Sustainable Water Resources Management (2021) 7:98
https://doi.org/10.1007/s40899-021-00584-y

ORIGINAL ARTICLE

A review on the applications of machine learning for runoff modeling

Babak Mohammadi[1]

Received: 2 August 2021 / Accepted: 13 October 2021 / Published online: 19 October 2021

Abstract

The growing menace of global warming and restrictions on access to water in each region is a huge threat to global hydrological sustainability. Hence, the perspective at which hydrological studies are currently being carried out across the world to quantify and understand the water cycle modeling requires a further boost. In the past few decades, the theoretical understanding of machine learning (ML) algorithms for solving engineering issues, and the application of this method to practical problems have made very significant progress. In the field of hydrology, ML has been using for a better understanding of hydrological complexities. Then, using ML-based approaches for hydrological simulation have been a popular method for runoff modeling in recent years; it seems necessary to understand the application of ML in runoff modeling fully. Current research seeks to have an overview for rainfall–runoff modeling using ML approaches in recent years, including integrated and ordinary ML techniques (such as ANFIS, ANN, and SVM models). The main hydrological topics in this review study include surface hydrology, streamflow, rainfall–runoff, and flood modeling via ML approaches. Therefore, in this study, the author has critically reviewed the characteristics of machine learning models in runoff simulation, including advantages and disadvantages of three widely used machine learning models.

Keywords Future research direction · Hydrology · Machine learning · Runoff simulation · Water resources engineering

Introduction

In recent years, methods involving data-driven models and machine learning (ML) have been developed to predict runoff (Nourani et al. 2011, 2021; Mohammadi et al. 2020a, c). The relationship between hydrologic cycle variables and runoff in theoretical system models is approached directly without considering the physical processes involved (Okkan et al. 2021; Alizadeh et al. 2021). Also, this type of black-box model may consider some unpredictable hydrological terms during the modeling process, and they can be understood as the hydrological phoneme in view of data-driven knowledge. Nonetheless, such ML (black-box) methods have been proved to have impressive accuracy in runoff simulating (Mohammadi and Mehdizadeh 2020; Sang 2013; Abbot and Marohasy 2012).

In hydrological modeling studies, accurate runoff modeling is the main research topic that affects water resources planning, including dam design, water resource allocation plans, catchment area management, and flood management (Nourani et al. 2009; Zhou et al. 2019; Chadalawada et al. 2020; Mohammadi et al. 2020b). It is scientifically proven that due to the physical processes and natural changes related to the river system, then the prediction of the river system and its runoff behavior is particularly difficult to analyze. In hydrological applications, the need to improve the reliability and accuracy of hydrological variable prediction has attracted much attention (Niu et al. 2019). In the process of the research plan carried out by hydrometeorological researchers, no one has been determined yet. Due to different physical phenomena, such as the pattern, periodicity or randomness in model input and target data, and natural randomness in general, a method that can usually be used to simulate hydrological processes under different conditions is ML approaches (Sharafati et al. 2020; Mohammadi et al. 2021a). Considering this point of view can also be assumed that there is no general model that performs better than other models under various hydrological conditions and different catchment characteristics (Adnan et al. 2021). Due to model instability and runoff dynamics, including extreme events in historical data, a large number of models cannot

✉ Babak Mohammadi
babak.mohammadi@nateko.lu.se

[1] Department of Physical Geography and Ecosystem Science, Lund University, Sölvegatan 12, SE-223 62 Lund, Sweden

make consistent predictions (Oppel and Schumann 2020). Because of these limitations, researchers prefer to study and develop more robust and general models to improve performance using available historical data. Besides, researchers must consider the benefits of complex and rapidly evolving computing power that can enhance modeling methods and threshold accuracy in hydrological forecasting applications. In addition, the researchers also applied complex modeling theorems and newly developed ML approaches (Oppel and Schumann 2020; Tripathy and Schwefel 1982).

