The Pharma Innovation Journal 2023; 12(2): 1651-1662

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277-7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2023; 12(2): 1651-1662

www.thepharmajournal.com
Received: 08-01-2022
Accepted: 12-02-2023

Avanish Yadav
Department of Soil and Water Conservation Engineering, Sam Higginbottom University of Agriculture, Technology and Sciences, Prayagraj, Uttar Pradesh, India

MA Alam
Department of Irrigation and Drainage Engineering, Sam Higginbottom University of Agriculture, Technology and Sciences, Prayagraj, Uttar Pradesh, India

Shakti Suryavanshi
Department of Soil and Water Conservation Engineering, Sam Higginbottom University of Agriculture, Technology and Sciences, Prayagraj, Uttar Pradesh, India

Corresponding Author:
Avanish Yadav
Department of Soil and Water Conservation Engineering, Sam Higginbottom University of Agriculture, Technology and Sciences, Prayagraj, Uttar Pradesh, India

Rainfall-runoff analysis using artificial intelligence couples with genetic algorithm

Avanish Yadav, MA Alam and Shakti Suryavanshi

Abstract
A daily rainfall runoff time-series estimate is crucial to water resource planning and development. The aim of this research to comparatively examine the applicability of Artificial Intelligence (AI) methods (i.e., multilayer perceptron (MLP), Support Vector Machine (SVM), multilayer perceptrons neural network coupled with genetic algorithms (MLP-GA) and the Support Vector Machine coupled with genetic algorithms (SVM-GA) model for estimating and simulating the daily runoff discharge. The models investigated at Jondhara stations, Seonath stream in Bilaspur district, Chhattisgarh, India. Three performance criteria, including correlation coefficient (CC), root mean square error (RMSE) and percent bias (PBIAS), were applied for model performance assessment. Based on the result findings, SVM-GA algorithms showed superior performance to other models for stations with low correlation coefficients (CC), RMSE and PBIAS (CC = 0.9751, RMSE = 0.3230, and PBAIS = -0.0080 in training data set and CC = 0.9575, = RMSE = 0.3709 and PBAIS = -0.0117 in testing data set in Q-4 model and CC = 0.9743, RMSE = 0.3354 and PBAIS = -0.0084 for training data set and CC = 0.9733, RMSE = 0.3390, and PBAIS = -0.0104 for testing data det in Q-16 model). As compared to statistical criteria all methods, specially SVR-GA performed exceptionally well. The study indicated that SVR-GA could handle and simulate daily runoff based on limited information.

Keywords: Hydrological process, multilayer perceptron neural networks, support vector machine, genetic algorithm, runoff forecasting, rainfall-runoff relationship

1. Introduction

The accurate forecasting of river flows/runoff/stream discharge and sediment yield are one of the most vital parts of achieving a sustainable water resource management and protecting the environment [1–3]. In order to design hydraulic structures, to calculate flood and drought patterns, to schedule irrigation, to develop reservoir operations, to make environmental planning decisions, and to analyse flood and drought patterns, it is critically important to produce precise short and long-term stream discharge forecasts [3, 4]. Water is primarily provided by runoff from rainfall, especially in arid and low-rainfall regions [5]. In areas with arid or semi-arid and dry climates, technical and scientific methods are increasingly essential to accumulate and control surface runoff to provide water storage for irrigation, domestic and dam projects purpose. Therefore, in order to improve the estimation of the runoff from catchment basins after rain, it is crucial that we obtain more accurate data of runoff from rainfall [1, 2, 6].

Several literature reviews have suggested that historic precipitation data set is often used to estimate surface runoff [4, 7]. The runoff is also affected by many meteorological variables such as evaporation, air temperature, net solar radiation, relative humidity and wind speed, etc., as well as basin-specific characteristics such as area, length, slope, distance from the basin's centroid to the basin outlet, etc [8, 9]. Over the past few decades, a number of research efforts have been conducted all over the world to try and predict more than just rainfall. In most of these studies, empirical, numerical, deterministic, and statistical methods were the main approaches used [10–12]. There have been many approaches developed since the middle of the previous century to address this issue, including the soil conservation curve number method (SCS-CN), developed to deal with this issue [13]. It's been a challenge to forecast rainfall-runoff with rainfall as an input in water resources management. Serval studies have shown that optimization techniques like Harris Hawks Optimization [14], Genetic Algorithm (GA) [15], Grey Wolf Optimization (GWO) [16], augmented grey wolf optimization (ANN-AGWO) [16], marine predators algorithm (ANN-MPA) [16] and Particle Swarm Optimization (PSO) [17], Neural Network-Based Forecasting Approaches [2, 10, 18], machine learning [19], and artificial intelligence [11, 20] can be appropriate solutions to solve complex complications related to water resources like, rainfall-runoff forecasting, meteorological parameters forecasting, mechanical and electrical engineering, coastal engineering, and safety in water supply networks and pipe networking etc [1, 21–28].

Studies have been conducted using the artificial intelligence model, and there have been no studies comparing much more more models. As a result, additional models should be examined and compared with each other because, depending on the conditions in a region, different models can produce different rainfall-runoff results. Past studies have not explored the use of two AI algorithms such as MLP, SVM and one optimizer technique such as genetic algorithm programming (GEP) methods together, which is explored in this study. The present research simulates rainfall-runoff in Jondhara station at Seonath stream in Bilaspur district, Chhattisgarh using multilayer perceptron (MLP), support vector machine (SVM), genetic algorithm-based hybrid multilayer perceptron (MLP-GA), and genetic algorithm-based hybrid support vector machine (SVM-GA) models., respectively.

2. Methodology

2.1 Study Area Description and Collection of Data

The origin of Seonath stream is Panabaras cluster in the Rajnandgaon, Chhattisgarh-India.

Basin is situated between latitude 20^0 16' N to 22^0 41' N & Longitude 80^0 25' E to $82^0$35' E. This basin of the river extends over an area of 30,860 square kilometers up to the confluence of the river with the Mahanadi. This river traverses a length of 380 kilometers during its journey. Its main tributaries include Kharun, Arpa, Hamp, Agar, and Maniyari Rivers. There is a wide variation in the amount of mean average rainfall received in the basin each year, ranging from 1005 mm to 1255 mm. The details of data used in model are given Table 1.

Table 1: List of hydrological gauge stations in Seonath Basin

Station	District	Latitude	Longitude	Stream	Data Available	Type of Site
Jondhara	Bilaspur	21°42'47"	82°21'30"	Seonath	2000-2020	GDSQ

2.2 Artificial intelligence models

2.2.1 Multilayer perceptron neural networks (MLP)

A neural network consists of layers of neurons arranged in parallel [29]. Neurons are located in the hidden layers of an ANN, which is composed of an input layer, a hidden layer, and an output layer. During the training phase, neurons are connected by weights to neurons in subsequent layers. There are two activation functions that can be used in order to analyse the features present in the input data. These are sigmoid activation functions and linear activation functions. These two functions are commonly used in the hidden, layer as well as output layer [30, 31]. A multilayer perceptron with a back propagation algorithm is considered to be the most common and popular kind of network, which will be the subject of this study. A number of prediction problems have been successfully tackled with the help of the backpropagation training algorithm [32–34]. Ghorbani *et al.* [35] elaborated more details about the MLP method.

2.2.2 Support vector machine (SVM)

The support-vector machine (SVM, also known as the support-vector network) is a programming model for supervised learning which consists of an associated learning algorithm to analyze the data in order to make classifications and regressions This technique was developed with his colleagues at AT&T Bell Laboratories by Vladimir Vapnik [36, 37]. SVMs can be considered to be one of the most reliable methods for predicting outcomes, since they are based on statistical learning frameworks or on the V-C theory proposed by Cortes and Vapnik [38]. As a result of a set of training examples that have each been marked as belonging to a specific category, the SVM training algorithm builds a model by assigning each new example to one of the two categories based on the mark on the training examples. It is therefore a non-probabilistic binary linear classifier (although methods like Platt scaling exist for implementing SVM in a probabilistic classification environment). This method of machine learning maps the training examples to points in space in such a manner as to increase the width of the gap between the two categories. After mapping the new examples into this space, the predictions are made to determine whether they belong to a specific category based on the side of the gap from which they come [39].

2.2.3 Hybrid artificial intelligence Algorithm Based on GA

GAs are stochastic optimization methods that don't require the use derivatives and are encouraged by natural assortment in the fields of genomics and evolution in biology. In many ways, it is finest to other optimization methods in terms of performance. In terms of continuous and discrete optimization, this method can be used to solve problems in both areas. Unlike the artificial intelligence method, the GA method is less likely to cause the user to get stuck in local least than the AI method. The computational model is based on a population-inspired approach. This is a population genetics-inspired algorithm for learning how to learn new things. Traditionally, it has been used primarily as an optimization function and that has been found to be a valuable global optimization method, particularly for multi-model and non-continuous processes that require global optimization. An overview of the suggested hybrid algorithms can be seen in Figures 1a and 2b which show a schematic representation of them. The proposed model proposes a hybrid AI (MLP and SVM) learning technique using the GA tool to optimise the hyperparameters of the network through integration with artificial intelligence. A chromosome of hyperparameters is encoded on an encoding matrix that allows the GA to tune each hyperparameter on the network. In the final step of the process, as a consequence of the GA procedure, the AI technique is used to train the network.

