



A Hybrid ANFIS-GA Approach for Estimation of Regional Rainfall Amount

M. Hanefi CALP^{1*} 

¹Management Information Systems, Faculty of Economics & Administrative Sciences, Karadeniz Technical University, Trabzon, Turkey

Article Info

Received: 16/03/2018
Accepted: 02/11/2018

Keywords

Rainfall Amount
Estimation
Artificial Intelligence
ANFIS
Genetic Algorithm

Abstract

Effective use and management of ever-diminishing water resources are critically important to the future of humanity. At this point, rainfall is one of the most important factors that supply water resources, but the fact that the rainfall higher is more than normal causes many disasters such as flood, erosion. Therefore, rainfall amount must be analyzed mathematically, statistically or heuristically in order to take precautions, in the region. In this study, an Adaptive Neuro Fuzzy Inference System - Genetic Algorithm (ANFIS-GA) based hybrid model was proposed for estimation of regional rainfall amount. Purpose of the study is to minimize the loss of life and goods for people of the region by estimating the amount of annual rainfall and ensuring effective management of water resources and allowing some evaluations and preparations according to possible climate changes. The estimation model was developed by coding in the MATLAB package program. In the development of the model, 3650 meteorological data from 2008-2018 years belonging to Basel, a Swiss city, were utilized. The real data were tested on both the Artificial Neural Network (ANN) and the hybrid ANFIS-GA model. The obtained results demonstrated that the training R-value of the suggested ANFIS-GA model was 0.9920, the testing R-value was 0.9840 and the error ratio was 0.0011. This clearly shows that predictive performance of the model is high and error level is low, and therefore that hybrid approaches such as ANFIS-GA can be easily used in predicting meteorological events.

1. INTRODUCTION

Climate is the average of one year's weather conditions for a country or region. The basic elements of climate are parameters such as temperature, rainfall, relative humidity, sunshine duration and intensity, pressure, wind speed and direction, evaporation. Prediction of rainfall within these parameters is very important in the prevention of floods, the engineering structures used and the design of urban drainage networks. However, since extreme events occur due to many influences, it is difficult to predict future amounts of these events. In this respect, it is important to determine the expected quantities of the hydrological variability to be considered in determining the project criteria of flood control structures at various recurrence times. This is also especially important for the efficient use of water resources [1,2].

The planning and management of water resources has become very important due to the rapid increase in water demand and it is obligatory to carry out more detailed and extensive researches on the development of water structures [3]. Many parameters (rainfall, flow, infiltration, evaporation, plant ordnance, etc.) need to be analyzed correctly so that water resources can be used correctly and water structures to be constructed can be projected and executed correctly. Rainfall is one of the most important of these parameters. The ability to accurately measure this parameter is crucial to the management and operation of water resources. With a temporary data collection system to be created in places where there is no measurement, several years data can be collected and past data can be obtained by integrating with other measurement stations in the same basin. This is why methods of estimating rainfall or completing missing data are important. As a result of an accurate analysis, all historical data with a few years of measurement can be completed with few errors. In addition, the lack of data that may arise due to any question in the future can be eliminated with rainfall estimation methods [4].

*Corresponding author, e-mail: mhcalp@ktu.edu.tr

Rainfall plays a very important role in human life and agriculture, and is very important for irrigation. Rainfall forecasting is also useful for sewer management, water management and flood forecasting. There are many factors that affect rainfall estimates such as temperature, humidity, wind speed, pressure, and dew point [5]. Future estimates based on historical records are needed to evaluate design criteria in the planning of water constructions. The determination of suitable production models for coming rainfall is an important prerequisite for the successful planning and management of water resources. In particular, missing data filling or data estimation can be provided by artificial intelligent modeling techniques such as GA, ANNs, ANFIS and Fuzzy Logic (FL), etc. [6]. At this point, rainfall forecasting is one of the most challenging tasks. Although many algorithms have been already proposed, it is still very difficult to accurately predict rainfall. A small fluctuation in seasonal rainfall can have devastating effects on the agricultural sector. Accurate rainfall forecasting provides a potential benefit in preventing by natural disasters and damages caused. In addition, accurate rainfall forecasting is useful for agriculture management, too [7]. Rainfall forecasting models can be collected under two main headings: experimental and dynamic. Experimental approaches are generally history-based methods with significant correlation between them. The most important of these are ANNs, stochastic models, fuzzy logic and group model with database. Dynamic approaches are based on sets of equations that are made up of climatic conditions and changes that occur in the atmosphere [4,8].

In this context, it is aimed to estimate the amount of regional rainfall by developing a hybrid model with ANFIS and GA methods in the study, thus minimizing both property and life losses that may occur in the region. The literature review was given in the second part of the study; all details of the study was expressed in the third part under the heading of material and method; findings and performance analysis were discussed in the fourth part, and finally, conclusions and recommendations were provided in the fifth part.

2. LITERATURE REVIEW

In recent years, different models have been utilized to study rainfall amount or meteorological forecasting using different modeling techniques. While some of the researchers use a single method to estimate their study, the vast majority prefer hybrid methods because of their high performance. For example, Sukanya and Prabha discussed data mining technique which is suitable for predicting rainfall in the study. Estimation was done using various classification algorithms such as Decision Tree (DT) and ANN. ANN increases the productivity of rainfall forecasting by analyzing historical and current facts to make accurate forecasts for the future. Various data mining techniques were used with ANN for rainfall forecasting and compared with many studies. As a result, CART, C4.5, ID3 and ANN algorithms are more effective and hybrid algorithms have more successful results [9]. Deo and Sahin introduced a simple, fast and efficient nonlinear algorithm known as the Extreme Learning Machine (ELM) Effective Drought Index (EDI) in eastern Australia. The data between 1957 and 2008 were trained to form the predictive model. Estimation variables for the proposed model are precipitation and mean, minimum and maximum air temperatures supported by major scale climate mode indices as regression variables. They performed a performance comparison in terms of learning speeds between the traditional ANN algorithm to show the effectiveness of the proposed data analysis model with the predictive capabilities and proposed ELM algorithm. The forecasting metrics showed excellent performance in the ANN model for overall test sites. Thus, the Mean Absolute Errors (MAE), Root-Mean Square Error (RMSE), Detection Coefficients and ELM model 32 times faster than the learning rate and 6.1 times faster than the ANN model. An ELM model was proposed to estimate the duration and severity of drought [10].

Wu et al. proposed an efficient hybrid optimization strategy using genetic algorithm (GA) and particle swarm optimization (PSO) methods. In addition, radial basis function neural networks (RBF-NN) was used to solve a specific problem about rainfall prediction. The aim is to automatically determine parameters such as the number of neurons, their centers and radius. In the study, the new generation of individuals was created with three approaches to improve the global optimization performance, which is the elitist strategy, PSO strategy and GA strategy. The performance of the application demonstrated that the hybrid strategy has a more effective global discovery capability compared to GA in the basis function neural networks design problems. Findings showed that the proposed hybrid optimization strategy can be used as an