For instance, Mohammadi et al. (2020a) showed ML models have excellent performance for simulating streamflow time series in four rivers in Canada and the United States. They implemented four different types of data-driven models by the name of bi-linear, multi-layer perceptron (MLP), MLP coupled by particle swarm optimization (MLP-PSO), and MLP-PSO coupled with the multi-verse optimizer (MLP-PSOMVO). Their results ranged R^2 between 0.90 to 0.99, and they resulted in ML models can understand SF phoneme, and then they can have a suitable runoff simulation. Tikhamarine et al. (2020a) compared some different types of ML models, including, MLP, and Least Squares Support Vector Machine (LSSVM), and MLP and LSSVM integrated with PSO and Harris Hawks Optimization (HHO) optimization algorithms. They presented the best results were related to LSSVM–HHO and LSSVM–PSO by NS = 0.737. Safari et al. (2020) employed Reproducing Kernel Hilbert Space (RRKHS), radial basic function (RBF), and Multivariate adaptive regression splines (MARS) approach for streamflow simulating in the Haldizen watershed (Turkey). They reported the RRKHS had the best performance by NS = 0.944 for runoff modeling. Some of reviewed articles provided by Table 1.

In the past two decades, the motivation for applying machine learning techniques to predict river flow (streamflow) has attracted significant attention to hydrology (Jothiprakash and Magar 2012; Kentel 2009; Terzi and Ergin 2014; Valipour and Montazar 2012a, b). Machine learning has made big changes in hydrological forecasting issues and handling the complexity of missing data issues in hydrological science (Wen et al. 2019). ML-based methods such as optimization algorithms, logical methods, classification methods, statistical learning methods and probability methods are widely used. The three subcategories of machine learning are particularly widely used in the hydrology and runoff fields: (i) adaptive neuro-fuzzy inference system (ANFIS) (Jang 1993), (ii) artificial neural networks (ANN) (Haykin 2004), and (iii) support vector machine (SVM) (Cortes and Vapnik 1995).

In hydrological research, the most widely used ML methods are ANFIS, ANN, and SVM models. The current study focuses on reviewing journal articles with high impact factors written about runoff modeling in different worldwide case studies. Also, this study seeks to provide the advantages and disadvantages of mentioned models in different regions. The schematic flowchart of current research is shown in Fig. 1.

Rainfall–runoff modeling via machine learning

Application of the ANFIS in runoff simulation

This method was developed first in 1993 by Jang (1993). Different researchers have developed various methods/models to simulate precipitation and runoff processes, so reliable models suitable for effective planning and management of catchment areas must be selected. ANFIS is one of these popular methods, and it is a type of artificial neural network based on the Takagi–Sugeno fuzzy system. Figure 2 shows the structure of the ANFIS model.

The aim of creating a model has always been for maximizing its application, for having a high accuracy result and overcoming the complexity of the modeling process. Generally considering the greater uncertainty is a reason for reducing the complexity of the model and increase the robustness of the model. Zadeh (1965) introduced a fuzzy set theory, the main advantage of applying this theory is that it allows having a minimum uncertainty of modeling process. This was done by looking at input variables related to preferences to make the modeling unique, or by looking at interval data rather than input variables in the form of complex data to make the modeling more explicit (Wedding 1997). They exist because of the ambiguity and inaccuracy of the system input data (Kreinovich et al. 2000).

The fuzzy theory has been widely used for improving accuracy of runoff modeling process in various studies. In another study by Chang and Chen (2001), they considered a type of the fuzzy network, which was a combined approach via fuzzy system and ANN (namely, CFNN). This model (CFNN) was employed for developing some hydrological models and it created a rainfall-based model for predicting the amount of streamflow. The form of triangular was used as a membership function for the original CFNN, which is replaced by the Gaussian function in this study. Understanding and predicting hourly runoff successfully was the biggest advantage of the proposed model (Chang and Chen 2001).

