



A Hybrid ANFIS-GA Approach for Estimation of Hydrological Time Series

Bulent Haznedar¹ · Huseyin Cagan Kilinc²

Received: 18 May 2022 / Accepted: 27 July 2022 / Published online: 6 August 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract

River flow modeling plays a leading role in the management of water resources and ensuring sustainability. The complex nature of hydrological systems and the difficulty in the application process have led researchers to seek more instantaneous methods for flow predictions. Therefore, with the development of artificial intelligence-based techniques, hybrid modeling has become popular among hydrologists in recent years. For that reason, this study seeks to develop a hybrid model that integrates an adaptive neuro-fuzzy inference system (ANFIS) with a genetic algorithm (GA) to predict river flow. Fundamentally, the performance of an ANFIS model depends on the optimum model parameters. Thus, it is aimed to increase the prediction performance by optimizing the ANFIS parameters with the population-based GA algorithm, which is a powerful algorithm. In this respect, the data gathered from Zamanti and K rk n Flow Measurement Stations (FMS) of Seyhan River, one of Turkey’s significant rivers, were employed. Besides, the proposed hybrid ANFIS-GA approach was compared to classical ANFIS model to demonstrate the improvement of its performance. Also, within the scope of simulation studies, the traditional artificial neural networks (ANN) and the long-short term memory (LSTM) method which is a quite popular in recent years were used to predict streamflow data. The estimation results of the models were evaluated with *RMSE*, *MAE*, *MAPE*, *SD*, and R^2 statistical metrics. In a nutshell, the outcomes indicated that the proposed ANFIS-GA method was the most successful model by achieving the highest values of R^2 (≈ 0.9409) and R^2 (≈ 0.9263) for the Zamanti and K rk n FMS data.

Keywords ANFIS · Genetic algorithm · Soft computing · Streamflow · Forecasting · River flow

✉ Bulent Haznedar
haznedar@gaziantep.edu.tr

Huseyin Cagan Kilinc
huseyncagankilinc@eseniyurt.edu.tr

¹ Department of Computer Engineering, Gaziantep University, Gaziantep, Turkey

² Department of Civil Engineering, Istanbul Esenyurt University, Istanbul, Turkey

1 Introduction

Human beings have endeavored to determine the features of water, bring it under control, define the laws governing its movement, prevent possible threats, and make the best use of water to survive in every phase of their lives (Kuru and Tezer 2020). The increase in water demand owing to global warming, drought, climate change, unplanned consumption, industrialization and agricultural use puts pressure on clean water resources. The way to reduce the threat is the efficacious use of available resources and water management. Sustainable water resources planning and effective river basin management present significant benefits for engineers and decision-makers. One of the critical points to ensure sustainability is the forward estimation of river flows (Kilinc and Haznedar 2022).

Long-term flow forecasting is required for optimal management of hydrological resources, ecological restoration, flood mitigation, and power generation (Ni et al. 2020). Numerous models are utilized to enhance the prediction accuracy of hydrological time series. Generally, these models are divided into physical-based and data-based models. Physical-based models require multi-source knowledge, powerful mathematical calculation tools, and a complex parameter optimization process. On the other hand, data-based models are an efficient alternative that establishes direct connections between input and output data without comprehending complex physical mechanisms (Chau et al. 2005).

Nowadays, computational issues have risen exponentially compared to previous years. Those growing issues have led to the emergence of too much data, which enabled the process of a large amount of data. In addition, it has become possible to predict a future situation or event from old data stacks. Machine learning emerged from these large data piles to solve certain problems or make predictions for certain situations (Dehghani et al. 2018).

Hydrological systems have a complex nature and are quite difficult to implement over large areas. Thus, researchers have turned to the use of relatively faster and more flexible artificial intelligence techniques based on simplifications to estimate the hydrological flow. Artificial intelligence techniques have added a new dimension to the theoretical modeling approach as they can solve various issues and provide successful results related to water resources and hydrological engineering. The hybrid and updated designs of the models mentioned in Artificial Neural Networks (ANN), Adaptive Neuro-Fuzzy Inference System (ANFIS), Gene Expression Programming (GEP), Long Short Term Memory (LSTM), and Gated Recurrent Units (GRU) are the main topics in the relevant literature (Gerger et al. 2021).

In recent years, the performance of models has been analyzed in scientific studies by integrating artificial intelligence techniques with various parameters such as temperature. Gerger et al. (2021) evaluated the relationship between estimation and flow measurement with the help of ANFIS, ANN and GEP methods by using the monthly average current values of two stations within the boundaries of the Dicle Basin. Kisi et al. (2014) modeled the rainfall-runoff relationship for a small basin in Turkey by conducting ANFIS, ANN and GEP methods with 4 years of data. These models were compared to the multi-linear regression (CDR) method. As a result, it has emerged that the gap method was successful in modeling the estimation-flow process. Moreover, it was thought the model mentioned could be considered as an alternative to other applied artificial intelligence models. Akrami et al. (2014) conducted a hybrid study on the connection of forecasting the flow in the Klang River basin with ANFIS and enhanced MANFIS. The outcomes of the study demonstrated that the MANFIS method contains a lower error and calculation complexity and higher precipitation prediction when MANFIS compared to the traditional ANFIS model. Yaseen

et al. (2017) determined the monthly flow estimation of the Pahang River in favor of a new approach of ANFIS, called ANFIS-FFA. The study displayed that FFA could improve the prediction accuracy of the model in a situation where the hybrid model was compared with traditional ANFIS. In addition to all these studies, Calp (2019) analyzed the amount of local precipitation with ANFIS-GA and ANN models which were optimized with the Genetic Algorithm (GA) by using 10 years of meteorological data from the Basel region of Switzerland. It has been revealed that the model could be efficiently used in the prediction of meteorological events since the predictive success of the results obtained was quite high. Zare and Koch (2018) have studied groundwater level fluctuations in the Miandarband plain of Iran. They applied the new hybrid Wavelet-ANFIS model with various combinations of input and main wavelets for the estimation of the ripple. The hybrid model performed better than other model variants. Seifi and Riahi-Madvar (2019) developed a hybrid digital intelligence model for distribution in channel flows. ANFIS, which was optimized by GA and intelligence based on ANN operated. The study revealed that the model presented better results than other classical methods. According to Azad et al. (2018), Particle Swarm Optimization (PSO) exhibited an improvement in the performance of classical ANFIS for all measurements in river flow estimation. In order to optimize the prediction performance, Qasem et al. (2017) analyzed the models of ANFIS-GA, ANFIS-DEC, and ANFIS-PSO. As a result of his studying, ANFIS-PSO was found to be the one that provided the best performance among evolutionary algorithms. Xu et al. (2020) applied the LSTM network, which targeted the time series data field in the process of predicting the river flow estimation. LSTM's performance was compared with Support Vector Regression (SVR) and Multilayer Perceptron Method (MLP). As a result of the performance, it turned out that LSTM provided more promising results. Likewise, Latifoglu et al. (2015) precipitated river flow estimation by using LSTM networks. He presented the performance analysis of the new model developed by Single Spectrum Analysis (TSA) for the forecast data. On the other hand, Kara (2019) proposed the LSTM network to create an efficient and accurate model for daily solar radiation estimation. The effectiveness of the proposed method was compared with machine learning algorithms. It has been suggested that the mentioned model has more pleasing performance than other machine learning models. Lian et al. (2022a, b) proposed a cause-driven runoff forecasting framework based on the LSTM hybrid model. The results showed that the hybrid model was superior to other models. Poursaeid et al. (2022) used different artificial intelligence models (extreme learning machine (ELM), least square support vector machine (LSSVM), adaptive neuro-fuzzy inference system (ANFIS), and multiple linear regression (MLR) model) to simulate and estimate the groundwater level. The results revealed that ELM model was more accurate than other methods. Wang et al. (2022) proposed a novel application of the wavelet decomposition-prediction-reconstruction (WDPRM) model. The model consisted of wavelet decomposition, a particle swarm optimization support vector machine (PSO-SVM), and an artificial bee colony algorithm-optimized BP neural network (ABC-BP). The results indicated that the WDPRM model has advantages against single prediction methods. Khosravi et al. (2022) hybridized convolutional neural network (CNN) using a metaheuristic bat-inspired algorithm to improve the test accuracy of streamflow forecasting. The model was compared with well-known benchmarks models (multilayer perceptron, adaptive neuro-fuzzy inference system, support vector regression and random forest). The evaluation metrics showed that the hybrid CNN-BAT algorithm outperformed the other algorithms. Lian et al. (2022a, b) proposed a climate-driven streamflow forecasting framework (CDSF) to improve runoff forecasting accuracy. This framework consisted of principal component analysis (PCA), long short term memory (LSTM), and bayesian optimization (BO). The

results revealed that the hybrid model was significantly better than other benchmark models. Hu et al. (2021) suggested a novel hybrid decompose-ensemble model that employed variational mode decomposition (VMD) and back-propagation neural networks (BPNN) in order to improve the accuracy and stability of daily streamflow estimation. The results confirmed that the VMD outperformed a single estimation model. The proposed model also had a better performance in estimating higher-magnitude flows.