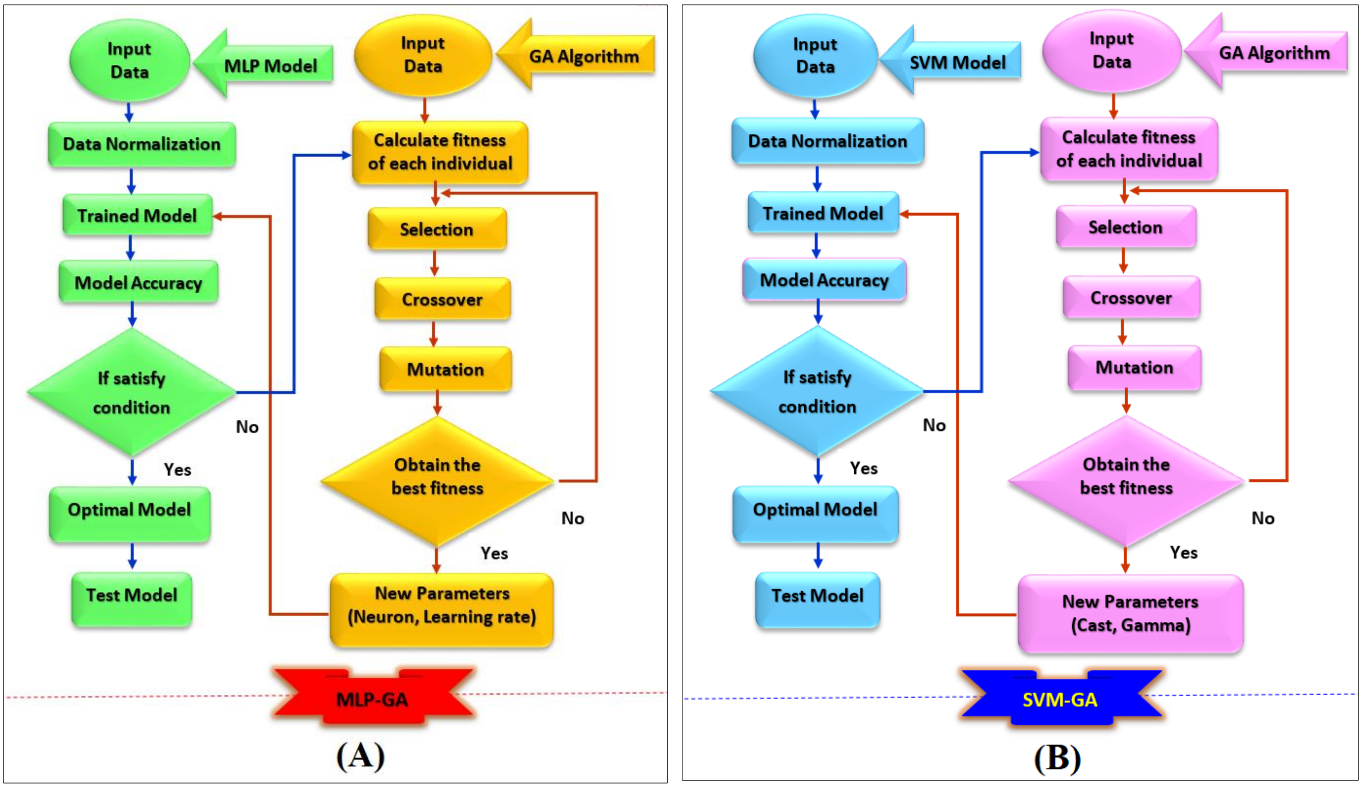


Fig 1: Flow diagram of hybrid MLP-GA and SVM-GA algorithm

2.3 Model development

To predict the daily runoff, it is a complex procedure, which is dependent on various parameters such as rainfall behavior, runoff, soil properties, and vegetative cover, etc. It is also dependent on a particular time lag for the purposes of modelling. Thus, for the purpose of developing rainfall runoff models, different combinations of runoff and rainfall were combined.

$$Q_{i,j} = f\left(Q_{i,j}, Q_{i,j-1}, Q_{i,j-2}, \dots \dots \dots Q_{i,j-m}, R_{i,j-1}, R_{i,j-2}, R_{i,j-2}, \dots . . R_{i,j-m}\right) \quad (1)$$

where, $Q_{i,j}$ is the runoff for j^{th} day of the i^{th} year, S_{ij} is the runoff for j^{th} day of the i^{th} year and m standards for time lag which is taken as three in present study. Hence, different combinations of runoff and rainfall need to be considered when developing runoff prediction models. Various input variables were employed to effectively predict the daily runoff discharge presented in Table 1 in accordance with the significant correlation between the inputs and output.

Table 1: Combination of input- output variable

Model	Output	Inputs variable
Q-1	Q_t	R_t, R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4}, R_{t-5}, R_{t-6}
Q-2	Q_t	R_t, R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4}, R_{t-5}
Q-3	Q_t	R_t, R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4}
Q-4	Q_t	R_t, R_{t-1}, R_{t-2}, R_{t-3}
Q-5	Q_t	R_t, R_{t-1}, R_{t-2}
Q-6	Q_t	R_t, R_{t-1}
Q-7	Q_t	R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4}, R_{t-5}, R_{t-6}
Q-8	Q_t	R_{t-2}, R_{t-3}, R_{t-4}, R_{t-5}, R_{t-6}
Q-9	Q_t	R_t, R_{t-1}, R_{t-2}, R_{t-5}, R_{t-6}
Q-10	Q_t	R_t, R_{t-2}, R_{t-3},
Q-11	Q_t	Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}, R_t, R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4}, R_{t-5}
Q-12	Q_t	Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, R_t, R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4}
Q-13	Q_t	Q_{t-1}, Q_{t-2}, Q_{t-3}, R_t, R_{t-1}, R_{t-2}, R_{t-3}
Q-14	Q_t	Q_{t-1}, Q_{t-2}, R_t, R_{t-1}, R_{t-2}, R_{t-3}
Q-15	Q_t	Q_{t-1}, Q_{t-2}, R_t, R_{t-1}, R_{t-2}
Q-16	Q_t	Q_{t-1}, Q_{t-2}, Q_{t-3}, R_t, R_{t-1}
Q-17	Q_t	Q_{t-1}, Q_{t-2}, Q_{t-3}, R_t,
Q-18	Q_t	Q_{t-1}, R_t, R_{t-1},
Q-19	Q_t	Q_{t-1}, R_t, R_{t-1}, R_{t-2}
Q-20	Q_t	Q_{t-1}, R_t, R_{t-1}, R_{t-2}, R_{t-3}

Where Q_t and R_t = Present day discharge (m^3/sec) and rainfall (mm).

2.4 Best input Variable selection: Gamma Test (GT)

In any hydrological environment, there are many complex, dynamic, and non-uniform processes occurring constantly. A Gamma test establishes an impartial and multi-objective way of determining each input parameter's significant potential. Scholars use a tedious and time-consuming trial-and-error method to determine the ideal input combination [1, 10]. Therefore, to resolve this problem, a novel approach Gamma Test, is used to evaluate the ideal input variables in a data set, introduced by Stefansson *et al.* [40]. It is competent enough to create a trustworthy and smooth model. The gamma test can be used to determine whether a continuous, nonlinear model has the least possible standard error for each set of input-output data by examining its variance [12, 20, 41–46]. The two-gamma test statistic, gamma value (Γ) and V-ratio, are used to select the number of input variables. The relationship between the inputs *(x)* and output *(y)* variables is determined by Eq. (2):

$$y = Gx + \Gamma \tag{2}$$

Where, G and Γ denote the gradient and intercept of the line of regression (x = 0), y describes the output. Another indicator, i.e., V-ratio (VR):

$$VR = \frac{\Gamma}{\sigma^2(y)} \tag{3}$$

Here, Γ = gamma function, and $\sigma^2(y)$ = output variance. All combinations of inputs could be tested using the Gamma test in order to discover the input combination with the lowest absolute Gamma value. When m scalar inputs are present, there are 2m - 1 possible combinations. We can produce a superior mathematical model if the gamma, standard error, and V-ratio are below zero; when the values of gamma, standard error, and V-ratio are lower, we have a higher chance of model consistency. Input pairings were selected from those that had the lowest gamma, standard error, and V-ratio values [1, 2, 11, 20, 47].

2.5 Model performance evaluation

A qualitative and quantitative evaluation of the developed models was conducted in order to check the predictive ability accuracy for the daily stream discharge of Jondhara stations.

The models were compared with predicted and observed daily runoff for the years 2000 and 2002 in order to assess their qualitative and quantitative performance. In order to this, the quantitative performance of the model, statistical and hydrological indices like Pearson correlation coefficient (CC), root mean square error (RMSE) and percent bias (PBIAS) were estimated. For this study, the acceptable threshold for correlation coefficient (CC) is 0.9 and above. A model between 0.8 and 0.9 is considered to be good, and one below 0.8 is regarded as unsatisfactory. For PBIAS, between 0 to 10 is considered to be very good, 10 to 15 is considered to be good, 15 to 25 is considered to be fair and more than 25 is regarded as unsatisfactory or inadequate. A better and excellence model is one that has the minimum value of root mean square error (RMSE) and percent bias (PBIAS) and higher values of coefficients of correlation (CC) [24, 25, 48–51].

$$CC = \frac{\sum_{i=1}^{N}\left(Q_{t_i}^{Obs} - \overline{Q_{t_i}^{Obs}}\right)\left(Q_{t_i}^{Pre} - \overline{Q_{t_i}^{Pre}}\right)}{\sqrt{\sum_{i=1}^{N}\left(Q_{t_i}^{Obs} - \overline{Q_{t_i}^{Obs}}\right)^2 \sum_{i=1}^{N}\left(Q_{t_i}^{Pre} - \overline{Q_{t_i}^{Pre}}\right)^2}} \tag{4}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Q_{t_i}^{Obs} - Q_{t_i}^{Pre}\right)^2} \tag{5}$$

$$RMSE = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Q_{t_i}^{Pre} - Q_{t_i}^{Obs}\right) \times 100}}{\sum_{i=1}^{N}\left(Q_{t_i}^{Obs}\right)} \tag{6}$$

Where, all parameters are indicated as follows: $Q_{t_i}^{Obs}$ is an observed or actual value of stream discharge, $Q_{t_i}^{Pre}$ is simulated or forecasted value of stream discharge, $\overline{S_{t_1}^{Pre}}$ and $\overline{S_{t_1}^{Obs}}$ are the mean value of observed or actual and simulated or forecasted value of stream discharge samples, respectively and N is the total number of data points.