alternative prediction approach for higher predictive accuracy and optimum generalization capability [11]. Plouffe et al. assessed rainfall data on an agricultural ecological survey to produce total monthly rainfall maps in Sri Lanka. For this, they compared four techniques known as spatial interpolation such as adverse distance weighting, thin-plate splines, ordinary and Bayesian kriging. According to results, Bayesian kriging and splines did best in low and high rainfall, respectively. In addition, the study revealed that there is a relationship between rainfall maps produced from the agro-ecological network and former studies in Sri Lanka [12]. Partal et al. used radial basis function, feed forward back propagation, generalized regression neural network and wavelet transform techniques for estimating daily rainfall in their study. Input combinations that is determined differently had been tested for rainfall prediction. As a result, optimal neural network model for each station has been determined. In addition, performance of the linear regression model was compared with wavelet neural networks models. Wavelet neural network model has been found to provide the best performance assessment criterion. Results shown that integrating wavelet with neural network can ensure important benefits for prediction [13]. Beheshti et al. implemented several meta-heuristic algorithms to increase the accuracy of the rainfall forecast. A multilayer perceptron (MLP) network as a feed-forward ANN was trained for rainfall estimation by centripetal accelerated particle swarm optimization, an imperialist competitive algorithm and a gravitational search algorithm. The next 5 and 10-year average monthly rainfall estimates are model predicted using data pre-processing with two original modes (without data pre-processing) and singular spectrum analysis. The suggested methods integrate both the authenticity and structure of ANN. Results showed that MLP's hybrid learning with centripetal accelerated PSO method provides higher rainfall prediction and higher classification accuracy, and lower error [14]. Tezel and Buyukyildiz proposed an artificial intelligence approach to predict monthly pan evaporation. For this, they utilized the ANNs (multilayer perceptron (MLP) and radial basis function network (RBFN)) and support vector regression (SVR) artificial intelligence methods. They determined temperature, relative humidity, wind speed and rainfall data from the Beysehir meteorological station between 1972-2005 as inputs, and pan evaporation values as output. The obtained results were compared with the evaporation data. Researchers used four different training algorithms. These algorithms are gradient descent with momentum and adaptive learning rule backpropagation, Levenberg-Marquardt (LVM), Scaled Conjugate Gradient (SCG), and Resilient BackPropagation (RBP) in MLP method. The models were created by 10-fold cross-validation, and performance of the proposed algorithm was analyzed using MAE, RMSE, and coefficient of determination (R^2). Obtained performance findings demonstrated that ANN algorithms and SVR had given close results. In addition, according to results obtained from the ANNs and SVR techniques, it was achieved preferable performance than the Romanenko and Meyer [15].

Akrami et al. developed a modified ANFIS (MANFIS) model to create an effective technique after observing the traditional ANFIS architecture. In the study, two types scenarios were presented. In the first scenario, monthly rainfall was used as input only for (t) time (t-4) in different time delays for traditional ANFIS and in the second scenario modified ANFIS was used to increase the rainfall forecasting performance. Conclusions demonstrated that the modified ANFIS model made better rainfall prediction performance. Compared to traditional ANFIS model, it has lower error and computational complexity [16]. Islam and Talukdar tried to increase the performance of statistical methods of time series estimates. For this, they specified that PSO is a suitable technique for accurately determining model parameters. This technique was developed to improve the performance of ANNs for time series estimation. In addition, they used the particle swarm optimization technique to determine the variables of an exponential autoregressive model for estimating time series. The model was applied for annual rainfall forecasting and performed very well compared to the statistical ARIMA model [17]. Qiu et. al. proposed a multitasking convolutional neural network model to automatically extract features from time series measured in observation regions and to utilize the correlation between multiple sites for weather forecasting via multi-tasking. They pointed out that estimation techniques for short-term rainfall quantities based on multi-domain characteristics were used for the first time. Experiments on data from the European Center for Medium Range Weather Forecasts (ECMWF) shown that the learned field correlations are understandable and proposed model is significantly superior [18].

Taormina and Chau proposed a new technique for Input Variable Selection (IVS). This approach utilizes Binary-coded separate Exactly Informed Particle Swarm optimization (BFIPS) and Extreme Learning Machines (ELM) to create IVS model with higher accuracy and lower error in their study. The performance

results of ELM-based techniques analyzed handling assessment steps and datasets recommended by Galelli et al. According to study, techniques suggested using four main IVS approaches have very high accuracy level [19]. Sethi and Garg stated that India is a farming country and that most of the Indian economy is dependent on agriculture. They also emphasized that rainfall has critical important in agriculture, therefore rainfall should be early estimated for Indian economy. Therefore, they utilized the widely used Regression analysis, clustering and ANN techniques to predict their study. At the same time, they emphasized the importance of multiple linear regression technique in early prediction of rainfall. The experimental results showed very high level success of the suggested model [20]. Kashani et al. have emphasized that creating simulations for real precipitation is effective in managing water resources. For this, in the study, he presented a Volterra model simulating rainfall flow process with ANN (IVANN) technique. Proposed integrated model includes Volterra and ANN models. The IVANN model was developed using thirteen-hour rain and flow data to examine the short-term response of a forest basin in northern Iran. The analysis results are compared with the integrated (IANN) model and the Volterra model. The Volterra model was applied as a nonlinear model (Second Order Volterra (SOV) model) and tested using the least squares method. Five criteria were determined to evaluate the performances of the models. These are the efficiency coefficient, the root mean square error, the total volume error, the relative error of top discharge and the time error that must reach peak. According to the results, the IVANN model performed better than other semi-dispersed and agglomerated models to simulate the precipitation-flow process. Also, it was found that the collected SOV model has less sensitivity to simulate the settling-flow process [21]. Shafaei and Kisi investigated the capabilities of wavelet ANN (WANN) for the estimation of short-time daily river flow. The WANN model was developed with two methods based on regression analysis, discrete wavelet transforms and ANNs, respectively. The suggested WANN models tested on daily flow data on the Ajichai River in the north-western part of Iran and compared with ANN and support vector machine (SVM) approaches. Mean Square Error (MSE), Mean Absolute Error (MAE) and correlation coefficient (R) statistics were utilized to evaluate the accuracy of the WANN, ANN and SVM models. The results demonstrated that the WANN model higher performance than the ANN and SVM techniques in the short-term daily river flow forecasting [22].

3. MATERIAL AND METHOD

In this section, all details of the proposed ANFIS-GA based hybrid model for the estimation of regional rainfall were given. The reason why the hybrid model is preferred that such models are significantly increase the accuracy of estimation, especially in non-linear problems. The proposed model aims to provide efficient management of water resources by estimating the amount of rainfall that can occur in the region, to allow some evaluations and preparations according to possible climate changes and thus to reduce the loss of life and goods of the people of the region. The prediction model was developed using the MATLAB package program. The ANFIS and GA techniques that are the subject of study before the details of the model are briefly explained.

3.1. Techniques Used in the Study

3.1.1. Adaptive Network Based Fuzzy Inference Systems (ANFIS)

ANFIS is an artificial intelligence method used to solve complex and nonlinear problems. While ANFIS integrates with both fuzzy inference systems and ANNs, it is effective in solving non-linear and complex problems within a frame. In this structure, Takagi-Sugeno type Fuzzy Inference System (FIS) and hybrid learning algorithm are used. The FIS framework has three main components, the fuzzy rule (if-then), the database and the mechanism of the arguments based on the if-then theory [23-25].

ANFIS creates a fuzzy inference system by regulating the membership function parameters by using the input/output dataset together with the backpropagation algorithm of ANN alone or with the least squares method. This arrangement enables the fuzzy system to learn the relevant system and update itself by using environmental information with the help of modeled data. It is adaptable for this reason. ANFIS learning algorithm is a complicated learning algorithm. This algorithm consists of least squares method and

backpropagation algorithm. This algorithm is based on error back propagation. The learning process consists of two stages. In the first step, input samples are generated and the best secondary parameters are determined by using the least mean square method, assuming that the precursor parameters are constant. In the second stage, the input samples are reproduced and the secondary parameters are assumed to be constant and the precursor parameters are changed by the gradient descent method. This process is repeated later [23,26,27]. ANFIS also includes the advanced data analysis techniques such as numerical grouping and rulemaking. In the study, the ANFIS structure (Figure 1), the node functions of each layer and the functioning of the layers will be as follows [28,29].