While ANN and ANFIS models were frequently studied in the relevant literature, it has been seen the LSTM method was conducted limitedly in river current estimates. In this study, the flow measurement station was located on the Zamanti River of the Seyhan Basin in the South of Turkey. In addition, flow estimation analysis was conducted to support the sustainable water potential of this basin. Thus, the prediction performance of the models was assessed with artificial intelligence methods, with the help of measured daily flow data at the flow measurement station (FMS) throughout the years. Studies on estimation were carried out by means of a hybrid approach similar to the training of the ANFIS model with the GA algorithm. The performance of the proposed hybrid ANFIS-GA approach was compared with the performances of classical ANFIS model. Also, within the scope of simulation studies, the traditional artificial neural networks (ANN) and long-short term memory (LSTM) method which is a quite popular in recent years were used to predict streamflow data. The classical optimizers as 'Adam', 'Adamax', 'Nadam' and 'RMSprop' were applied in the training process of the LSTM method. In this respect, the success of ANFIS-GA, ANFIS, ANN and LSTM methods in the process of determining river flow estimations was combined with comparative analysis. Thus, the relevancy of statistical analysis in forecasting the flow was enhanced. The outcomes indicate that the proposed ANFIS-GA model enhances prediction accuracy more than the others and is viable for predicting river flow.

Some points that show the originality of this study are as follows. The performance of the proposed hybrid ANFIS-GA method has been compared with LSTM, which is a very popular and effective deep learning method in recent years. In other respects, the LSTM method was conducted limitedly in real-time river flow prediction. In addition, a limited number of studies are found for the prediction of river flow in the Seyhan basin in the literature. The remaining article was structured as follows: In Sect. 2, the data set and the components of the applied method were specified. In Sect. 3, the gathered outcomes from the research were presented comparatively and a detailed evaluation was provided. Eventually, the outcomes of the study were summarized in Sect. 4.

2 Materials and Methods

In this study, a hybrid ANFIS-GA model was conducted for the Zamanti River, which was constructed as a result of optimizing the initial and final parameters gathered in the ANFIS network with GA, thus, the river flow rate was estimated.

2.1 Study Area and Dataset

The Seyhan Basin is located in the west of Turkey and extends in a wedge shape from Çukurova to the north. The upper part is located in Central Anatolia, the middle and lower parts are located in the Mediterranean Region. The basin, which contains the catchment areas of the Seyhan River and Göksu and Zamanti branches, is between $36^{\circ} 33'$ - $39^{\circ} 12'$ North and $34^{\circ} 24'$ - $36^{\circ} 56'$ east latitude and longitude. The basin is 2.213 km^2 wide and

covers 2.82% of Turkey. The Seyhan Basin holds a total rainfall area of 20,731 km², and the average annual rainfall is 624 mm; the average annual flow is 211.07 m³/s. The average annual yield of the basin is 10.18 L/s km³, and the ratio of the flow to precipitation is 0.51. Regarding such factors as the fact that the entire Seyhan Basin is located within the country's borders, instantaneous high flow changes, and the area of sea-connected basin exits, it is essential to study the Seyhan Basin. Furthermore, the catchment potential, sustainable water resources, and accurate river flow forecasts for watershed management are among the essential aspects to be considered (Kilinc 2022).

The research area covers the Zamanti and Körkün FMS located in the upper part of the Seyhan River Basin, as shown in Fig. 1. The Seyhan River, which carries its source from the Taurus Mountains, empties into the sea west of the Çukurova Delta. The size of the river's basin is approximately 21,700 km². There is the Ceyhan River basin in the east of the Seyhan Basin and the Euphrates River basin in the northeast. It is bounded to the west by the closed basin of Sultan Marshes, to the northwest by the Kızılırmak River basin and to the southwest by the Eastern Mediterranean basin. The lower territories of the Seyhan River basin are under the influence of the Mediterranean climate, while the continental climate is effective in the middle and upper parts. The average annual rainfall is about 700 mm in coastal areas, while in mountainous areas, the value has risen to 1000 mm, and in the north and inland, it falls to 400 mm.

The Seyhan basin, one of the largest rivers of Turkey pouring into the Mediterranean, is one of the leading basins that will be significantly affected by drought due to climate change. It is predicted that there will be significant decreases up to 30% in surface water resources,

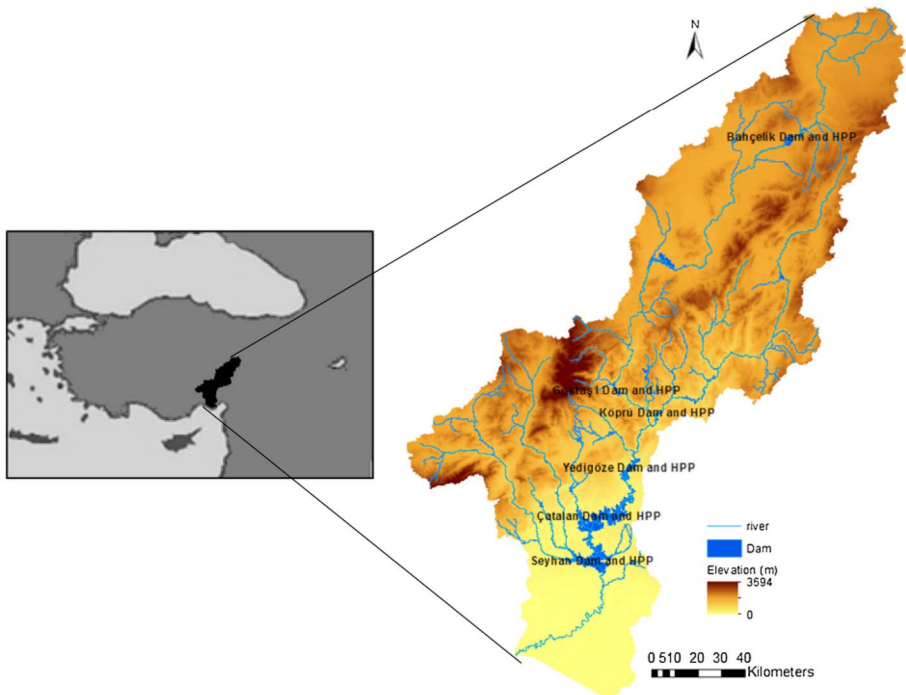


Fig. 1 Location of Seyhan River Basin

snow storage, and groundwater potential in this basin (Acikel and Ekmekci 2021). This situation reveals the importance of the basin in future planning. In addition to climate change, the increasing effect of drought on the basin day by day will significantly affect the economic activities of the basin and the management of water resources. Information on the yields of streams in areas with arid and semi-arid climates is important in terms of utility and irrigation water, determination of irrigation time and reservoir operation. The change in precipitation in the river basin will directly affect the river flows. This situation reveals the importance of analyzing the flow data in different time periods in terms of monitoring the drought of the basin. The fact that the Seyhan Basin has a high sensitivity to climate changes and that precipitation and temperature changes significantly affect both the surface water potential and groundwater recharge of the basin have been effective in considering this region as a study area.

In the study, sample data whose positions are given in Fig. 2 and flow characteristics are shown in Figs. 3 and 4 are indicated. The data were obtained from Zamanti Ergenüsağı (E18A026) and Körkün Hocali FMS (E18A020) of the Seyhan River, which is one of the significant rivers of Turkey. The flow values between October 2000 and September 2011 are used as sample data. The number of river flow measurements belonging to Zamanti and Körkün FMS is 4383 and 3653 per day, respectively. The number of measurements, the continuity of the data among the measurements and the homogeneous distribution of the data were the determining factors in using the flow data of these stations in the study. Within the scope of the study, 80% of the total data set was operated for training and the remaining 20% for testing. Additionally, the statistical parameters of the flow measurement data of Zamanti and Körkün FMSs on the Seyhan River are given in Table 1.

2.2 Long Short Term Memory Network (LSTM)

LSTM networks are primarily an effective method utilized to clarify and predict future information by merging past information with current information. As shown in Fig. 5, LSTM blocks have three gates that perform the cell's write, read and reset operations. All cells are managed by these three gates (Tasabat and Aydin 2021). The entrance gate controls the input information to the open cell, determines how much information is transferred to new data using the past gate, and how much information is used when calculating the output using the exit gate.

$$f_t = \sigma(W_{f,x} * X_t + W_{f,h} * h_{t-1} + b_f) \quad (1)$$

$$i_t = \sigma(W_{i,x} * X_t + W_{i,h} * h_{t-1} + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_{c,x} * X_t + W_{c,h} * h_{t-1} + b_c) \quad (3)$$