3. Results and discussion

3.1 Best Input Selection

The selection of appropriate input parameters for daily stream discharge and sediment yield forecasts is challenging. A tool like this eliminates unnecessary input parameters that don't contribute to predicting the outcome. The most relevant inputs are selected to avoid complexity. Once the models have been developed, they can easily be interpreted and analyzed. A Gamma test was used to select suitable input variables for MLP, SVM, MLP-GA and SVM-GA models. The selection of the optimal input variables is a crucial stage in modeling for the best result of the chosen models. Various input combinations of stream discharge (m^3/sec) and rainfall (mm) with multi lag were used to determine the best input combination for present day stream discharge prediction. Various input variables were employed to effectively predict the daily stream discharge and daily sediment yield presented in Table 3 in accordance with the significant correlation between the inputs and output, which is illustrated above.

Table 3: Selection of Input variable based on Gamma Test

Model	Input variable	Jondhara	
		Gamma	SE
Q-1	Q_t	0.5520	0.0595
Q-2	Q_t	0.5562	0.0588
Q-3	Q_t	0.5990	0.0566
Q-4	Q_t	0.5122	0.0538
Q-5	Q_t	0.5892	0.0541
Q-6	Q_t	0.5204	0.0595
Q-7	Q_t	0.4552	0.0581
Q-8	Q_t	0.5505	0.0584
Q-9	Q_t	0.5704	0.0575
Q-10	Q_t	0.4534	0.0575
Q-11	Q_t	0.3679	0.0481
Q-12	Q_t	0.5565	0.0578
Q-13	Q_t	0.5359	0.0564
Q-14	Q_t	0.5494	0.0575
Q-15	Q_t	0.3588	0.0477
Q-16	Q_t	0.3364	0.0458
Q-17	Q_t	0.3475	0.0463
Q-18	Q_t	0.4883	0.0590
Q-19	Q_t	0.5416	0.0571
Q-20	Q_t	0.5416	0.0576

Table 3 lists the values of the two gamma test indicators mask, gamma value along with the standard error for each input pairings for Jondhara station for present day discharge prediction. The lower gamma test statistics show that an input combination performs better. Out of 20 feasible combinations, for Jondhara station, model numbers Q-4 (R_t, R_{t-1}, R_{t-2}, R_{t-3}) were selected best input for rainfall lags only and second-best input combinations which is combination of stream discharge and rainfall, Q-16 (Q_{t-1}, Q_{t-2}, Q_{t-3}, R_t, R_{t-1}) were selected for Jondhara station further study to keep climate change influence.

3.2 Artificial intelligence-based rainfall – runoff modeling at jondhara station

The daily runoff in a watershed which is a function of not only the daily rainfall, but also the runoff from the previous day. A watershed's characteristics, such as its size, shape, slope, soil type, type of subsoil, and others, influence the runoff of the following day as a result of the previous day's rainfall and runoff. The relationship between the input and output variables is presented in Tables 2 of the present study, respectively. In order to develop the Rainfall – runoff model, a combination of present-day rainfall, one lag day, two lag days, and three lag days of rainfall (i.e., model Q-4 (R_t, R_{t-1}, R_{t-2}, R_{t-3})) for Jondhara station was used as inputs, and the current runoff, as outputs, as result Gamma test shown in Table 3. As mentioned, daily runoff in a watershed which is a function of not only the daily rainfall, but also the runoff from the previous day thus, in order to develop another Rainfall – runoff model for climatic change, a combination of previous-day, two day and tree days lag runoff, and present-day rainfall, one lag days of rainfall, was used as inputs (i.e., Model Q-16 (Q_{t-1}, Q_{t-2}, Q_{t-3}, R_t, R_{t-1})).

The Optimal setting parameters (Architecture) of MLP, SVM, MLP-GA and SVM-GA models algorithms for tuning AI models shown in Table 4, were trained and the quantitative performance using developed AI models was evaluated based on various performance indicators such as correlation coefficient (CC), root mean square error (RMSE) and percent bias (PBIAS) as shown in Table 4.

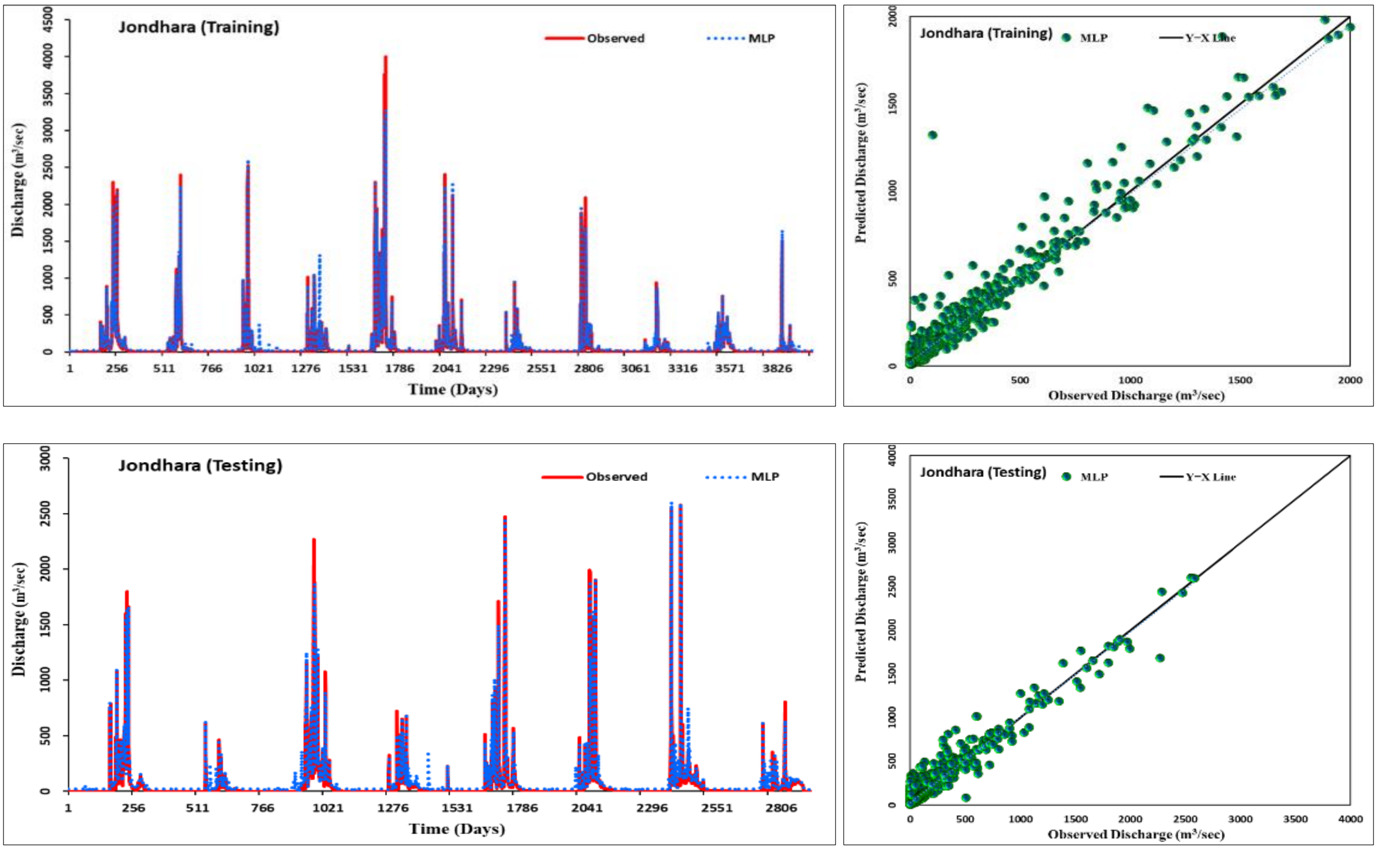
Table 4: Results of Rainfall – runoff modelling at Jondhara

Input	Algorithms	Architecture	Training			Testing		
			CC	RMSE	PBIAS	CC	RMSE	PBIAS
Q-4	MLP	4-31-1	0.9843	0.2609	-0.0065	0.8877	0.5064	-0.0156
	SVM	Cast: 17 Gamma: 0.20	0.9310	0.5093	-0.0127	0.9465	0.4230	-0.0130
	MLP-GA	4-21-13-1	0.9685	0.4161	-0.0104	0.9562	0.3765	-0.0119
	SVM-GA	Cast: 15 Gamma: 0.20	0.9751	0.3230	-0.0080	0.9575	0.3709	-0.0117
Q-16	MLP	5-27-1	0.9740	0.3416	-0.0085	0.9555	0.3893	-0.0120
	SVM	Cast: 9 Gamma: 0.20	0.9742	0.3385	-0.0084	0.9559	0.3879	-0.0119
	MLP-GA	5-37-1	0.9745	0.3323	-0.0083	0.9565	0.3851	-0.0118
	SVM-GA	Cast: 12 Gamma: 0.25	0.9743	0.3354	-0.0084	0.9733	0.3390	-0.0104