The ANFIS model is also widely used for runoff modeling; for example: El-Shafie et al. (2007) developed the ANFIS model. It is recommended to forecast the monthly runoff (El-Shafie et al. 2007). The characteristic of the ANFIS model is that it can handle the inaccuracy and uncertainty in the input of the streamflow database, because the input data can be split by a fuzzy subspace and into a linear function for predicting streamflow. They

Table 1 Information of literature reviewed about black-box models (ML-based models) in this study

Reference	Models used	Timescale	Study area	Suggested model
Abudu et al. (2010)	Artificial neural networks (ANN), autoregressive integrated moving average (ARIMA), seasonal ARIMA (SARIMA)	Monthly	Kizil River, China	Similar results were reported
Adamowski et al. (2012)	Multivariate adaptive regression splines (MARS), ANN, Wavelet-ANN	Daily	The mountainous watershed of Sainji in the Himalayas	MARS, Wavelet-ANN
Mohammadi et al. (2020c)	Adaptive neuro-fuzzy inference system (ANFIS), ANFIS–Shuffled frog leaping algorithm (SFLA)	Monthly	Vu Gia Thu Bon basin, Vietnam	ANFIS–SFLA
Humphrey et al. (2016)	ANN	Monthly	South Australia	Hybrid ANN approaches
Kisi et al. (2012)	ANN, ANFIS, Support vector machine (SVM), Local linear regression (LLR), Dynamic LLR (DLLR)	Daily	Ergene River and Seytan Stream, Turkey	ANN, ANFIS
Liu et al. (2014)	Support vector regression (SVR), W-SVR	Daily and monthly	Fork White River and Eel River, United State of America (USA)	W-SVR
Uysal et al. (2016)	MLP, RBF	Daily	Karasu River, Turkey	MLP
Abdollahi et al. (2017)	ANN, Gene expression programming (GEP), W-ANN, W-GEP	Daily	Behesht-Abad and Joneghan river, Iran	W-ANN
Siqueira et al. (2018)	Extreme learning machine (ELM), Echo state network (ESN), MLP, Partial autoregressive (PAR)	Monthly	Brazilian hydroelectric plants	ESN
Hadi and Tombul (2018)	ANN, Multigene genetic programming (MGGP), W-ANN, W-MGGP	Monthly	Goksu-Gokdere Basin, Turkey	MGGP
Tongal and Boiij (2018)	ANN, SVM, Random forests (RF)	Daily	North Fork, Chehalis, Carson and Sacramento rivers	ANN, RF
Al-Sudani et al. (2019)	Multivariate adaptive regression splines (MARS), LSSVR, MARS–Differential evolution (DE)	Monthly	Tigris River, Iraq	MARS–DE
Tikhamarine et al. (2020b)	ANN, SVR, Multiple linear regression (MLR), ANN–GWO, SVR–Gray wolf optimization (GWO), MLR–GWO	Monthly	Aswan High Dam, Iraq	SVR–GWO
Qu et al. (2021)	Regularized extreme learning machine (RELM), SVR, Gray wolf optimizer, particle swarm optimization, and genetic algorithm coupled with regularized extreme learning machine (BGWO/BPSO/GA-RELM)	Monthly	Two basins in the USA	BGWO–RELM
Parisouj et al. (2020)	ANN, ELM, and SVR	Monthly/ daily	Four rivers in the USA	SVR
Mohammadi et al. (2021b)	Two hydrological models, ANFIS, SVM, and group method of data handling (GMDH)	Monthly	Four rivers in Indonesia	Hydrological model coupled by machine learning model
Parvinizadeh et al. (2021)	MLP and supervised brain emotional learning (SBEL)	Daily	Dez Dam watershed (Iran)	MLP

Table 1 (continued)

Reference	Models used	Timescale	Study area	Suggested model
Siddiqi et al. (2021)	ANN and ELM	Monthly	Tarbela Dam (Indus River basin, Pakistan)	ELM