$$C_t = C_{t-1} * f_t + i_t * \tilde{C}_t \quad (4)$$

$$o_t = \sigma(W_{o,x} * X_t + W_{o,h} * h_{t-1} + b_o) \quad (5)$$

$$h_t = o_t * \tanh(C_t) \quad (6)$$

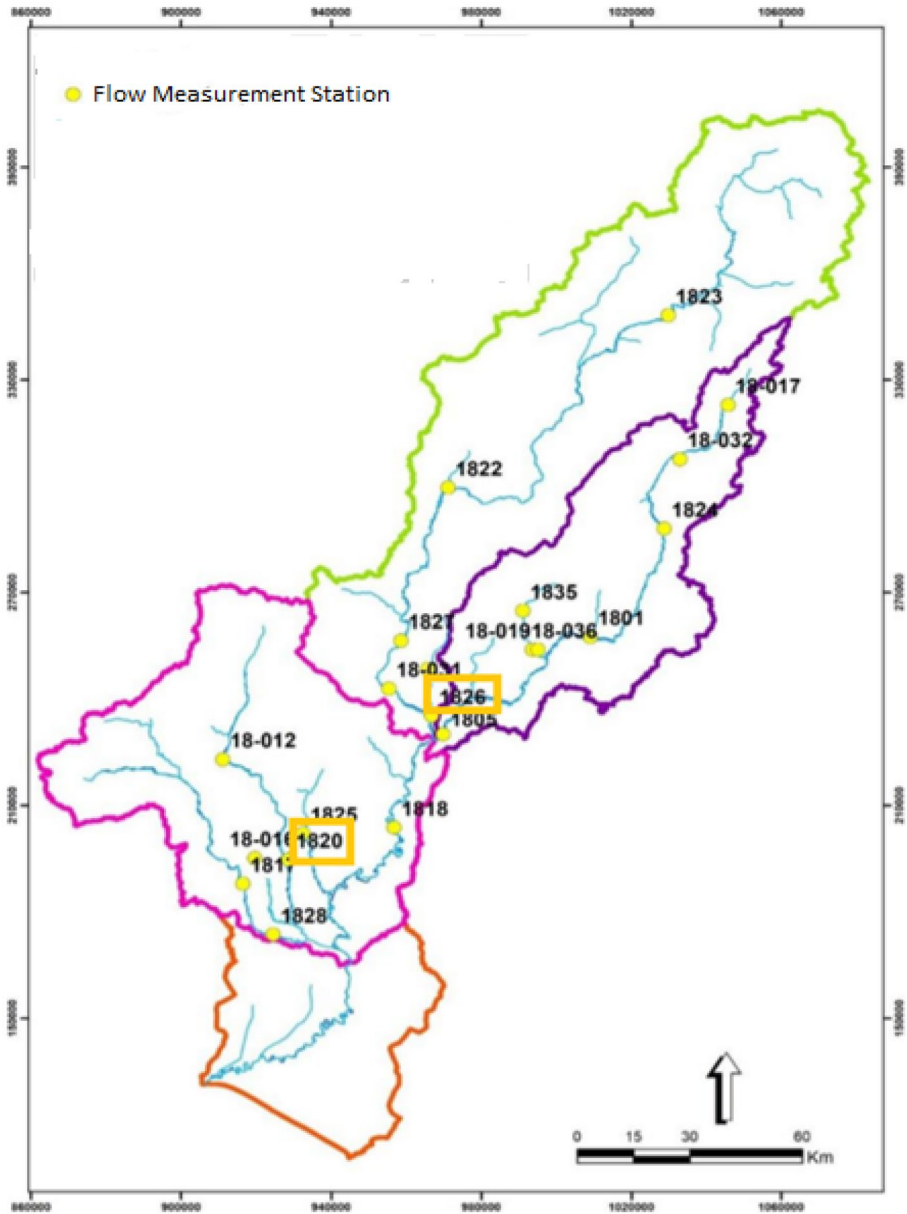


Fig. 2 Location of Zamanti and Körkün (FMS) on Seyhan River

In a situation where σ represents a logistic sigmoid function, W are the input weights of different gates. LSTM has occurred from the entrance gate (i_t), the entrance modulation gate (\tilde{C}_t), forget gate (f_t) and exit gate (o_t). Also, (b) bias vector, (C_t) current cell state and (h_t) hidden state have been stated in this network. All these controllers

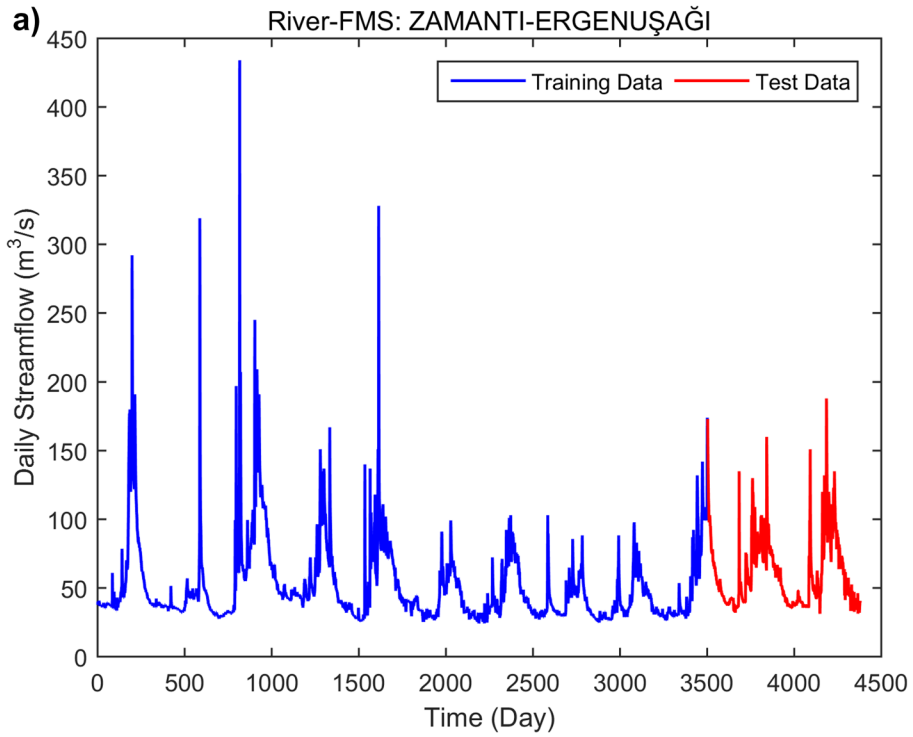


Fig. 3 The approximate 20-year river flow measurement of Zamanti-FMS

determine the amount of information gathered from the previous cycle and the information transferred to the new one.

2.3 Adaptive Neuro-Fuzzy Inference System (ANFIS)

ANFIS is an artificial system developed by Jang and inspired by the Takagi–Sugeno fuzzy model (Haznedar et al. 2021). It is capable of back-propagation learning of artificial neural networks with the ability to draw conclusions from fuzzy systems. Both methods are combined, and thus a hybrid artificial intelligence model called neuro-fuzzy network is created. ANFIS calculates the output value with a certain systematic, which makes it possible by distributing the input data performed with membership functions over the neural network with IF–THEN rules. This process provides inference ability and is less dependent on experience. As a result, the success of the ANFIS model is quite high in forecasting problems. ANFIS is mainly composed of five layers. Figure 6 shows a basic ANFIS structure consisting of two inputs and one output (Haznedar et al. 2017).

1st Layer

This layer is called the fuzzification layer. Nodes in this layer are adaptive so that they can be optimized. At this stage, the output signal for each node is generated, which relies on the input values and the type of conducted membership function. The

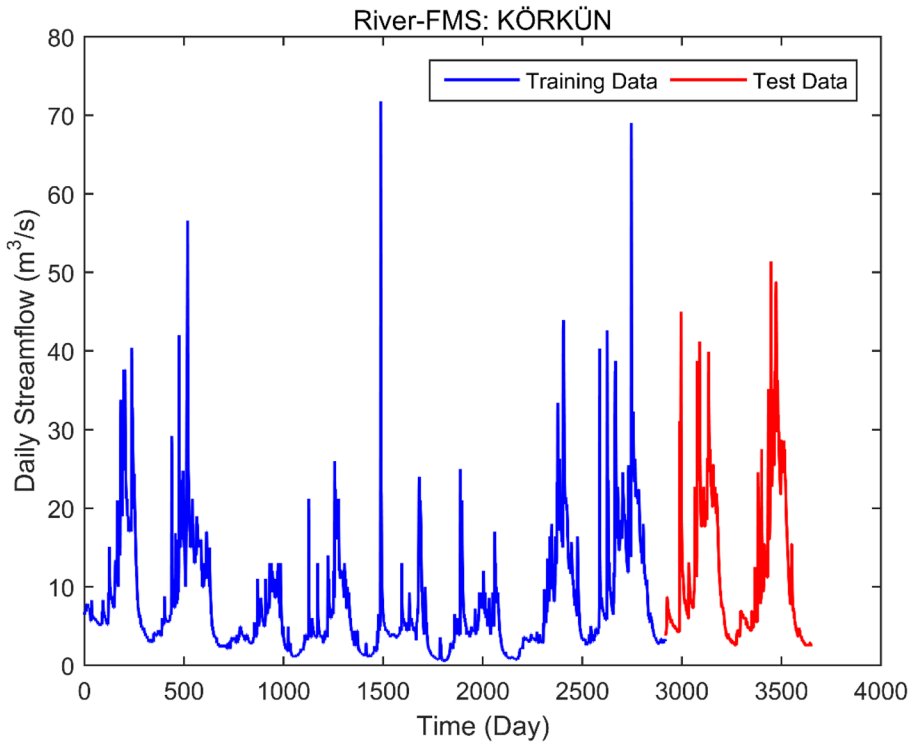


Fig. 4 The approximate 20-year river flow measurement of Korkün FMS

outputs of the nodes in this layer (O_{li}) are indicated in Eqs. (7) and (8) (Karaboga and Kaya 2019).

$$O_{li} = \mu A_i(x) \quad i = 1, 2 \tag{7}$$

$$O_{li} = \mu B_{i-2}(x) \quad i = 3, 4 \tag{8}$$

A_i and B_i represent any membership function and μA_i and μB_i illustrate the membership degree. In a situation where the bell curve is operated as a membership function, μA_i is calculated with Eq. (9). Here, a_i is the sigma of the membership function, b_i is the slope of the membership function and c_i is the center of the membership function.

