SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ DERGİSİ

Sakarya University Journal of Science
SAUJS

ISSN 1301-4048 e-ISSN 2147-835X Period Bimonthly Founded 1997 Publisher Sakarya University
http://www.saujs.sakarya.edu.tr/

Title: Application of Soft Computing Techniques in River Flow Modeling

Authors: Sefa Nur YESİLYURT, Hüseyin Yildirim DALKILIÇ, Pijush SAMUI

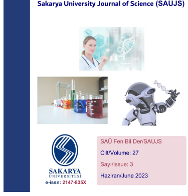
Recieved: 2022-08-10 00:00:00

Accepted: 2023-03-22 00:00:00

Article Type: Research Article

Volume: 27
Issue: 3
Month: June
Year: 2023
Pages: 643-659

How to cite
Sefa Nur YESİLYURT, Hüseyin Yildirim DALKILIÇ, Pijush SAMUI; (2023), Application of Soft Computing Techniques in River Flow Modeling. Sakarya University Journal of Science, 27(3), 643-659, DOI: 10.16984/saufenbilder.1152982
Access link
https://dergipark.org.tr/en/pub/saufenbilder/issue/78131/1152982



Sakarya University Journal of Science 27(3), 643-659, 2023

SAKARYA ÜNİVERSİTESİ

Application of Soft Computing Techniques in River Flow Modeling

Sefa Nur YESILYURT [*1], Huseyin Yildirim DALKILIC[1], Pijush SAMUI[2]

Abstract

Modeling of data is critical in the analysis and evaluation of hydrological behavior. River flow data is one of the most important data in explaining hydrology. Management of water resources; It takes place in the literature as an area that needs to be investigated in order to provide early warning for undesirable situations such as floods and drought. For this reason, it is of important to develop different techniques for the estimation and modeling of river flow or to make comparisons between techniques. In this study, the flow data of fourteen stations located in the Euphrates-Tigris basin between 1981 and 2010 were used. Adaptive Network Based Fuzzy Inference Systems (ANFIS), Support Vector Machine (SVM) techniques that are frequently used in the literature, and newly introduced Gaussian Process Regression (GPR), Extreme Learning Machine (ELM) and Emotional Neural Network (ENN) artificial intelligence techniques are compared. In addition, considering all performance indices, it was determined which technique gave better results with rank analysis. Although all models worked well, it was seen that the methods were ranked as ELM, GPR, ENN, SVM and ANFIS starting from the best. This has shown that ELM, GPR and ENN methods, which have been used recently in flow modeling, give better results than traditional methods with complex structures. In addition, flow values were used in the whole study and these values were examined in 3 different combinations. It was seen that the model structure that gave the best performance was the model structure that used the flow data from one, two and three days ago as an estimator. The results were analyzed with a Taylor diagram and time series graphs.

Keywords: ANFIS, ELM, ENN, GPR, SVM

1. INTRODUCTION

Water, the main component of life, is an indispensable resource used to provide the energy required for living things to survive. In cases such as efficient use of existing water resources and development of water structures, basin modeling and estimation are required. At the same time, the fact that there are many unknown factors in the occurrence of

---

* Corresponding author: sefa.yesilyurt@erzincan.edu.tr (S.N. YESİLYURT)
[1] Erzincan Binali Yildirim University
[2] National Institute of Technology Patna
E-mail: hydalkilic@erzincan.edu.tr, pijush@nitp.ac.in
ORCID: https://orcid.org/ 0000-0001-6173-3038, https://orcid.org/ 0000-0002-4405-9341, https://orcid.org/ 0000-0003-2906-6479

hydrological events, instabilities in the work field, irregularities in river systems and flow data have made it necessary for researchers to create models and make estimations for the future. These estimations, which can be made through some mathematical methods, provide more successful results through artificial intelligence techniques and fuzzy logic methods, and thus, can be modeled within a shorter time. When the studies on this subject are examined, it is seen that Zhou et al. used Genetic Algorithm (GA)- Least Square Estimator (GL) and adaptively developed Adaptive Network Based Fuzzy Inference Systems (ANFIS); R (ANFIS) and R-ANFIS (GL) for river flow modeling [1].

It has been observed that the R-ANFIS (GL) model gives better results when time series are used by Zhou et al. [1]. In another study by He et al., Artificial Neural Network (ANN), ANFIS and Support Vector Machine (SVM) based on three different data were used for river flow estimation. It was seen that the SVM method gave better results and it was stated that these methods could be used in regions with complex topography [2]. In their study Yaseen et al. improved the Extreme Learning Machine (ELM) model (improved ELM) as EELM, compared it with SVM, and found that the improved ELM model gives much better results [3]. According to Yasin et al. Emotional Neural Network (ENN), Multivariate Adaptive Regression Splines (MARS), Minimax Probability Regression (MPMR) and Fitness Vector Machine (RVM) methods were used for hourly river flow modeling. He also proposed the ENN algorithm for the first time for river flow modeling. It was also stated that the ELM model was found to be far superior to other models [4].

Wang et al. analyzed the deterministic and stochastic components of the modeling at the same time and examined four different station data between 1971–2010. They conducted a performance comparison between hybrid models and found that SETAR (Self-Exciting Threshold Autoregressive) model gave better results [5]. Sun et al. made river flow estimations using Gaussian Process Regression (GPR) for MOPEX basins. They observed that the GPR model worked well for long-term flow data [6]. Yaseen et al. analyzed the artificial intelligence models for flow estimations and determined the advantages and disadvantages of the models. They examined the possible applications of artificial intelligence and conducted comprehensive literature research [7]. Khadangi et al. applied ANFIS and Radial Base Function (RBF) methods for daily river flow modeling in their study and found that ANFIS provided much better performance [8].

The ELM construct created to eliminate the need for iterative tuning of hidden neuron parameters in traditional models was proposed by Huang et al. The model was first used by Siqueira et al. for river flow modeling. They observed that the model is suitable for hydraulic power plants in Brazil and for river flow studies. At the same time, the ELM structure was developed within time and different ELM structures were created for different studies [9, 10]. Yaseen et al. used ELM and ANFIS to estimate river flows in their studies and observed that the improved ELM gave better results when compared to these techniques. In addition, in a study conducted in Iran, a semi-arid region, Generalized Regression Neural Network (GNRR), SVM, and ELM were compared and it was concluded that ELM gave better results [11].