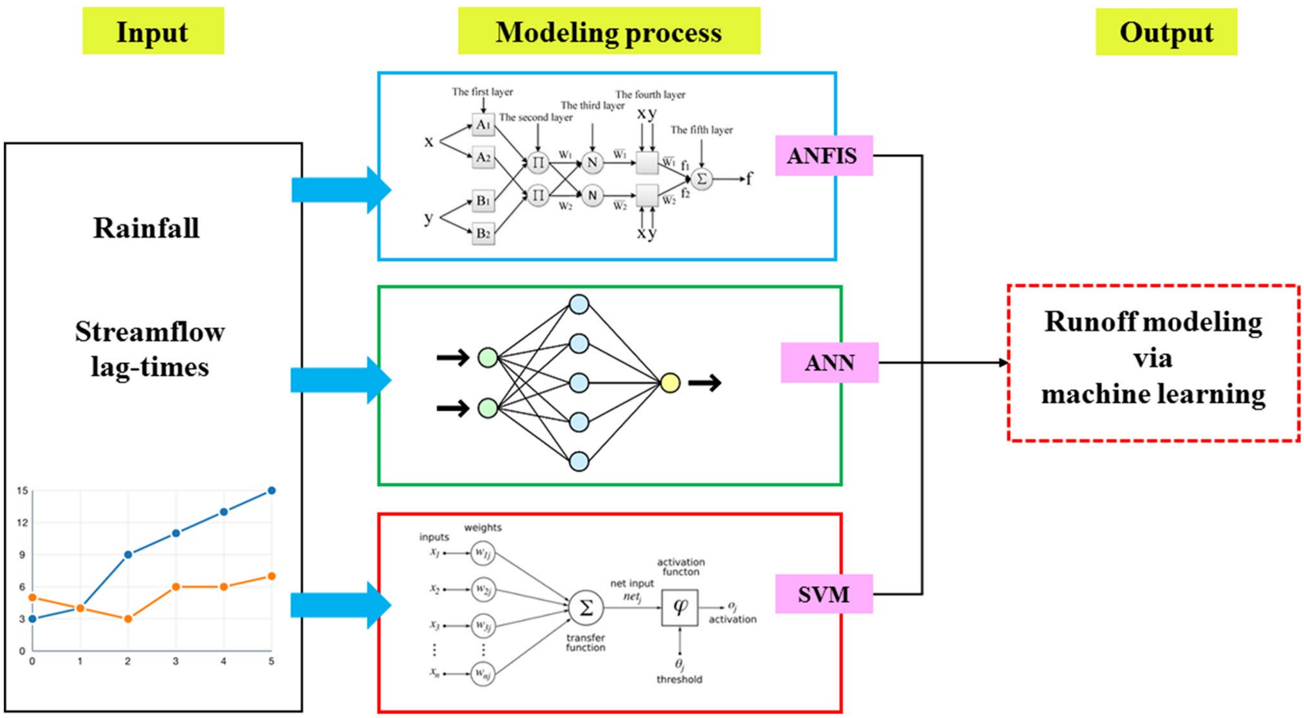


Fig. 1 The schematic diagram of current study

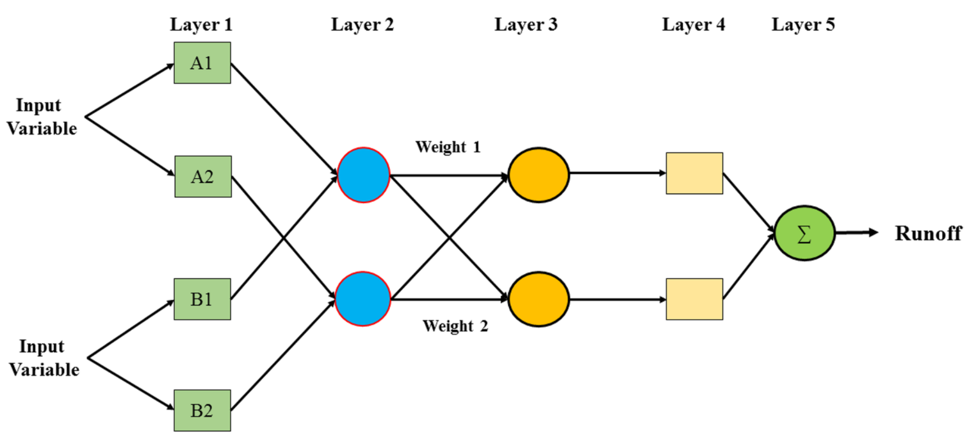


Fig. 2 The structure of ordinary the ANFIS model

used 130 years of monthly inflow historical database for training the ANFIS approach and testing the performance of ANFIS for runoff simulation. Finally, they compared ANFIS's result with the MLP model; the ANFIS showed consistently high accuracy in predicting runoff events, and its accuracy in predicting extreme streamflow event was significantly higher than the MLP model. Reliable performance in runoff prediction showed identification and application of the effective input patterns for model training can increase accuracy of runoff simulation (El-Shafie et al. 2007).