Table 1 The statistical characteristics of river flow data

River-FMS	Dataset	Count	Minimum (m³/s)	Maximum (m³/s)	Median (m³/s)	Mean (m³/s)
Zamanti	Total	4378	24.20	434	42.80	53.26
	Training	3502	24.20	434	41.20	51.45
	Testing	876	31.30	188	52.60	60.65
Körkün	Total	3648	0.0006	71.80	5.60	8.57
	Training	2918	0.0006	71.80	5.20	7.67
	Testing	730	2.58	51.40	7.80	12.14

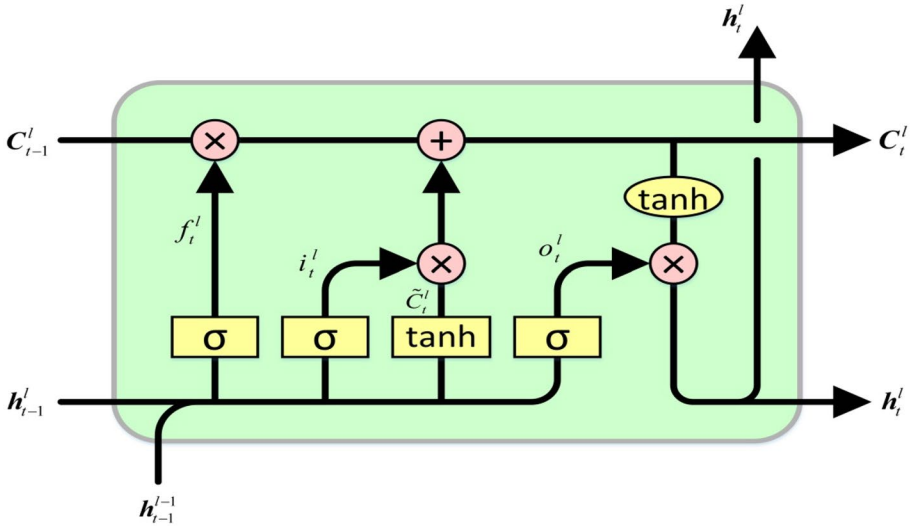


Fig. 5 The structure of the LSTM model cell

$$\mu A_i = \frac{1}{1 + \left[\left(\frac{x-c_i}{a_i} \right)^2 \right]^{b_i}} \quad i = 1, 2 \tag{9}$$

2nd Layer

The second layer is known as the rule layer, and each node in the layer specifies the rules of the fuzzy system and the number of rules. As a result, the firing strength (w_i) of each rule is calculated by multiplying the membership degrees coming from the fuzzification layer.

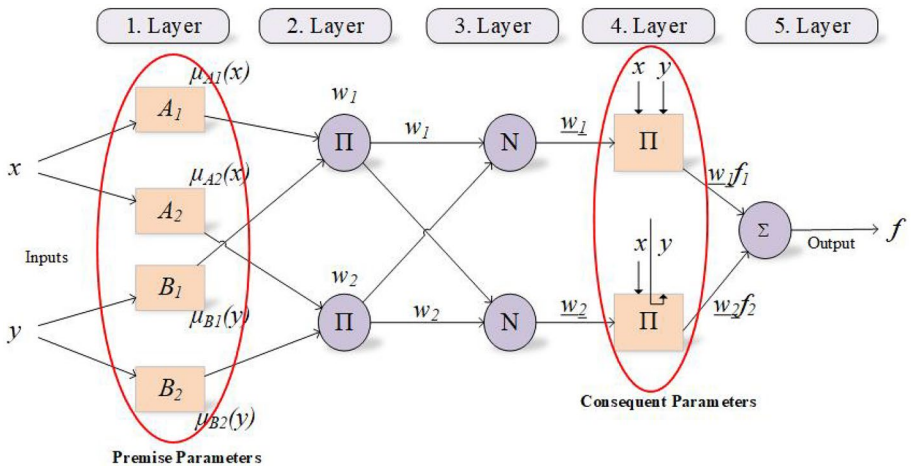


Fig. 6 A basic structure of ANFIS that has two inputs and one output

$$O_{2i} = w_i = \mu A_i(x) \cdot \mu B_i(y) \quad i = 1, 2 \tag{10}$$

3rd Layer

The third layer, in which the firing strength of each rule is normalized, is called the normalization layer.

$$O_{3i} = \bar{w}_i = \frac{w_i}{w_1 + w_2} \quad i = 1, 2 \tag{11}$$

4th Layer

This layer is the defuzzification layer. The weighted output value given in Eq. (12) is calculated as a result of multiplying the normalized firing strength value of each rule coming from the normalization layer with the first-order polynomial.

$$O_{4i} = \bar{w}_i \cdot f_i = \bar{w}_i \cdot (p_i x + q_i y + r_i) \quad i = 1, 2 \tag{12}$$

5th Layer

The fifth layer, known as the output layer, collects the weighted output value of each node from the fourth layer, as stated in Eq. (13). Thus, the actual output value of the ANFIS system is found.

$$O_{5i} = f = \sum \bar{w}_i \cdot f_i = \frac{\sum \bar{w}_i \cdot f_i}{\sum w_i} \quad i = 1, 2 \tag{13}$$

ANFIS has two fuzzy logic system parameters, the premise-consequent parameters, which merge the fuzzy rules. One of the main objectives of ANFIS is to optimize these parameters with a learning algorithm. To optimize the parameters, it is provided by using the input–output data set. Additionally, parameter optimization is performed in order to minimize the error value between the actual output and the target output. In this context, the traditional derivative-based back propagation (BP) algorithm is commonly used to optimize the premise-consequent parameters of the classical ANFIS model developed by Jang.

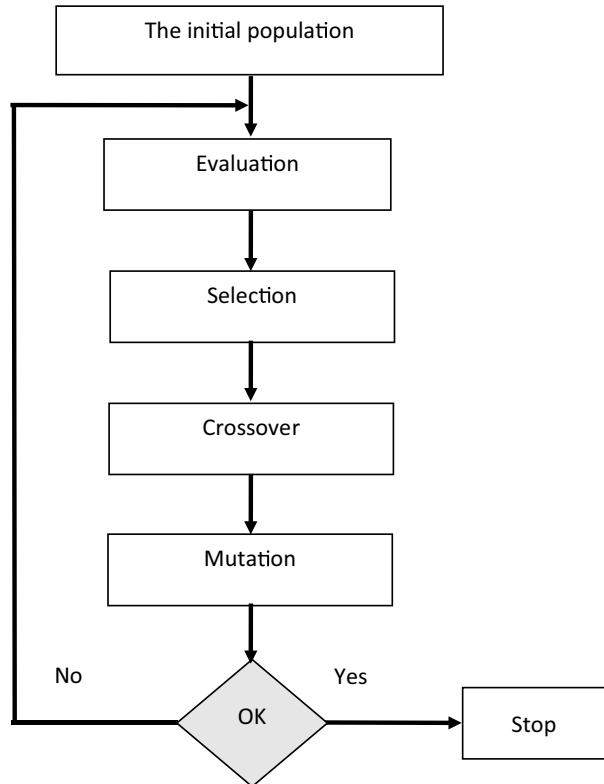
2.4 Genetic Algorithm (GA)

GA was developed in the 1970s by John Holland, whose basic principles were laid down and inspired by the mechanisms of evolutionary development of living things (Holland 1975). Furthermore, GA is a heuristic algorithm designed to locate the global optimum in the search space. Due to the fact that it performs with a set of solutions, it holds the ability to access the region where the global optimum is located rapidly. For this reason, the success of obtaining optimal results is high even in multidimensional problems with a large search space and a large number of variables. Moreover, these results appear within acceptable periods. Operators such as crossover, mutation, inheritance and selection in the flow diagram are used to produce the most appropriate new solutions for the cost function of the problem. The flow diagram is demonstrated in Fig. 7.

Initial Populations

The initial population is created using randomly selected candidate solutions. Although there are not any specific criteria for the number of selected candidate solu-

Fig. 7 The flow chart of GA



tions, it is considered that selection of large number of individuals in the initial phase did not influence the goodness of the solution and solution time.

Evaluation of Fitness Function

The fitness function value represents the solution quality of each individual and determines the individuals which are transferred to the next generation. It is obtained by subjecting each chromosome into fitness function.

Selection

Roulette Wheel and Tournament Methods are often used for the selection criteria. In the Roulette Wheel, parent chromosomes are selected from chromosomes having high fitness values rather than the ones having relatively low fitness values. On the other hand, in the tournament, the best chromosomes are selected as the parent chromosomes among randomly selected 'n' chromosomes from the population (Haznedar et al. 2017).

Crossover

Crossover is usually carried out by genes from two parent chromosomes. From the first gene to predetermine the number of a gene is taken from the first parent chromo-

some while the rest of them are taken from the second parent chromosome. Crossover operator provides better solutions by combining the good features of individuals.

Mutation

The mutation process produces new solution candidates by increasing the diversity of chromosomes in the population. In this method, more than one gene is changed to generate a different individual. The mutation rate is usually between 0.5% to 5%. The role of the mutation operator is that it prevents the algorithm from sticking to a local optimum since the mutation operator randomly selects a gene on a chromosome.