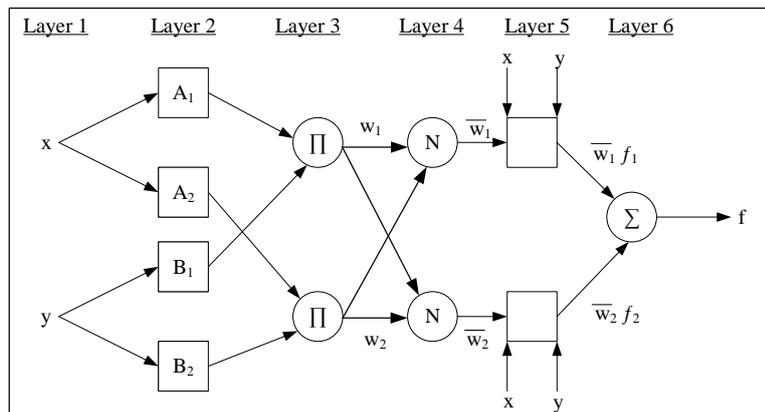


Figure 1. Adaptive network based fuzzy logic inference system (ANFIS)

Layer 1: Input layer

Input signals obtained from nodes in this layer are transferred to the other layers, and are provided the next layer's input values.

Layer 2: Fuzzification layer

In this layer, the output of each node consists of the membership values. This values are about the input values and the membership function. Jang's ANFIS model use the generalized Bell activation function in the form of membership functions for distinguish input values from blurred clusters, [24].

Layer 3: Rule layer

Each node in this layer represent the number and number of rules created by the Takagi-Sugeno fuzzy logic derivation system developed by Jang. The output of each rule node μ_i is multiplied by the membership grades from the Layer 2. If the values of μ_i are obtained, (Equation 1) [24]. $\mu_{A_i}(x)$ and $\mu_{B_i}(y)$: Fuzzy clusters,

$$y_i^3 = w_i = \mu_{A_i}(x) \times \mu_{B_i}(y) = \mu_i, (i=1 \dots n) \quad (1)$$

Here, each node i is a square node, x and y represents the input variable, A_i and B_i represents the fuzzy cluster represented by this node, and w_i represents the membership function for A_i and B_i .

y_i^3 is the output values of layer 3; n is the number of nodes in this layer. Figure 2 shows the Sugeno model.

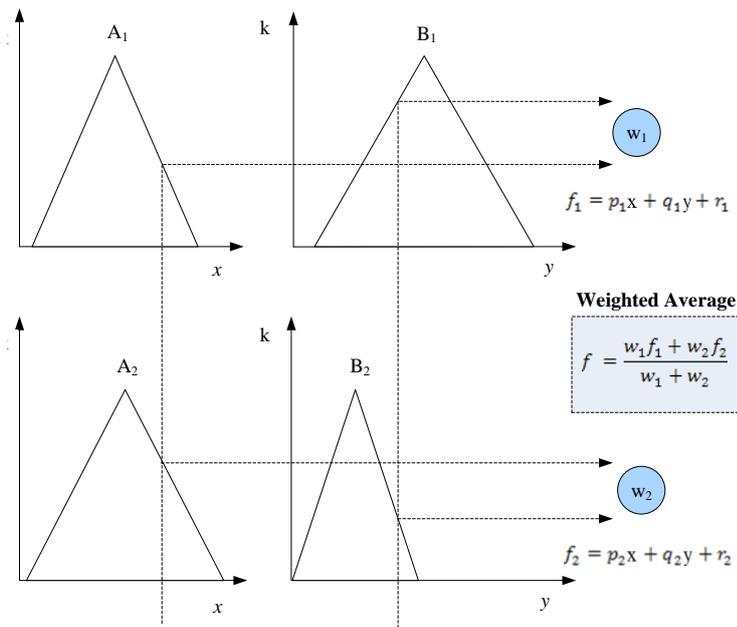


Figure 2. Sugeno model

Layer 4: Normalization layer

Each node in this layer accept all nodes coming from the rule layer as input values and the normalized firing level of each rule is calculated. The calculation formula of the normalized ignition level (μ_i) is given in Equation 2.

$$y_i^4 = N_i = \frac{\mu_i}{\sum_{i=1}^n \mu_i} = \bar{\mu}_i, \quad (i = 1, n) \tag{2}$$

Layer 5: Defuzzification layer

The weighted values of the rules given in each node are calculated. The rules for an ANFIS model with two inputs (x, y) and one output (f) can be expressed as:

Rule 1:

If X= A₁ and Y=B₁ then $f_1 = p_1x + q_1y + r_1$ (3)

Rule 2:

If X= A₂ and Y=B₂ then $f_2 = p_2x + q_2y + r_2$ (4)

In here, A and B are fuzzy sets of inputs, and f (x, y) is a polynomial of p, q and r constants. In a nutshell, the output value of the i. node is given in Equation 5.

$$y_i^5 = \bar{\mu}_i [p_ix + q_iy + r_i], \quad (i = 1, n) \tag{5}$$

Layer 6: Summation layer

The output value of each node is summed up and, as a result, the true value of the ANFIS system is obtained. The calculation of y that is the output value of the system is according to Eq. 6.

$$y = \sum_{i=1}^n \bar{\mu}_i [p_ix + q_iy + r_i] \tag{6}$$

3.1.2. Genetic Algorithm (GA)

GA is the first artificial intelligence technique inspired by the basic elements and phenomena of the Theory of Evolution brought to the scientific world by Charles Robert Darwin. In general, a typical GA is based on subjecting a variety of genetic manipulations according to fitness values of the particles (individuals) represented by the chromosome, which are to be initially used for solution and are represented as chromosomes in a selected coding. According to the results of the initial population, the processes such as crossing and mutation are carried out. The new individuals are obtained and improved through the relatively better individuals until the desired result is obtained or a certain stop criterion is obtained. Genetic manipulations such as crossover and mutation are used to the coded individuals in the GA technique solution process, in order to genetically alter and improve the identified populations and thus to proceed in the positive direction of the solving process. The crossover process is to reproduce the pre-determined individuals to form a new individual. For this, new individuals are obtained by reciprocal gene - code exchange according to the GA's relevant parameters and the determined crossing pattern. In the mutation process, mutation processing is also applied to increase the potential of successful individuals to reach more successful solutions within the algorithm. In the selected individuals for mutation, the process of changing the codes is performed again in the same way as the GA's related parameters. The simplest example of this process is to make some 0's 1 and some 1's 0's in an individual coded with 0's and 1's [30,31].

In Table 1, the general pseudo-code of the GA is given [31-34].

Table 1. Pseudo-code of the Genetic Algorithm

- | |
|--|
| <p>Step 1. Generate initial population from random values</p> <p>Step 2. Calculate the fitness values of the chromosomes</p> <p>Step 3. Continue until conditions are met (Eligibility value, transaction time, etc.)</p> <ul style="list-style-type: none"> (i) Identify chromosomes with bad conformity in the population (ii) Determine parent chromosomes for child chromosomes (iii) Generate child chromosomes from parent chromosomes by crossover (iv) Child chromosomes mutate (v) Calculate the fitness values of newly included chromosomes in the population |
|--|

3.2. Heuristic Algorithms Used in ANFIS Training

When the literature is searched, it is seen that all or some part of ANFIS is trained with heuristic algorithms. The updating and training of the ANFIS model is basically the process of determining the most suitable values for the input and output parameters of ANFIS. When using derivative based optimization algorithms in ANFIS training, there is a local minimum risk. However, for derivative based algorithms, the calculation of the gradient at each step is relatively difficult and can be caused by a local minimum of probing, which must be used with the chain rule. In addition, the convergence of the parameters in the slope method is quite slow [35,36].

For this reason, heuristic algorithms are widely used in ANFIS training. Standard variants and heuristic algorithms are used for ANFIS training. The number of heuristic algorithms used in ANFIS training has increased rapidly in recent years. Heuristic algorithms; (GA), particle swarm optimization (PSO), artificial bee colony (DE), differential evolution (DE), Harmony search, CS (Cuckoo search), FA (Firefly algorithm), SA (Simulated annealing) MBA (Mine blast algorithm) and AIS (Artificial immune system). With the intuitive algorithms, ANFIS education has been popular recently, and it is emphasized that the ANFIS models, especially GA-supported models, are very effective, in the literature [28,36,37].

