorithm minimizes the error rate and is proposed to be used for temperature prediction. Statistical parameters such as the determination coefficient (R^2: 0,992) and the Mean Absolute Error (MAE: 0,03) obtained because of this study show the accuracy of the prediction.

Acknowledgments: The authors thank General Directorate of Meteorology for the observed monthly total precipitation and temperature data provided the Editor and the anonymous reviewers for their contributions on the content and development of this paper.

References

1. Şen, Z. *Artificial neural networks principles*. Water Foundation, **2004**.
2. Pielke, R.A.; Cotton, W.R.; Walko, R.E.A.; Tremback, C.J.; Lyons, W.A.; Grasso, L.D.; ... and Copeland, J.H. A comprehensive meteorological modeling system—RAMS. *Meteorology and atmospheric Physics* **1992**, *49*(1-4), 69-91.
3. Güç, R. Solar energy analysis and temperature forecast with artificial neural networks for bilecik province, Bilecik Şeyh Edebali University, Institute of science and technology, Bilecik, **2016**.
4. Sanikhani, H.; Deo, R. C.; Samui, P.; Kisi, O.; Mert, C.; Mirabbasi, R.; ... & Yaseen, Z. M. Survey of different data-intelligent modeling strategies for forecasting air temperature using geographic information as model predictors. *Computers and Electronics in Agriculture* **2018**, *152*, 242-260.
5. Vakili, M.; Sabbagh-Yazdi, S. R.; Khosrojerdi, S.; & Kalhor, K. Evaluating the effect of particulate matter pollution on estimation of daily global solar radiation using artificial neural network modeling based on meteorological data. *Journal of cleaner production* **2017**, *141*, 1275-1285.
6. Behmanesh, J; Mehdizadeh, S. Estimation of soil temperature using gene expression programming and artificial neural networks in a semiarid region. *Environmental Earth Sciences* **2017**, *76*(2), 76.
7. Zhu, S.; Heddam, S.; Nyarko, E. K.; Hadzima-Nyarko, M.; Piccolroaz, S.; Wu, S. Modeling daily water temperature for rivers: comparison between adaptive neuro-fuzzy inference systems and artificial neural networks models. *Environmental Science and Pollution Research* **2019**, *26*(1), 402-420.
8. Taşar, B.; Üneş, F.; Demirci, M.; Kaya, Y. Z. Evaporation amount estimation using artificial neural networks method, *Dicle University Journal of Engineering* **2018**, vol. 9, no. 1, pp. 543-551.
9. Rahman, S.A.; Chakrabarty, D. Sediment transport modelling in an alluvial river with artificial neural network. *Journal of Hydrology* **2020**, *588*, 125056.
10. Yıldıran A.; Kandemir, S.Y. Estimation of Rainfall Amount with Artificial Neural Networks, *Bilecik Şeyh Edebali University Journal of Science* **2018,** vol. 5, no. 2, pp. 97-104.
11. Afzaal, H.; Farooque, A. A.; Abbas, F.; Acharya, B.; Esau, T. Groundwater estimation from major physical hydrology components using artificial neural networks and deep learning. *Water* 2020, *12*(1), 5.
12. Dalkiliç, H. Y.; Hashimi, S. A. Prediction of daily streamflow using artificial neural networks (ANNs), wavelet neural networks (WNNs), and adaptive neuro-fuzzy inference system (ANFIS) models. *Water Supply* **2020**, *20*(4), 1396-1408.
13. Kızılaslan, M.; Sağın, F.; Doğan, E.; Sönmez, O. Estimation of lower Sakarya River flow using artificial neural networks," *Sakarya University Journal of the Institute of Science* **2014** vol. 18, no. 2, pp. 99-103.
14. Minns, A.; Hall, M. Artificial neural networks as rainfall-runoff models, *Hydrological sciences journal* **1996**, vol. 41, no. 3, pp. 399-417.
15. Bishop, C.M. Neural networks and their applications, *Review of scientific instruments* **1994**, vol. 65, no. 6, pp. 1803-1832.
16. Campolo, M.; Andreussi, P; Soldati, A. River flood forecasting with a neural network model, *Water resources research* **1999** vol. 35, no. 4, pp. 1191-1197.
17. Ilie, C.; Ilie, M.; Melnic, L.; Topalu, A.-M. Estimating the Romanian Economic Sentiment Indicator Using Artificial Intelligence Techniques. *Journal of Eastern Europe Research in Business & Economics* **2012**, 1.
18. Hocking, R.R. A Biometrics invited paper. The analysis and selection of variables in linear regression. *Biometrics*, **1976**, 32(1), 1-49.

19. Lindley, D.V. Regression and correlation analysis. In *Time Series and Statistics* Palgrave Macmillan, London. **1990**; pp. 237-243.

20. Dombaycı, Ö. A.; and Gölcü, M. Daily means ambient temperature prediction using artificial neural network method: A case study of Turkey. *Renewable Energy*, **2009**, *34*(4), 1158-1161.

21. Akyüz, A. Ö.; Kumaş, K.; Ayan, M,; Güngör, A. Antalya İli Meteorolojik Verileri Yardımıyla Hava Sıcaklığının Yapay Sinir Ağları Metodu ile Tahmini. *Gümüşhane Üniversitesi Fen Bilimleri Enstitüsü Dergisi* **2020**, *10*(1), 146-154.