As mentioned earlier, GA can be described as a population-based and artificial intelligence-based heuristic optimization algorithm. The term for candidate solutions forming the algorithm's population is chromosomes. Chromosomes are transformed into solutions that represent better results through various evolutionary processes. This process persists until stopping criteria are met. For example, it continues until an acceptable fitness value is reached or criteria such as a predetermined processing time or several generations are completed. The fundamental steps of the Genetic Algorithm are given below.

1. Generate the initial population.
2. Calculate the fitness value for each solution in the population.
3. Repeat the following steps until a new population is brought into existence.
 - (a) Select two parental chromosomes from the population, depending on the fitness values.
 - (b) Apply crossover to n parents depending on the probability of a crossover to generate a new product, which is also known as a child.
 - (c) Mutate the new product at the rate of the probability of mutation.
4. Use the newly produced population in the algorithm's next run.
5. If the stop criterion is provided, STOP.
6. GO to Step 2.

2.5 Estimation of Streamflow Data Based on ANFIS-GA (The Proposed Method)

As mentioned previously, ANFIS is a hybrid AI method that is generated by combining the neural networks' learning ability with the inference of fuzzy logic (Simon 2002). ANFIS is trained by utilizing the current input and output data set to solve prevailing problems. This section clarifies the components of the offered innovative approach for training the ANFIS model with the GA algorithm. ANFIS has two fuzzy logic system parameters, the premise-consequent parameters, which need to be optimized. The premise parameters pertain to the membership functions provided in Eq. (9). They are operated to place input values in the fuzzification layer. The consequent parameters are the p , q , and r parameters given in Eq. (12) belonging to the linear functions in the defuzzification layer. Generally, traditional derivative-based algorithms are operated to optimize the parameters of ANFIS. Nevertheless, for reasons such as slow convergence of derivative-based algorithms, local minimization, largely depending on the initial values, and so on, optimizing ANFIS parameters has

turned into one of the primary concerns. Population-based algorithms, performing with a set of solutions, often allow them to access the region rapidly where the global optimum is located. ANFIS is a challenging model for optimizing parameters. Hence, a population-based GA algorithm, which is a mighty algorithm, was employed to optimize the parameters and eliminate the drawbacks. The *RMSE* function provided in Eq. (14) is applied to calculate the fitness value of the solutions. GA tries to find the best premise and consequent parameters with the chromosome with acceptable fitness value. Considering the training process of ANFIS, the block diagram of the proposed method is given in Fig. 8.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (F(i) - F_d(i))^2}{N}} \tag{14}$$

The *F* and *F_d* in Eq. (14) demonstrate the value estimated through ANFIS and the actual output value of the data. In addition, *N* represents the total number of samples in the training dataset. Optimizing the premise and consequent parameters minimizes the *RMSE* function presented in Eq. (14). The most acceptable *RMSE* value is obtained when the actual and the predicted value are closest to each other. GA seeks to discover the most appropriate *RMSE* value until the stop criterion is formed. In this study, the number of iterations is determined as the stopping criterion.

3 Comparative Results

In this study, solutions to the hydrological time series issues of Zamanti and Körkün FMS, as presented in Figs. 3 and 4, were searched by optimizing the parameters of ANFIS structures using GA. While forecasting, data sets were organized by designing 5-offset scenarios for the time delay. For the data set of the 5 time-shifted scenarios, the daily flow values starting from 5 days before and up to the day before being used as input in the 6th

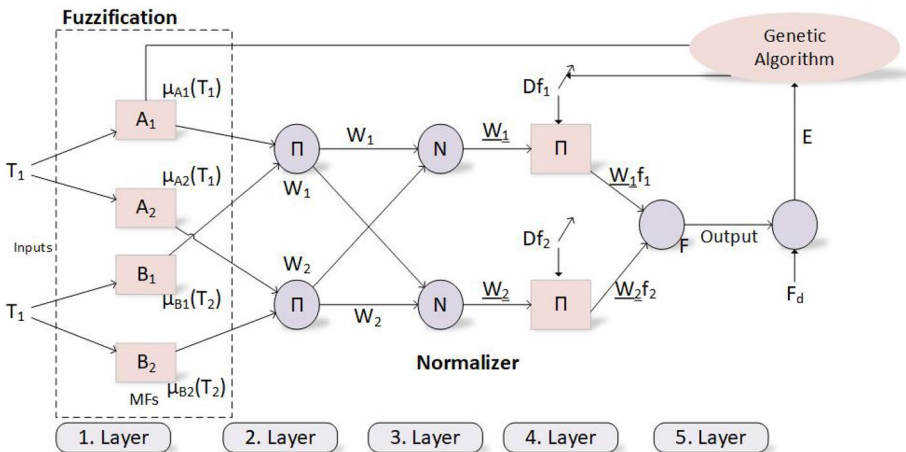


Fig. 8 Schematic representation of the proposed model

Table 2 Performance Measurements for Zamanti FMS

Modeller	RMSE (m ³ /s)	MAE (m ³ /s)	MAPE (m ³ /s)	SD	R ²
LSTM-RMSPROP	18.1011	14.8973	27.5110	0.1760	0.9221
LSTM-ADAM	17.3525	14.5524	27.0175	0.1562	0.9374
LSTM-ADAMAX	17.2549	14.3717	26.6680	0.1597	0.9338
LSTM-NADAM	16.7175	13.8637	25.8264	0.1530	0.9329
ANN	20.2975	10.4173	13.1626	0.2936	0.7684
ANFIS	8.3413	4.0947	6.3468	0.1231	0.9038
ANFIS-GA	6.6517	2.7875	4.3105	0.1056	0.9409

day forecasting. Numerous attempts were made to determine the GA algorithm’s initial parameter values. As a consequence of the attempts, it was uncovered that the number of iterations was 1000, the population size was 50, the crossover rate was 0.8, and the mutation rate was 0.01. The performance of the proposed hybrid ANFIS-GA approach was compared with the performances of the classical ANFIS model. Thus, the classical ANFIS network is additionally trained with the traditional BP algorithm to estimate the river flow data. To serve and achieve this purpose, the number of iterations was determined as 1000, the learning coefficient was 0.2, and the momentum ratio was 0.4.

Moreover, the performance of the presented ANFIS-GA approach was compared with the performance of deep learning methods owing to the fact that deep learning has become relatively widespread and has been involved in numerous various areas. In this respect, the Long Short Term Memory (LSTM) network, one of the deep learning models, was involved in the aforementioned time series issues. The LSTM network consists of 1 input layer, 4 hidden layers and 1 output layer. The number of iterations is established as 1024 for the 5-offset scenario. ‘Adam’, ‘Adamax’, ‘Nadam’ and ‘RMSprop’ were applied as optimization functions and the loss function was determined as ‘MAE’. In order to show the performance of traditional methods for the current study, the ANN method was also applied in the estimation studies.

The performance measures of the methods were as indicated; average square root error (RMSE), mean absolute error (MAE), average absolute percentage error (MAPE), determination coefficient (R²), and standard deviation (SD). The evaluation metrics is

Table 3 Performance Measurements for Körkün FMS

Models	RMSE (m ³ /s)	MAE (m ³ /s)	MAPE (m ³ /s)	SD	R ²
LSTM-RMSPROP	2.9209	1.1447	8.2874	0.2830	0.9106
LSTM-ADAM	2.7797	1.3773	10.3454	0.2613	0.9197
LSTM-ADAMAX	2.7543	1.1898	7.7377	0.2854	0.9188
LSTM-NADAM	2.8400	1.2087	6.9898	0.2776	0.9144
ANN	3.8249	1.4305	7.4485	0.3513	0.8514
ANFIS	2.9103	1.2521	8.8225	0.2916	0.9100
ANFIS-GA	2.6433	1.0308	6.6980	0.2764	0.9263

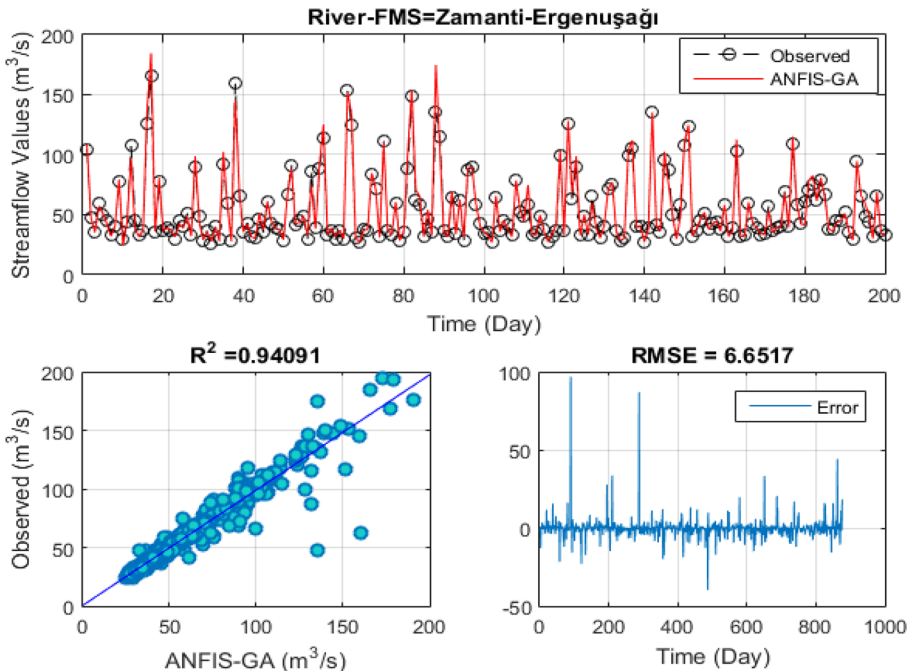


Fig. 9 Comparative plots of the observed and predicted flow of the ANFIS-GA model during the testing phase for Zamanti FMS

among the measurement tools widely operated in multiple studies to predict daily flow values and determine effectiveness. (Abyaneh et al. 2011; Arslan and Sekertekin 2019). Performance measurements of ANFIS-GA, ANFIS, ANN, and LSTM methods are shown in Tables 2 and 3. It was detected that the performance of the hybrid ANFIS-GA model was relatively prospering compared to the ANFIS, ANN and LSTM network. Furthermore, the methods applied in the test data obtained from Zamanti and Körkün FMSs and the scatter charts are given in Figs. 9, 10, 11, 12, 13, and 14, respectively. In the graphs demonstrated, the 'ADAM' optimizer function provided the LSTM network's most satisfactory results compared with the obtained results. The scatter graph indicates that the results of the hybrid ANFIS-GA method are closer to the actual river current values. In addition, it has been revealed to hold the highest success prediction with an R^2 value of 0.9409 for the Zamanti River and an R^2 value of 0.9263 for the Körkün River.