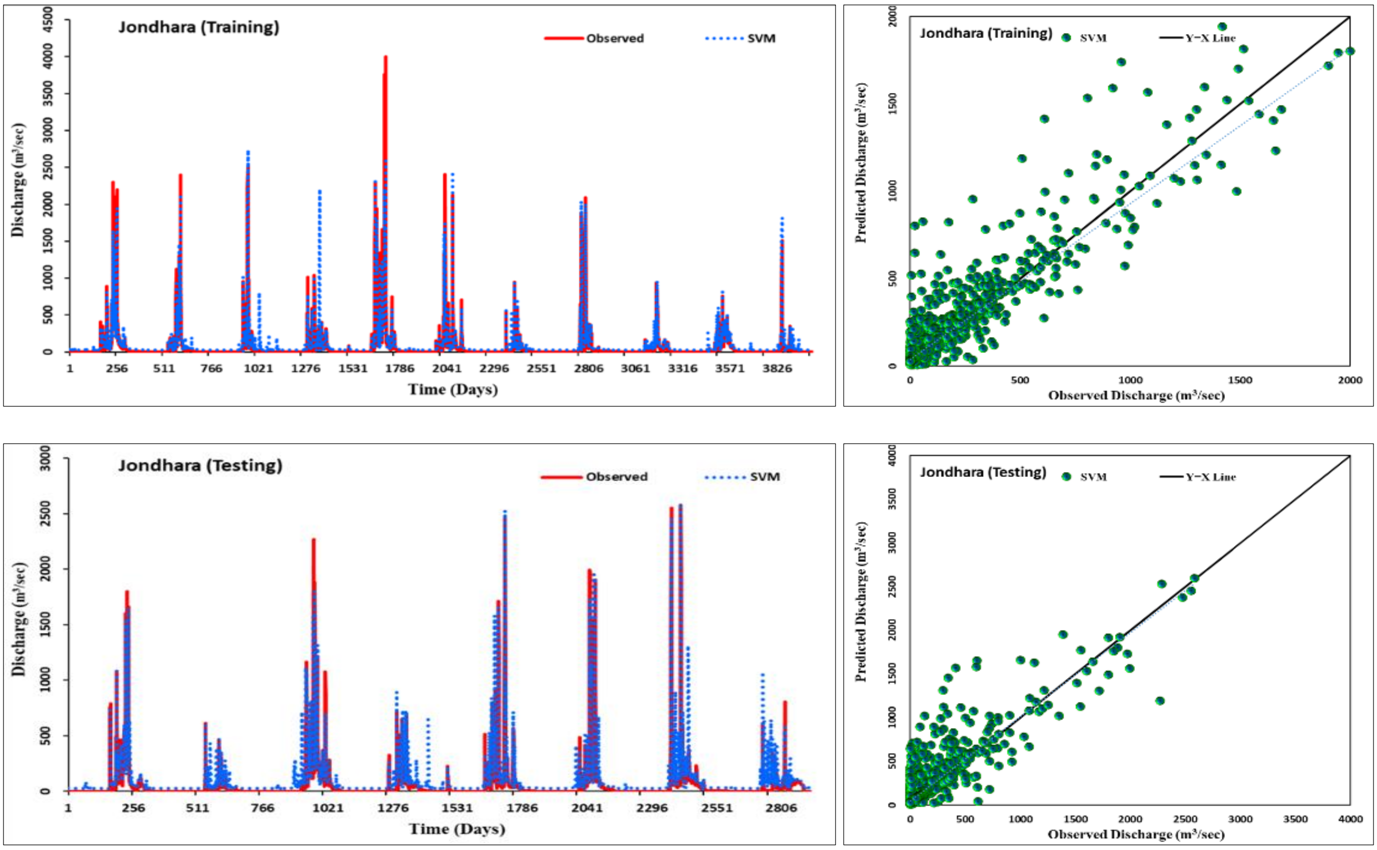
It is apparent from Table that CC, RMSE and PBIAS values of developed MLP Q-4 (Architecture 4-31-1) models were found 0.9843, 0.2609 (m^3/sec), and -0.0065 (%) during the training period, and 0.8877, 0.5064 (m^3/sec) and -0.0156 (%) during testing period, respectively. For SVM Q-5 (Architecture- Cast: 17 and Gamma: 0.20), the CC, RMSE and PBIAS values were found 0.9310, 0.5093 (m^3/sec) and -0.0127 (%) during training period and 0.9465, 0.4230 (m^3/sec) and -0.0130 (%) during testing period respectively; For MLP-GA Q-5 (Architecture- 4-21-13-1), the CC, RMSE and PBIAS values were found 0.9685, 0.4161 (m^3/sec) and -0.0104 (%) during training period and 0.9562, 0.3765 (m^3/sec) and -0.0119 (%) during testing period respectively; and For SVM-GA Q-4 (Architecture- Cast: 15 and Gamma: 0.20), the CC, RMSE and PBIAS values were found 0.9751, 0.3230 (m^3/sec) and -0.0080 (%) during training period and 0.9575, 0.3709 (m^3/sec) and -0.0117 (%) during testing period respectively. However, in terms of quantitative values of statistical metrics presented in Table 4.5, the SVM-GA was found to perform better compared to MLP, SVM and MLP-GA at Jondhara stations in most of the three statistical metrics. During the training period, the MLP model had a good performance, but during the testing period, it failed to perform as well.

The estimation of daily stream discharge using MLP, SVM, MLP-GA and SVM-GA models was done on the basis of CC, RMSE and PBIAS during training testing period for Rainfall – runoff model for climatic change at Sigma station with consideration of rainfall. The values of CC, RMSE and PBIAS during the testing period for MLP, SVM, MLP-GA and SVM-GA models are summarized in Table 4. As evident from Table 4, the SVM-GA model provided the best accuracy (CC = 0.9743, RMSE = 0.3354 (m^3/sec) and PBAIS = -0.0084 (%) for training period and CC= 0.9733, RMSE = 0.3390 (m^3/sec) and PBIAS = -0.0104 (%) for testing period) with Cast: 12 and Gamma: 0.25 architecture. The MLP-GA with 5-37-1 architecture was found to be suitable for daily stream discharge estimation with CC = 0.9745, RMSE = 0.3323 (m^3/sec) and PBAIS = -0.0083 (%) for training period and CC = 0.9565, RMSE = 0.3851 (m^3/sec) and PBIAS = -0.0118 (%) for testing period. Similarly, the optimal MLP and SVM models gave CC = 0.9740 and 0.9742, RMSE = 0.3416 and 0.3385 (m^3/sec), and PBIAS = -0.0085 and -0.0084 for training period respectively, and CC = 0.9555 and 0.9559, RMSE = 0.3893 and 0.3879 (m^3/sec), and PBIAS = -0.0120 and -0.0119 (%) for testing period, respectively with 5-27-1 and Cast: 9, Gamma: 0.20 architecture, respectively.

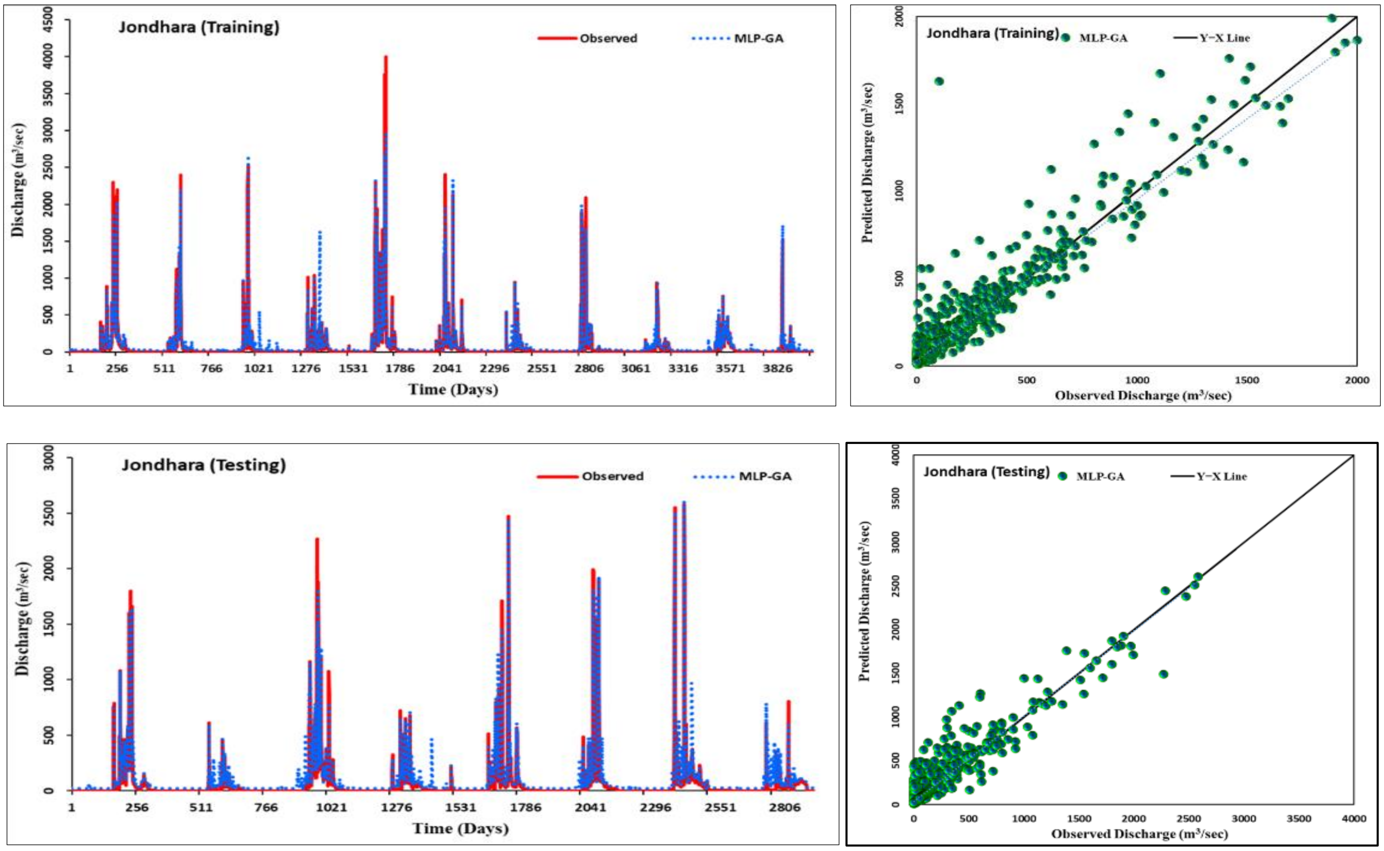
As a first step in the process of analyzing the AI-based models that were developed, a qualitative analysis was conducted using the interpretation of their graphs and scatter diagrams in order to get a better understanding of the models' characteristics. In order to verify and validate the equivalence between models and observations, actual stream discharges were compared with predicted model runoffs to obtain daily stream discharges. The daily stream discharge values during the training period and the subsequent testing period along with the scatter plots for MLP, SVM, MLP-GA and SVM-GA models are shown in Figure 2 and 3 along with their temporal variations as well as their scatter plots for Q-4 model and Fig. 3 for Q-16 model, respectively.



