



# Utilizing deep learning machine for inflow forecasting in two different environment regions: a case study of a tropical and semi-arid region

Saad Mawlood Saab<sup>1</sup> · Faridah Othman<sup>1</sup> · Chee Ghuan Tan<sup>1</sup> · Mohammed Falah Allawi<sup>2</sup> · Mohsen Sherif<sup>3</sup> · Ahmed El-Shafie<sup>1,3</sup>

Received: 2 July 2022 / Accepted: 17 October 2022 / Published online: 2 November 2022  
© The Author(s) 2022

## Abstract

Reservoir inflow ( $Q_{\text{flow}}$ ) forecasting is one of the crucial processes in achieving the best water resources management in a particular catchment area. Although physical models have taken place in solving this problem, those models showed a noticeable limitation due to their requirements for huge efforts, hydrology and climate data, and time-consuming learning process. Hence, the recent alternative technology is the development of the machine learning models and deep learning neural network (DLNN) is the recent promising methodology explored in the field of water resources. The current research was adopted to forecast  $Q_{\text{flow}}$  at two different catchment areas characterized with different type of inflow stochasticity, (semi-arid and tropical). Validation against two classical algorithms of neural network including multilayer perceptron neural network (MLPNN) and radial basis function neural network (RBFNN) was elaborated and discussed. The research was further investigated the potential of the feature selection algorithm “genetic algorithm (GA)”, for identifying the appropriate predictors. The research finding confirmed the feasibility of the developed DLNN model for the investigated two case studies. In addition, the DLNN model confirmed its capability in solving daily scale  $Q$  more accurately in comparison with the monthly scale. The applied GA as feature selection algorithm was reduced the dimension and complexity of the learning process of the applied predictive model. Further, the research finding approved the adequacy of the data span used in the current investigation development of computerized ML algorithm.

**Keywords** Reservoir inflow · Deep learning · Lead time · Tropical region · Semi-arid region

✉ Ahmed El-Shafie  
elshafie@um.edu.my

Saad Mawlood Saab  
s2003992@siswa.um.edu.my

Faridah Othman  
faridahothman@um.edu.my

Chee Ghuan Tan  
tancg@um.edu.my

Mohammed Falah Allawi  
mohammed.falah@uoanbar.edu.iq

Mohsen Sherif  
MSherif@uaeu.ac.ae

<sup>1</sup> Civil Engineering Department, Faculty of Engineering, Universiti Malaya, 50603 Kuala Lumpur, Malaysia

<sup>2</sup> Dams and Water Resources Engineering Department, College of Engineering, University of Anbar, Ramadi 31001, Iraq

<sup>3</sup> National Water and Energy Center, United Arab Emirate University, P.O. Box, 15551 Al Ain, United Arab Emirates

## Introduction

One of the useful and most direct ways of guiding reservoir operation and management is reservoir inflow ( $Q_{\text{flow}}$ ) prediction; it is also useful for flood control, reservoir operation, drought management, irrigation water management, and reservoir operation (Rezaeianzadeh et al. 2016; Xu et al. 2021). Using the forecasted  $Q_{\text{flow}}$  as an input information, the delicacy management of water resources at a reservoir is strongly reliant on precise  $Q_{\text{flow}}$  predictions (Herbert et al. 2021). In most parts of the world, accurate and real-time daily or monthly prediction of  $Q_{\text{flow}}$  remains a difficult challenge due to the nonlinearity and non-stationarity of the associated real hydrological data (Kim et al. 2019; Lee et al. 2020). Hence, this research topic has received much attention by the water engineers and decision makers.

Reservoir inflow prediction has become a major topic in hydrologic time series over the last few decades (Esmaeilzadeh et al. 2017; Bashir et al. 2019; Allawi et al. 2019a).

However, the paucity of information about physical concepts while studying the relationships between variables has necessitated the use of data-driven models in hydrological forecasting as an alternative to knowledge-driven methods. Two types of hydrological models are currently available to find solutions to nonlinear complex problems; these are the standard statistical methods and the physical-based methods (Tran et al. 2021). The standard statistical methods can not accurately capture nonlinear patterns while the physical process-based methods cannot sufficiently capture information to characterize basins based on hydrologic parameters (Yaseen et al. 2016b). While both methods depend on historical data to forecast the future, the underlying complexity of hydrologic input–output interactions necessitates a model that is strong enough to discern nonlinear patterns without sacrificing accuracy (Petty and Dhingra 2018). The new adopted computer-based approaches, such as "machine learning (ML) algorithms," have shown capacity in capturing complex nonlinear data patterns that would have required extrapolation (Nearing et al. 2021; Zounemat-Kermani et al. 2021); hence, they are considered an alternative method to the existing prediction methods in several fields of hydrology such as river flow forecasting (Yaseen et al. 2016a; Osman et al. 2020; Afan et al. 2020), rainfall forecasting (Tao et al. 2018; Ali et al. 2020), reservoir operation (Hossain and El-shafie 2014; Ehteram et al. 2018), evaporation process simulation (Allawi et al. 2019b; Salih et al. 2019), surface water quality prediction (Yaseen et al. 2018; Yahya et al. 2019), geo-science related problems (Mukhlisin et al. 2012; Alizamir et al. 2020), drought detection (Alamgir et al. 2020; Singh et al. 2021), and several others (Raghavendra and Deka 2014a; Zounemat-Kermani et al. 2021).

In both direct and multi-step scenarios, ML algorithms were used to predict reservoir inflow, resulting in more reliable predictions of extreme inflows. It was first published in Coulibaly et al. (2000), where the authors used a feed-forward neural network (FFNN) that was trained by using an early stopping approach for the prediction of real-time inflow with lead periods ranging from one to seven days. An enhanced version of the ML model for the prediction of daily inflow which employs a robust weighted-average ensemble that combines 3 different frameworks (a physical model, nearest neighbors, and an artificial neural network (ANN) was reported by Coulibaly et al. (2005). The use of least squares support vector machine (Bai et al. 2015), ensemble ML models (Ahmed et al. 2015), fuzzy logic models (El-Shafie et al. 2007), classical artificial neural network-based radial basis function (El-Shafie et al. 2009), support relevance vector model (Liu et al. 2016), random forest (Liu et al. 2017), were noticed over the literature to predict inflow. Readers are encouraged to go more into the associated literature for reservoir inflow prediction using ML models by

looking at the survey studies reported by Choong and El-Shafie (2015), Mosavi et al. (2018) and Wee et al. (2021).

A recurrent neural network (RNN) model was utilized to predict daily flow data by Apaydin et al. (2020). To check the predictive model, the performance of the RNN model is compared with the ANN model. Based on several indicators, the research concluded that the RNN model is more accurate than the ANN in predicting reservoir flow records.

The effectiveness of well-known computation methods, including MLP, ANN, and SVM, to predict inflow data was examined by Lee et al. (2020). The coefficient S, NSE, and other indices were used to assess these forecasting techniques. The study demonstrated that the models that were created could be a useful tool to predict reservoir inflow records.

The possible use of an ANN model in predicting reservoir inflow data was examined by Hadiyan et al. (2020). The research gave helpful data to simulate the inflow for the reservoir Sefidroud, Iran. Allawi et al. (2017) employed the Coactive Neuro-Fuzzy Inference System (CANFIS) method to forecast reservoir inflow. The proposed model succeeded in providing high accuracy prediction results.

The use of ML algorithms to predict reservoir inflow is inconsistent, making it difficult to determine which strategy is preferable. In addition, the artificial intelligence models, such as ANN models, have several drawbacks, such as generalizing performance and learning divergence shortfalls, local minimum entrapment, and over-fitting issues (Ghimire et al. 2018). While the support vector machine model appears to overcome some of the shortcomings of ANN, it does so at the expense of a long simulation time due to the kernel function (penalty factor and kernel width) (Raghavendra and Deka 2014a, b). As a result, ML algorithms may not efficiently learn all the conditions if there is high data complexity. In the field of hydrology, the search for new and more reliable ML algorithms is still underway. New deep learning-based ML models have been recently developed for inflow simulation. These models, such as Deep Learning Neural Network [DNNN], Long Short-Term Memory [LSTM], and Convolutional Neural Network [CNN] have been created and are frequently employed in the prediction of hydrological time-series. These DL models have advantages such as the ability to handle highly stochastic data and the ability to extract the internal physical mechanism (Hrnjica and Mehr 2020).

Although there have been several researches conducted on the inflow forecasting over the literature (Bai et al. 2016, 2018; Aljanabi et al. 2017; Allawi et al. 2018; Herbert et al. 2021), limitations are still existed and motivated the hydrological scientists to further study this essential problem. For instance, the robustness of the ML model, classical models such as ANN, SVM, ANFIS, etc., has demonstrated limitation in the learning process of the network and thus