Among the results obtained, it is noticed that the hybrid ANFIS-GA method offers quite a success in river flow estimation compared to LSTM, ANN and ANFIS methods. The ANFIS-GA method exhibited the highest R^2 (≈ 0.9409) values with minimum RMSE (≈ 6.6517 m³/s), MAE (≈ 2.7875 m³/s), MAPE (≈ 4.3105 m³/s) and SD (≈ 0.1056) for the FMS test set over time. In addition, it was the most successful model by achieving the minimum RMSE (≈ 2.6433 m³/s), MAE (≈ 1.0308 m³/s), MAPE (≈ 6.6980 m³/s), SD

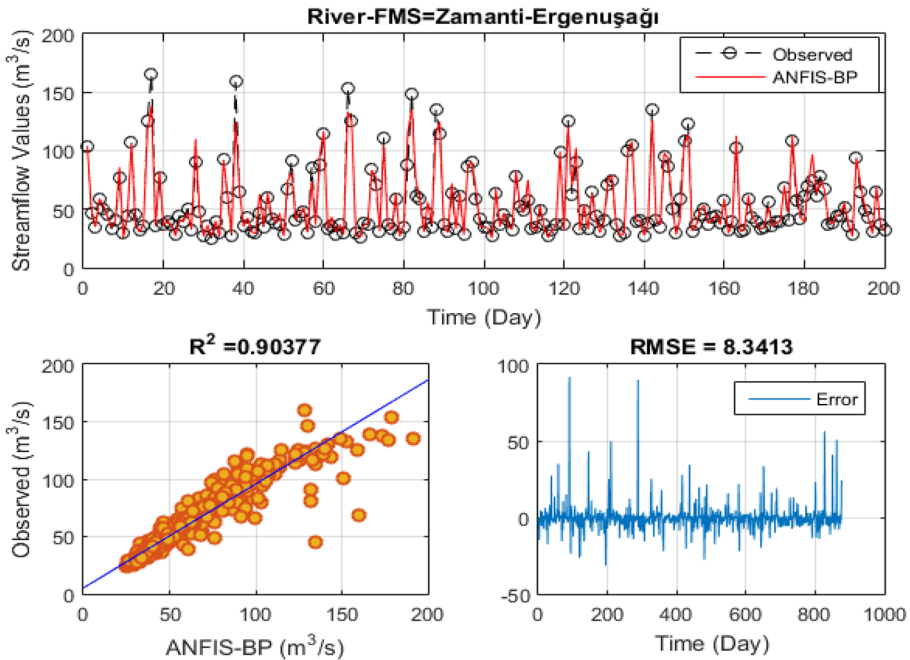


Fig. 10 Comparative plots of the observed and predicted flow of the classical ANFIS model during the testing phase for Zamanti FMS

(≈ 0.2764), and the highest R^2 (≈ 0.9263) values for the K rk n FMS test set with different characters in the same basin.

The success of the presented method is the result of the evolutionary approach that the elitism conservation strategy of the GA algorithm can evaluate and select the optimal solution from 1000 solutions in each generation. Furthermore, by overcoming the minimum disruption problem of the population-based GA and derivative-based BP algorithm, optimal solutions were achieved in the global region, and thus the ANFIS model was optimized more successfully. The results of various optimization functions of the LSTM network are also displayed in Table 1 even though the findings indicated the LSTM network revealed a superior R^2 value than the ANFIS model, it was encountered that it lagged behind the ANFIS models in the RMSE, MAE and MAPE measurements. The LSTM network is a fairly popular and frequently used method in recent years. Nowadays, the LSTM network is quite prevalent and a frequently conducted method. Nevertheless, it is not able to perform more pleasingly than the traditional ANFIS method, and it lags behind the hybrid ANFIS-GA model, including the coefficient of determination (R^2) in the study. In addition, as depicted in Tables 1 and 2, traditional ANN method performance for the current study lagged behind all other methods with R^2 (≈ 0.7684) and R^2 (≈ 0.8514), respectively. However, results in terms of other

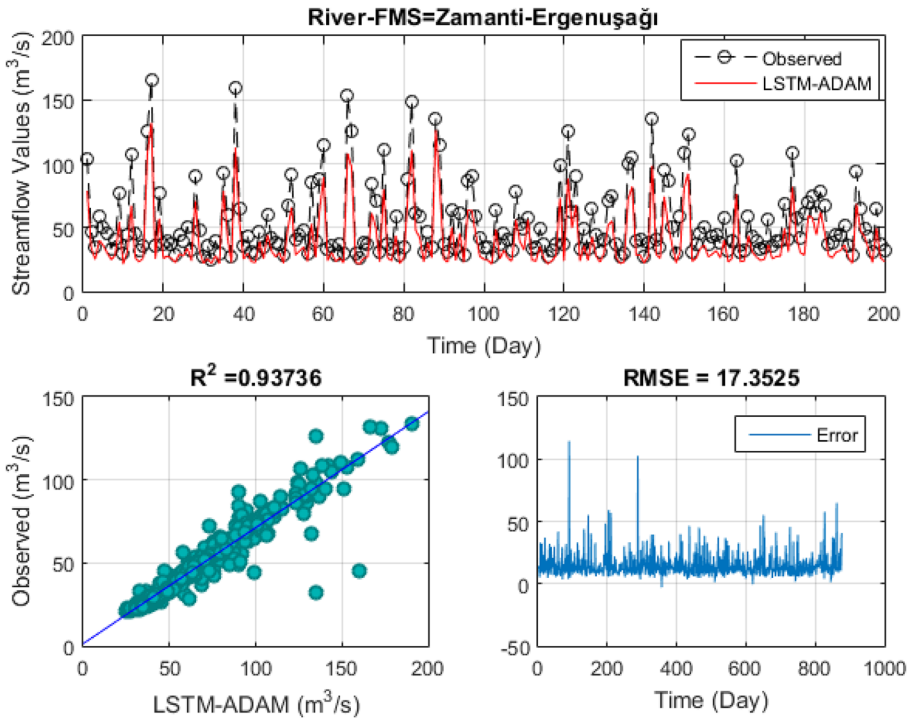


Fig. 11 Comparative plots of the observed and predicted flow of the LSTM-ADAM model during the testing phase for Zamanti FMS

measurements are similar, which indicates the fact that the generalization ability of the traditional ANN method is poor.

All evaluation criteria confirm that the hybrid ANFIS-GA model is quite successful. The performance of the LSTM model remains inadequate compared to other models. In addition, the correlation of river currents carries LSTM networks one step further. One of the most significant aspects causing the matter is the volatilities noticed in the flow rates between the two stations. The distinctions in the instantaneous variability of the currents in the Zamanti River are instances of the matter. Following the trend line at Körkün station, the current values make the prediction capacity higher than anticipated. Even though LSTM is approaching ANFIS at this station, the success of the hybrid ANFIS-GA model remains at the forefront.