3.3. Proposed ANFIS-GA Model

This section contains full details of the proposed hybrid ANFIS-GA model, which was developed for the estimation of regional rainfall quantities.

3.3.1. Collection and Preparation of the Dataset

Within the scope of the study, total 3650 meteorological data from 2008-2018 of Basel, a Swiss city were used to construct the proposed model [38].

Data were given in Table 2 as training and testing.

Table 2. Part of the data used for the experiment (Normalized)

Data Type	Inputs					Output
	Temperature daily mean [2 m above gnd]	Relative Humidity daily mean [2 m above gnd]	Mean Sea Level Pressure daily mean	Sunshine Duration daily sum [sfc]	Wind Speed daily mean [10 m above gnd]	Total Rainfall daily sum [sfc]
Training Data	0,183825	0,626351	0,580044	0,274171	0,232694	0,1750
	0,237829	0,724155	0,42622	0,145596	0,322769	0,1852
	0,234371	0,736149	0,471231	0,129132	0,228524	0,1300
	0,447193	0,577027	0,489731	0	0,263136	0,5036
	0,367385	0,909291	0,512454	0,131128	0,241034	0,2446
Testing Data	0,440809	0,660135	0,808012	0,033721	0,511259	0,8993
	0,558127	0,692568	0,648507	0,172342	0,767515	0,2210
	0,350093	0,66723	0,68842	0,215494	0,550042	0,1922
	0,370843	0,237162	0,709687	0	0,359883	0,1125
	0,41261	0,168919	0,585142	0,51552	0,276897	0,1654

3.3.2. Creating of the Model

In the study, an ANFIS-GA was developed by integrating a GA technique in order to estimate the amount of regional rainfall with high accuracy. GA was utilized to optimize ANFIS parameters. GA was used because of it is stuck parameters in the local optimal have high probability and the ANFIS structure not reaching global optimization. GA technique in the developed ANFIS-GA model provides a closer relationship between the output and input of the model. First, input and output parameters were determined for creating the model. The parameters that make up the dataset were determined by taking into account the relationships between input and output. The inputs of the model are Temperature daily mean [2 m above gnd], Relative Humidity daily mean [2 m above gnd], Mean Sea Level pressure daily mean, Sunshine Duration daily sum [sfc] and Wind Speed daily mean [10 m above gnd]. At this point, inputs were determined according to the parameters that are directly related to the output. The output is Total Rainfall daily sum [sfc]. Particular attention was paid to the selection of the parameters that directly affect the amount of rainfall while the input parameters of the model were determined. In creation of the proposed model, MATLAB package program was used and Regression values (R) and MSE were taken into account. Flow chart of ANFIS-GA model was given in Figure 3.

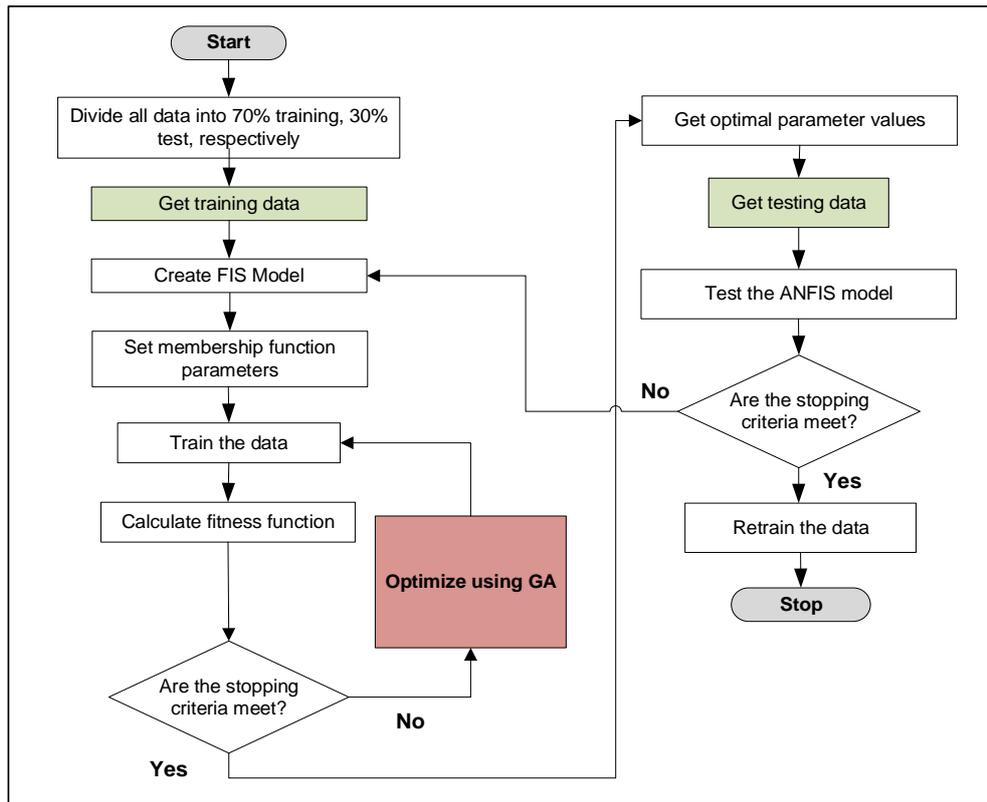


Figure 3. Flowchart of the proposed ANFIS-GA model

Modeling of ANFIS

Before ANFIS model was created, noisy (irrelevant or too far from normal value) data was cleared. The obtained data were normalized to the range [0-1] by the min-max method (See Table 2). Here, v_R is the real value of the input, v_{min} is the minimum input value, and v_{max} is the maximum input value (Equation 7) [39,40].

$$V_n = \frac{V_R - V_{min}}{V_{max} - V_{min}} \quad (7)$$

The data was first converted to fuzzy values to create the FIS part of the ANFIS system after the normalization process. Fuzzification process is the process of transforming control input information received from the system into symbolic values, which are linguistic qualifiers. Fuzzy set and membership degree to which the input information belongs are determined and the numerical value inputted is assigned the linguistic variable values such as the small, medium and high by using the membership function. Fuzzy sets in different shapes (triangle, trapezoid, bell curve etc.) can be selected to ensure efficient operation of the system [41,42].

In this study, a triangular membership function was chosen for each input and output variable, linguistic term clusters were created and weights were taken equal. The effects and variation intervals of the linguistic terms for input and output variables were given in Figure 4.

In the study, the linguistic terms for both input and output were determined as low, medium and high, and all details were given in Table 3.

Table 3. Parameters of the FIS model

No	Parameters	Types
1	FIS Type	Sugeno
2	MF Type	Triangle
3	Output MF	Triangle
4	Number of inputs	5
5	Number of outputs	1
6	And method	Prod
7	Or method	probar
8	Implication	Min
9	Aggregation	max
10	Defuzzification	wtaver
11	Number of Fuzzy Rule	243

Then, fuzzy inference process was applied. Expressions obtained by applying fuzzy logic on fuzzy rules are called fuzzy inference [41]. In the study, an inference was made considering the rule base when Sugeno method was applied to the numerical values obtained from the input variables (Fig. 5). Sugeno Fuzzy Inference System was produced by a direct transformation of the Mamdani Fuzzy [42].