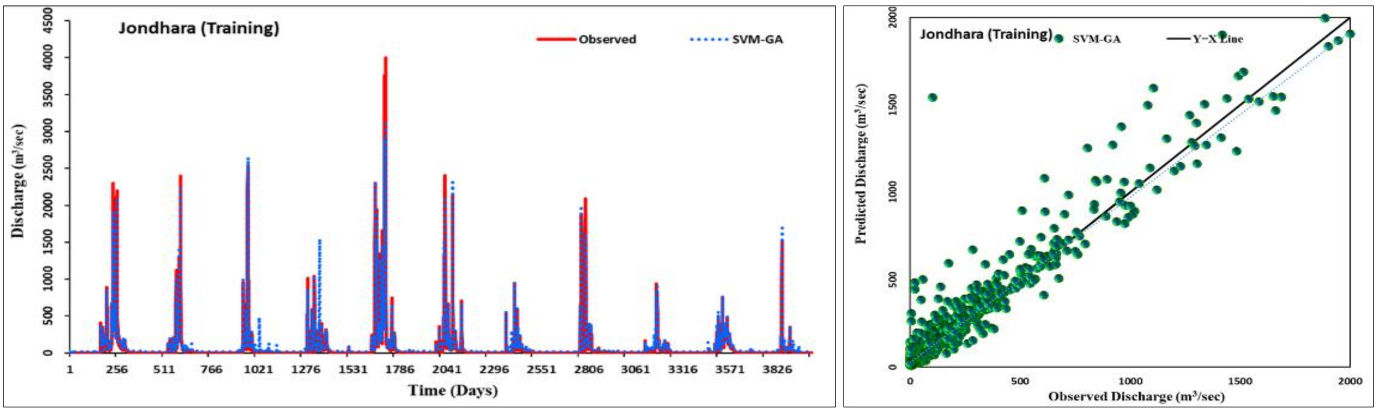
A



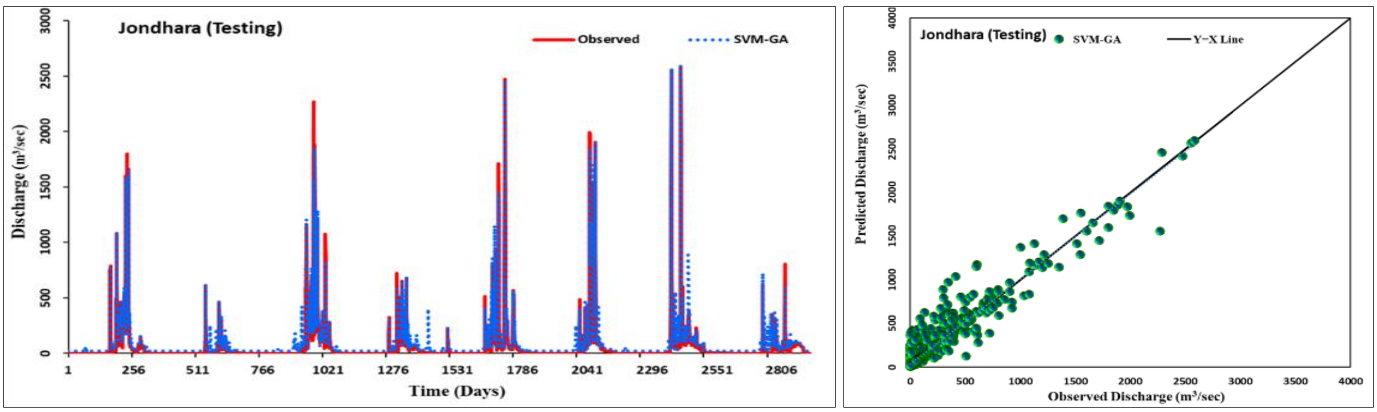
B



C

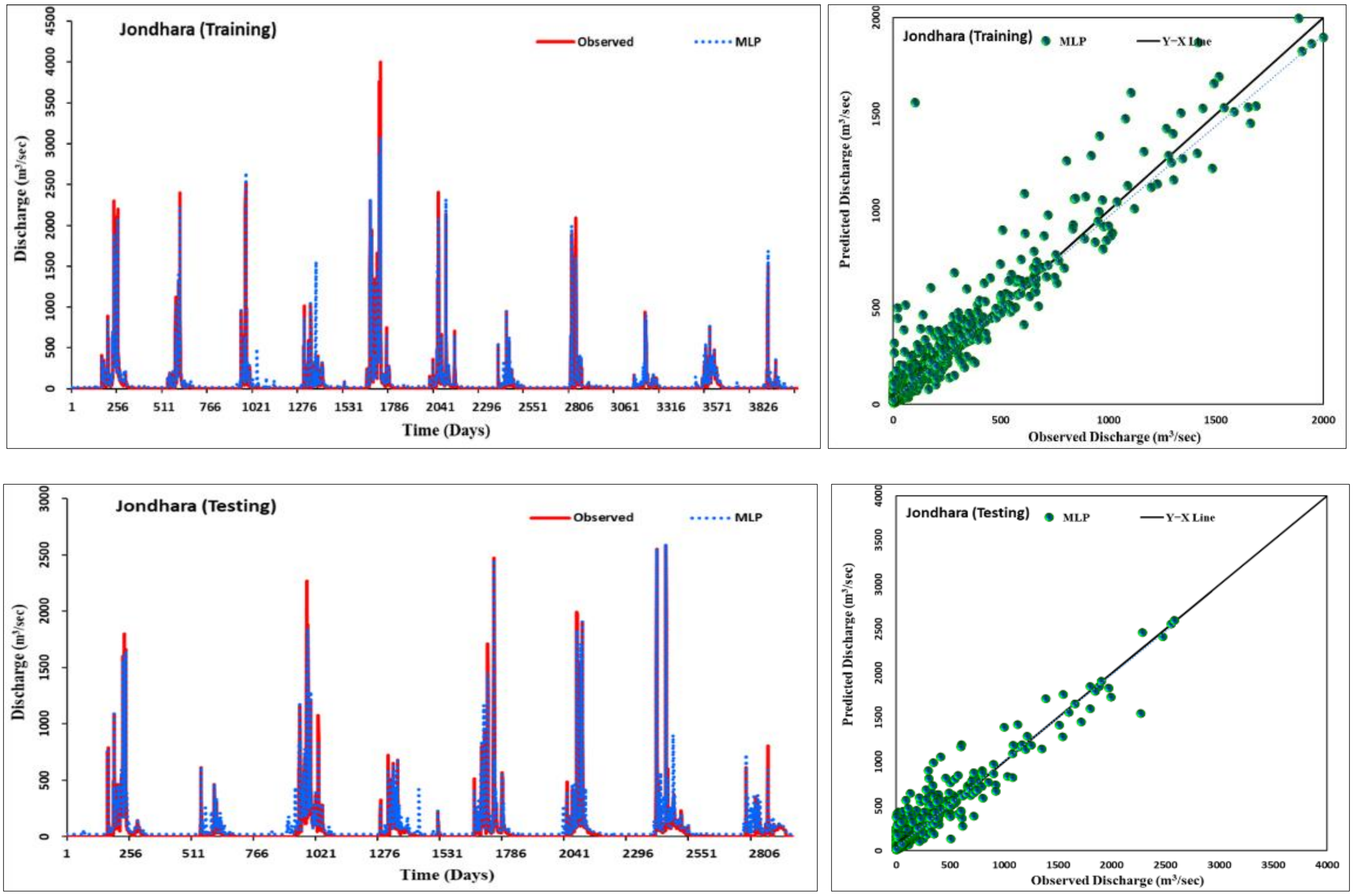


G

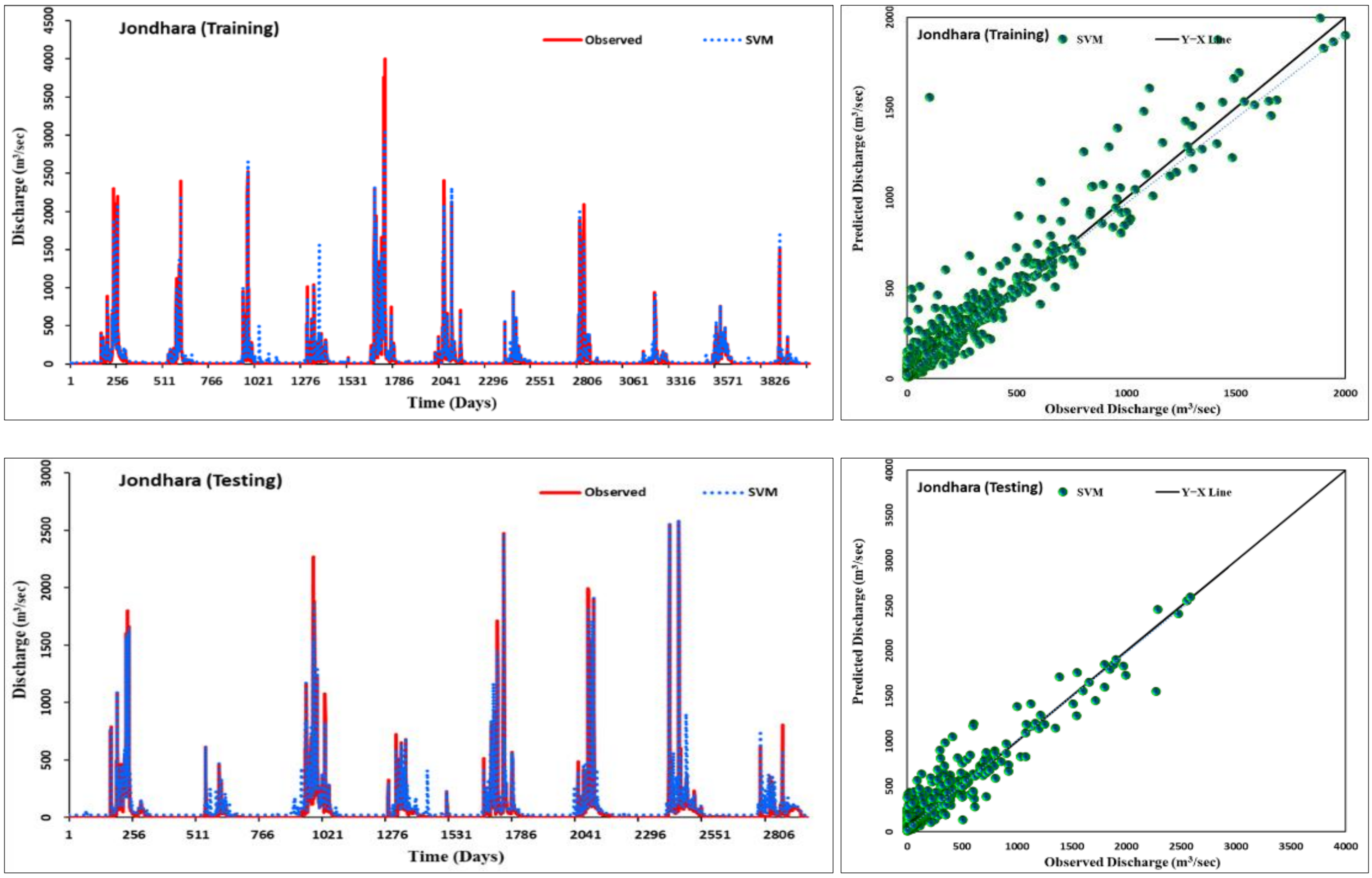


D

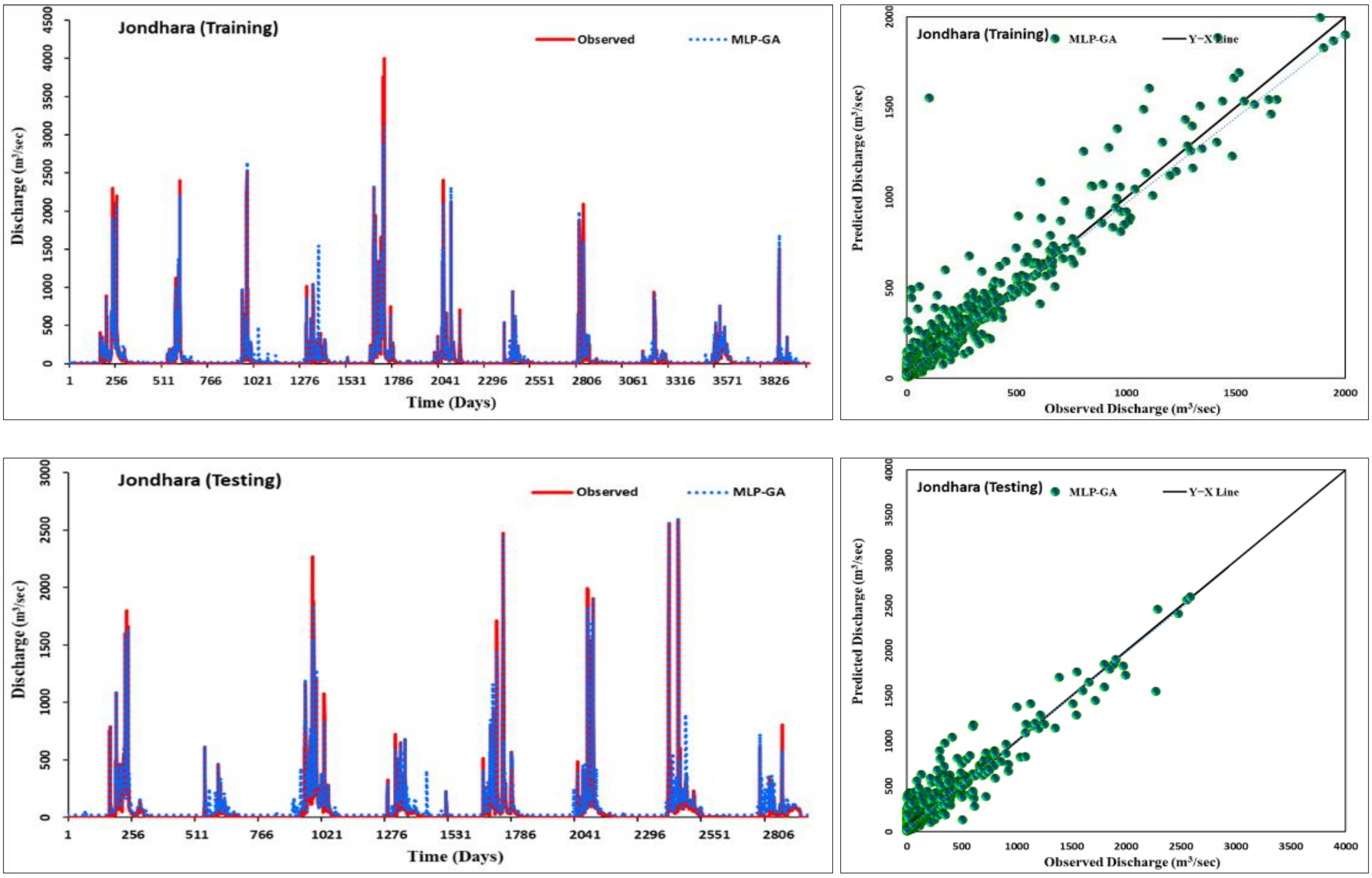
Fig 2: Line and scatter diagram of (a) MLP, (b) SVM, (c) MLP-GA and (d) SVM-GA models model (Q-4) during training and testing period


Jondhara (Training)
Observed
MLP
Discharge (m³/sec)
Time (Days)
Y-X Line
Predicted Discharge (m³/sec)
Observed Discharge (m³/sec)
Jondhara (Testing)