Nayak et al. (2007) enhanced two different machine learning approaches (i.e., ANN and ANFIS) to simulate the process of rainfall–runoff effectively. The results showed their proposed approach (namely, hyper-ellipsoid fuzzy clustering method (HIS)) that HIS can be selected as an alternative rainfall–runoff method, because HIS proved it can be implemented by minimum required parameters in the minimum time (Nayak et al. 2007). Özger (2009) simulated the runoff time series with Takagi Sugeno Fuzzy Inference System (TS). The TS rule was based on a set of linear functions for runoff forecasting. All the uncertainty and complexity of the proposed model were considered in the TS relationship function, the correlation between the observation and the prediction values was acceptable (Özger 2009).

Because of the complexity and the non-linearity behavior of the runoff phenomenon and also, due to the lack of suitable historical data in all regions, it is difficult to model the required runoff with physics-based models. Pramanik and Panda (2009) studied two machine learning methods (ANFIS and ANN) that use upstream flow data to estimate downstream flows. For evaluating the performance of ANFIS and ANN, the daily runoff from the reservoir upstream of the dam was used. Two methods are used to evaluate models with different input combinations to obtain the best accuracy of runoff modeling. The importance of two upstream tributaries in assessing dam runoff was also evaluated. Studies have shown that the performance of the conjugate neural gradient network is better than the Levenberg–Marquardt and gradient descent algorithms and ANFIS showed it could have more accurately runoff estimation in outlier data conditions (Pramanik and Panda 2009).

Katambara and Ndiritu (2010) reported a hybrid concept fuzzy inference model to simulate the streamflow in the South Africa. The development of the fuzzy concept hybrid model successfully applied for simulating dynamic behavior of streamflow. The study described developing a hybrid model of fuzzy calibration concepts and examined its ability to reproduce natural and human processes. The performance of this model proved a satisfactory result about modeling of hydrological system complexity and its impact on daily streamflow. The performance of streamflow simulation in the downstream direction was improved, and an independent process fuzzy model was successfully implemented. The conclusion showed that for complex river systems with a lack of data, the fuzzy concept hybrid model can be used as an capable machine learning model for reliable streamflow simulation and operation analysis (Katambara and Ndiritu 2010).

Sanikhani and Kisi (2012) developed two different ANFIS models for simulating monthly streamflow values. First, two types of ANFIS models were proposed in the mentioned study, namely ANFIS with sub-clusters (ANFISSC) and ANFIS with separated grids (ANFISGP). Both proposed approaches were used to predict the flow rate 1 month in advance, and the impact of periodicity on the model's prediction performance is examined. Another step of this study evaluated the effectiveness of the ANFIS method in assessing the flow rate. The results show that the ANFISSC model is slightly better in predicting rivers. ANFISGP model (Sanikhani and Kisi 2012). Greco (2012) studied the gradual pattern of the spread of runoff process on a daily scale. The mentioned study employed a hybrid of the autoregressive (AR) model and via the fuzzy inference system. The AR model is specifically used to identify the mainstream, and a set of fuzzy rules is determined based on the knowledge of the basic physical characteristics of the rainfall process, which limits the number of relevant parameters of the model. The daily inflow into the catchment area after 5 days is calculated based on the weighted average of the precipitation data of six rain gauges distributed in the catchment area, which are collected every day or more than 5 days. The missing values for precipitation time series data were filled by resetting the precipitation recorded during several observation months, resulting in wrong runoff peak times. The results showed that the introduced approach had a suitable performance in runoff simulating for both minimum and maximum water level conditions. The results prove that it is not a residual analysis of white noise, indicating that the model does not fully identify the causal relationship between rainfall and runoff (Greco 2012).