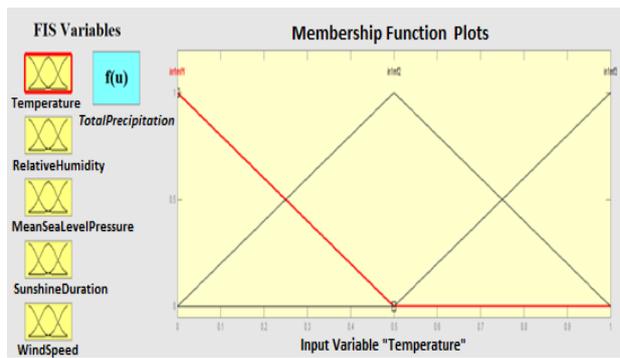


Figure 4. Effects and variation intervals of linguistic terms for input and output variables

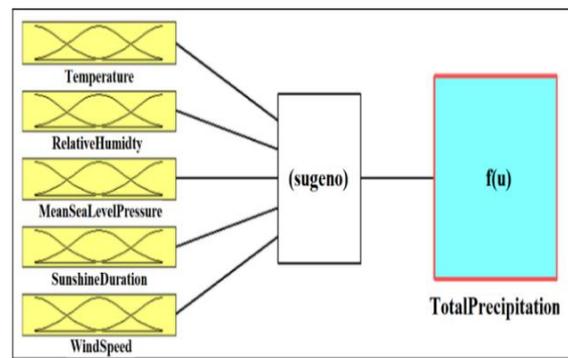


Figure 5. Sugeno FIS model

The model was created after the membership function and rules for input and output were determined. Figure 6 shows that there are 5 inputs in the first layer of the proposed model and a hidden layer in the second layer in which three (3) neurons corresponding to input membership functions (low, medium, high) the total number of inputs is 5), there are $3^5=243$ rules (the blue color shows that 'AND' is the aggregation operator used.), and finally shows that there is only one neuron for inference or output.

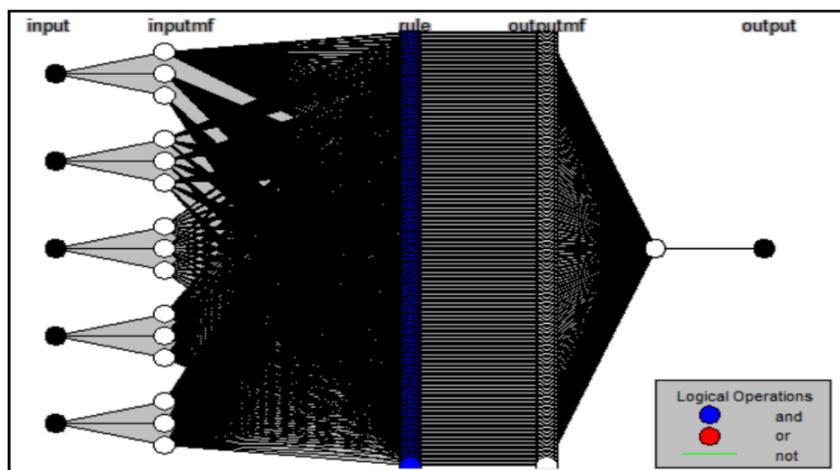


Figure 6. Proposed ANFIS model

Modeling of the GA

The genetic algorithm process, which is based on the natural selection and evolution approach and which is a search algorithm, is explained according to GA steps in the study. Firstly, GA parameters that determined and applied were given in Table 4. GA parameters were determined by taking widely used studies with successful results in the literature into consideration.

Table 4. GA parameters

No	Parameters	Types
1	Optimization method	Genetic algorithm
2	GA population size	3650
3	Selection	Roulette Wheel
4	Crossover percentage	One point
5	Crossover rate	0,7
6	Mutation rate	0,2
7	Elitism rate	0,1

Initialization of Population

Initialization of Population is known as the first step of genetic algorithm applications and means the total population. In the proposed model, the total population obtained was taken. So, there are 3650 populations in the model. This number refers to the data giving optimum results according to the fitness function.

Fitness Function

Fitness Function is utilized to measure the desired state, the threshold level accepted in the solution of the problem, and the quality of the problem [43,44]. The fitness function of the problem was calculated by taking into account the MSE between the real and the predicted data (Equation 8-9).

$$Fitness_Function = \frac{1}{1 + MSE} \quad (8)$$

$$MSE = \frac{\sum (y_t - \hat{y}_t)^2}{T} \quad (9)$$

(y_t = Real values, \hat{y}_t = Estimated values, T = Number of estimate) [40,45,46].

Selection

It is necessary to select new chromosomes after the fitness function has been identified. There are many selection methods in the literature such as stochastic universal sampling, Boltzmann, random, tournament, rank and Roulette wheel selection. The Roulette Wheel technique, which is widely used in literature, was used in the study [47,48].

Crossover and mutation

Chromosomes with the best fitness value are selected by the roulette wheel method, then crossed with other chromosomes to produce new chromosomes, and mutated using mutation. Thus, new generations are created. Crossover is the procedure of barter of genes between parents. Single-Point Crossover method was used in the proposed model (Figure.7).

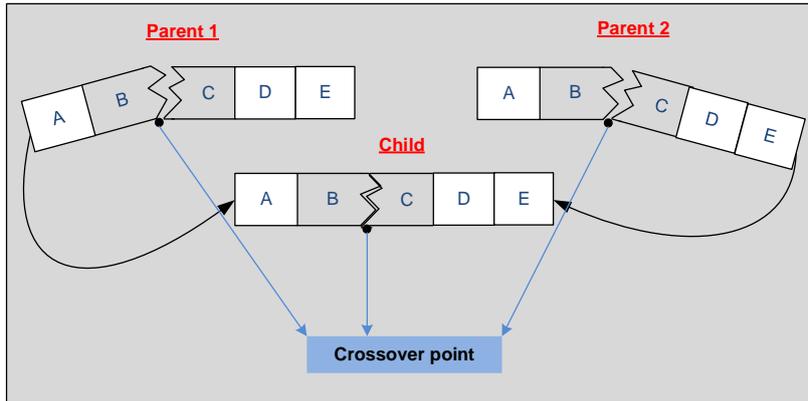


Figure 7. Crossover used in the ANFIS-GA

Then, mutation process was applied. The aim here is to prevent local optimal solution in the search process and to provide genetic diversity from the new generation by producing different experiences on optimization [49,50]. The mutation rate in study was set at 0.2.

Elitism

In the elitism mode, the strongest individuals remain. That is, when individuals with the best fitness function value are reached and the number of specified iterations (number of generations) is reached, the rate of elitism goes back to the other individual [31,51,52].

In the study, the number of chromosomes in rate (0,1) determined within the generated population was chosen as the best solution set and the others were discarded. After all these steps, the GA process was finalized.

4. FINDINGS AND PERFORMANCE ANALYSIS

In this section, the results of the experiment and the performance analysis of the study are included. The training, testing and verification R and MSE values of both the ANFIS model and the proposed hybrid ANFIS-GA model were shown in Table 5. Both ANFIS models were created using the same parameters. In both models, the dataset created in the study was utilized. In addition, the algorithms were run 100 times and the best results were given in Table 5. Learning algorithm used in the standard ANFIS is the backpropagation algorithm. Table 5 shows that the ANFIS-GA model developed by the hybrid approach has a much better performance than the ANFIS model. At this point, the optimization by using GA of the inputs used for training was effective.

Table 5. Features of the ANFIS and ANFIS-GA model

Method	Dataset	Performance Criteria				
		MSE	RMSE	R	Time (s)	Epochs
ANFIS	Training	0.021	0,1449	0.9112	33,5	112
	Testing	0.026	0,1612	0.8973	26,3	87
ANFIS-GA	Training	0.0011	0,0331	0.9920	12,6	39
	Testing	0.0014	0,0374	0.9840	9,4	27

Figure 8 shows the training and testing error graphs of the proposed ANFIS-GA approach.