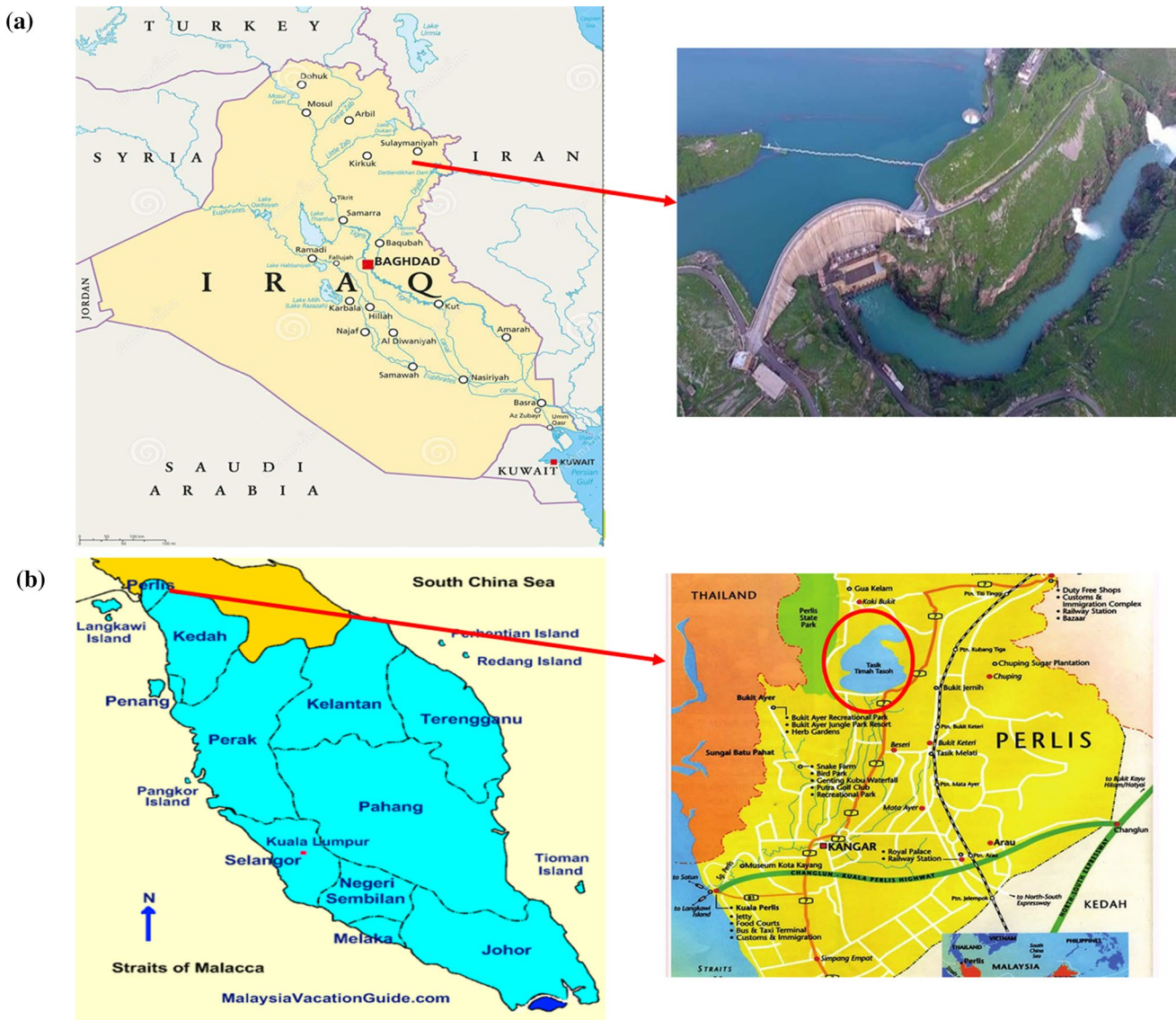


Fig. 1 Case Studies maps a Dukan Dam and b Timah Tasoh Dam

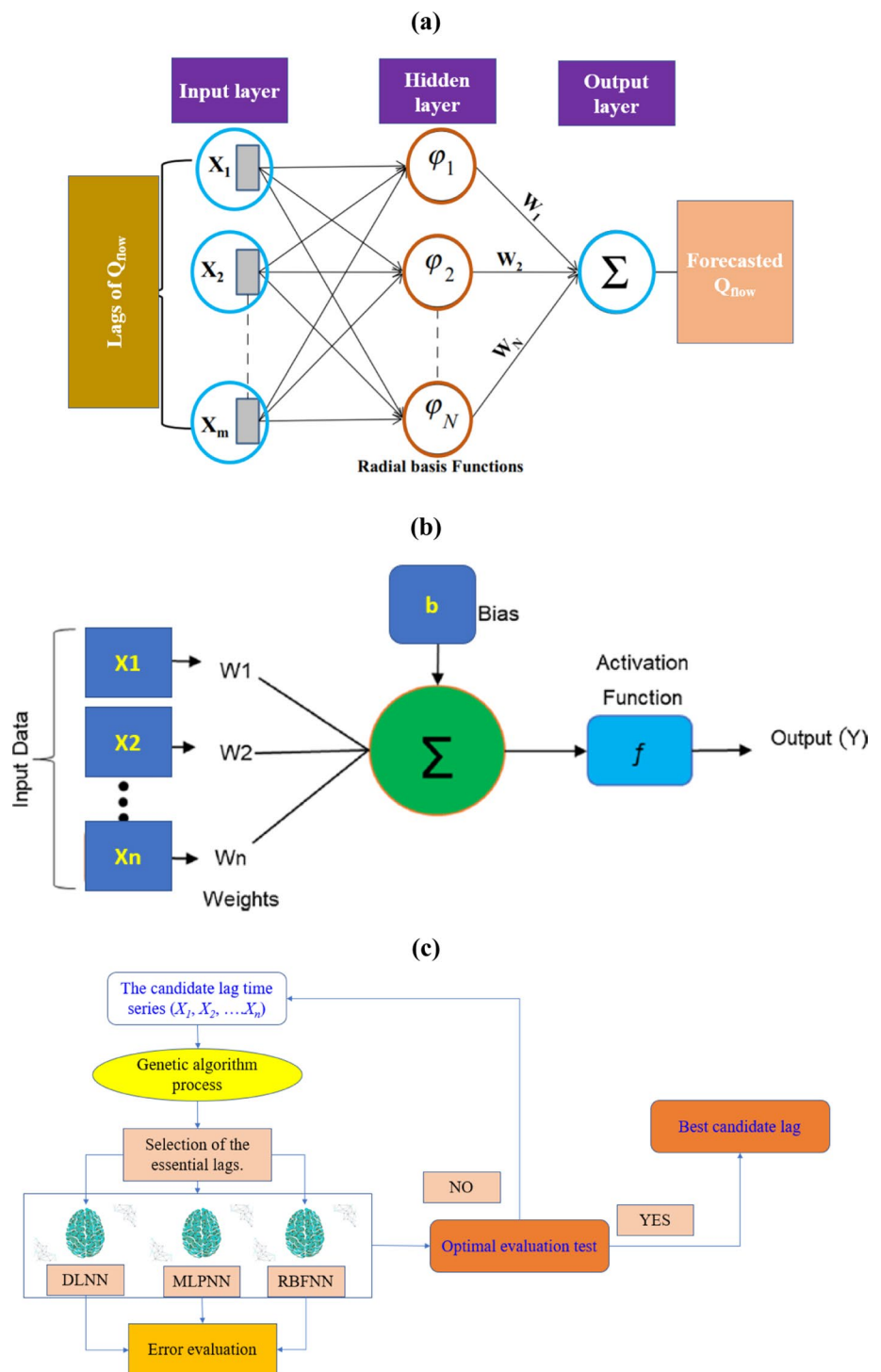
exploring new version of ML such as deep learning can contribute to overcome the classical ML models. The recognition of the appropriate features in order to construct the learning process of the ML model has been observed to be serious element in the computer aid models development and thus the reliable nature inspired optimization called genetic algorithm has been integrated as feature selection for the proper lead time reservoir inflow forecasting. As a matter of fact, the stochastic variation varies from one dam to another, thus, implementation of the methodology based on two different flow mechanisms can be tested where the generalization method for this hydrology can be examined.

### Case studies

#### Semi-arid region

The first case study used in the current research is Dukan reservoir. It is located around 67 km north of Sulaimani City in northern Iraq. The dam is adjacent to the city of Ranya and is located at 35°57'13.24" N and 44°57'11.61" E. It has a total capacity of 6.8 km<sup>3</sup> and is situated near Latitude 35°57'13.24" N and Longitude 44°57'11.61" E. It is a reservoir that was created during Dukan Dam construction on the small Zab River. This multipurpose dam

**Fig. 2** **a** The structure of the RBFNN model, **b** the structure of the MLPNN model, **c** the GA mechanism procedure



was constructed between 1954 and 1959 to provide water to farmers and to supply hydro plants for power generation. The dam is a concrete arch dam with gravity monoliths abutting it. It measures 360 m (1180 ft) in length and 116.5 m (382 ft) in height. It measures 32.5 m (107 ft) wide at the bottom and 6.2 m (20 ft) wide at the top. The dam's total maximum discharge is around 4300 m<sup>3</sup>/s (150,000 ft<sup>3</sup>/s). This

is partitioned between a spillway tunnel with 3 radial gates and an emergency bell mouth glory hole spillway that can discharge 2440 m<sup>3</sup>/s (86,000 ft<sup>3</sup>) and 1860 m<sup>3</sup>/s (66,000 ft<sup>3</sup>) per second, respectively. There are also 2 irrigation outlets that can co-discharge 220 m<sup>3</sup>/s (7,800 ft<sup>3</sup>/s) per second, but they haven't been used in 10 years. There is a powerhouse of 5 Francis units, each with an output of 80 MW, emitting

**Table 1** The statistical indicators whiling testing phase for three methods “Semi-arid case study”

Models	RMSE	MAE	MBE	NSE	SI	BIAS	d	CI
MLPNN1	81.032	46.983	0.109	0.813	0.598	− 5.570	0.896	0.728
MLPNN2	53.188	36.794	0.164	0.919	0.393	17.141	0.948	0.871
MLPNN3	66.727	45.720	0.245	0.873	0.494	17.276	0.927	0.809
MLPNN4	71.482	47.623	0.229	0.854	0.529	14.492	0.919	0.785
MLPNN5	68.598	45.949	0.226	0.866	0.508	20.233	0.924	0.800
RBFNN1	72.505	40.174	0.014	0.850	0.535	24.549	0.916	0.778
RBFNN2	52.193	33.347	− 0.061	0.922	0.386	21.283	0.950	0.876
RBFNN3	71.537	47.540	0.148	0.854	0.530	30.092	0.921	0.787
RBFNN4	59.987	41.991	0.199	0.897	0.444	16.716	0.938	0.842
RBFNN5	65.299	44.097	0.230	0.878	0.484	16.682	0.930	0.816
DLNN1	48.763	27.322	0.052	0.932	0.360	9.766	0.954	0.889
DLNN2	39.627	23.678	0.125	0.955	0.293	2.578	0.967	0.923
DLNN3	45.795	27.632	0.169	0.940	0.339	3.010	0.958	0.901
DLNN4	53.427	34.548	0.140	0.918	0.396	9.788	0.948	0.871
DLNN5	57.845	36.355	0.155	0.904	0.428	15.894	0.941	0.851

between 110 and 550 m<sup>3</sup>/s (3900 and 19,000 ft<sup>3</sup>/s) of water. The lake has a surface size of 270 km<sup>2</sup>. The reservoir's capacity is 6.8 km<sup>3</sup> in normal operation, with a maximum capacity of 8.3 km<sup>3</sup>. The surface elevation is 515 m above sea level. The surface elevation of the dam must be within 469 and 511 m to operate the power station. The Dukan Dam's drainage basin spans 11,700 km<sup>2</sup>, with part of it in Iraq and the rest in Iran. The main source of water is the Zab River. The daily inflow to the reservoir over 11 years (January 2010–December 2020) is the only available data record. A Google map of this reservoir is shown in Fig. 1a.