According to a review article on using the ANFIS methods to predict runoff, the fuzzy inference system is used because it can handle missing data and complex data that characterize the runoff time series. It is difficult to describe accurately, so an approximation method (fuzzy set) was proposed for obtaining a reasonable result in runoff modeling. In addition, several studies stated that the advantages of ANFIS allowed them to have a high accuracy result for runoff modeling in different time scales.

Application of ANN in runoff simulation

The ANN is a large-scale distributed parallel information analyzing theory, which has some performance characteristics similar to the human brain biological neural network (Haykin 2004). It is inspired by human cognition and neurobiology by a mathematical model, ANNs are technologically advanced, and they can do lots of huge computing in minimum time. The base of ANN's structure follows some rules: (i) exchanging of information occurs in the independent elements by the name of neurons, (ii) signals are transmitted between neurons via the transfer functions, (iii) every transform function corresponds to the weight representing its adhesive strength, and (iv) each neuron usually employees for a non-linear transfer function to its network input for determining the output. Generally, an artificial neural network consists of three parts: (a) an input layer containing multiple input nodes, (b) one or more hidden layers containing trigger functions, and (c) multiple output layer nodes. The current is modeled using forward feedback (FFBP), RBF, and generalized regression neural network (GRNN) algorithms. FFBP is arguably the most widely used ANN for engineering problems regarded as non-linear general approximations (Hornik et al. 1989). Figure 3 shows the ordinary ANN's structure.

A new dynamic ANN method developed by Chiang et al. (2004) simulated rainfall–runoff by the impact of time dimension on the dataset. The proposed method has a profound effect on network learning. They compared the evaluation results of the dynamic ANN with an ordinary static ANN. The proposed method showed a more stable input current prediction and positive performance than static ANN. Furthermore, the repetitive real-time learning algorithm helped for updating the ANN again and again for the training phase, which had advantages when recording time changes in the process of rainfall–runoff modeling (Chiang et al. 2004). Cigizoglu (2005) reported an investigation on the effectiveness of GRNN for daily runoff modeling. Cigizoglu used GRNN as a boosting tool for enhancing ability of ordinary FFBP. GRNN can handle the local minimum problem, and it was a suitable boosting approach for improving the accuracy of runoff prediction. This is because GRNN predictions are limited to the extreme values of the observation, which prevents the training of network for providing predictions that are physically impossible (Cigizoglu 2005).

To study new measures for improving the precision of machine learning-based runoff models, Hu et al. 2005 developed the ANN by the name of target programming neural network for simulation of the streamflow phenomenon and it has a successful result. They did three fundamental improvements: (a) Clearly integrating the previous hydrological knowledge into the training of the neural network; (b) A modification on the objective function of ANN; and (c) Reducing the network's sensitivity to input variables errors (Hu et al. 2005).

Wu et al. (2005) predicted runoff in the river basin by application of a multi-layer neural network. Two models have been developed: (i) four steps ahead or 1 hour ahead (with a resolution interval of 15 min) for streamflow forecasting and (ii) flood forecasting in advance times using upstream station's maximum streamflow data. They used a data set such as the precipitation with seven lag times and the streamflow data with three lag-time values to predict the runoff in four steps (1 h with a resolution of 15 min). However, it is found that the model's accuracy gradually decreases as the number of prediction steps increases. Therefore, the result of one-step prediction is more accurate than the result of two-step prediction. In addition, research showed that the proposed technique effectively solved runoff peak time prediction, especially in predicting flow and water volume in near real time (Wu et al. 2005).