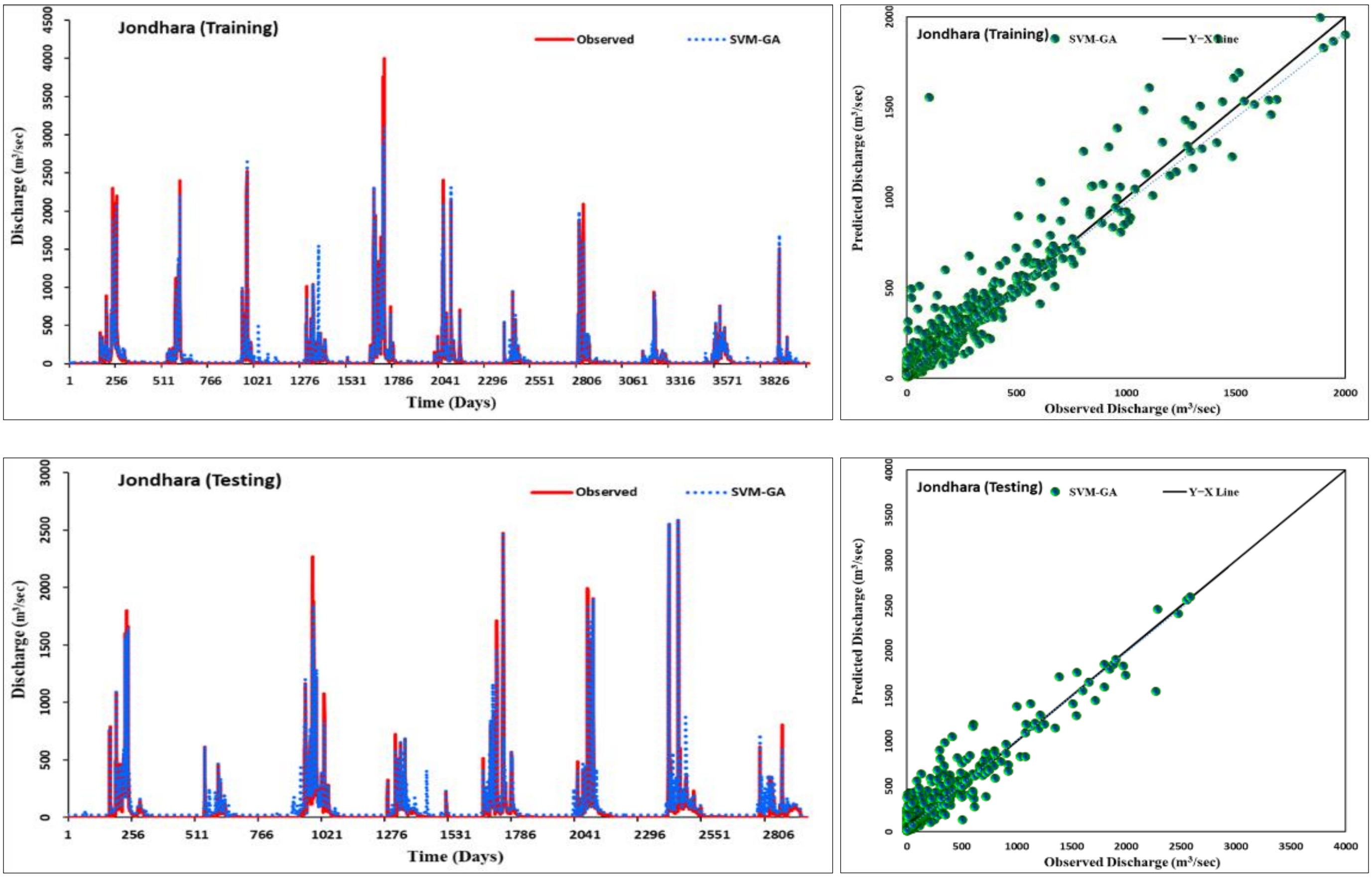
A


Jondhara (Training)
Observed
SVM
Discharge (m³/sec)
Time (Days)
Y-X Line
Predicted Discharge (m³/sec)
Observed Discharge (m³/sec)
Jondhara (Testing)

B



C



D

Fig 3: Line and scatter diagram of (a) MLP, (b) SVM, (c) MLP-GA and (d) SVM-GA models model ((Q-16) during training and testing period

Figs. 2 and 3 illustrate the estimated and observed values of stream discharge using various developed AI models for Jondhara station during the training and testing period. The graphs and scatterplots show that the developed models generally slightly underestimate daily stream discharge. Considering the particular conditions of the station Jondhara, the qualitative performance of the developed models with regards to predicting the daily runoff has been found to be satisfactory. As compared to other models, the regression line and the best fit (1:1) line are often close to one another. For example, the line of regression is nearly the same for all models, but the best fit (1:1) line is closer for the SVM-GA model, since it has fewer scattered estimates than other models. Based on the regression results from MLP, SVM, MLP-GA, and SVM-GA models for Jondhara, it seems that all the models are slightly inaccurate in predicting the daily stream discharge values at the station, which indicates that all the models under-predict the stream discharge values by a very slight degree.

4. Conclusion

The Jondhara station at Seonath River in Chhattisgarh, India was examined to assess the capabilities and application of artificial intelligence models. A hybrid AI algorithm based on GA (Genetic Algorithm), MLP-GA, and SVM-GA, was employed in this study to develop models of daily discharge based on MLP, and SVM. In this study, various type architectures of MLP, SVM, MLP-GA, and SVM-GA models were developed, train the and tested (validation) the model for the checking the accuracy the developed model. To select the best network structure among MLP, SVM, MLP-GA and SVM-GA models, a trial-and-error methodology was used because there is no specific rule for choosing the best structure. In order to evaluate the quantitative performance, correlation coefficients (CC), root mean square errors (RMSE) and percent biases (PBIAS) were estimated. Performance evaluation indices were used to select models for analysis based on their performance during training and testing. Taking this study's results into account, the following conclusions can be drawn:

1. For future forecasting of daily discharge and sediment yield, best input parameter play an important role, thus for this goal, gamma test was used to identify the best input combination.
2. Based on the performance indices, among artificial neural networks SVM-GA model (Cast: 15, Gamma: 0.20) for Q-4 model and SVM-GA model (Cast: 12, Gamma: 0.25) for Q-16 model outperformed than MLP, SVM, MLP-GA models in daily discharge prediction for Jondhara station.

5. References

1. Singh AK, Kumar P, Ali R, Al-Ansari N, Vishwakarma DK, Kushwaha KS, *et al*. An Integrated Statistical-Machine Learning Approach for Runoff Prediction. Sustainability. 2022;14(13):8209. https://doi.org/10.3390/su14138209
2. Shukla R, Kumar P, Vishwakarma DK, Ali R, Kumar R, Kuriqi A. Modeling of stage-discharge using back propagation ANN-, ANFIS-, and WANN-based computing techniques. Theor. Appl. Climatol. 2021; Aug 17:1-23. https://doi.org/10.1007/s00704-021-03863-y
3. Srinivasulu S, Jain A. A comparative analysis of training methods for artificial neural network rainfall–runoff models. Appl. Soft Comput. 2006;6(3):295-306. https://doi.org/10.1016/j.asoc.2005.02.002
4. Ditthakit P, Pinthong S, Salaeh N, Weekaew J, Thanh Tran T, Bao Pham Q. Comparative study of machine learning methods and GR2M model for monthly runoff prediction. Ain Shams Eng. J. 2023;14(4):101941. https://doi.org/10.1016/j.asej.2022.101941
5. Sayed BT, Al-Mohair HK, Alkhayyat A, Ramírez-Coronel AA, Elsahabi M. Comparing machine-learning-based black box techniques and white box models to predict rainfall-runoff in a northern area of Iraq, the Little Khabur River. Water Sci. Technol. 2023;87(3):812-822. https://doi.org/10.2166/wst.2023.014
6. Yavari F, Salehi Neyshabouri SA, Yazdi J, Molajou A, Brysiewicz A. A Novel Framework for Urban Flood damage Assessment. Water Resour. Manag. 2022;36(6):1991-2011. https://doi.org/10.1007/s11269-022-03122-3
7. Razavi T, Coulibaly P. Streamflow Prediction in Ungauged Basins: Review of Regionalization Methods. J. Hydrol. Eng. 2013;18(8):958-975. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000690
8. Adnan RM, Petroselli A, Heddam S, Santos CAG, Kisi O. Short term rainfall-runoff modelling using several machine learning methods and a conceptual event-based model. Stoch. Environ. Res. Risk Assess. 2021;35(3):597-616. https://doi.org/10.1007/s00477-020-01910-0
9. Alizadeh MJ, Kavianpour MR, Kisi O, Nourani V. A new approach for simulating and forecasting the rainfall-runoff process within the next two months. J. Hydrol. 2017;548:588-597. https://doi.org/10.1016/j.jhydrol.2017.03.032
10. Markuna S, Kumar P, Ali R, Vishwkarma DK, Kushwaha KS, Kumar R, *et al*. Application of Innovative Machine Learning Techniques for Long-Term Rainfall Prediction. Pure Appl. Geophys. 2023 Jan 18:1-29.
11. Singh VK, Panda KC, Sagar A, Al-Ansari N, Duan HF, Paramaguru PK, *et al*. Novel Genetic Algorithm (GA) based hybrid machine learning-pedotransfer Function (ML-PTF) for prediction of spatial pattern of saturated hydraulic conductivity. Eng. Appl. Comput. Fluid Mech. 2022;16(1):1082-1099. https://doi.org/10.1080/19942060.2022.2071994
12. Singh VK, Kumar D, Kashyap PS, Kisi O. Simulation of suspended sediment based on gamma test, heuristic, and regression-based techniques. Environ. Earth Sci. 2018;77:708. https://doi.org/10.1007/s12665-018-7892-6
13. Stewart D, Canfield E, Hawkins R. Curve Number Determination Methods and Uncertainty in Hydrologic Soil Groups from Semiarid Watershed Data. J. Hydrol. Eng. 2012;17(11):1180-1187. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000452
14. Heidari AA, Mirjalili S, Faris H, Aljarah I, Mafarja M, Chen H. Harris hawks optimization: Algorithm and applications. Futur. Gener. Comput. Syst. 2019;97:849-872. https://doi.org/10.1016/j.future.2019.02.028
15. Ding S, Su C, Yu J. An optimizing BP neural network algorithm based on genetic algorithm. Artif. Intell. Rev. 2011;36:153-162. https://doi.org/10.1007/s10462-011-9208-z
16. Rabbani A, Samui P, Kumari S. A novel hybrid model of augmented grey wolf optimizer and artificial neural