### Tropical region

Timah Tasoh Dam (TTD) construction began in 1987 and was finished in 1992 in Perlis, Malaysia (6°36' N; 100°14' E). TTD is an essential hydraulic construction within Peninsular Malaysia and its  $Q_{\text{flow}}$  patterns operation and quantification is highly important for the water resources management of that region. In fact, the high variance and nonlinearity seen in the  $Q_{\text{flow}}$  of the tropical zone frequently include a high stochastic pattern that contributes to the complexity of the dam's reservoir systems. This case study will necessitate the development of a new method for evaluating the offered models. As a result, a thorough comparison of the existing and proposed operating procedures is required. The reservoir system has a total surface area of over 13.3 km<sup>2</sup>. The reservoir's overall capacity is around 40 million cubic meters (MCM). With an entry average runoff of over 100 MCM, the reservoir water storage has two major zones: a dead zone of 6.7 MCM and a live zone of 33.3 MCM. The reservoir could be classified as a shallow reservoir, with a maximum depth of 10 m. The reservoir's position was chosen to receive water from two main rivers in Perlis State:

The Tasoh and Perlarit Rivers. The TTD provides irrigation water for 3100 ha at a rate of roughly 55 MCM per year. Furthermore, it delivers around 55\*10<sup>3</sup> m<sup>3</sup> of water each day for home consumption. Dams are built to regulate and avoid floods that are expected during the rainy season. The location of the Timah Tasoh Dam is displayed in Fig. 1b.

## Applied machine learning

### Deep learning neural network

Deep learning (DL) has emerged as a new branch of ANN research that is altering different scientific disciplines in the modern day (Goodfellow et al. 2016). The term “deep” in this method refers to a connection of layers that allows the translation of data representation from one to another. A deep net (DN) is a type of ANN that has numerous hidden layers, an input layer, and an output layer (Lecun et al. 2015). In comparison with traditional machine learning methods, a DL-based model necessitates a huge amount of training data in order to comprehend the underlying data patterns increases in the network depth (i.e. number of layers) allowing the extraction of the most appropriate data hierarchical representations using a proper data transformation (Schmidhuber 2015). In recent years, DL has found use in remote sensing, hydrological prediction, and image processing. Although DL has different versions adopted over the literature, in the current study, the long short-term memory (LSTM) is conducted for the reservoir inflow forecasting. The design of the LSTM model is having feedback connection with the learning layers that support the concept of complete input sequences. The LSTM model is established to fit the pattern of the inflow based on lag times

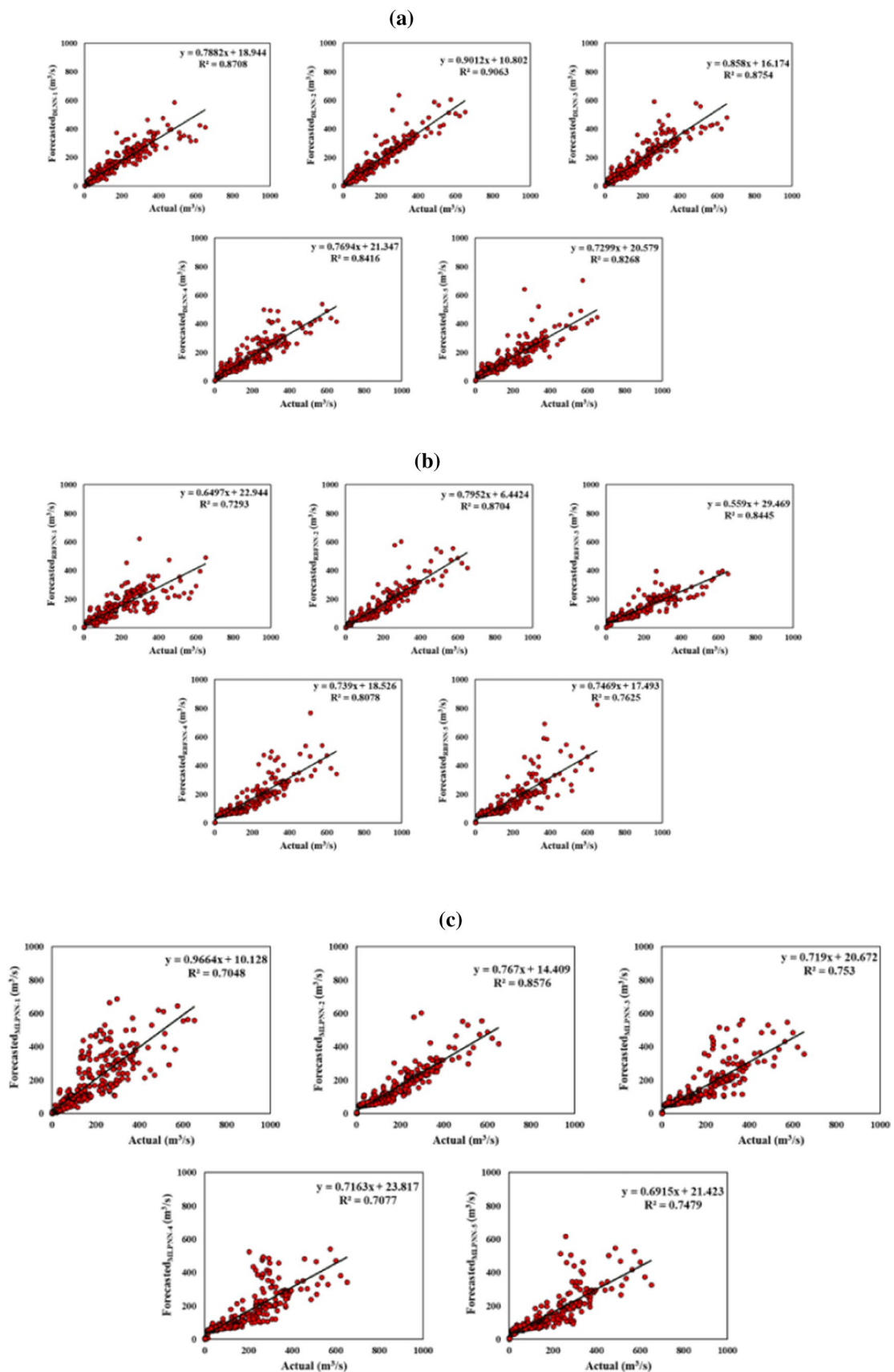


Fig. 3 Scatter Plots between actual and predicted  $Q_{flow}$  using standalone ML models and for the five input combinations “Semi-arid case study”

**Table 2** The statistical indicators whiling testing phase for three methods “Tropical case study”

Models	RMSE	MAE	MBE	NSE	SI	BIAS	d	CI
MLPNN1	6.062	4.271	0.230	0.844	0.584	0.342	0.916	0.773
MLPNN2	7.895	5.144	0.521	0.734	0.767	0.119	0.876	0.643
MLPNN3	5.661	3.676	0.048	0.859	0.567	1.567	0.922	0.792
MLPNN4	7.346	5.363	0.790	0.738	0.769	− 1.466	0.865	0.639
MLPNN5	7.020	4.711	0.433	0.760	0.741	0.687	0.883	0.671
RBFNN1	7.215	5.039	0.375	0.779	0.695	0.976	0.891	0.695
RBFNN2	5.160	3.084	0.093	0.887	0.501	0.826	0.934	0.828
RBFNN3	6.306	4.340	0.421	0.824	0.632	− 0.433	0.903	0.744
RBFNN4	6.544	4.758	0.667	0.792	0.685	− 1.004	0.890	0.705
RBFNN5	7.196	3.398	0.198	0.747	0.760	0.068	0.878	0.656
DLNN1	6.930	5.069	0.366	0.796	0.667	1.236	0.899	0.716
DLNN2	4.699	2.899	0.007	0.906	0.456	1.528	0.944	0.855
DLNN3	4.833	2.867	0.258	0.897	0.484	− 0.020	0.939	0.842
DLNN4	5.381	3.786	0.423	0.859	0.564	0.461	0.921	0.791
DLNN5	6.756	4.672	0.433	0.777	0.713	1.128	0.889	0.691

of inflow. Conceptually, it is a version of recurrent neural network “Cell construction model”. Every cell consists of three gates that presented input gate, forgetting gate and the model output gate. In addition, there is a vector that deals with the long-term memory of the forgoing gates. Owing to this, the input lags can be added/deleted due to the gates setting. Worth to highlight here, the gradient disappearance can be resolved based on the potential if the last two gates on forgetting the past information. The WekaDeeplearning4j tool which provides a graphical user interface (GUI) was used in this work to train and test the DL models in Weka software. For more detail on this instrument, we refer the interested reader to the article by Lang et al. (2019).

### Artificial neural network

#### Multiple linear perceptron

A feed-forward network is a multilayer perceptron NN (MLPNN) with numerous layers; in this network, the output of one neuron serves as the input to the next neuron layer. Figure 2b depicts the MLPNN model. The input layer nodes in the MLPNN can only forward the input values of the first hidden layer's node. The input–output correlation of each node can be displayed in the hidden layers as follows:

$$y = f\left(\sum_j w_j x_j + b\right) \tag{1}$$

where  $x_j$  is the output that corresponds to the  $j$  node of the previous layer,  $w_j$  is the weight that connects the  $j$  node and the current node,  $b$  is the bias value at the current node, and  $f$  is a sigmoid-like transfer function with nonlinear attributes.