In addition, all evaluation criteria for the two stations have indicated that the hybrid ANFIS-GA model has achieved considerable progress in percentage terms. The values with the highest R^2 and lowest standard deviation are indicated in the Zamanti station in the general evaluation. Zamanti Ergenuşığı station is located on the transition level of the Ceyhan water collection basin includes an association with proximity to the upstream point. On the other hand, the Körkün station is placed on the same line as the final hydrological station on the Seyhan River and manages the river's flow to the sea. Elements such as the capacities of various precipitation areas and the climate characteristics of the two stations differ when compared to each other. These elements impact the correlation between observed and estimated values. In the light of all the details, it has

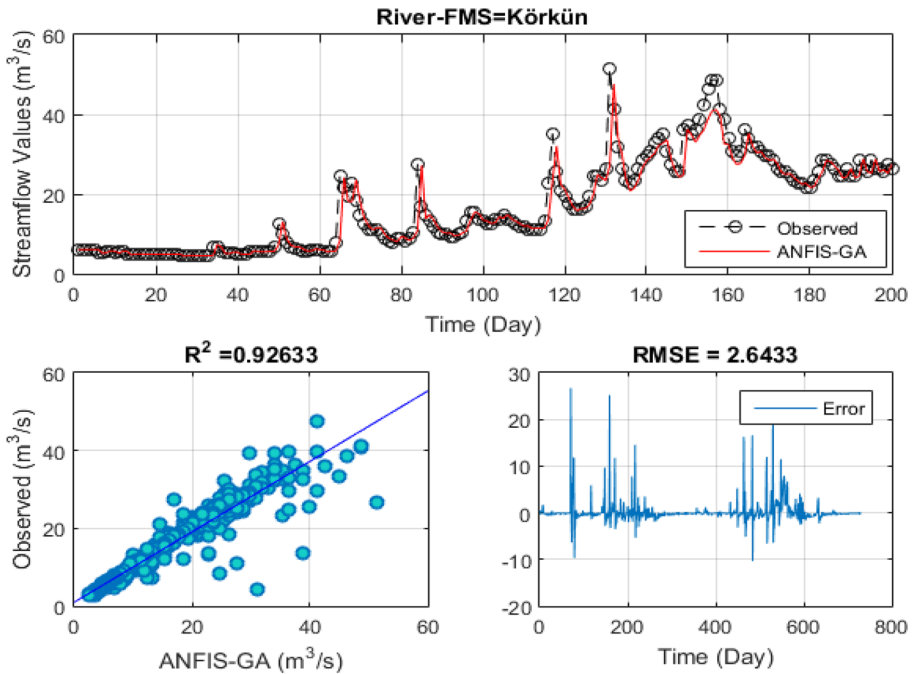


Fig. 12 Comparative plots of the observed and predicted flow of the ANFIS-GA model during the testing phase for Korkun FMS

been recognized that it is essential to transform the results into LSTM parameters and optimize the learning rate and the number of hidden neurons, which are its two primary parameters.

4 Discussions

The training of the ANFIS parameters is one of the main problem. The GA algorithm was used to overcome this challenge in this study. It is seen that the proposed hybrid ANFIS-GA has high accuracy in forecasting streamflow. In addition, the proposed ANFIS-GA method's results compared with the results of many existing studies in the literature using the ANFIS based hybrid models to predict different time series problems. These comparisons showed that artificial intelligence algorithms successfully optimize ANFIS parameters.

Adnan et al. (2022) developed a novel hybrid model where gradient-based optimization (GBO) algorithm was employed to adjust adaptive neuro-fuzzy systems (ANFIS) hyperparameters. Several benchmark methods for optimizing ANFIS parameters were compared, which included particle swarm optimization (PSO), grey wolf optimization (GWO), and colony optimization (ACO). The results contained that all ANFIS hybridized model successfully enhanced the prediction accuracy. Kayhomayoon et al. (2022)

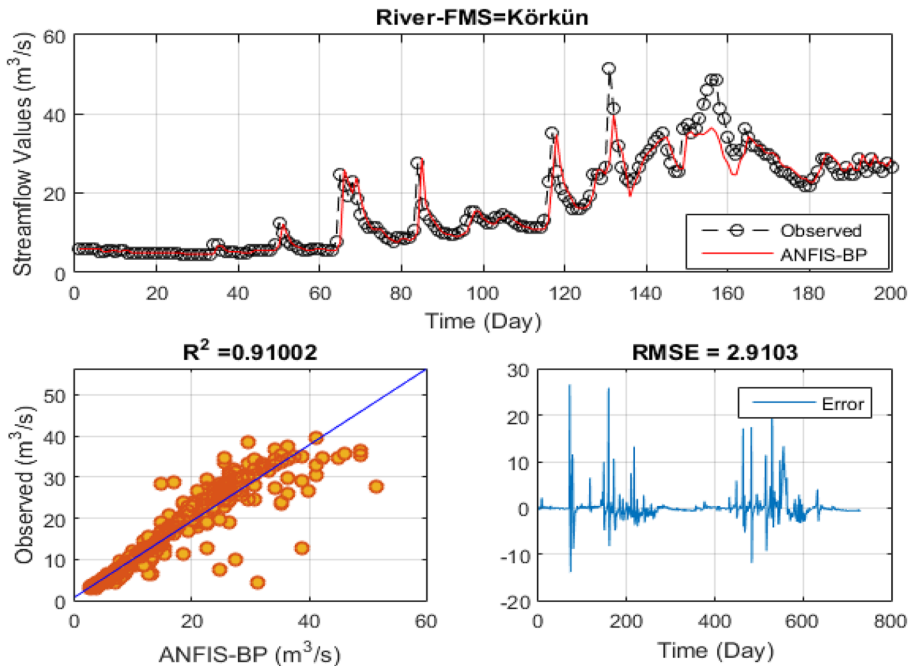


Fig. 13 Comparative plots of the observed and predicted flow of the classical ANFIS model during the testing phase for Körkün FMS

built a hybrid ANFIS model based on several metaheuristic algorithms such as genetic algorithm (GA), particle swarm optimization (PSO), and colony optimization for continuous domains (ACOR). All hybridized models compared with benchmark models. The results revealed that the best prediction was achieved by ANFIS-ACOR model. Also, the findings of the study displayed that metaheuristic algorithms could significantly improve the performance of the ANFIS model. Fadaee et al. (2020) investigated the capability of butterfly optimization algorithm (BOA) and genetic algorithm (GA) integrated with ANFIS, ANN and multiple linear regression (MLR) models. The results indicated that the performances of metaheuristic algorithms increased the accuracy of the ANFIS model. Karami et al. (2022) combined the ANFIS model with GA, PSO, ACO, and differential evolution (DE) algorithms. The comparison results showed that hybrid ANFIS-ACO extended to continuous domain algorithm with the best prediction. Rathnayake et al. (2022) assessed the capacity of the novel Cascaded-ANFIS algorithm and compared this benchmark model with the gated recurrent unit (GRU), LSTM and recurrent neural network (RNN) models. The results clearly revealed that Cascaded-ANFIS outperformed better than single models.

The proposed hybrid ANFIS-GA method successfully predicted the daily flow rate in flow measurement station data of two different hydrological conditions. In addition, the study demonstrated the success of the hybrid model in predicting the optimal level of streamflow when compared with the hybrid models and linear regression models in the literature.

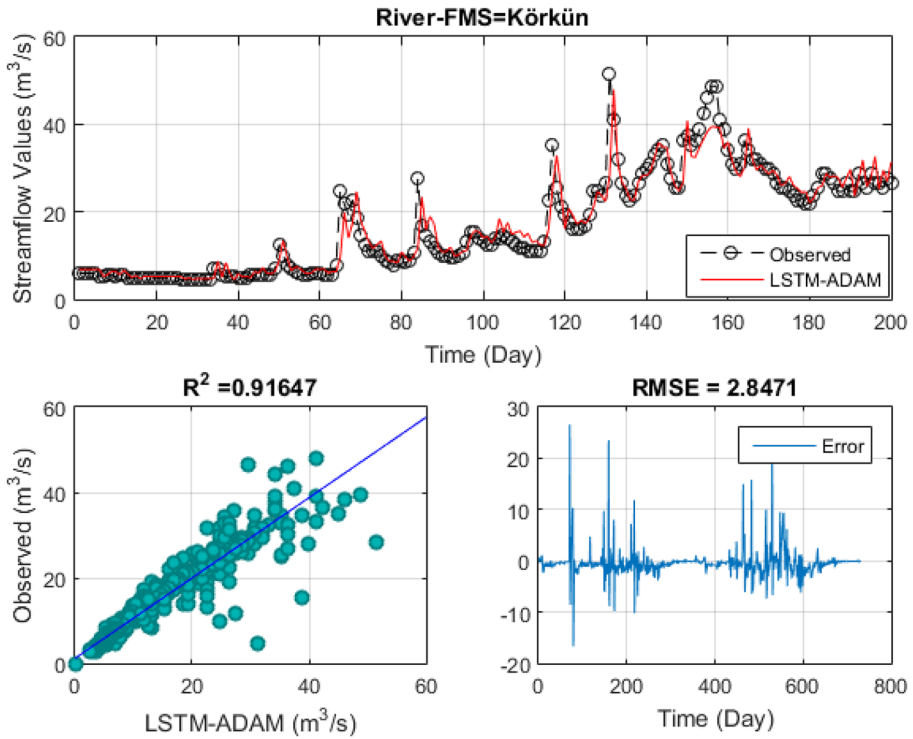


Fig. 14 Comparative plots of the observed and predicted flow of the LSTM-ADAM model during the testing phase for Körkün FMS

5 Conclusions

In conclusion, the performance of ANFIS-GA, ANFIS, ANN and LSTM models for Seyhan river flow forecasting was examined. Therefore, the current values which were obtained from Zamanti Ergenüşağı and Körkün current stations between the dates of October 2000 and September 2011 were operated as sample data. In the study, 3502 and 2918 data, which were 80% of the total data set, were used for training, and 876 and 730 data, the remaining 20%, were used for testing. Derivative-based algorithms, such as BP, are widely used in training parameters of the ANFIS model. However, there are difficulties such as slope calculation in derivative-based algorithms, as well as causing problems such as local minimum. The aforementioned situations also reveal fundamental problems in training and updating ANFIS with derivative-based algorithms. In this context, the most appropriate premise and consequent parameters for the river current velocity estimation of the ANFIS network were determined using the population-based GA algorithm, which is a powerful algorithm to overcome the aforementioned problems. The hybrid ANFIS-GA model's performance was compared with the classic ANFIS model, the traditional ANN and deep learning LSTM network models that have been pretty prevalent recently. RMSE, MAE, MAPE, STD and R^2 different statistical indicators were conducted to compare the

models. The results of the performance indicators in Tables 2 and 3 presented the suggested hybrid ANFIS-GA model can deliver more accurate and reliable estimates than other methods. The results demonstrate that the ANFIS-GA model may suit current river estimation.