In another study Adnan et al., in which ELM was used as Optimally Pruned ELM (OP-ELM), ANFIS-PSO (Particle Swarm Optimization), MARS, and M5 model tree (M5Tree) techniques were compared by cross-validation and it was concluded that OP-ELM method could be used successfully in daily stream flow estimation [12]. The ENN structure, which takes emotional parameters into account in addition to other models that simulate the brain structure in modeling studies, was developed by Rumelhart, and was used in river flow modeling for the first time by Yaseen et al., ENN was used in the study to create an hourly river flow model, it was compared with other well-structured machine learning methods, and it was found that ENN performed better. In a study in which SVM, developed by Rumelhart, was used for river flow modeling, ELM was compared with Artificial Intelligence (AI),

Genetic Programming (GP) and SVM and it was observed that ELM method gave faster and better results in river flow forecasting than the other methods [4, 13, 14]. Sun et al. studied the monthly estimation of GPR, compared GPR with Autoregressive Moving Average with Exogenous variables (ARMAX) and multilayer perceptron (MLP), used for more than four hundred stations in the USA, and concluded that GPR performed better [6].

In this study, conducted for river flow forecasting and modeling, which is of great importance for water resources engineering, it is aimed to find the best results in the shortest time in river flow modeling by comparing widely used methods such as ANFIS and SVM with the rarely used ones such as ELM, GPR and ENN methods, to find the membership functions in traditional methods by trial and error, and to eliminate undesirable conditions.

2. MATERIALS AND METHODS

There are many methods for flow modeling which have various advantages and disadvantages. In the current study, ANFIS and SVM methods, which are known to give good results and are frequently used in flow modeling, have been compared to more recent methods. The advantages and disadvantages of these methods have been taken into account. Data from fourteen flow observation stations located in the Euphrates-Tigris Basin were used in the study. ELM, which has good learning capacity and generalization performance and is much faster than traditional algorithms, ENN, which has high application convenience and efficiency, and GPR, which can seamlessly integrate various machine learning tasks such as hyperparameter estimation, model training and uncertainty estimation, have been preferred.

2.1. Adaptive Neuro-Fuzzy Inference System (ANFIS)

ANFIS, based on Takagi-SugenoKang inference system, was developed by Jang to model nonlinear functions, determine nonlinear components in the control system, and predict the chaotic time series [15-16]. The fuzzy logic inference system evaluated in ANFIS is transformed into adaptive networks and the most suitable condition is created through a learning algorithm. Neural adaptive learning techniques develop a model that “learns” the related system by using the data set selected for the fuzzy modeling.

In other words, ANFIS uses the input/output data set and the backpropagation algorithm used in artificial neural networks alone or together with the least-squares method, and thus, by regulating the membership functions parameters, creates a Fuzzy İnference System (FIS). This regulation allows the fuzzy system to learn the relevant system with the help of the data that it has modeled. Namely, it customizes/adapts itself to the data that will be modeled. Thanks to this structure, ANFIS has both used the environmental information about the system and gained the ability to update itself using the input and output data related to the system [17] (Figure 1).

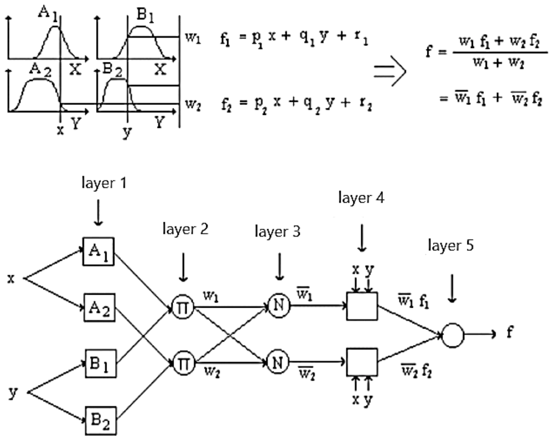


Figure 1 ANFIS Structure [18]

While the biggest advantages of the ANFIS model can be regarded as its efficiency in mathematical analysis, success in adaptation and successful conclusion in numerical data, too much human intervention can be supposed as a disadvantage since the training of ANFIS parameters takes quite a long time and the model has a structure with many rules [18].

2.2. Extreme Learning Machine (ELM)

ELM is a fully connected artificial neural network model developed by Huang et al. and consists of an input layer, a hidden layer and an output layer [9]. Unlike the commonly used gradient-based network structures, ELM, whose input weights and threshold values are randomly generated but output weights are analytically generated, creates an analytical equation of the model beyond finding the model weights, and thus, it prevents error clogging at a local point and removes the problem of learning process that takes a long time as in the other methods. In this way, it provides better performance compared to other methods and speeds up the model production process.

At the same time, other learning algorithms sometimes have to apply procedures such as stopping the training process of the model earlier, adding regulation parameters, breaking weights or using validity sets as they may encounter undesirable situations such as improper learning rate, excessive learning and memorization, and stuck in local minimums, whereas ELM reaches the solution directly without any intermediate processing. In addition to all these advantages, the structure of the ELM method, which offers the possibility to use many activation functions which can be derivative, underivative or discrete, consists of the input layer where the data is read, the output layer where the classes are determined and the hidden layer where the intermediate operations are conducted, as shown in Figure 2.

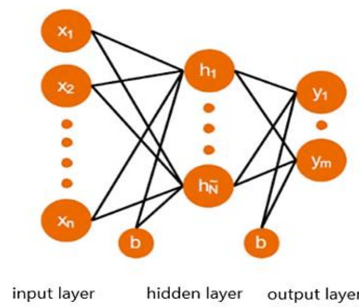


Figure 2 Algorithm of Extreme Learning Machine [19]

The ELM structure can calculate the output weight with the Moore-Penrose generalized inverse latent matrix without any need for iterative optimization. If L stands for the hidden node, β_j symbolizes the output neurons, and jth is symbolized as the weight value connecting the hidden neurons, then the ELM structure can be expressed as;

$$\sum_{j=1}^{L} \beta_j h_j(x_i) = y_i, \qquad i = 1, \dots, N, \qquad (1)$$

Mapping the properties for J_{th} hidden node output $h_j(x_i)$; is

$$h_j(x_i) = \frac{1}{1+exp\left(-\left(w_j^T x_i + b_j\right)\right)} \qquad (2)$$

w_j refers to the weight vector connecting input neurons used in this equation,

$W_j = [W_{j1}, \ldots, W_{jD}]^T \in \mathbb{R}^D$ and J_{ht} hidden neuron, and b_j is expressed as trend (deviation) term.

$$H\beta = y, \beta = [\beta_1, \ldots, \beta_L]^T \in \mathbb{R}^L, \ y = [y_1, \ldots, y_N]^T \in \mathbb{R}^N \tag{3}$$

$$H(w_1, \ldots, w_L, b_1, \ldots, b_L, x_1, \ldots, x_N) = \begin{bmatrix} h_1(x_1) & \ldots & h_L(x_1) \\ \vdots & \ddots & \vdots \\ h_1(x_N) & \ldots & h_L(x_N) \end{bmatrix} \in \mathbb{R}^{N*L} \tag{4}$$