Kişi (2007) used different ANN algorithms to examine short-term daily runoff forecasts. Four different ANN algorithms were applied on the streamflow time series data, namely backpropagation, Levenberg–Marquardt, cascade correlation, and conjugate gradient, algorithms. The results showed that the Levenberg–Marquardt algorithm requires only a small part of the time required (by minimum data requirement) for the other three algorithms to train the ANN, then LM provided a more accurate result for runoff time series prediction (Kişi 2007). Jain and Kumar (2007) developed a new coupled ANN model for having the better training for the ordinary ANN. The proposed method includes a general modeling framework, which was a hybridization

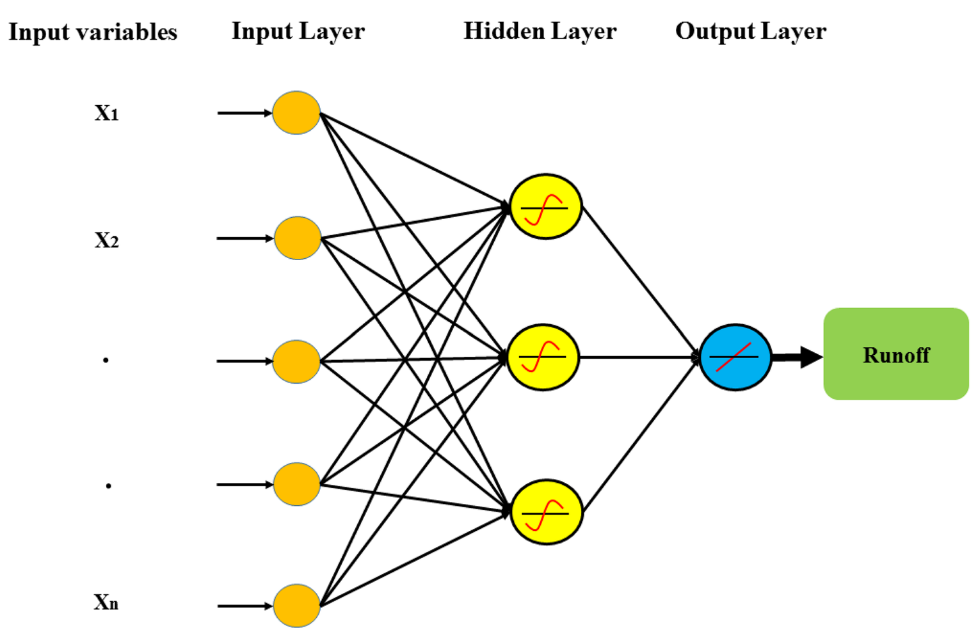


Fig. 3 The structure of the ordinary ANN model

of traditional methods and ANN methods (Jain and Kumar 2007).

Sedighi et al. (2016) employed ANN and SVM for runoff simulating in a snow-covered watershed (in Iran). First, they showed the machine learning model can be used in snow-covered regions by acceptable accuracy for runoff modeling. Second, they resulted in the best condition ANN simulated runoff by the coefficient of determination equal 0.77 for validation section (Sedighi et al. 2016). Another study was provided by Toth and Brath (2007) on the runoff real-time prediction capabilities by ANN models. Results of two runoff real-time simulations showed the yield can be achieved by increasing the lead time and analyzing the impact of the modeling calibration process. The results show that if there is a large amount of hydrometeorological data available for analysis, the neural network has proven to be an excellent approach for rainfall–runoff simulating in a continuous period of time (including low, medium and peak runoff). Compared with data-driven methods that focus on flood forecasting, conceptual formulas can significantly improve forecasting, especially when the availability of calibration data is limited (Toth and Brath 2007).

Mutlu et al. (2008) employed two different types of ANN models, namely MLP and RBF, to predict the streamflow of four different stations. Different lag times were considered as input of models and compared based on their ability to predict river flow. These models performed satisfactorily in predicting the streamflow of several discharge stations. However, the MLP model is better than the RBF model (Mutlu et al. 2008). Kagoda et al. (2010) used RBF in 2010 to generate a one-day runoff forecast. Because some river basins may not always have the data needed to apply many complex machine learning models successfully. Researchers have shown that depending on the situation, RBF can more accurately predict the time curve area by selecting the objective function; for example, when predicting small currents is important. The results show that artificial neural networks can do a lot in predicting rivers (Kagoda et al. 2010).