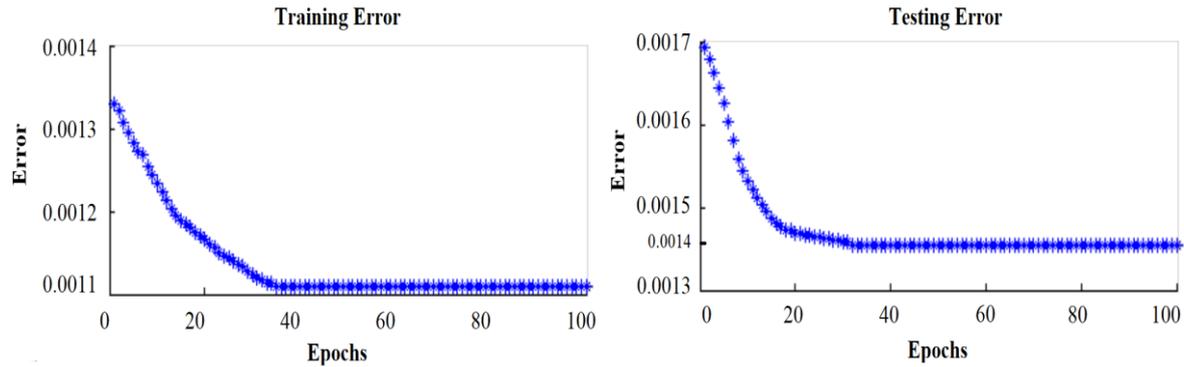


Figure 8. Error plot of the ANFIS-GA model

In order to measure the performances of ANN models, RMSE (Equation 10), Mean Absolute Percentage Error-MAPE (Equation 11), MSE (Equation 12), and Absolute Change Percentage- Equation 13) are used in the literature. In this study, it was preferred to get the lowest possible value from Equation 10-12, and the highest value from Equation 13 [45,46,53].

$$RMSE = \sqrt{\frac{\sum (y_t - \hat{y}_t)^2}{T}} \quad (10) \quad MSE = \frac{\sum (y_t - \hat{y}_t)^2}{T} \quad (12)$$

$$MAPE = \frac{1}{T} \sum \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100 \quad (11) \quad R^2 = 1 - \left(\frac{\sum (y_t - \hat{y}_t)^2}{\sum (\hat{y}_t)^2} \right) \quad (13)$$

(y_t = Real values, \hat{y}_t = Estimated values, T = Number of estimate)

The performance level of the suggested model was tested on real data using the formulas given in Equation 10-13, and the obtained data were given in Table 6. When Table 6 is examined, it was seen that the values obtained with ANFIS-GA model have lower error rates and higher success rates than the results obtained with standard ANFIS. According to the mean values of the results obtained in both models, ANFIS-GA model has lower RMSE, MAPE and MSE values and higher R^2 .

Table 6. Performance analysis of the proposed ANFIS-GA model

No	Real Value	Estimated Value		RMSE		MAPE		MSE		R ²	
	ANN and ANFIS-GA	ANN	ANFIS-GA	ANN	ANFIS-GA	ANN	ANFIS-GA	ANN	ANFIS-GA	ANN	ANFIS-GA
1	0,17750	0,21120	0,18520	0,03370	0,00770	2,86200	0,65393	0,00114	0,00006	97,13800	99,34607
2	0,13280	0,20520	0,14250	0,07240	0,00970	6,39124	0,85629	0,00524	0,00009	93,60876	99,14371
3	0,18270	0,12910	0,17190	0,05360	0,01080	4,53200	0,91316	0,00287	0,00012	95,46800	99,08684
4	0,34180	0,38230	0,35100	0,04050	0,00920	3,01833	0,68565	0,00164	0,00008	96,98167	99,31435
5	0,31300	0,37542	0,29520	0,06242	0,01780	4,75400	1,35567	0,00390	0,00032	95,24600	98,64433
6	0,16140	0,09860	0,15620	0,06280	0,00520	5,40727	0,44774	0,00394	0,00003	94,59273	99,55226
7	0,21520	0,29210	0,22280	0,07690	0,00760	6,32818	0,62541	0,00591	0,00006	93,67182	99,37459
8	0,71640	0,65321	0,72480	0,06319	0,00840	3,68154	0,48940	0,00399	0,00007	96,31846	99,51060
9	0,01800	0,03835	0,02230	0,02035	0,00430	1,99902	0,42240	0,00041	0,00002	98,00098	99,57760
10	0,82560	0,91587	0,83520	0,09027	0,00960	4,94468	0,52585	0,00815	0,00009	95,05532	99,47415
11	0,95130	0,89100	0,94510	0,06030	0,00620	3,09025	0,31774	0,00364	0,00004	96,90975	99,68226
12	0,24590	0,18990	0,23560	0,05600	0,01030	4,49474	0,82671	0,00314	0,00011	95,50526	99,17329
13	0,23320	0,32854	0,25470	0,09534	0,02150	7,73111	1,74343	0,00909	0,00046	92,26889	98,25657
14	0,89590	0,99925	0,91010	0,10335	0,01420	5,45124	0,74898	0,01068	0,00020	94,54876	99,25102
15	0,35730	0,45850	0,36490	0,10120	0,00760	7,45598	0,55994	0,01024	0,00006	92,54402	99,44006
16	0,20951	0,27440	0,22590	0,06489	0,01639	5,36503	1,35514	0,00421	0,00027	94,63497	98,64486
17	0,09992	0,11230	0,13450	0,01238	0,03458	1,12509	3,14341	0,00015	0,00120	98,87491	96,85659
18	0,54749	0,50365	0,62010	0,04384	0,07261	2,83322	4,69186	0,00192	0,00527	97,16678	95,30814
19	0,45412	0,49580	0,47100	0,04168	0,01688	2,86661	1,16111	0,00174	0,00029	97,13339	98,83889
20	0,56885	0,60030	0,55330	0,03145	0,01555	2,00448	0,99134	0,00099	0,00024	97,99552	99,00866
21	0,17892	0,22570	0,18940	0,04678	0,01048	3,96802	0,88893	0,00219	0,00011	96,03198	99,11107
22	0,49363	0,46790	0,51010	0,02573	0,01647	1,72295	1,10238	0,00066	0,00027	98,27705	98,89762
23	0,65397	0,52540	0,62930	0,12857	0,02467	7,77319	1,49132	0,01653	0,00061	92,22681	98,50868
24	0,40839	0,42250	0,40580	0,01411	0,00259	1,00162	0,18413	0,00020	0,00001	98,99838	99,81587
25	0,99622	0,94690	0,98410	0,04932	0,01212	2,47077	0,60725	0,00243	0,00015	97,52923	99,39275
26	0,38349	0,42560	0,40010	0,04211	0,01661	3,04356	1,20040	0,00177	0,00028	96,95644	98,79960
27	0,65397	0,72420	0,67220	0,07023	0,01823	4,24641	1,10245	0,00493	0,00033	95,75359	98,89755
28	0,59354	0,66310	0,60140	0,06956	0,00786	4,36521	0,49332	0,00484	0,00006	95,63479	99,50668
29	0,72953	0,78810	0,70890	0,05857	0,02063	3,38667	1,19262	0,00343	0,00043	96,61333	98,80738
30	0,82341	0,98310	0,83370	0,15969	0,01029	8,75784	0,56439	0,02550	0,00011	91,24216	99,43561
Mean of Results				0,06171	0,01487	4,23574	1,04474	0,00485	0,00038	95,76426	98,95526

Figure 9 shows the comparative display of real and forecasted values obtained from the ANN and ANFIS-GA models. Based on the results obtained using real data, the proposed hybrid ANFIS-GA model performed higher than the ANN model predictive data. It is also clear that proposed Hybrid ANFIS-GA model overlaps with the real data in a significant manner, the accuracy level is high and the error rate is very low.