H used in these equations denotes the hidden layer output matrix. ELM chooses the case with the minimum error and the lowest output weight among different traditional learning algorithms. Randomly initialized w_j hidden node parameters and bj is (j = 1; . . .; L) and the least squares solution of equation 1 is as follows;

$$\beta = H^{\dagger} \tag{5}$$

Here † Moore-Penrose denotes the generalized opposite. Decision function to be created to write $\hat{x}$, which is the new test example of ELM structure, can be expressed as [19];

$$\hat{y} = sign\ (h(\hat{x})\beta) \tag{6}$$

2.3. Emotional Neural Network Algorithm (EmNN-ENN)

This section describes the emotional neural network algorithm (EmNN). EmNN is based on the emotional back-propagation algorithm (EmBP- emotional backpropagation), which is a modified version of the traditional back-propagation algorithm (BP-back propagation). As stated by David and James, the BP method is often preferred because of its simplicity of application and its rapid operation, especially when it has a sufficient database. EmBP is described according to the information flow layers of the three-layer EmNN algorithm [20]. Layers of the EmNN algorithm are called as follows:

i: input layer with neurons

h: hidden layer with neurons

j: output layer with neurons

(Figure 3) shows the process for EmNN feed-forward calculation [21–22];

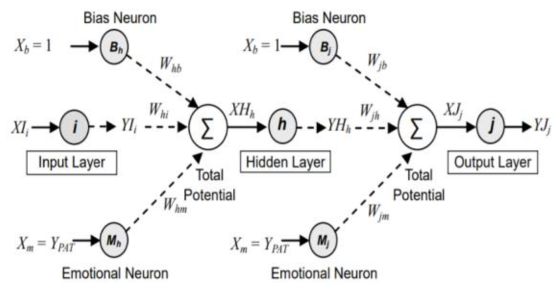


Figure 3 Process for EmNN feed-forward calculation [21-22]

Emotional parameters are used in conjunction with the current learning coefficient (η) and α momentum ratio. (μ) is defined as the anxiety coefficient and k is defined as the confidence coefficient, and it is observed how these two parameters act when learning each new task. Anxiety level decreases as confidence level increases. Both coefficients have normalized values between 0 and 1. The level of anxiety depends on the mean value of the input pattern and the error indicator for each period. The average input value used here must always be positive because the pixel values are normalized to values between 0-1. At the same time, the error indication may provide negative feedback if an unstable condition exists there. In this case, the heuristic network will be unreliable and unstable, similar to traditional networks. Therefore, three parameters are arranged until stable learning is found. These three parameters stand for the learning rate, momentum ratio and the count of hidden neurons. Therefore, as learning progresses, the anxiety rate decreases and the value of the confidence coefficient increases [12].

The anxiety coefficient can be defined as follows:

$$\mu = Y_{AvPAT} + E \tag{7}$$

Y_{AvPAT} is defined as the mean value of the patterns presented in the EmNN algorithm.
If p represents the pattern index, N is the total number of patterns presented in a period, and E is the feedback error, then;

$$Y_{AvPAT} = \frac{\sum_{p=1}^{N_P} Y_{PAT}}{N} \tag{8}$$

$$E = \frac{\sum_{j=1}^{N_j} (T_j - Y_{Jj})^2}{N_p . N_j} \tag{9}$$

k confidence coefficient;

$$k = \mu_0 - \mu_i \tag{10}$$

μ_0: the value of anxiety coefficient at the end of the first iteration

μ_i : coefficient of anxiety at the end of subsequent iterations

2.4.Support Vector Machine (SVM)

SVM, an algorithm based on optimization, was designed by Vapnik as a classification algorithm that minimizes the error [23]. Later, the algorithm started to be used for regression purposes with the name SVR. Since SVM depends on core functions, it is considered a nonparametric technique. SVM, created by including the maximum value in the structure, has become more efficient than other regression models. When the weight vector in the structure is expressed as w and the error value as ε, the minimization process is expressed based on the following equations;
$\min 1/2\|w\|^2$

$$y_i - (w, x_i + b) \leq \varepsilon \,, \; (w, x_i + b) - y_i \leq \varepsilon \tag{11}$$

When x is a point on the hyperplane and b is called a bias, then the constraint equation is as follows;

$$f(x) = y_i(w, x_i + b) \tag{12}$$

If the model margin value is wanted to be calculated to keep all data in it, minimization is used. However, it is not possible to use all values in this way. In this case, slack variables are used (ξi, $\xi i*$).

$$\min 1/2\|w\|^2 + C \sum_{i=1}^{N} (\xi i + \xi i^*) \tag{13}$$

Equation is formed depending on the $y_i - (w, x_i + b) \leq \varepsilon + \xi i$ ve $(w, x_i + b) - y_i \leq \varepsilon + \xi i$ equations. C>0 constant is used and values where equation f is greater than ± ε are tolerated as shown in Figure 4 [24-28].

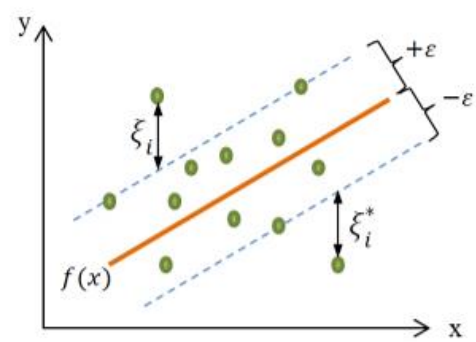


Figure 4 An example for SVM model structure [29]

SVM, which is widely preferred due to its ease of application and compatibility with both linear and nonlinear data, also has disadvantages such as difficulties in interpreting model parameters and long duration of model training [30-31].

2.5. Gauss Process Regression (GPR)

GPR, a non-parametric model suitable for use in solving nonlinear regression problems, is based on the conversion of prior functions to posterior functions in Gaussian distribution GP describes the probability distribution on functions and when M (x) refers to mean, K(x,x') refers to covariance function, then the

$$f(x) \sim GP(m(x), K(x,x')) \quad (14)$$

equation is formed. In this equation m(x) and K(x,x') are expressed as follows;

$$m(x) = \mathbb{E}[f(x)] \quad (15)$$

$$K(x,x') = \mathbb{E}[(f(x) - m(x))((f(x') - m(x'))^T] \quad (16)$$

If θ_f indicates x- scaling (amplitude) and θ_t indicates y- scaling (length), then the covariance function is expressed with the equation below [17-26-32-33];

$$K =. \theta_f^2 \exp\left(-\frac{1}{\theta_t^2}\|x - x'\|^2\right) \quad (17)$$

The covariance matrix is as follows;

$$K = Ƙ((x_1, \ldots, x_n), (x_1, \ldots, x_n)) = \begin{bmatrix} Ƙ(x_1, x_1) & Ƙ(x_1, x_2) & \ldots & Ƙ(x_1, x_n) \\ Ƙ(x_2, x_1) & Ƙ(x_2, x_2) & \ldots & Ƙ(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ Ƙ(x_n, x_1) & Ƙ(x_n, x_2) & \ldots & Ƙ(x_n, x_n) \end{bmatrix} \quad (18)$$