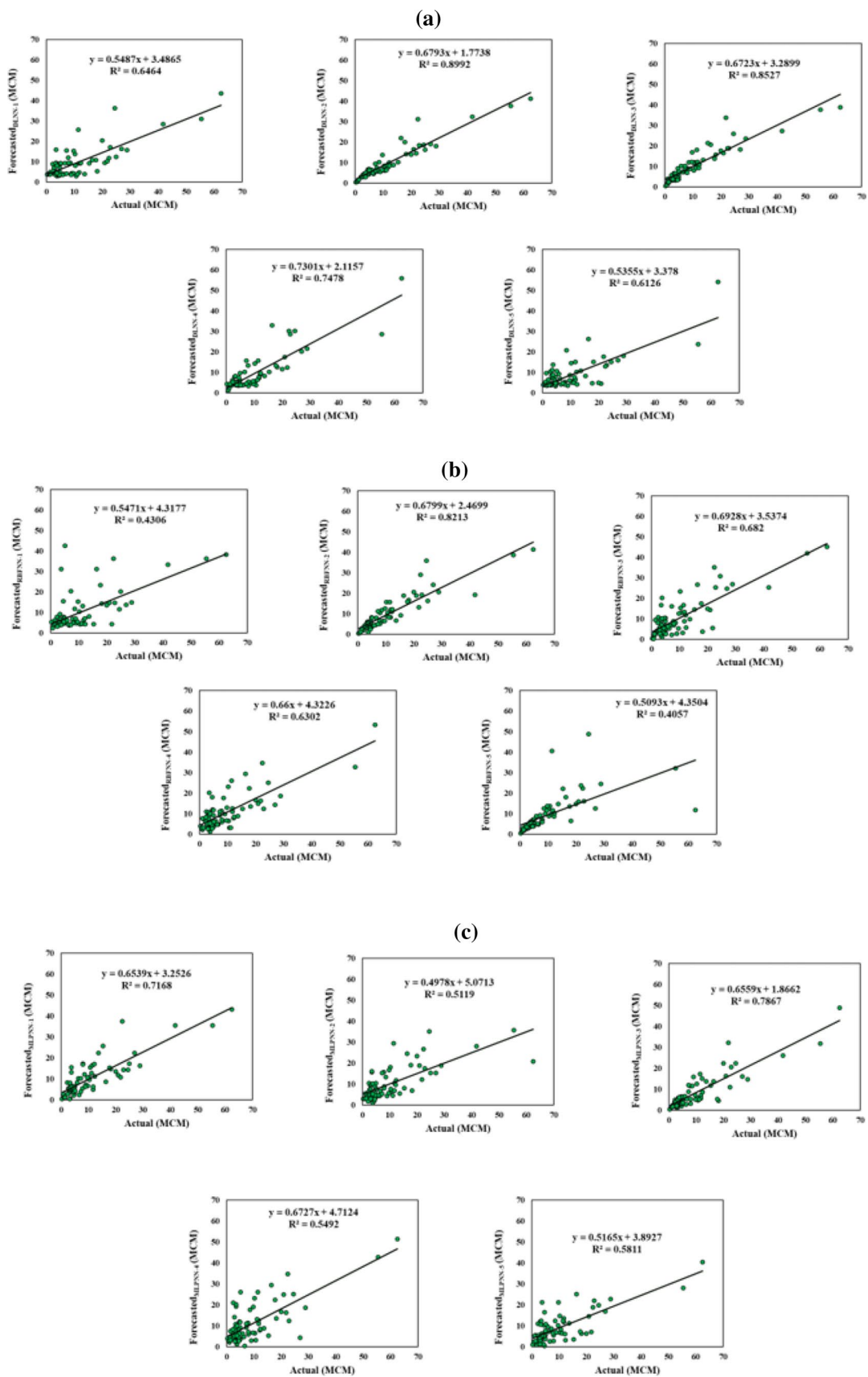
$$f(z) = \frac{1}{1 + \exp(z)} \tag{2}$$

where  $z$  is the weighted inputs aggregate, while  $f(z)$  is the neuron’s output.

The unit description of an MLPNN is an architecture that allows the computation of a nonlinear function using the scalar product of the weight and input vectors. The network architecture determines the efficiency of MLPNN models. It contains the hidden layer count, the neurons specific to each layer, as well as the form of computation employed by each neuron.

#### Radial basis function neural network (RBFNN)

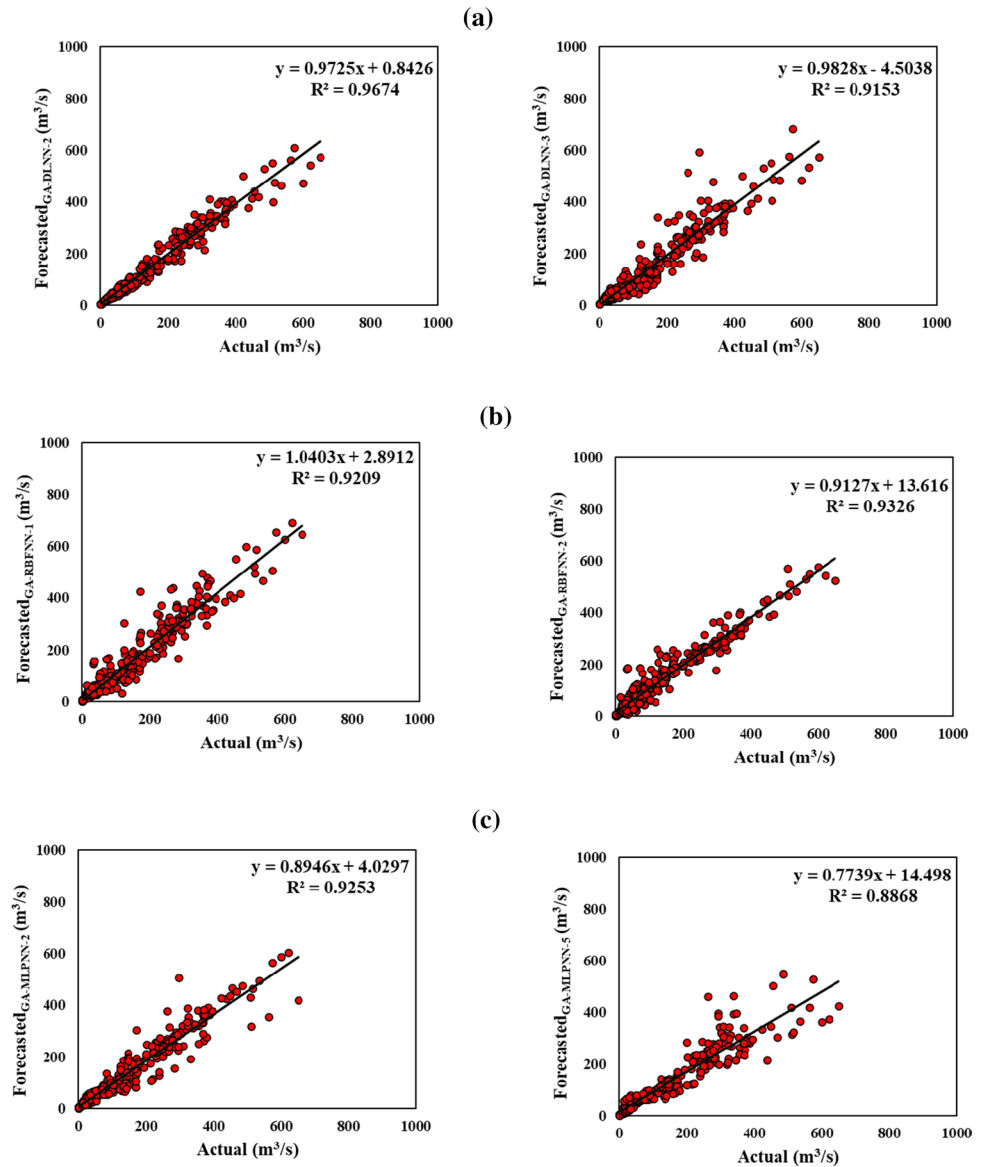
RBFNN is a function approximation variation of the standard ANN model that has a faster learning capacity (Cotar and Brilly 2008). The model structure has one input layer and one output layer, as well as a single hidden layer; it uses Gaussian functions as the basis and the least-square criterion as the objective function (Talukdar et al. 2020). In the hidden layer, the Gaussian functions give a significant response to the input boost when the network input falls within a restricted region of the input space. The RBF is presented as  $\varphi$  which is also know the hidden later function, whereas the hidden space is state in the following form  $\{\varphi_i(x)\}_{i=1}^N$ . In the forgoing function, the number of the basis is less than input data observation, typically. Hence, the role of the Gaussian is the player for the solution of the one-dimension problem that is explained as  $\varphi(x, \mu) = e^{-\frac{\|x-\mu\|^2}{2d^2}}$ .  $\mu$  is the center value of the Gaussian function.  $d$  is the radius “distance” from the input value  $x$  to the  $\varphi(x, \mu)$ , that indicates the measure of the spread of the Gaussian curve. Because of this mathematical mechanism, the RBFNN model is sometimes known as a



**Fig. 4** Scatter Plots between actual and predicted  $Q_{flow}$  using integrated ML models, **a** DLNN, **b** RBFNN, **c** MLPNN “Tropical case study”



**Fig. 5** Scatter Plots between actual and predicted  $Q_{flow}$  using integrated ML models, **a** GA-DLNN, **b** GA-RBFNN, **c** GA-MLPNN “Semi-arid region”



localized receptive field network. The convex form of the error function of BFNN allows for fast convergence to global optima (Bagtzoglou and Hossain 2009). The RBFNN framework was chosen in this work using a trial-and-error method and four distinct learning algorithms (Elzwayie et al. 2016).