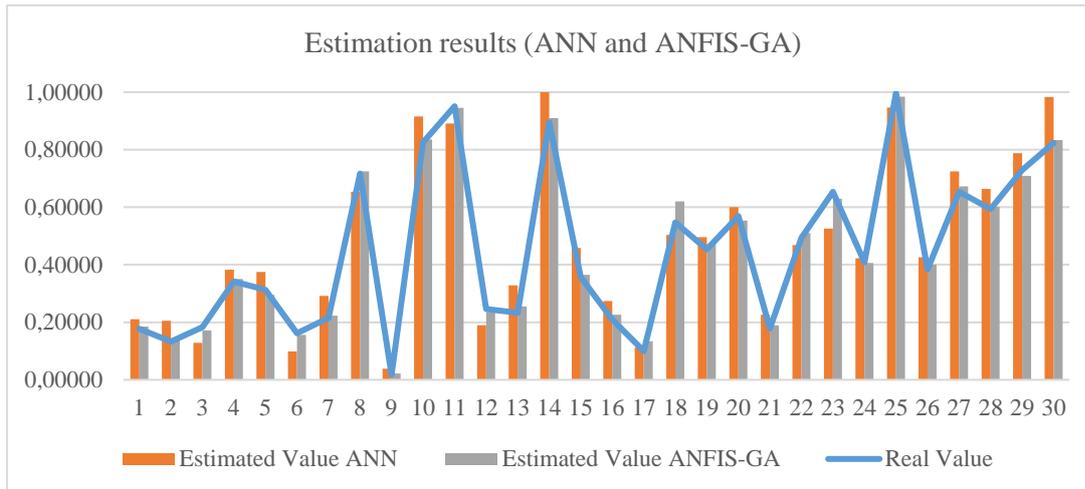


Figure 9. The comparison of real and estimated results for ANN and ANFIS-GA

The regression curve in Figure 10 demonstrated that there is a very high correlation between the estimated values obtained from the model and the real values with the ratio of 0.9955.

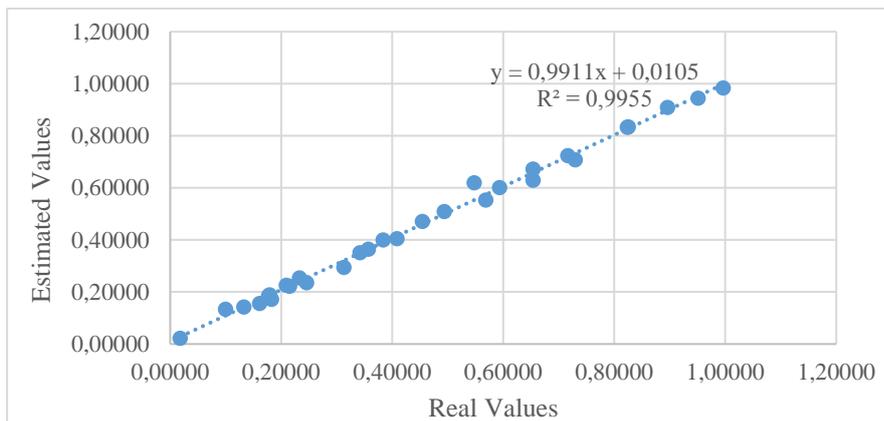


Figure 10. Regression graph of real and estimated values

5. CONCLUSION AND RECOMMENDATIONS

In this study, an ANFIS-GA based hybrid model was proposed to estimate regional annual precipitation. The input values of the ANFIS model was optimized using the GA technique. The experimental findings of both normal ANFIS and proposed hybrid ANFIS-GA models were compared and it was obtained that the hybrid ANFIS-GA model had higher performance (lower error rate, higher accuracy level in prediction). According to the results, the training R-value of the proposed ANFIS-GA model, the testing R-value, and the error rate were obtained as 0.9920, 0.9840 and 0.0011, respectively. Because it has a high predictive rate, it is clear that the suggested hybrid ANFIS-GA model can be used easily in meteorological matters such as rainfall amount. In addition, this study shows that the development and use of ANN models in a hybrid way, supported by optimization methods, affects the prediction performance of the model positively. At this point, the results of the experiment can be compared by testing with other optimization techniques (particle swarm optimization, simulated annealing, hybrid algorithms and differential evolution etc.), while achieving high-performance results with ANFIS-GA. In addition, if it is thought that sensitivity of the study will affect positively, the sub-factors affecting the amount of rainfall can be considered as an input. Finally, the study can be used as an important support tool for estimating the amount of annual rainfall, ensuring effective management of water resources, and conducting some assessments and preparations based on possible climate changes. In the next phase of the study, a web-based system is planned to present statistical and graphical results of the proposed model.

CONFLICT OF INTEREST

No conflict of interest was declared by the authors.

REFERENCES

- [1] Anlı et al., "Regional Frequency Analysis of the Annual Maximum Precipitation Observed In Trabzon Province", *Journal of Agricultural Sciences*, 15(3), 240-248, (2009).
- [2] Tonkaz, T., "An Assessment of Monthly Total Precipitation Characteristics in GAP Area and Generation of Synthetic Series of Monthly Precipitation Data", *Journal of Agricultural Sciences*, 13(1):29-37, (2007).
- [3] Turhan, E., Çağatay, H. Ö., and Çetin, A., "Modelling of Rainfall-Runoff Relation with Artificial Neural Network Methods for Lower Seyhan Plain Sub-Basin and Assessment in Point of Rainy-Droughty Terms", *Çukurova University Journal of the Faculty of Engineering and Architecture*, 31(2), pp. 227-241, December (2016).
- [4] Saplıoğlu, K. and Çimen, M., "Predicting Of Daily Precipitation Using Artificial Neural Network", *Journal of Engineering Science and Design*, Vol:1 No:1 pp.14-21, (2010).
- [5] Sharma, A. and Nijhawan, G. "Rainfall Prediction Using Neural Network", *IJCST* 3.3, 65-69, (2015).
- [6] Taylan, E. D., "Precipitation Prediction Model with Genetic Evaluatory Programming", *SDU International Journal of Technological Science* 7.1 (2015).
- [7] Shoba, G. and Shobha, G., "Rainfall prediction using Data Mining techniques: A Survey". *Int. J. of Eng. and Comput. Sci*3, no. 5: 6206-6211, (2014).
- [8] Zaw, W. T., and Naing, T. T. "Empirical statistical modeling of rainfall prediction over Myanmar", *World Academy of Science, Engineering and Technology* 2, no. 10: 500-504, (2008).
- [9] Sukanya, R. and Prabha, K., "Comparative Analysis for Prediction of Rainfall using Data Mining Techniques with Artificial Neural Network." Volume-5, Issue-6, Page 288-292, (2017).
- [10] Deo, R. C., and Şahin, M. "Application of the extreme learning machine algorithm for the prediction of monthly Effective Drought Index in eastern Australia", *Atmospheric Research*, 153, 512-525, (2015).
- [11] Wu, J., Long, J., and Liu, M., "Evolving RBF neural networks for rainfall prediction using hybrid particle swarm optimization and genetic algorithm", *Neurocomputing*, 148, 136-142, (2015).
- [12] Plouffe, C. C., Robertson, C., and Chandrapala, L., "Comparing interpolation techniques for monthly rainfall mapping using multiple evaluation criteria and auxiliary data sources: A case study of Sri Lanka", *Environmental Modelling & Software*, 67, 57-71, (2015).
- [13] Partal, T., Cigizoglu, H. K., and Kahya, E., "Daily precipitation predictions using three different wavelet neural network algorithms by meteorological data", *Stochastic environmental research and risk assessment*, 29(5), 1317-1329, (2015).
- [14] Beheshti, Z., Firouzi, M., Shamsuddin, S. M., Zibarzani, M., and Yusop, Z., "A new rainfall forecasting model using the CAPSO algorithm and an artificial neural network", *Neural Computing and Applications*, 27(8), 2551-2565, (2016).
- [15] Tezel, G., and Buyukyildiz, M., "Monthly evaporation forecasting using artificial neural networks and support vector machines. Theoretical and applied climatology", 124(1-2), 69-80, (2016).
- [16] Akrami, S. A., El-Shafie, A., and Jaafar, O., "Improving rainfall forecasting efficiency using modified adaptive Neuro-Fuzzy Inference System (MANFIS)", *Water resources management*, 27(9), 3507-3523, (2013).