2.6.Model development-evaluation criteria and rank analysis

Four different performance indices were used to evaluate the performance of the developed models. These indices, called correlation coefficient (R), Root Mean Squared Error (RMSE), Mean Squared Error (MSE), Mean Absolute Error (MAE), can be obtained with the help of the following equations;

$$R = \sqrt{1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2}} \quad (19)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2} \quad (20)$$

$$MSE = RMSE^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2 \quad (21)$$

$$MAE = \frac{1}{N}\sum_{j=1}^{N}|y_i - y_j| \quad (22)$$

y refers to the measured value, $\bar{y}$ refers to the average of measured values, and N refers to the total number of data. R-value may have the best value of 1 and RMSE, MAE and MSE may have the best value of 0 [30, 34-36]. The performance criteria have been chosen in a way that best suits the model structure and data. It has been aimed that the results could be easily interpreted and the data could be fully handled. The failures in the model which may be caused by the outlier structure of the data have been ignored by the MAE, and the case that the loss functions create has been taken into account calculating the RMSE and MSE. Considering that the data are not linear, rank analysis has been applied to evaluate both cases.

Although there are many statistical methods, rank analysis has been preferred because of its suitability, easy application and interpretation. All model evaluation criteria have affected the final result and the models have been fully evaluated thanks to rank analysis. Rank analysis is a method applied to determine the best-performing model among the models by considering all evaluation criteria. This method, aiming to determine the performance evaluation score of the models and to find the model that gives the best result, is performed by assigning a rank to the models according to their proximity to the best value for each data set, and collecting and comparing the scores for all data sets. If R_i is represented as the rank value in the selected model of each data set and n is the number of models, the total rank value is determined by the equation that follows [37].

$$Modal\ Total\ Ranke = \sum_{i=1}^{n} R_i \quad (23)$$

3. STUDY AREA AND DATA

The Euphrates-Tigris basin, consisting of the Euphrates and Tigris Rivers in eastern Turkey, has a total river basin area of 184,914 km^2, of which 127,300 km^2 of the Euphrates Basin and 57,614 km^2 of the Tigris Basin (Figure 5). Examination of this basin, which is the largest drainage area in Turkey and consisting of the Euphrates River, the longest river in Western Asia, and the Tigris River, the second largest river in Western Asia, is of great importance since its average annual flow value is 52.94 km^3, its average annual output is 21.4l sec / km^2 and its annual average energy generation potential is 54.7 GWh. At the same time, the Euphrates-Tigris basin is also very important for the riparian countries.

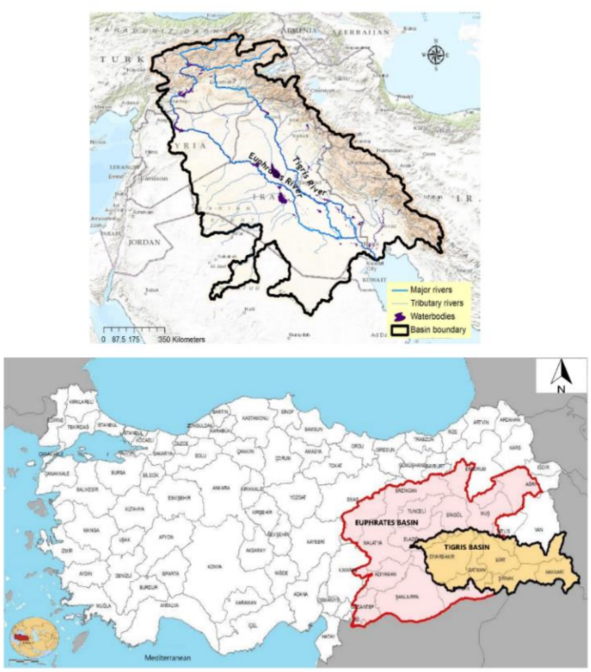


Figure 5 Euphrates-Tigris River Basin bordering on Turkey and riparian countries and the part of the basin in Turkey (examined in this study) [38]

Euphrates-Tigris Basin, in addition to these important features, has the most complete daily stream data of all the basins in Turkey. This is crucial for getting better and more reliable results with more data. Among the many stations, 14 were selected to standardize global assessment and climate monitoring studies, and the stream data averages, standard deviation values, minimum and maximum values of those stations between 1981–2010 are shown in a Table (Directorate General for S Hydraulic Works (dsi) (Turkey) (Table 1) (Figure 6) [39]. Intergovernmental Panel on Climate Change–IPCC projection reference interval has been determined to be between 1981–2010. As the data after 2010 have been thought to provide similar results, the data between the years 1981–2010 have been selected. The data have been arranged as 70% training, 30% testing, taking into account the rates at which the best result was achieved in the experiments.