**Genetic algorithm**

Similar to several introduced feature selection, GA is one of the robust one introduced in the domain of hydrology (Kamp and Savenije 2006; Moreno and Paster 2019). Numerous benefits and disadvantages of the optimization algorithms have been reported. The most popular methods used in optimization algorithms come from evolutionary computation,

a branch of computational intelligence. A genetic algorithm (GA) is a good example of the concept of evolutionary computation. This algorithm is based on the generation of a population that mimics natural evolution, selection, and natural genes (Zou et al. 2007; Sreekanth and Datta 2010; Lee and Tong 2011; Olyaie et al. 2017).

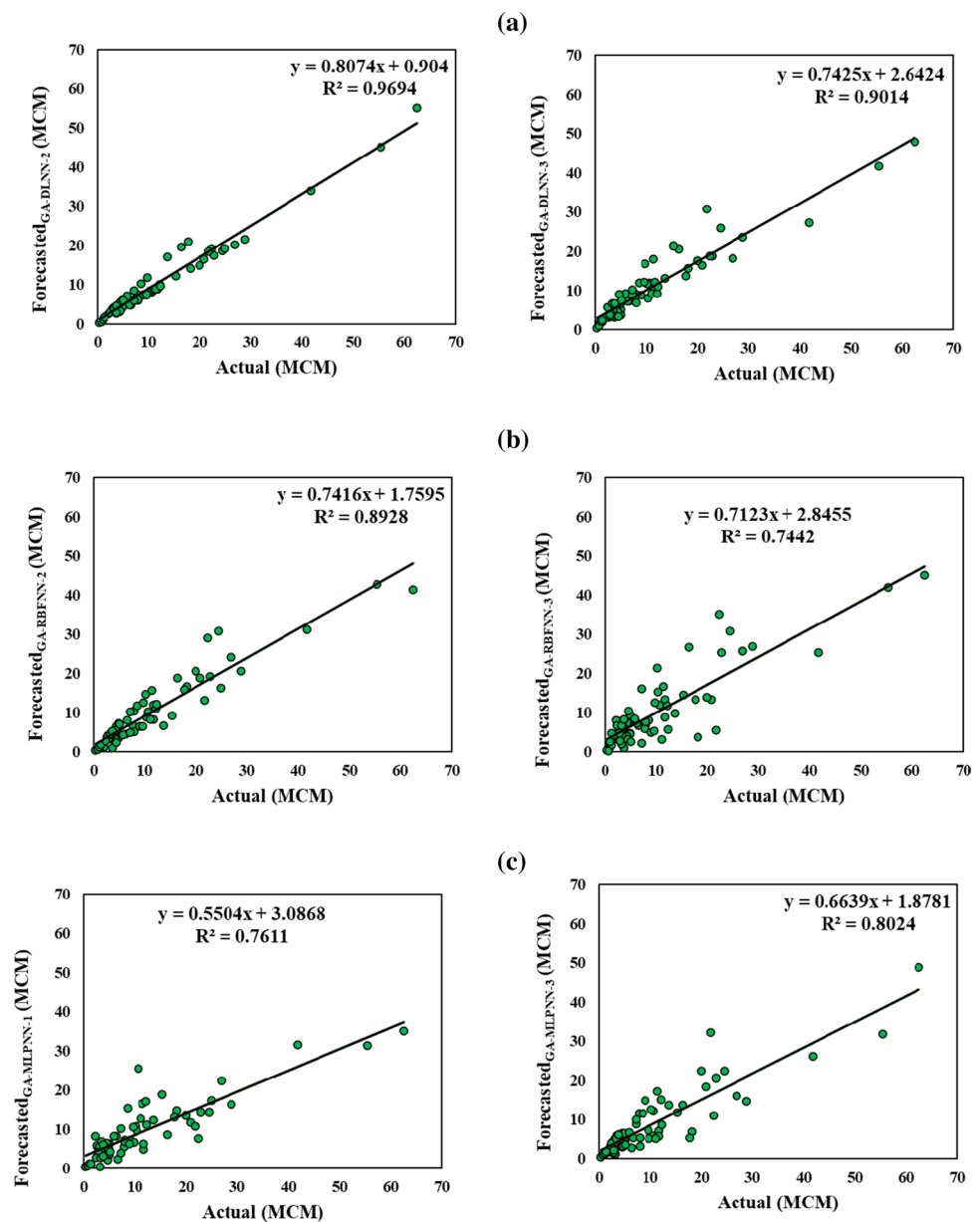
The framework of the GA is reported in Fig. 2c. The selection of the optimal lags is conducted simultaneously and determined based on the minimal error metric (i.e., root-mean-square error). The procedure is adopted due to the satisfaction of the fitness function of the GA approach (Chang et al. 2019). Worth to highlight, the GA approach is worked based on the three-optimization processes including selection, crossover and mutation.

## Implementation results and analysis

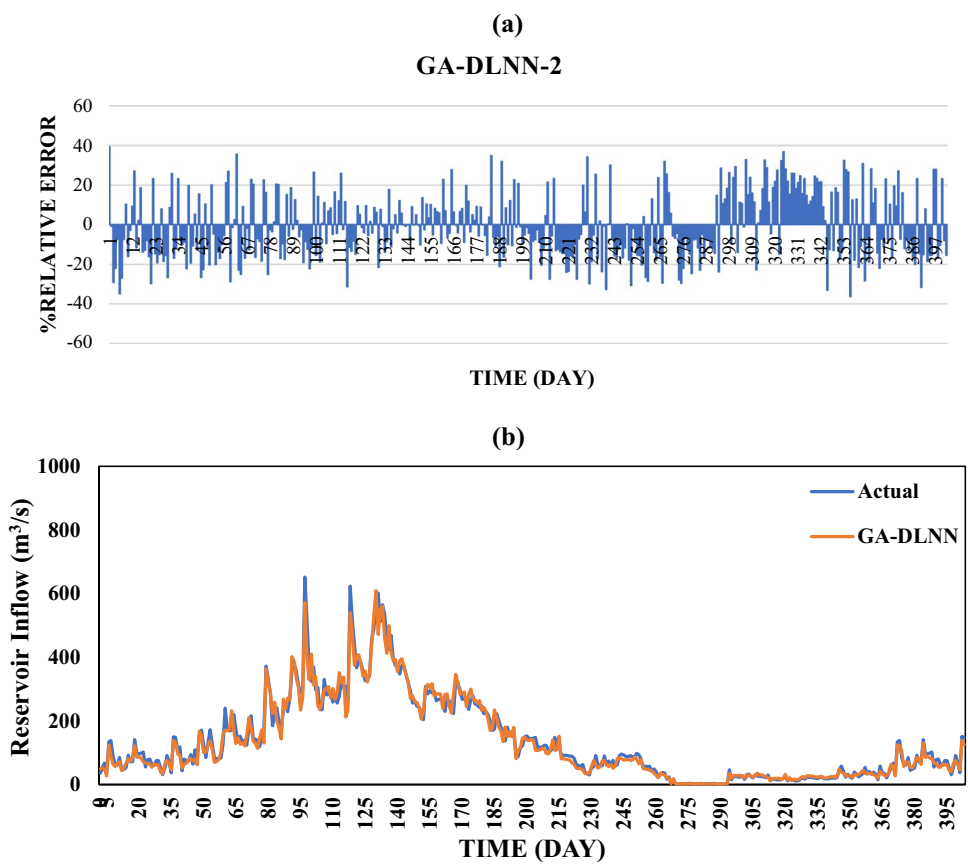
The research was adopted on the development of new machine learning model for forecasting  $Q_{\text{flow}}$  at two different regions located at semi-arid and tropical (Iraq and Malaysia), respectively. The proposed model was emphasized from the latest version of deep learning and was validated against two classical ANN algorithms. The modelling structure was initiated based on univariate modelling where only lead time of previous records was used for the initial development of the learning algorithms, where correlated lags were used as predictors for the prediction matrix. Worth to highlight, that forecasting  $Q_{\text{flow}}$  using only lag times is a distinguished modeling scheme where the merit of the machine learning models take place in mimicking the complex relationship

between the predictors and predicated. As this research was conducted in different climatic zones, this section will cover two subsections elaborating the modeling results of the developed deep learning model and its validation classical neural network algorithms. The third subsection is focused on the feasibility of integrating feature selection algorithm prior the forecasting process. Several metrics were calculated for the prediction evaluation that present the best-fit-goodness [i.e., Nash–Sutcliffe efficiency (NSE), Willmott index (d)], absolute error indicators [i.e., root-mean-square error (RMSE), mean absolute error (MAE), Nash], and scatter index (SI), BIAS, MBE; readers are advised to refer to the following literature for the reference of the mathematical expression (Yaseen 2021).