The ANFIS-GA model turned out to provide promising results in river flow forecasts. However, there are some limitations in the study: only streaming data was conducted as input. However, while flow time series are stochastic, they are not linear in nature. In addition, the data rely on various meteorological parameters such as evaporation, humidity, snowmelt and temperature. Consequently, the presented study can be extended in the future by gathering more input data. Decomposition techniques can be incorporated into the model to taken the flow data's motion and nonlinearity. In the following studies, offered model's performance can be considered at shorter time intervals such as hourly, 30 min, 10 min for the predictions. Hydrological variables such as evaporation, precipitation, and sediment can be involved in hydrology to examine and develop the application of the suggested model since the hybrid model's contribution to the study is promising.

Overall, the study of population-based algorithms like GA with a set of candidate solutions usually allows them to quickly access the region of the global optimum. However, negative aspects have arisen due to many parameters and the network's training formed from a random point. For population-based algorithms, it usually carries a long computational time to find the optimal solution in the region where they are located, which is due to high probability-based search strategies. This case is the most obvious limitation of the proposed ANFIS-GA method. In this respect, forthcoming studies are arranged to be carried out to train the ANFIS network by utilizing different meta-heuristic search techniques while examining its implementation on related problems. In contrast, new regions with various flow aspects can be analysed with the help of the ANFIS-GA model. Hence, data with various aspects are monitored, and forthcoming research paths can be chosen. In addition, the GA algorithm used in the proposed hybrid ANFIS-GA method can be involved in the training process of the LSTM network and so, new studies can be carried out to increase the prediction performance of the studied basin.

Author Contribution All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by all authors.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Availability of Data and Materials The data will be available upon reasonable request.

Declarations

Ethics Approval The authors undertake that this article has not been published in any other journal and that no plagiarism has occurred.

Consent to Participate The authors agree to participate in the journal.

Consent to Publish The authors agree to publish in the journal.

Competing Interests The Authors declare no conflict of interests.

References

- Abyaneh HZ, Nia AM, Varkeshi MB, Marofi S, Kisi O (2011) Performance evaluation of ANN and ANFIS models for estimating garlic crop evapotranspiration. *J Irrig Drain Eng* 137:280–286
- Acikel S, Ekmekci M (2021) Distinction of multiple groundwater systems in a coastal karst spring zone in SW Turkey by hydrochemical and isotopic characteristics. *Bull Eng Geol Env* 80:5781–5795
- Adnan RM, Mostafa RR, Elbeltagi A et al (2022) Development of new machine learning model for streamflow prediction: Case studies in Pakistan. *Stoch Environ Res Risk Assess* 36:999–1033
- Akrami SA, Nourani V, Hâkim SJS (2014) Development of nonlinear model based on wavelet-ANFIS for rainfall forecasting at Klang Gates Dam. *Water Resour Manage* 28(10):2999–3018
- Arslan N, Sekertekin A (2019) Application of long short-term memory neural network model for the reconstruction of MODIS land surface temperature images. *J Atmos Sol Terr Phys* 194:105100
- Azad A, Karami H, Farzin S et al (2018) Prediction of water quality parameters using ANFIS optimized by intelligence algorithms (Case Study: Gorganrood River). *KSCE J Civ Eng* 22:2206–2213
- Calp MH (2019) A hybrid ANFIS-GA approach for estimation of regional rainfall amount. *Gazi Univ J Sci* 32(1):145–162
- Chau R et al (2005) Benchmarking nanotechnology for high-performance and low-power logic transistor applications. *IEEE Trans Nanotechnol* 4(2):153–158
- Dehghani M, Seifi A, Madvar HR (2018) Novel forecasting models for immediate-short-term to long-term influent flow prediction by combining ANFIS and greywolf optimization. *J Hydrol* 576:698–725
- Fadaee M, Mahdavi-Meymand A, Zounemat-Kermani M (2020) Suspended sediment prediction using integrative soft computing models: on the analogy between the butterfly optimization and genetic Algorithms. *Geocarto Int* 37(4):1–14
- Gerger R, Gumus V, Dere S (2021) The evaluation of different artificial intelligence methods in determination of Tigris basin's rainfall runoff relationship. *BSEU J Sci* 8(1):300–311
- Haznedar B, Arslan MT, Kalinli A (2017) Training ANFIS structure using genetic algorithm for liver cancer classification based on microarray gene expression data. *Sakarya Univ J Sci* 21(1):54–62
- Haznedar B, Arslan MT, Kalinli A (2021) Optimizing ANFIS using simulated annealing algorithm for classification of microarray gene expression cancer data. *Med Biol Eng Comput* 59:497–509
- Holland JH (1975) *Adaptation in natural and artificial systems*. University of Michigan Press, Ann Arbor, MI, USA, p 183
- Hu H, Zhang J, Li TA (2021) Novel hybrid decompose-ensemble strategy with a vmd-bpnn approach for daily streamflow estimating. *Water Resour Manage* 35:5119–5138
- Kara A (2019) Global solar irradiance time series prediction using long short-term memory network. *Gazi Univ J Sci* 4:882–892
- Karaboga D, Kaya E (2019) Adaptive network based fuzzy inference system (ANFIS) training approaches: a comprehensive survey. *Artif Intell Rev* 52(4):2263–2293
- Karami H, DadrasAjirlou Y, Jun C, Bateni SM, Band SS, Mosavi A, Moslehpour M, Chau KW (2022) A novel approach for estimation of sediment load in dam reservoir with hybrid intelligent algorithms. *Front Environ Sci* 10:821079
- Kayhomayoon Z, Babaeian F, Ghordoyee Milan S, Arya Azar N, Berndtsson RA (2022) Combination of metaheuristic optimization algorithms and machine learning methods improves the prediction of groundwater level. *Water* 14:751
- Khosravi K, Golkarian A, Tiefenbacher JP (2022) Using optimized deep learning to predict daily streamflow: a comparison to common machine learning algorithms. *Water Resour Manage* 36:699–716
- Kilinc HC (2022) Daily streamflow forecasting based on the hybrid particle swarm optimization and long short-term memory model in the Orontes basin. *Water* 14(3):490
- Kilinc HC, Haznedar B (2022) A hybrid model for streamflow forecasting in the basin of Euphrates. *Water* 14(1):80
- Kisi O, Latifoglu L, Latifoglu F (2014) Investigation of empirical mode decomposition in forecasting of hydrological time series. *Water Resour Manage* 28:4045–4057
- Kuru A, Tezer A (2020) New approach to determine the protection zones for drinking water basins: the case study of Kırklareli dam. *J Faculty Eng Architect Gazi Univ* 35(1):519–536
- Latifoglu L, Kisi O, Latifoglu F (2015) Importance of hybrid models for forecasting of hydrological variable. *Neural Comput Appl* 26:1669–1680
- Lian Y, Luo J, Wang J et al (2022a) Climate-driven model based on long short-term memory and bayesian optimization for multi-day-ahead daily streamflow forecasting. *Water Resour Manage* 36:21–37

- Lian Y, Luo J, Xue W et al (2022b) Cause-driven streamflow forecasting framework based on linear correlation reconstruction and long short-term memory. *Water Resour Manage* 36:1661–1678
- Ni L, Wang D, Singh VP, Wu J, Wang Y, Tao Y, Zhang J (2020) Streamflow and rainfall forecasting by two long short term memory-based models. *J Hydrol* 583(2):124296
- Poursaeid M, Poursaeid AH, Shabanlou SA (2022) Comparative study of artificial intelligence models and a statistical method for groundwater level prediction. *Water Resour Manage* 36:1499–1519
- Qasem SN, Ebtehaj I, Madavar HR (2017) Optimizing ANFIS for sediment transport in open channels using different evolutionary algorithms. *J Appl Res Water Wastewater* 4(1):290–298
- Rathnayake N, Rathnayake U, Dang TL, Hoshino YA (2022) Cascaded adaptive network-based fuzzy inference system for hydropower forecasting. *Sensors* 22:2905
- Seifi A, Riahi-Madvar H (2019) Improving one-dimensional pollution dispersion modeling in rivers using ANFIS and ANN-based GA optimized models. *Environ Sci Pollut Res* 26:867–885
- Simon D (2002) Training fuzzy systems with the extended Kalman Filter. *Fuzzy Sets Syst* 32:189–199
- Tasabat S, Aydin O (2021) Using long-short term memory networks with genetic algorithm to predict engine condition. *Gazi Univ J Sci* 35(1)
- Wang Y, Liu J, Li R et al (2022) Medium and long-term precipitation prediction using wavelet decomposition-prediction-reconstruction model. *Water Resour Manage* 36:971–987
- Xu W, Jiang Y, Zhang X, Li Y, Zhang R, Fu G (2020) Using long short-term memory networks for river flow prediction. *Hydrol Res* 51(6):1358–1376
- Yaseen ZM et al (2017) Novel approach for stream flow forecasting using a hybrid ANFIS-FFA model. *J Hydrol* 554:263–276
- Zare M, Koch M (2018) Ground water level fluctuations simulation and prediction by ANFIS-and hybrid Wavelet-ANFIS / Fuzzy C-Means (FCM) clustering models: Application to the Miandarband plain. *J Hydro-Environ Res* 18:63–76

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.