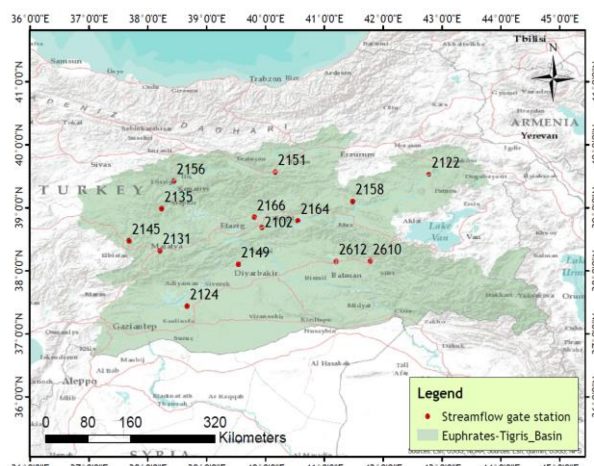


Figure 6 Selected stream observation stations in the Euphrates-Tigris Basin

Table 1 Selected stream stations in the Euphrates-Tigris Basin

Station number	Name	Longitude-latitude	Mean (flow) (m^3/sn)	Max (flow) (m^3/s)	Min (flow) (m^3/s)	Standard deviation (flow)
2102	Murat Rıver - Palu	(39° 56' 22" E - 38° 41' 49" N)	179,23	997	12,1	207,606
2122	Murat Rıver- Tutak	(42° 46' 49" E - 39° 32' 19" N)	47,48	821	1,97	73,041
2124	Tohma Bourn - Yazıkoy	(37° 26' 33" E - 38° 40' 21" N)	6,605	59,8	0,425	3,855
2131	Bey Stream - Kılayık	(38° 12' 36" E - 38° 19' 21" N)	1,343	38,8	0,11	1,894
2135	Bulam Stream - Fatopasa	(38° 14' 13" E - 37° 59' 38" N)	3,624	27,3	0,844	2,438
2145	Tohma Bourn - Hısarcık	(37° 41' 08" E - 38° 28' 32" N)	20,019	251	5,53	13,285
2149	Munzur Bourn - Mıskısag	(39° 32' 35" E - 39° 06' 29" N)	24,714	274	5,53	23,045
2151	Fırat Rıver - Demirkapı (Sansa)	(40° 10' 05" E - 39° 34' 41" N)	58,863	712	4,07	74,378
2156	Karasu - Asagıkagdarıc	(38° 26' 55" E - 39° 25' 57" N)	150,9272	980	54,8	116,844
2158	Bingöl Stream - Abdurrahman paşa Brıdge	(41° 29' 14" E - 39° 06' 30" N)	18,4965	338	1,3	29,181
2164	Goynuk Stream - Çayagzı	(40° 33' 17" E - 38° 48' 06" N)	32,497	630	0,45	56,143
2166	Perı Bourn - Logmar	(39° 48' 50" E - 38° 51' 31" N)	76,742	967	0,55	96,458
2610	Bıtlıs Stream - Baykan	(41° 46' 57" E - 38° 09' 41" N)	17,969	420	1,95	24,602
2612	Batman stream - Malabadı Brıdge	(41° 12' 16" E - 38° 09' 16" N)	112,848	990	0,015	150,300

4. RESULTS

For flow data, it was obtained from the Flow Observation Yearbooks published as open access by the Turkish State Hydraulic Works. In this study conducted with the data of 14 stations in the Euphrates-Tigris Basin, daily stream data were divided into two as 70% training and 30% testing. While choosing the training/testing ratio, the tests conducted with limited data had been taken into consideration and the most successful ratio has been used. In addition, the studies that used a low number of variables with a large amount of data were examined and it was found that 70% training and 30% testing rates were successful [40–43].

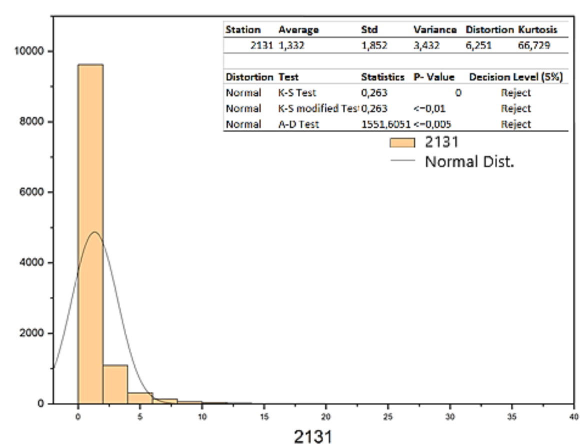


Figure 7 Normal distribution graph of station 2131

In addition, in order to examine the correlation status of the flow data, firstly, the compatibility of the data with the normal distribution was examined. It was seen that the data did not fit the normal distribution and as an example, the distribution graph of station 2151 is given in Figure 7.

Since it was seen that the data did not fit the normal distribution, input selection was made by Spearman's correlation analysis. As an example, the correlation matrix of station 2131 is given in Figure 8.

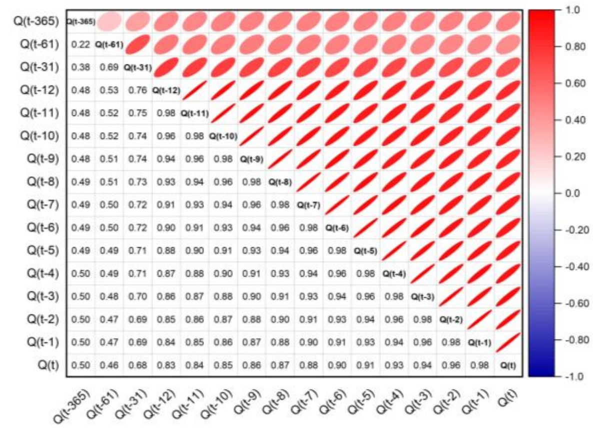


Figure 8 Correlation matrix of station 2131

Table 2 Model Results of Murat River- Palu (2131) station.

		TRAIN								TEST									
	M	**RMSE**	*Rn*	**MSE**	*Rn*	**R**	*Rn*	**MAE**	*Rn*	**RMSE**	*Rn*	**MSE**	*Rn*	**R**	*Rn*	**MAE**	*Rn*	**TP**	**TP**
ELM	M1	0,627	4	0,393	4	0,953	4	0,000	5	0,662	1	0,438	1	0,844	1	0,016	5	25	
	M2	0,620	4	0,384	4	0,957	4	0,000	5	0,640	2	0,410	2	0,856	2	0,020	3	26	81
	M3	0,631	3	0,398	4	0,951	3	0,000	5	0,639	4	0,408	4	0,861	3	0,019	4	30	
ANFIS	M1	0,635	3	0,404	3	0,952	3	0,006	4	0,622	5	0,386	5	0,867	4	0,020	3	30	
	M2	0,683	1	0,466	1	0,942	1	0,013	4	0,613	5	0,375	5	0,873	4	0,035	2	23	68
	M3	0,767	1	0,589	1	0,931	1	0,011	4	0,658	2	0,433	2	0,850	2	0,042	2	15	
SVM	M1	0,645	1	0,416	1	0,949	2	0,160	2	0,633	3	0,401	3	0,863	3	0,024	2	17	
	M2	0,646	2	0,418	2	0,949	3	0,160	2	0,639	3	0,409	3	0,864	3	0,018	4	22	60
	M3	0,640	2	0,410	2	0,949	2	0,157	2	0,642	3	0,412	3	0,862	4	0,023	3	21	
GPR	M1	0,528	5	0,279	5	0,964	5	0,140	3	0,657	2	0,432	2	0,858	2	0,019	4	28	
	M2	0,249	5	0,062	5	0,990	5	0,068	3	0,678	1	0,460	1	0,842	1	0,011	5	26	80
	M3	0,177	5	0,031	5	0,995	5	0,042	3	0,683	1	0,466	1	0,846	1	0,012	5	26	
ENN	M1	0,644	2	0,415	2	0,948	1	0,173	1	0,632	4	0,400	4	0,886	5	0,170	1	20	
	M2	0,633	3	0,401	3	0,950	2	0,167	1	0,625	4	0,390	4	0,888	5	0,167	1	23	71
	M3	0,624	4	0,390	3	0,951	4	0,165	1	0,634	5	0,402	5	0,884	5	0,167	1	28	

The first of these input combinations uses the stream data from a month ago (Q(t-1)) as input, and includes the current stream data as output (Q(t)), the second combination comprises Q(t-2)+Q(t-1) input data and Q(t) output data, and the third combination contains Q(t-3),Q(t-2),Q(t-1) input data and Q(t) output data. Modeling results of station 2131 made through these combinations are given in Table 2. As can be seen in the Table, according to the results of R, RMSE, MSE and MAE, rank analysis was performed both between models and between data set combinations, and it was observed that the ELM model gave the best results for station 2131, while the best result among data set combinations was found to be input Q(t-2)+Q(t-1)/ output Q(t) combination (Q (t-*i*): flow data *i* days ago). Owing to the large number of data used in the study, limited data sets have been tried first. Experimental models have been created with the previous period data used annually, monthly and daily. Since daily data and combinations used in the study have given better results in the experiments, the models have been created with these data. Due to different structures of performance evaluation criteria and no superiority among them, the final decision had been taken by considering all performance evaluation indexes (Table 2, Table 3).