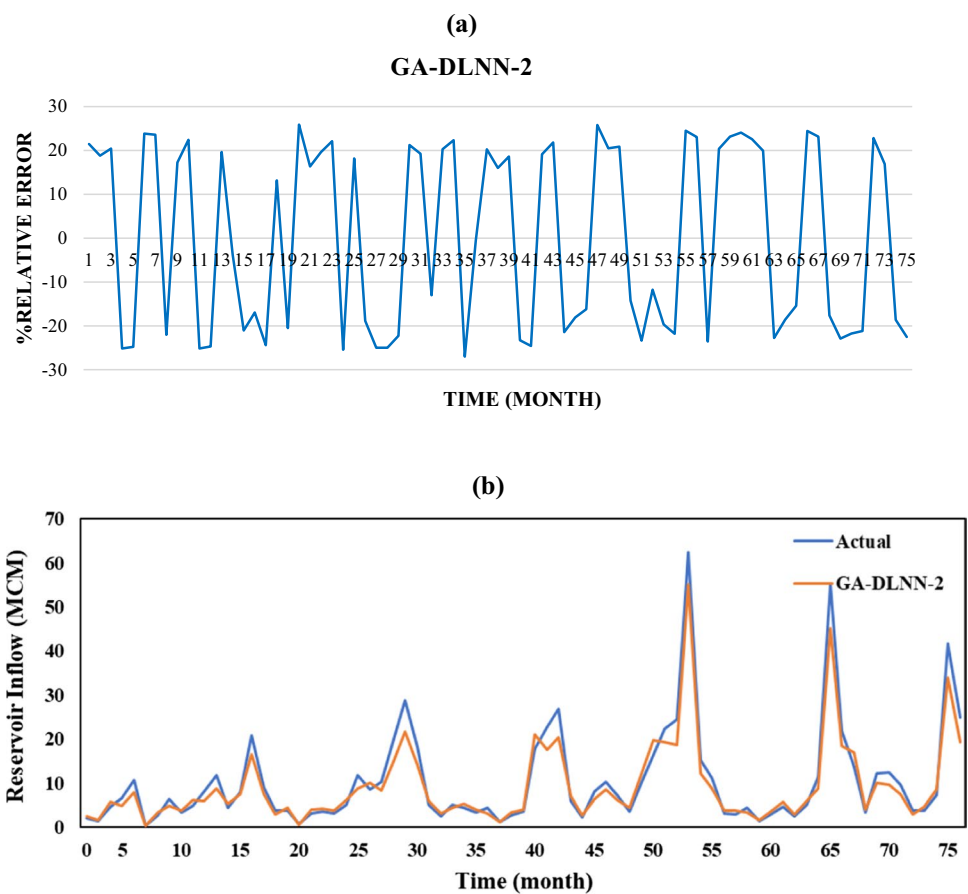
**Fig. 6** Scatter Plots between actual and predicted  $Q_{\text{flow}}$  using integrated ML models, **a** GA-DLNN, **b** GA-RBFNN, **c** GA-MLPNN “Tropical case study”

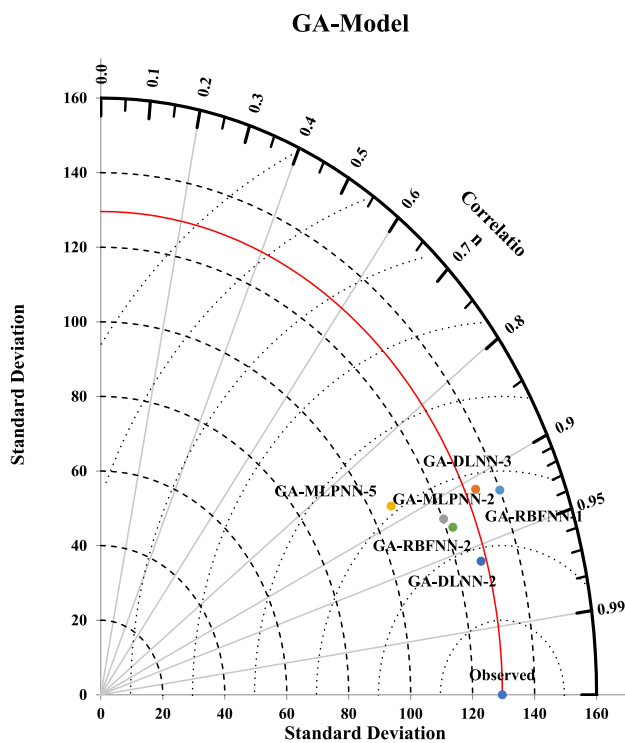


**Fig. 7** **a** The relative error percentage for the integrative GA-DLNN model for the Semi-arid case study, **b** The actual and predicted for the best results of the integrative GA-DLNN model for the Semi-arid case study

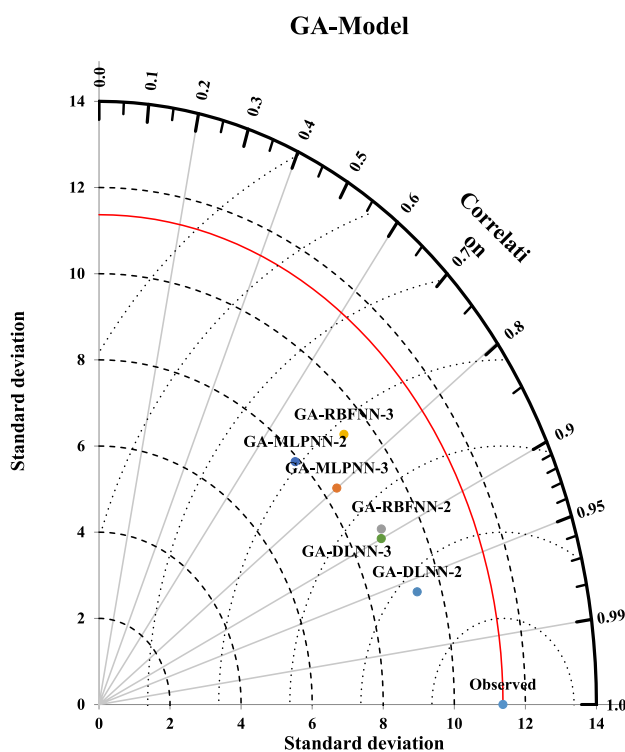


**Fig. 8** **a** The relative error percentage for the integrative GA-DLNN model for the Tropical case study, **b** the actual and predicted for the best results of the integrative GA-DLNN model for the Tropical case study





**Fig. 9** Taylor diagram for GA-model: GA-MLPNN, GA-RBFNN and GA-DLNN “Semi-arid case study”



**Fig. 10** Taylor diagram for GA-model: GA-MLPNN, GA-RBFNN and GA-DLNN “Tropical case study”

## Semi-arid case study

Table 1 tabulates the statistical results of the testing phase for the semi-arid region case study. The introduced DLNN model reported the superior results in comparison with the two classical ANN algorithms. It can be observed that MLPNN and RBFNN attained almost similar forecasting results. In quantitative explanation, DLNN model attained minimum RMSE and MAE (39.62 and 23.67) and maximum  $d$  and NSE (0.96 and 0.95). With respect to the benchmark models MLPNN and RBFNN, the statistical indicators showed much lower prediction results MLPNN attained (RMSE = 53.18, MAE = 36.79,  $d$  = 0.94, NSE = 0.91) and RBFNN reported (RMSE = 52.19, MAE = 33.34,  $d$  = 0.95, NSE = 0.92). It can be noted here that all models for the case of the semi-arid region, the second lag time series provided the best forecasting results. Although the correlation was determined for the five lags using the auto-correlation statistics, this gives the credit that the applied ML models reported a homogeneous mechanism in abstracting the essential information from the memorial time series.

Figure 3a, b and c explains the deviation from the identical line in the form of scatter plots for the applied ML models (i.e., DLNN, MLPNN and RBFNN) and for the five-input combination configured at the first place. The maximum determination coefficient uses the second input combination for the DLNN model ( $R^2 = 0.90$ ), whereas the comparable models attained MLPNN ( $R^2 = 0.85$ ) and RBFNN ( $R^2 = 0.87$ ). It can be observed from Fig. 3 presentation, the models in general performed well. Particularly the DLNN attained identical prediction for the whole range of the data minimum and maximum  $Q_{\text{flow}}$  data.

## Tropical case study

The statistical results over the testing phases for the tropical region case study are reported in Table 2. Apparently, the developed DLNN model attained the best prediction results with values of (RMSE = 4.69, MAE = 2.89,  $d$  = 0.94, NSE = 0.90), whereas MLPNN attained (RMSE = 5.66, MAE = 3.67,  $d$  = 0.92, NSE = 0.85) and RBFNN attained (RMSE = 5.16, MAE = 3.08,  $d$  = 0.93, NSE = 0.88). The best results indicated that the best results were achieved using the second lags incorporating two months of previous inflow to forecast one step ahead inflow for the DLNN and RBFNN models. On the other hand, the MLPNN showed that including three lags is the best scenario for the forecasting process. The superiority of the DLNN clearly explained the prediction performance enhancement. Also, this is elaborating the merit of the DLNN in better understanding the complicated relationship using the feasibility of the deep learning processes executed using multiple layers learning over the classical introduced ML algorithms over the literature.

**Table 3** The statistical indicators whiling testing phase for the integrative models “Semi-arid case study”

Models	RMSE	MAE	MBE	NSE	SI	BIAS	d	CI
GA-MLPNN1	49.972	32.194	0.144	0.929	0.369	− 0.503	0.951	0.883
GA-MLPNN2	36.986	21.717	− 0.006	0.961	0.273	10.238	0.970	0.932
GA-RBFNN1	40.561	25.641	0.160	0.953	0.299	− 8.349	0.965	0.920
GA-RBFNN2	33.680	21.886	0.093	0.968	0.249	− 1.799	0.974	0.943
GA-DLNN2	23.493	15.556	− 0.008	0.984	0.174	2.882	0.987	0.972
GA-DLNN3	39.304	24.247	− 0.031	0.956	0.291	6.834	0.967	0.924

Graphical results were confirmed the presented predictability performance displayed in Table 2, based on the scatter plots generated in Fig. 4; the ideal correlation was observed for the DLNN using the second lag with max determination coefficient ( $R^2=0.89$ ). In comparison with the benchmark models, MLPNN was attained max determination coefficient equal to ( $R^2=0.78$ ); RBFNN was given max determination coefficient equal to ( $R^2=0.82$ ).

**Integrative predictive model results**

Modeling  $Q_{flow}$  based on univariate modeling where only historical data of inflow used for the learning process is somehow a complex hydrological problem. Hence, reducing the dimension of the prediction matrix though the integration of the feature selection can participate essentially on providing a reliable and robust predictive model. Hence, the results of the hypothesized integration of GA as selection algorithm are presented in this subsection.