- [17] Islam, S., and Talukdar, B., "Performance Improvement of a Rainfall Prediction Model using Particle Swarm Optimization", *International Journal of Computational Engineering Research*, 6(7), 39-42, (2016).
- [18] Qiu, M., Zhao, P., Zhang, K., Huang, J., Shi, X., Wang, X., and Chu, W. "A Short-Term Rainfall Prediction Model using Multi-Task Convolutional Neural Networks", In *Data Mining (ICDM), 2017 IEEE International Conference on* (pp. 395-404). IEEE. (2017, November).
- [19] Taormina, R., and Chau, K. W. "Data-driven input variable selection for rainfall-runoff modeling using binary-coded particle swarm optimization and extreme learning machines", *Journal of Hydrology*, 529, 1617-1632. (2015).
- [20] Sethi, N., and Garg, K., "Exploiting data mining technique for rainfall prediction", *International Journal of Computer Science and Information Technologies*, 5(3), 3982-3984. (2014).
- [21] Kashani, M. H., Ghorbani, M. A., Dinpashoh, Y., and Shahmorad, S., "Integration of Volterra model with artificial neural networks for rainfall-runoff simulation in forested catchment of northern Iran", *Journal of Hydrology*, 540, 340-354. (2016).
- [22] Shafaei, M., and Kisi, O., "Predicting river daily flow using wavelet-artificial neural networks based on regression analyses in comparison with artificial neural networks and support vector machine models", *Neural Computing and Applications*, 28(1), 15-28. (2017).
- [23] Cuevas, E., Díaz, P., Avalos, O., Zaldívar, D., and Pérez-Cisneros, M. "Nonlinear system identification based on ANFIS-Hammerstein model using Gravitational search algorithm", *Applied Intelligence*, 48(1), 182-203. (2018).
- [24] Jang, J. S., "ANFIS: Adaptive-Network-Based Fuzzy Inference System", *IEEE Transactions on Systems, Man and Cybernetics*, 23(3), 665-685, (1993).
- [25] Prasad, K., Gorai, A. K., Goyal, P., "Development of ANFIS models for air quality forecasting and input optimization for reducing the computational cost and time", *Atmos Environ*, 128, 246-262, (2016).
- [26] Demuth, H., Beale, M. "ANFIS and the ANFIS Editor GUI, Fuzzy logic toolbox for use with MATLAB", *User's Guide Version 4*, 2:104-130, MA, (2000).
- [27] Zare, M. and Koch, M., "Groundwater level fluctuations simulation and prediction by ANFIS-and hybrid Wavelet-ANFIS/Fuzzy C-Means (FCM) clustering models: Application to the Miandarband plain", *Journal of Hydro-environment Research*, 18, 63-76. (2018).
- [28] Karaboga, D. and Kaya, E. "Adaptive network based fuzzy inference system (ANFIS) training approaches: a comprehensive survey", *Artificial Intelligence Review*, 1-31. (2018).
- [29] Takagi, T., Sugeno, M., "Fuzzy identification of systems and its applications to modelling and control", *IEEE Transactions on Systems, Man and Cybernetics*, 15: 116-132, (1985).
- [30] Kramer, O., "Genetic Algorithm Essentials", Springer, (2017).
- [31] Dener, M., Calp, M. H., "Solving the Exam Scheduling Problems in Central Exams with Genetic Algorithms", *Mugla Journal of Science and Technology*, 4(1), 102-115, (2018).
- [32] Haznedar, B., Arslan, M. T., and Kalınlı, A., "Training ANFIS structure using genetic algorithm for liver cancer classification based on microarray gene expression data", *Sakarya University Journal of Science*, 21(1), 54-62, (2017).
- [33] Calp, M. H., and Akcayol, M. A., "Optimization of Project Scheduling Activities in Dynamic CPM and PERT Networks Using Genetic Algorithms", *Süleyman Demirel University Journal of Natural and Applied Sciences (SDU J Nat Appl Sci)*, 22(2), 615-627, (2018).
- [34] Calp, M. H., "Medical Diagnosis with a Novel SVM-CoDOA Based Hybrid Approach", *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, 9(4), 6-16, (2018).

- [35] Haznedar, B. and Kalinli, A., "Training ANFIS structure using simulated annealing algorithm for dynamic systems identification", *Neurocomputing* 302: 66-74, (2018).
- [36] Kacher, O., Boltnev, A., Kalitkin, N., "Speeding up the convergence of simple grained methods", in: *Proceedings of the 10th International Conference*, pp. 349-354, (2005).
- [37] Simon, D. J., "Training fuzzy systems with the extended Kalman filter", *Fuzzy sets and systems* 132(2), 189, (2002).
- [38] Meteoblue Weather,
https://www.meteoblue.com/en/weather/archive/export/basel_switzerland_2661604, Access Date: 07.04.2018.
- [39] Bayram, S., Kaplan, K., Kuncan, M., Ertunç, H. M., "Ball Bearings space of time Statistical Feature Extraction and Neural Networks with Error Estimation Method Size", *Automatic Control National Meeting, TOK2013, Malatya, 26-28 September*, (2013).
- [40] Calp, M. H., "An estimation of personnel food demand quantity for businesses by using artificial neural networks", DOI: 10.2339/politeknik.444380, (2019). (Accepted - In Press).
- [41] Özkan, M. T., Eldem, C., and Köksal, E., "Notch Sensitivity Factor Determination with Artificial Neural Network for Shafts under the Bending Stress", *Pamukkale University Journal of Engineering Sciences*, 19(1), 24-32, (2013).
- [42] Gandomi, A. H. and Roke, D. A., "Assessment of artificial neural network and genetic programming as predictive tools", *Advances in Engineering Software*, 88, 63-72, (2015).
- [43] Elmas, Ç., "Artificial Intelligence Applications", Seçkin Publishing, First Edition, November, (2007).
- [44] Calp, M. H., Şahin, İ., "The Determination by Using Fuzzy Expert System of the Usability Level of Website User Interface Design", *International Journal of Human Sciences*, Volume: 10 Special Issue, 141-150, (2013).
- [45] Isiaka, R.M., Omidiora, E.O, Olabiyisi, S.O, and Okediran, O.O., "Mamdani fuzzy model for learning activities evaluation", *International Journal of Applied Information Systems (IJ AIS)*. Foundation of Computer Science FCS. New York, USA. 7(3): p. 1-8, (2014).
- [46] Pereira, L. F., Patil, S. A., and Mahadeshwar, C. D., "Artifact removal from EEG using ANFIS-GA", In *Green Engineering and Technologies (IC-GET)*, 2016 Online International Conference on (pp. 1-6). IEEE, (2016, November).
- [47] Wahyuni, I. and Mahmudy, W. F., "Rainfall Prediction in Tengger-Indonesia Using Hybrid Tsukamoto FIS and Genetic Algorithm", Submitted to *J. ICT Res. Appl.*, pp. 1–8, (2016).
- [48] Moradi, S.T. and Nikolaev, N.I., "Optimization of Cement Spacer Rheology Model Using Genetic Algorithm", *IJE Trans. A Basics*, 29(1), pp. 127-131, (2016).
- [49] Lu, J. T., Chang, Y. C., and Ho, C. Y., "The optimization of chiller loading by adaptive neuro-fuzzy inference system and genetic algorithms," *Mathematical Problems in Engineering*, (2015).
- [50] Sarkheyli, A., Zain, A. M., and Sharif, S., "A multi-performance prediction model based on ANFIS and new modified-GA for machining processes", *Journal of Intelligent Manufacturing*, 26(4), 703-716, (2015).
- [51] Wahyuni, I., Mahmudy, W. F., and Iriany, A., "Rainfall Prediction using Hybrid Adaptive Neuro Fuzzy Inference System (ANFIS) and Genetic Algorithm", *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, 9(2-8), 51-56, (2017).
- [52] Lee, Y.G., "Genetic Optimization Applied to Larry Connors Trading System in Two Markets", *Review of Computational Science and Engineering* 3, no. 3:29, (2017).
- [53] Qasem, S. N., Ebtehaj, I., & Riahi Madavar, H., "Optimizing ANFIS for sediment transport in open channels using different evolutionary algorithms", *Journal of Applied Research in Water and Wastewater*, 4(1), 290-298, (2017).