Table 3 Rank value table for the 14 stations.

TOTAL RANK (Evaluation According To The Method)															
Station	**Q(t–1)_–Q(t)**					**Q(t–1)+Q(t–1)_–Q(t)**					**Q(t–1)+Q(t–2)+Q(t–1)_–Q(t)**				
	ELM	**ANFIS**	**SVM**	**GPR**	**ENN**	**ELM**	**ANFIS**	**SVM**	**GPR**	**ENN**	**ELM**	**ANFIS**	**SVM**	**GPR**	**ENN**
2102	31	28	19	25	17	32	11	27	29	21	26	12	26	30	26
2122	27	31	16	25	21	34	14	20	27	25	34	14	20	28	24
2124	19	31	23	26	21	33	16	26	22	23	26	16	25	23	30
2131	25	30	17	28	20	26	23	22	26	23	30	15	21	26	28
2135	24	24	21	30	21	30	21	20	27	22	25	15	27	23	30
2145	28	26	18	25	23	29	12	26	28	25	27	12	26	28	27
2149	30	24	18	29	19	33	12	21	28	26	31	12	21	31	25
2151	24	31	18	25	22	34	12	22	27	25	35	13	19	30	23
2156	30	21	18	25	26	32	14	22	27	25	33	11	23	28	25
2158	26	29	20	26	19	35	20	17	26	22	34	11	21	28	26
2164	31	29	16	26	18	35	22	14	26	23	33	11	21	31	24
2166	25	29	24	26	16	36	24	18	25	17	34	15	24	26	21
2610	32	28	17	25	18	34	24	14	26	22	33	13	19	29	26
2612	24	26	23	25	22	27	24	23	24	22	26	26	22	24	22
Total	376	387	268	366	283	450	249	292	368	321	427	196	315	385	357
Comb. Total	**Total ELM**		1253	**Total ANFIS**		832	**Total SVM**		875	**Total GPR**		1119	**Total ENN**		961

TOTAL RANK (Evaluation by Data Combination)			
Station	**Q(t–1)_–Q(t)**	**Q(t–1)+Q(t–1)_–Q(t)**	**Q(t–1)+Q(t–2)+Q(t–1)_–Q(t)**
2102	73	95	72
2122	66	83	91
2124	64	86	90
2131	75	87	78
2135	72	79	89
2145	68	87	85
2149	63	89	88
2151	61	75	104
2156	66	81	93
2158	67	82	91
2164	72	72	96
2166	83	73	84
2610	67	75	98
2612	69	81	90
Total	**966**	**1145**	**1249**

In addition, rank values of all stations are given in Table 3, out of five points for five methods and out of three points for three data combinations.

Taking the 14 stations selected for the Euphrates-Tigris basin into consideration, it is seen that the model performance ranking appears to be ELM, GPR, ENN, SVM, ANFIS, which proves the eligibility of ELM, GPR and ENN techniques, which are rarely used for river flow. At the same time, this result shows that the problems and uncertainties in commonly used ANFIS and SVM models are solved in ELM, GPR and ENN models. Moreover, it is seen that the best data set combination is the one that takes the stream data from 1, 2 and/or 3 days ago as input and the current stream data as output. In order to better understand the results, the Taylor diagram of station 2131, which is taken as an example, is presented in Figure 9.

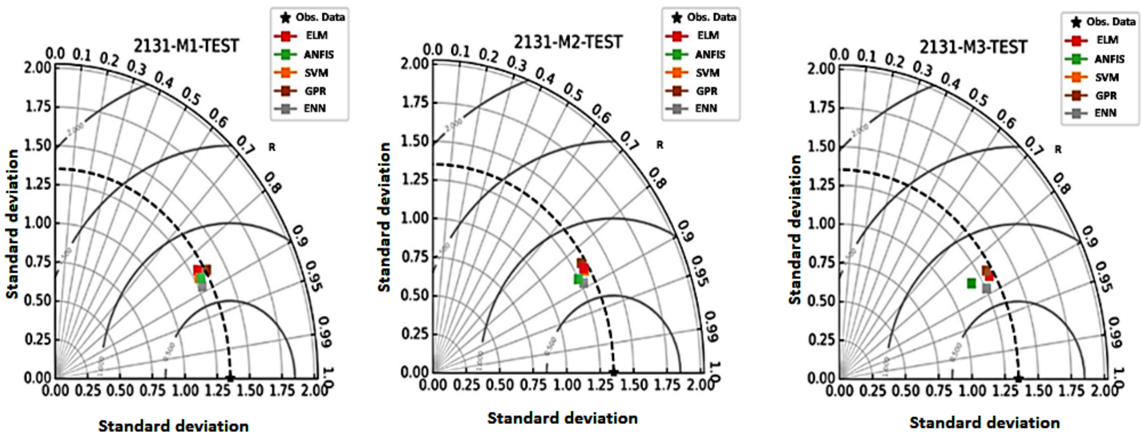


Figure 9 Taylor diagram of station 2131

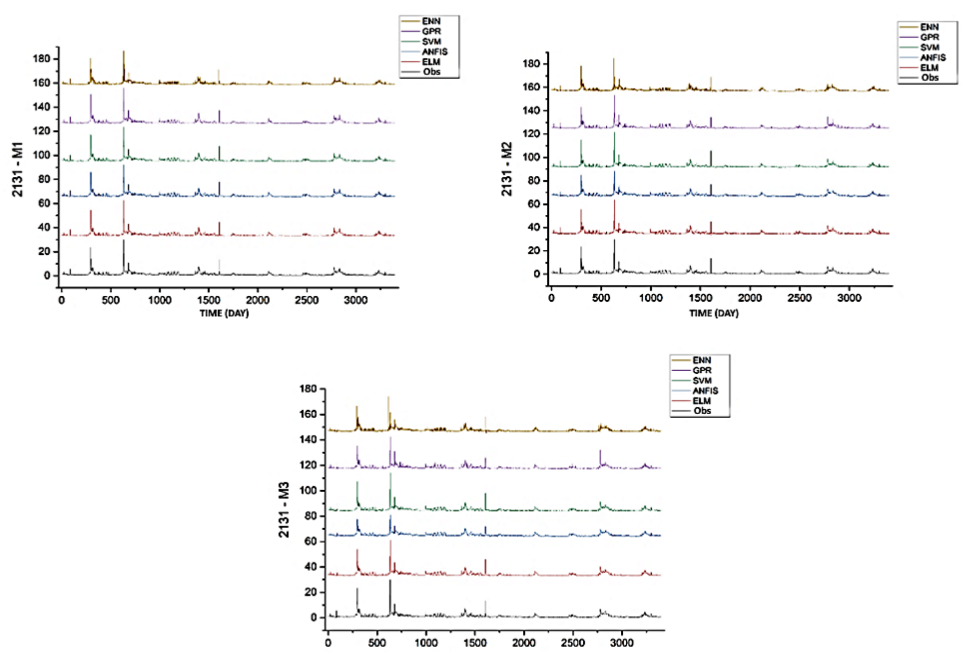


Figure 10 Time series graph of station 213

When the Taylor diagrams are examined, it is seen that the performance values are very close, but the ENN, ELM and GPR models give better results. When the time series given in Figure 10 are examined, it is seen that the model outputs produce outputs close to the observed data. It is also seen that the extreme situations experienced in the data can be represented in the models. This shows that the models can model the flow data with high performance.