The results of the semi-arid indicated the variation between the best selected lags between the applied ML models. The best results configured using GA feature selection using second and third lags with best statistical (i.e., GA-DLNN-2: (RMSE = 23.49, MAE = 15.55,  $d=0.98$ , NSE = 0.98) and GA-DLNN-3: (RMSE = 39.30, MAE = 24.24,  $d=0.96$ , NSE = 0.95)). However, lower prediction results uses the first two lags. The results of the Tropical case study revealed similar results with respect to the optimal lags, GA-MLPNN and GA-RBFNN best results the first two lags. GA-BLNN best results using the second and third lags. In quantitative results, [i.e., GA-DLNN-2: (RMSE = 2.92, MAE = 2.06,  $d=0.97$ , NSE = 0.96) and GA-DLNN-3: (RMSE = 3.99, MAE = 2.57,  $d=0.95$ , NSE = 0.93)].

**Table 4** The statistical indicators whiling testing phase for the integrative models “Tropical case study”

Models	RMSE	MAE	MBE	NSE	SI	BIAS	d	CI
GA-MLPNN1	6.342	3.949	0.080	0.829	0.611	1.583	0.912	0.756
GA-MLPNN3	5.483	3.428	0.015	0.867	0.550	1.475	0.926	0.803
GA-RBFNN2	4.216	2.612	0.019	0.924	0.410	0.901	0.952	0.880
GA-RBFNN3	5.691	3.939	0.276	0.857	0.570	0.025	0.919	0.788
GA-DLNN2	2.922	2.063	0.001	0.964	0.284	1.079	0.974	0.939
GA-DLNN3	3.993	2.574	0.214	0.930	0.400	− 0.073	0.955	0.887

Graphical result is based on scatter plots for the two cases in Figs. 5 and 6. The max determination coefficient (GA-DLNN:  $R^2=0.96$ ) was attained for both case studies. The relative error percentage and actual/forecasted time series graphics were calculated and presented in Figs. 7 and 8 for semi-arid and tropical, respectively. It can be observed that the relative error percentage ranged between  $\mp 30\%$  for the semi-arid case study, while the tropical case study achieved even lower relative error percentage ranged between  $\mp 20\%$ . Final graphical presentation tested for the research results model is the Taylor diagram (Taylor 2001). Figures 9 and 10 present the two dimensions of the Taylor diagram for the conducted integrative ML models for both cases. Clearly, the GA-DLNN model showed nearer coordinate to the observed record of  $Q_{flow}$ . The results for both cases confirmed crucial findings: (i) the feasibility of the GA feature selection algorithm for reducing the dimension of the predictors and facilitate more reliable prediction matrix for the learning process, (ii) the adopted integrative GA-DLNN model confirmed its capability in modeling different time scales  $Q_{flow}$  day scale “semi-arid region” and monthly scale “tropical region”, and (iii) GA-DLNN model could comprehend the actual mechanism that interconnect the predictors and predictand with more robust manner for both regions stochasticity (Tables 3, 4).

**Discussion, limitation and future research**

This current research is similar to several adopted related literature on reservoir inflow forecasting. However, the main contribution that is worth to highlight here is the potential of introducing new version of machine learning for better prediction accuracy. In addition, the capability to merge the mask input selection automated algorithm for selecting the

relevant predictors. The finding of the research is totally supporting the research hypothesized assigned in the first section of this article. The limitation of the current research is the large data span that is possibility to be incorporated in the model's construction. In addition, the whole process is an offline modeling for reservoir inflow forecasting. Therefore, for better practicality, an active learning can be adopted for such a kind of simulation and that can be adopted in future studies.

## Conclusion

The main objective of the current study was to roll out a new robust and reliable predictive model to forecast reservoir inflow data in two different climatic regions (semi-arid and tropical). The development and validation of the deep learning predictive model started using two algorithms of the ANN model (MLPNN and RBFNN). In addition, transcriptomes of certified ML models were tested where the genetic algorithm was incorporated as an approach to selecting reliable input variables.

The results of the prediction accuracy indicated the potential of the DLNN model over the benchmark models and for both investigated case studies. Also, it was observed that the DLNN model confirmed its capability in solving daily scale reservoir inflow more accurately in comparison with the monthly scale. Further, the research finding approved the adequacy of the data span used in the current investigation development of computerized ML algorithm. The capacity of the GA approach had reduced the dimension of the prediction matrix and provided the learning process of the ML models with more informative historical memorial time series data.

**Authors contribution** The first draft of the manuscript was written by the first author. The manuscript version has been edited and revised by fourth and fifth authors. Supervising and reviewing have been made by the second, third and sixth authors.

**Funding** The research has no funding.

**Availability of data and materials** Not applicable.

## Declarations

**Conflict of interest** The authors declare no conflict of interest.

**Ethical approval** We acknowledge that the current research has been conducted ethically and the final shape of the research has been agreed by all authors.

**Consent to participate** The authors consent to participate in this research study.

**Consent to publish** The authors consent to publish the current research in AWSC journal.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Afan HA, Allawi MF, El-Shafie A et al (2020) Input attributes optimization using the feasibility of genetic nature inspired algorithm: application of river flow forecasting. *Sci Rep* 10:1–15. <https://doi.org/10.1038/s41598-020-61355-x>
- Ahmed S, Coulibaly P, Tsanis I (2015) Improved spring peak-flow forecasting using ensemble meteorological predictions. *J Hydrol Eng.* [https://doi.org/10.1061/\(asce\)he.1943-5584.0001014](https://doi.org/10.1061/(asce)he.1943-5584.0001014)
- Alamgir M, Khan N, Shahid S et al (2020) Evaluating severity–area–frequency (SAF) of seasonal droughts in Bangladesh under climate change scenarios. *Stoch Environ Res Risk Assess.* <https://doi.org/10.1007/s00477-020-01768-2>
- Ali M, Prasad R, Xiang Y, Yaseen ZM (2020) Complete ensemble empirical mode decomposition hybridized with random forest and kernel ridge regression model for monthly rainfall forecasts. *J Hydrol.* <https://doi.org/10.1016/j.jhydrol.2020.124647>
- Alizamir M, Kisi O, Ahmed AN et al (2020) Advanced machine learning model for better prediction accuracy of soil temperature at different depths. *PLoS ONE.* <https://doi.org/10.1371/journal.pone.0231055>
- Aljanabi QA, Chik Z, Allawi MF et al (2017) Support vector regression-based model for prediction of behavior stone column parameters in soft clay under highway embankment. *Neural Comput Appl.* <https://doi.org/10.1007/s00521-016-2807-5>
- Allawi MF, Jaafar O, Mohamad Hamzah F et al (2019a) Forecasting hydrological parameters for reservoir system utilizing artificial intelligent models and exploring their influence on operation performance. *Knowl-Based Syst* 163:907–926. <https://doi.org/10.1016/J.KNOSYS.2018.10.013>
- Allawi MF, Jaafar O, Mohamad Hamzah F et al (2017) Reservoir inflow forecasting with a modified coactive neuro-fuzzy inference system: a case study for a semi-arid region. *Theor Appl Climatol.* <https://doi.org/10.1007/s00704-017-2292-5>
- Allawi MF, Jaafar O, Mohamad Hamzah F et al (2018) Operating a reservoir system based on the shark machine learning algorithm. *Environ Earth Sci* 77:366. <https://doi.org/10.1007/s12665-018-7546-8>
- Allawi MF, Othman FB, Afan HA et al (2019b) Reservoir evaporation prediction modeling based on artificial intelligence methods. *Water (switzerland).* <https://doi.org/10.3390/w11061226>
- Apaydin H, Feizi H, Sattari MT et al (2020) Comparative analysis of recurrent neural network architectures for reservoir inflow forecasting. *Water* 12:1500. <https://doi.org/10.3390/w12051500>
- Bagtzoglou AC, Hossain F (2009) Radial basis function neural network for hydrologic inversion: an appraisal with classical and spatio-temporal geostatistical techniques in the context of site characterization. *Stoch Environ Res Risk Assess* 23:933–945