network for predicting shear strength of soil. Model. Earth Syst. Environ. 2022 Dec 17:1-21. https://doi.org/10.1007/s40808-022-01610-4
17. Kennedy J, Eberhart R. Particle swarm optimization. In: Proceedings of ICNN'95 - International Conference on Neural Networks. IEEE. 1995, 1942–1948.
18. Riad S, Mania J, Bouchaou L, Najjar Y. Rainfall-runoff model usingan artificial neural network approach. Math. Comput. Model. 2004;40(7-8):839-846. https://doi.org/https://doi.org/10.1016/j.mcm.2004.10.012
19. Adnan RM, Petroselli A, Heddam S, Santos CAG, Kisi O. Comparison of different methodologies for rainfall–runoff modeling: machine learning vs conceptual approach. Nat. Hazards. 2021;105:2987-3011. https://doi.org/10.1007/s11069-020-04438-2
20. Kumar A, Singh VK, Saran B, Al-Ansari N, Singh VP, Adhikari S, *et al*. Development of Novel Hybrid Models for Prediction of Drought- and Stress-Tolerance Indices in Teosinte Introgressed Maize Lines Using Artificial Intelligence Techniques. Sustainability. 2022;14(4):2287. https://doi.org/10.3390/su14042287
21. Nourani V, Molajou A, Najafi H, Danandeh Mehr A. Emotional ANN (EANN): A New Generation of Neural Networks for Hydrological Modeling in IoT. In: Al-Turjman, F. (ed.) Artificial Intelligence in IoT. Transactions on Computational Science and Computational Intelligence. Springer International Publishing, Cham, 2019, 45–61.
22. Shahmirnoori A, Saadatpour M, Rasekh A. Using mobile and fixed sensors for optimal monitoring of water distribution network under dynamic water quality simulations. Sustain. Cities Soc. 2022;82:103875. https://doi.org/10.1016/j.scs.2022.103875
23. Elbeltagi A, Raza A, Hu Y, Al-Ansari N, Kushwaha NL, Srivastava A, *et al*. Data intelligence and hybrid metaheuristic algorithms-based estimation of reference evapotranspiration. Appl. Water Sci. 2022;12(7):152. https://doi.org/10.1007/s13201-022-01667-7
24. Vishwakarma DK, Kumar R, Kumar A, Kushwaha NL, Kushwaha KS, Elbeltagi A. Evaluation and development of empirical models for wetted soil fronts under drip irrigation in high-density apple crop from a point source. Irrig. Sci. 2022 Sep 25:1-24. https://doi.org/10.1007/s00271-022-00826-7
25. Kushwaha NL, Rajput J, Elbeltagi A, Elnaggar AY, Sena DR, Vishwakarma DK, *et al*. Data Intelligence Model and Meta-Heuristic Algorithms-Based Pan Evaporation Modelling in Two Different Agro-Climatic Zones: A Case Study from Northern India. Atmosphere (Basel). 2021;12(12):1654. https://doi.org/10.3390/atmos12121654
26. Elbeltagi A, Kumar M, Kushwaha NL, Pande CB, Ditthakit P, Vishwakarma DK, *et al*. Drought indicator analysis and forecasting using data driven models: case study in Jaisalmer, India. Stoch. Environ. Res. Risk Assess. 2022-2023 Jan;37(1):113-31.. https://doi.org/10.1007/s00477-022-02277-0
27. Vishwakarma DK, Ali R, Bhat SA, Elbeltagi A, Kushwaha NL, Kumar R, *et al*. Pre- and post-dam river water temperature alteration prediction using advanced machine learning models. Environ. Sci. Pollut. Res. 2022 Nov;29(55):83321-46.. https://doi.org/10.1007/s11356-022-21596-x
28. Achite M, Elshaboury N, Jehanzaib M, Vishwakarma DK, Pham QB, Anh DT, *et al*. Performance of Machine Learning Techniques for Meteorological Drought Forecasting in the Wadi Mina Basin, Algeria. Water. 2023;15(4):765. https://doi.org/10.3390/w15040765
29. Kawamoto AH, McClelland JL, Rumelhart DE. Explorations in Parallel Distributed Processing: A Handbook of Models, Programs, and Exercises. Am. J. Psychol. 1989;102:435. https://doi.org/10.2307/1423065
30. Deo RC, Samui P, Kim D. Estimation of monthly evaporative loss using relevance vector machine, extreme learning machine and multivariate adaptive regression spline models. Stoch. Environ. Res. Risk Assess. 2016;30:1769-1784. https://doi.org/10.1007/s00477-015-1153-y
31. Lima AR, Cannon AJ, Hsieh WW. Forecasting daily streamflow using online sequential extreme learning machines. J. Hydrol. 2016;537:431-443. https://doi.org/10.1016/j.jhydrol.2016.03.017
32. Mohanty S, Jha MK, Raul SK, Panda RK, Sudheer KP. Using Artificial Neural Network Approach for Simultaneous Forecasting of Weekly Groundwater Levels at Multiple Sites. Water Resour. Manag. 2015;29:5521-5532. https://doi.org/10.1007/s11269-015-1132-6
33. Günay ME. Forecasting annual gross electricity demand by artificial neural networks using predicted values of socio-economic indicators and climatic conditions: Case of Turkey. Energy Policy. 2016;90:92-101. https://doi.org/10.1016/j.enpol.2015.12.019
34. Kawamoto AH. Application of artificial neural network for predicting hourly indoor air temperature and relative humidity in modern building in humid region. Am. J. Psychol. 2023;102:435-438. https://doi.org/10.2307/1423065
35. Ghorbani MA, Khatibi R, Hosseini B, Bilgili M. Relative importance of parameters affecting wind speed prediction using artificial neural networks. Theor. Appl. Climatol. 2013;114:107-114. https://doi.org/10.1007/s00704-012-0821-9
36. Vapnik V. Statistical learning theory. John Wiley & Sons, Inc. Oxford, UK; c1998.
37. Vapnik V. The nature of statistical learning theory. Springer science & business media; c1999.
38. Cortes C, Vapnik V. Support-vector networks. Mach. Learn. 1995;20:273-297. https://doi.org/10.1007/BF00994018
39. Ma Y, Liu H, Zhai G, Huo Z. Financial Risk Early Warning Based on Wireless Network Communication and the Optimal Fuzzy SVM Artificial Intelligence Model. Wirel. Commun. Mob. Comput, 2021, 1-20. https://doi.org/10.1155/2021/7819011
40. Stefánsson A, Končar N, Jones AJ. A note on the Gamma test. Neural Comput. Appl. 1997;5(3):131-133. https://doi.org/10.1007/BF01413858
41. Seifi A, Riahi H. Estimating daily reference evapotranspiration using hybrid gamma test-least square support vector machine, gamma test-ANN, and gamma test-ANFIS models in an arid area of Iran. J. Water Clim. Chang. 2020;11(1):217-240. https://doi.org/10.2166/wcc.2018.003
42. Singh VK, Singh BP, Kumar A, Vivekanand. A Comparative Study of Artificial Intelligence and

Conventional Techniques for Rainfall-Runoff Modeling. Int. J Agric. Eng. 2017;10(2):441–449. https://doi.org/DOI: 10.15740/HAS/IJAE/10.2/441-449
43. Malik A, Kumar A, Kim S, Kashani MH, Karimi V, Sharafati A, *et al* Modeling monthly pan evaporation process over the Indian central Himalayas: application of multiple learning artificial intelligence model. Eng. Appl. Comput. Fluid Mech. 2020;14(1):323-338. https://doi.org/10.1080/19942060.2020.1715845
44. Kumar M, Kumar P, Kumar A, Elbeltagi A, Kuriqi A. Modeling stage–discharge–sediment using support vector machine and artificial neural network coupled with wavelet transform. Appl. Water Sci. 2022;12(5):87. https://doi.org/10.1007/s13201-022-01621-7
45. Bajirao TS, Kumar P, Kumar M, Elbeltagi A, Kuriqi A. Potential of hybrid wavelet-coupled data-driven-based algorithms for daily runoff prediction in complex river basins. Theor. Appl. Climatol. 2021;145(3-4):1207–1231. https://doi.org/10.1007/s00704-021-03681-2
46. Kuriqi A, Ali R, Pham QB, Montenegro Gambini J, Gupta V, Malik A, *et al*. Seasonality shift and streamflow flow variability trends in central India. Acta Geophys. 2020;68:1461-1475. https://doi.org/10.1007/s11600-020-00475-4
47. Piri J, Amin S, Moghaddamnia A, Keshavarz A, Han D, Remesan R. Daily Pan Evaporation Modeling in a Hot and Dry Climate. J. Hydrol. Eng. 2009;14(8):803-811. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000056
48. Vishwakarma DK, Kumar R, Pandey K, Singh V, Kushwaha KS. Modeling of Rainfall and Ground Water Fluctuation of Gonda District Uttar Pradesh, India. Int. J. Curr. Microbiol. Appl. Sci. 2018;7(5):2613-2618. https://doi.org/10.20546/ijcmas.2018.705.302
49. Vishwakarma DK, Pandey K, Kaur A, Kushwaha NL, Kumar R, Ali R, Elbeltagi A, *et al*. Methods to estimate evapotranspiration in humid and subtropical climate conditions. Agric. Water Manag. 2022;261:107378. https://doi.org/10.1016/j.agwat.2021.107378
50. Sagar A, Hasan M, Singh DK, Al-Ansari N, Chakraborty D, Singh MC, *et al*. Development of Smart Weighing Lysimeter for Measuring Evapotranspiration and Developing Crop Coefficient for Greenhouse Chrysanthemum. Sensors. 2022;22(16):6239. https://doi.org/10.3390/s22166239
51. Kumar R, Manzoor S, Vishwakarma DK, Al-Ansari N, Kushwaha NL, Elbeltagi A, *et al*. Assessment of Climate Change Impact on Snowmelt Runoff in Himalayan Region. Sustainability. 2022;14(3):1-23. https://doi.org/10.3390/su14031150