5. CONCLUSION

In this study, five different artificial intelligence techniques were used for daily river stream estimation and it was aimed to find the best technique. As the first step of the study; The data were analyzed and the distributions and correlations of the data were determined. Considering the correlation conditions, the second stage, the modeling stage, was started. In this context, river stream estimations were made using ANFIS, ELM, ENN, SVM, GPR techniques and daily stream data from the climate reference periods between 1981-2010. Rank analysis was applied to decide the best model and it was observed that the method with the highest rank value was ELM. In addition, the performance ranking was observed to be ELM, GPR, ENN, SVM, and ANFIS respectively. These results show that ELM, GPR and ENN give much better results when compared with traditional artificial intelligence techniques such as ANFIS and SVM. This shows that these techniques are also reliable models for river stream modeling, and the problems seen in traditional methods can be solved, and these models can be applied more quickly. When the evaluation was made on the basis of the data combination, it was observed that the best combination was the one created with Q(t–3), Q(t–2), Q(T–1) inputs and Q(t) output. In this way, more than one data set type was examined and it was found that the results given by the models for different input numbers were consistent. This means that these methods are reliable for flow modeling. They are thought to be influential in solving the problems such as the complex structures in traditional methods, a large number of membership functions, the increase in the number of rules when the number of entries increases, too much human intervention, and issues in interpreting model parameters. It is also thought that these models can be applied faster when compared with the models used commonly. The study is expected to encourage the use of these uncommon methods in river flow modeling. In addition, it is hoped that ELM, ENN and GPR methods, which are rarely used in hydrology, will lead up to hyperparameter optimization or hybrid model use in future studies.

Acknowledgments
T.R. We thank the General Directorate of State Hydraulic Works (DSI) for the data provided. https://www.dsi.gov.tr/Sayfa/Detay/744

Funding
The author (s) has no received any financial support for the research, authorship or publication of this study.

Authors' Contribution
The authors contributed equally to the study.

The Declaration of Conflict of Interest/ Common Interest
No conflict of interest or common interest has been declared by the authors.

The Declaration of Ethics Committee Approval
This study does not require ethics committee permission or any special permission.

The Declaration of Research and Publication Ethics
The authors of the paper declare that they comply with the scientific, ethical and quotation rules of SAUJS in all processes of the paper and that they do not make any

falsification on the data collected. In addition, they declare that Sakarya University Journal of Science and its editorial board have no responsibility for any ethical violations that may be encountered and that this study has not been evaluated in any academic publication environment other than Sakarya University Journal of Science.

REFERENCES

[1] Y. Zhou, S. Guo, F. Chang, "Explore an Evolutionary Recurrent ANFIS for Modelling Multi-Step-Ahead Flood Forecasts", Journal of Hydrology, vol. 570, pp. 343-355, 2019.

[2] Z. He, X. Wen, H. Liu, J. Du, "A Comparative Study Of Artificial Neural Network, Adaptive Neuro Fuzzy İnference System And Support Vector Machine For Forecasting River Flow İn The Semiarid Mountain Region", Journal of Hydrology, vol. 509, pp. 379-386, 2014.

[3] Z. Yaseen, S. Sulaiman, R. Deo, K. Chau, "An Enhanced Extreme Learning Machine Model For River Flow Forecasting: State-Of-The-Art, Practical Applications İn Water Resource Engineering Area And Future Research Direction", Journal of Hydrology, vol. 569, pp. 387-408, 2019.

[4] Z. Yaseen, S. R. Naganna, Z. Sa’adi, P. Samui, M.A. Ghorbani, S. Shahid, "Hourly River Flow Forecasting: Application Of Emotional Neural Network Versus Multiple Machine Learning Paradigms", Water Resources Management, vol. 34, no. 3, pp. 1075-1091, 2020.

[5] Z. Wang, N. F. Attar, K. Khalili, J. Behmanes, S. S. Band, A. Mosavi, K. K. Chau, "Monthly Streamflow Prediction Using A Hybrid Stochastic-Deterministic Approach For Parsimonious Non-Linear Time Series Modeling", Engineering Applications of Computational Fluid Mechanics, vol. 14, no. 1, pp. 1351-1372, 2020.

[6] A. Sun, D. Wang, X. Xu, "Monthly Streamflow Forecasting Using Gaussian Process Regression", Journal of Hydrology, vol. 511, pp. 72-81, 2014.

[7] Z. Yaseen, A. El-shafie, O. Jaafar, H. Afan, K. Sayl, "Artificial İntelligence Based Models For Stream-Flow Forecasting: 2000–2015", Journal of Hydrology, vol. 530, pp. 829-844, 2015.

[8] E. Khadangi, H. R. Madvar, M. M. Ebadzadeh “Comparison Of ANFIS And RBF Models İn Daily Stream Flow Forecasting”. In 2009 2nd International Conference on Computer, control and Communication (pp. 1-6). IEEE.

[9] G. Huang, Q. Zhu, C. Siew, "Extreme Learning Machine: Theory And Applications", Neurocomputing, vol. 70, no. 1-3, pp. 489-501, 2006.

[10] H. Sıqueıra, L. Boccato, R. Attux, C. Lyra, "Unorganızed Machınes For Seasonal Streamflow Serıes Forecastıng", International Journal of Neural Systems, vol. 24, no. 03, p. 1430009, 2014.

[11] Z. Yaseen, O. Jaafar, R. C. Deo, O. Kisi, J. Adamowski, J. Qulity, A. El-Shafie, "Stream-Flow Forecasting Using Extreme Learning Machines: A Case Study İn A Semi-Arid Region in Iraq", Journal of Hydrology, vol. 542, pp. 603-614, 2016.

[12] R. Adnan, Z. Liang, S. Trajkovic, M. Zounemat-Kermani, B. Li, O. Kisi,

"Daily Streamflow Prediction Using Optimally Pruned Extreme Learning Machine", Journal of Hydrology, vol. 577, p. 123981, 2019.