- Bai Y, Chen Z, Xie J, Li C (2016) Daily reservoir inflow forecasting using multiscale deep feature learning with hybrid models. *J Hydrol* 532:193–206. <https://doi.org/10.1016/j.jhydrol.2015.11.011>
- Bai Y, Sun Z, Zeng B et al (2018) Reservoir inflow forecast using a clustered random deep fusion approach in the three gorges reservoir. *China J Hydrol Eng* 23:4018041. [https://doi.org/10.1061/\(asce\)he.1943-5584.0001694](https://doi.org/10.1061/(asce)he.1943-5584.0001694)
- Bai Y, Wang P, Xie J et al (2015) additive model for monthly reservoir inflow forecast. *J Hydrol Eng* 20:4014079. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001101](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001101)
- Bashir A, Shehzad MA, Hussain I et al (2019) Reservoir inflow prediction by ensembling wavelet and bootstrap techniques to multiple linear regression model. *Water Resour Manag*. <https://doi.org/10.1007/s11269-019-02418-1>
- Chang LJ, Kuo CM, Tseng HW, Yu PS (2019) Application of multi-objective genetic algorithm on parameter optimization of DHSVM: a case study in shihmen reservoir catchment. *J Taiwan Agric Eng*. [https://doi.org/10.29974/JTAE.201903\\_65\(1\).0002](https://doi.org/10.29974/JTAE.201903_65(1).0002)
- Choong S-M, El-Shafie A (2015) State-of-the-art for modelling reservoir inflows and management optimization. *Water Resour Manag* 29:1267–1282
- Cotar A, Brilly M (2008) Use of normalized radial basis function in hydrology. In: AIP conference proceedings
- Coulibaly P, Anctil F, Bobée B (2000) Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. *J Hydrol* 230:244–257. [https://doi.org/10.1016/S0022-1694\(00\)00214-6](https://doi.org/10.1016/S0022-1694(00)00214-6)
- Coulibaly P, Haché M, Fortin V, Bobée B (2005) Improving daily reservoir inflow forecasts with model combination. *J Hydrol Eng*. [https://doi.org/10.1016/\(ASCE\)1084-0699\(2005\)10:2\(91\)](https://doi.org/10.1016/(ASCE)1084-0699(2005)10:2(91))
- Ehteram M, Mousavi SF, Karami H, et al (2018) Reservoir operation based on evolutionary algorithms and multi-criteria decision-making under climate change and uncertainty. *J Hydroinformat*. <https://doi.org/10.2166/hydro.2018.094>
- El-Shafie A, Abdin AE, Noureldin A, Taha MR (2009) Enhancing inflow forecasting model at aswan high dam utilizing radial basis neural network and upstream monitoring stations measurements. *Water Resour Manag* 23:2289–2315. <https://doi.org/10.1007/s11269-008-9382-1>
- El-Shafie A, Taha MR, Noureldin A (2007) A neuro-fuzzy model for inflow forecasting of the Nile river at Aswan high dam. *Water Resour Manag* 21:533–556. <https://doi.org/10.1007/s11269-006-9027-1>
- Elzwayie A, El-shafie A, Yaseen ZM et al (2016) RBFNN-based model for heavy metal prediction for different climatic and pollution conditions. *Neural Comput Appl*. <https://doi.org/10.1007/s00521-015-2174-7>
- Esmailzadeh B, Sattari MT, Samadianfard S (2017) Performance evaluation of ANNs and an M5 model tree in Sattarkhan Reservoir inflow prediction. *ISH J Hydraul Eng*. <https://doi.org/10.1080/09715010.2017.1308277>
- Ghimire S, Deo RC, Downs NJ, Raj N (2018) Self-adaptive differential evolutionary extreme learning machines for long-term solar radiation prediction with remotely-sensed MODIS satellite and reanalysis atmospheric products in solar-rich cities. *Remote Sens Environ* 212:176–198. <https://doi.org/10.1016/j.rse.2018.05.003>
- Goodfellow I, Bengio Y, Courville A (2016) *Deep learning*. MIT Press
- Hadiyan PP, Moeini R, Ehsanzadeh E (2020) Application of static and dynamic artificial neural networks for forecasting inflow discharges, case study: Sefidroud Dam reservoir. *Sustain Comput Informatics Syst* 27:100401. <https://doi.org/10.1016/j.suscom.2020.100401>
- Herbert ZC, Asghar Z, Oroza C (2021) Long-term reservoir inflow forecasts: enhanced water supply and inflow volume accuracy using deep learning. *J Hydrol* 601:126676. <https://doi.org/10.1016/j.jhydrol.2021.126676>
- Hossain MS, El-shafie A (2014) Performance analysis of artificial bee colony (ABC) algorithm in optimizing release policy of Aswan High Dam. *Neural Comput Appl* 24:1199–1206. <https://doi.org/10.1007/s00521-012-1309-3>
- Hrnjica B, Mehr AD (2020) Energy demand forecasting using deep learning. In: *Smart cities performability, cognition, and security*. Springer, pp 71–104
- Kamp RG, Savenije HHG (2006) Optimising training data for ANNs with genetic algorithms. *Hydrol Earth Syst Sci*. <https://doi.org/10.5194/hess-10-603-2006>
- Kim T, Shin JY, Kim H et al (2019) The use of large-scale climate indices in monthly reservoir inflow forecasting and its application on time series and artificial intelligence models. *Water (switzerland)*. <https://doi.org/10.3390/w11020374>
- Lang S, Bravo-Marquez F, Beckham C et al (2019) Wekadeeplearning4j: a deep learning package for weka based on deeplearning4j. *Knowl-Based Syst* 178:48–50
- Lecun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444. <https://doi.org/10.1038/nature14539>
- Lee D, Kim H, Jung I, Yoon J (2020) Monthly reservoir inflow forecasting for dry period using teleconnection indices: a statistical ensemble approach. *Appl Sci* 10:3470. <https://doi.org/10.3390/app10103470>
- Lee Y-S, Tong L-I (2011) Forecasting time series using a methodology based on autoregressive integrated moving average and genetic programming. *Knowl-Based Syst* 24:66–72. <https://doi.org/10.1016/J.KNOSYS.2010.07.006>
- Liu Y, Sang Y-F, Li X et al (2016) Long-term streamflow forecasting based on relevance vector machine model. *Water* 9:9. <https://doi.org/10.3390/w9010009>
- Liu Y, Sang YF, Li X et al (2017) Long-term streamflow forecasting based on relevance vector machine model. *Water (switzerland)*. <https://doi.org/10.3390/w9010009>
- Moreno Z, Paster A (2019) A genetic algorithm for stochastic inversion in contaminant subsurface hydrology. *Groundwater*. <https://doi.org/10.1111/gwat.12863>
- Mosavi A, Ozturk P, Chau KW (2018) Flood prediction using machine learning models: literature review. *Water* 10(11):1536. <https://doi.org/10.3390/w10111536>
- Mukhlisin M, El-Shafie A, Taha MR (2012) Regularized versus non-regularized neural network model for prediction of saturated soil-water content on weathered granite soil formation. *Neural Comput Appl*. <https://doi.org/10.1007/s00521-011-0545-2>
- Nearing GS, Kratzert F, Sampson AK et al (2021) What role does hydrological science play in the age of machine learning? *Water Resour Res*. <https://doi.org/10.1029/2020WR028091>
- Olyaie E, Zare Abyaneh H, Danandeh Mehr A (2017) A comparative analysis among computational intelligence techniques for dissolved oxygen prediction in Delaware River. *Geosci Front*. <https://doi.org/10.1016/j.gsf.2016.04.007>
- Osman A, Afan HA, Allawi MF et al (2020) Adaptive Fast Orthogonal Search (FOS) algorithm for forecasting streamflow. *J Hydrol* 586:124896. <https://doi.org/10.1016/j.jhydrol.2020.124896>
- Petty TR, Dhingra P (2018) Streamflow hydrology estimate using machine learning (SHEM). *JAWRA J Am Water Resour Assoc* 54:55–68
- Raghavendra NS, Deka PC (2014a) Support vector machine applications in the field of hydrology: a review. *Appl Soft Comput* 19:372–386. <https://doi.org/10.1016/j.asoc.2014.02.002>
- Raghavendra S, Deka PC (2014b) Support vector machine applications in the field of hydrology: a review. *Appl Soft Comput J* 19:372–386. <https://doi.org/10.1016/j.asoc.2014.02.002>
- Rezaeianzadeh M, Stein A, Cox JP (2016) Drought forecasting using Markov chain model and artificial neural networks.

- Water Resour Manag 30:2245–2259. <https://doi.org/10.1007/s11269-016-1283-0>
- Salih SQ, Allawi MF, Yousif AA et al (2019) Viability of the advanced adaptive neuro-fuzzy inference system model on reservoir evaporation process simulation: case study of Nasser Lake in Egypt. *Eng Appl Comput Fluid Mech* 13:878–891. <https://doi.org/10.1080/19942060.2019.1647879>
- Schmidhuber J (2015) Deep learning in neural networks: an overview. *Neural Netw* 61:85–117
- Singh TP, Nandimath P, Kumbhar V et al (2021) Drought risk assessment and prediction using artificial intelligence over the southern Maharashtra state of India. *Model Earth Syst Environ* 7:2005–2013
- Sreekanth J, Datta B (2010) Multi-objective management of saltwater intrusion in coastal aquifers using genetic programming and modular neural network based surrogate models. *J Hydrol*. <https://doi.org/10.1016/j.jhydrol.2010.08.023>
- Talukdar S, Pal S, Chakraborty A, Mahato S (2020) Damming effects on trophic and habitat state of riparian wetlands and their spatial relationship. *Ecol Indic*. <https://doi.org/10.1016/j.ecolind.2020.106757>
- Tao H, Sulaiman SO, Yaseen ZM, et al (2018) What is the potential of integrating phase space reconstruction with SVM-FFA data-intelligence model? Application of rainfall forecasting over regional scale. *Water Resour Manag*. <https://doi.org/10.1007/s11269-018-2028-z>
- Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. *J Geophys Res Atmos* 106:7183–7192. <https://doi.org/10.1029/2000JD900719>
- Tran TD, Tran VN, Kim J (2021) Improving the accuracy of dam inflow predictions using a long short-term memory network coupled with wavelet transform and predictor selection. *Mathematics* 9:551. <https://doi.org/10.3390/math9050551>
- Wee WJ, Zaini NB, Ahmed AN, El-Shafie A (2021) A review of models for water level forecasting based on machine learning. *Earth Sci Informatics*. <https://doi.org/10.1007/s12145-021-00664-9>
- Xu S, Chen Y, Xing L, Li C (2021) Baipenzhu reservoir inflow flood forecasting based on a distributed hydrological model. *Water* 13:272
- Yahya ASA, Ahmed AN, Othman FB et al (2019) Water quality prediction model based support vector machine model for ungauged river catchment under dual scenarios. *Water (switzerland)*. <https://doi.org/10.3390/w11061231>
- Yaseen ZM (2021) An insight into machine learning models era in simulating soil, water bodies and adsorption heavy metals: Review, challenges and solutions. *Chemosphere* 130126. <https://doi.org/10.1016/j.chemosphere.2021.130126>
- Yaseen ZM, Allawi MF, Yousif AA et al (2016a) Non-tuned machine learning approach for hydrological time series forecasting. *Neural Comput Appl*. <https://doi.org/10.1007/s00521-016-2763-0>
- Yaseen ZM, Ehteram M, Sharafati A et al (2018) The integration of nature-inspired algorithms with least square support vector regression models: application to modeling river dissolved oxygen concentration. *Water* 10:1124. <https://doi.org/10.3390/w10091124>
- Yaseen ZM, Kisi O, Demir V (2016b) Enhancing long-term streamflow forecasting and predicting using periodicity data component: application of artificial intelligence. *Water Resour Manag* 30:4125–4151. <https://doi.org/10.1007/s11269-016-1408-5>
- Zou R, Lung WS, Wu J (2007) An adaptive neural network embedded genetic algorithm approach for inverse water quality modeling. *Water Resour Res*. <https://doi.org/10.1029/2006WR005158>
- Zounemat-Kermani M, Batelaan O, Fadaee M, Hinkelmann R (2021) Ensemble machine learning paradigms in hydrology: a review. *J Hydrol*. <https://doi.org/10.1016/j.jhydrol.2021.126266>

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