[13] D. E. Rumelhart, G.E. Hinton, R. J. Williams "Learning Internal Representations By Error Propagation", Parallel Distributed Processing.

[14] M. Atiquzzaman, J. Kandasamy, "Robustness of Extreme Learning Machine İn The Prediction of Hydrological Flow Series", Computers & Geosciences, vol. 120, pp. 105-114, 2018.

[15] D. Kukolj, "Design of Adaptive Takagi–Sugeno–Kang Fuzzy Models", Applied Soft Computing, vol. 2, no. 2, pp. 89-103, 2002.

[16] Jang, "ANFIS: Adaptive-Network-Based Fuzzy Inference System", IEEE Transactions on Systems, Man, and Cybernetics, vol. 23, no. 3, pp. 665-685, 1993.

[17] A. Mosavi, P. Ozturk, K. Chau, "Flood Prediction Using Machine Learning Models: Literature Review", Water, vol. 10, no. 11, p. 1536, 2018.

[18] H. B. Demuth, "Fuzzy Logic Toolbox for Use With MATLAB", User's Guide Version 4. MA, 2000.

[19] Z. Jin, G. Zhou, D. Gao, Y. Zhang, "EEG Classification Using Sparse Bayesian Extreme Learning Machine For Brain–Computer Interface", Neural Computing and Applications, vol. 32, no. 11, pp. 6601-6609, 2018.

[20] E. R. David, L. M. James, "Learning Internal Representations By Error Propagation. In Parallel Distributed Processing: Explorations in The Microstructure Of Cognition: Foundations", MITP, pp. 318-362,1987.

[21] A. Khashman, "Application of An Emotional Neural Network to Facial Recognition", Neural Computing and Applications, vol. 18, no. 4, pp. 309-320, 2008.

[22] A. Khashman, "Blood Cell Identification Using Emotional Neural Networks", Journal of Information Science & Engineering, 2009, vol. 25,6.

[23] V. N. Vapnik, "Statistical Learning Theory", New York: Wiley, 1998, ISBN: 978-0-471-03003-4

[24] C. J. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition". Data mining and knowledge discovery, 1998, Vol. 2(2), pp. 121-167.

[25] F. Fotovatikhah, M. Herrera, S. Shamshirband, K. Chau, S. Faizollahzadeh Ardabili, M. Piran, "Survey of Computational İntelligence As Basis to Big Flood Management: Challenges, Research Directions And Future Work", Engineering Applications of Computational Fluid Mechanics, vol. 12, no. 1, pp. 411-437, 2018.

[26] A. Mehr, A. Gandomi, "MSGP-LASSO: An Improved Multi-Stage Genetic Programming Model for Streamflow Prediction", Information Sciences, vol. 561, pp. 181-195, 2021.

[27] U. Okkan, Z. Serbes, "Rainfall-Runoff Modeling Using Least Squares Support Vector Machines", Environmetrics, vol. 23, no. 6, pp. 549-564, 2012.

[28] N. Ceryan, U. Okkan, P. Samui, S. Ceryan, "Modeling of Tensile Strength of Rocks Materials Based on Support Vector Machines Approaches", International Journal for Numerical and Analytical Methods in Geomechanics, vol. 37, no. 16, pp. 2655-2670, 2012.

[29] R. Chanklan, N. Kaoungku, K. Suksut, K. Kerdprasop, N. Kerdprasop, "Runoff Prediction With a Combined Artificial Neural Network and Support Vector Regression", International Journal of Machine Learning and Computing, vol. 8, no. 1, pp. 39-43, 2018.

[30] S. Shamshirband, S. Hashemi, H. Salimi, S. Samadianfard, E. Asadi, S. Shadkani, K. Kargar, A. Mosavi, N. Nabipour, k. W. Chau, "Predicting Standardized Streamflow Index For Hydrological Drought Using Machine Learning Models", Engineering Applications of Computational Fluid Mechanics, vol. 14, no. 1, pp. 339-350, 2020.

[31] C. Wu, K. Chau, "Prediction of Rainfall Time Series Using Modular Soft Computing Methods", Engineering Applications of Artificial Intelligence, vol. 26, no. 3, pp. 997-1007, 2013.

[32] K. P. Murphy, “Machine Learning: A Probabilistic Perspective”. MIT press, 2012.

[33] R. Richardson, M. Osborne, D. Howey, "Gaussian Process Regression for Forecasting Battery State of Health", Journal of Power Sources, vol. 357, pp. 209-219, 2017.

[34] Z. Zhang, Q. Zhang, V. Singh, "Univariate Streamflow Forecasting Using Commonly Used Data-Driven Models: Literature Review and Case Study", Hydrological Sciences Journal, vol. 63, no. 7, pp. 1091-1111, 2018.

[35] A. Danandeh Mehr, V. Nourani, E. Kahya, B. Hrnjica, A. Sattar, Z. Yaseen, "Genetic Programming in Water Resources Engineering: A State-Of-The-Art Review", Journal of Hydrology, vol. 566, pp. 643-667, 2018.

[36] R. Taormina, K. Chau, "ANN-Based İnterval Forecasting of Streamflow Discharges Using The LUBE Method and MOFIPS", Engineering Applications of Artificial Intelligence, vol. 45, pp. 429-440, 2015.

[37] Y. Zhang, H. Yang, H. Cui, Q. Chen, "Comparison of The Ability of ARIMA, WNN and SVM Models for Drought Forecasting in The Sanjiang Plain, China", Natural Resources Research, vol. 29, no. 2, pp. 1447-1464, 2019.

[38] Z. Chen, M. Kavvas, N. Ohara, M. Anderson, J. Yoon, "Coupled Regional Hydroclimate Model and Its Application to the Tigris-Euphrates Basin", Journal of Hydrologic Engineering, vol. 16, no. 12, pp. 1059-1070, 2011.

[39] DSI:https://web.archive.org/web/200508130 23730/http://www.dsi.gov.tr/topraksu.htm ((2020, 07 16).

[40] C. C. Aggarwal, “Data Mining: The Textbook”, Springer, New York, 2015.

[41] G. B. Guacho, S. Abdali, N. Shah, E. E. Papalexakis, “Semi-Supervised Content-Based Detection of Misinformation Via Tensor Embeddings”, In 2018 IEEE/ACM international conference on advances in social networks analysis and

mining (ASONAM), 2018, pp. 322-325. IEEE.

[42] Y. Khan, S. S Chai, “Ensemble of ANN and ANFIS for Water Quality Prediction And Analysis-A Data Driven Approach”, Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 2017, Vol. 9(2-9), pp. 117-122.

[43] S. Samantaray, P. Dash, "Decision Tree Based Discrimination Between Inrush Currents And Internal Faults İn Power Transformer", International Journal of Electrical Power & Energy Systems, vol. 33, no. 4, pp. 1043-1048, 2011.