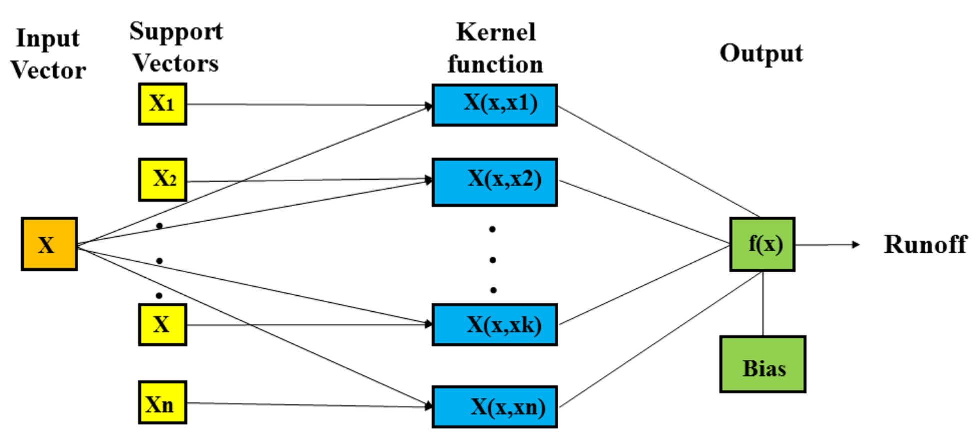
The ANN model and its implementation in river prediction are summarized by the literature review mentioned above. ANNs have some obvious shortcomings and limitations, such as local minimums, learning rates process, over-fitting problems, and trivial manual interventions such as learning. However, the researchers by considering some ANN settings can fix all mentioned issues and have a high accuracy in the runoff modeling process.

Application of SVM in runoff simulation

Recently, many researchers have explored the ability of SVM in the runoff modeling process. Dibike et al. (2001) used the SVM for rainfall–runoff simulation, they used the daily rainfall, evapotranspiration, and streamflow data from three different catchments with different precipitation rates to obtain appropriate data formats for SVM and ANN. Three kinds of kernel functions are used, namely polynomial kernel, RBF kernel, and neural network kernel. In the defect detection process, the core parameters such as the parameter ε and the capacity factor C corresponding to the defect dead zones are set to the optimal value. During the review period, using the average SVM method, the accuracy of runoff estimation was 15% higher than that of the ANN model. In short, they emphasized the difficulty in determining the optimal value of the parameter C, called it a "heuristic process", and suggested automating this process (Dibike et al. 2001). Figure 4 shows the structure of the ordinary SVM model.

Bray and Han (2004) emphasized the use of SVM to determine the appropriate model structure and related parameters to simulate runoff. Their training and testing data were compiled using rainfall and river flow datasets from the Bird Creek catchment area. They used scaling factors for precipitation and streamflow dataset, due the different units and values in the used data. They provided a flowchart for the model selection and modification of LIBSVM software to study the relationship between different model structures, kernel functions (linear, polynomial, radial, and sigmoidal), scaling factors, model parameters (C and epsilon), and the

Fig. 4 The structure of the ordinary SVM model



composition of the input vector (Bray and Han 2004). The SVM was demonstrated for statistically reducing error of rainfall–runoff modeling on the various time scales. This method has been used by the Indian Meteorological Department (IMD) and tested its effectiveness. The SVM-based seasonal downscaling (SD) model of high precipitation is developed for each IMD, using principal components extracted from predictors as input, and simultaneous observation of precipitation in IMD as output. The performance of SD is better than the traditional reduction model (Tripathi et al. 2006). Then, SVM-based SD is used to derive IMD's future precipitation forecast, which uses the second-generation coupled global climate model (CGCM2) for statistical downscaling of artificial neural networks for climate impact researches. They concluded that SVMs were ideal for downscaling problems because they have good generalization performance in capturing the non-linear regression relationship between measured values and predicted values, even though SVMs do not have any physically understanding about the hydrological phenomenon. Researchers have been developing many methods to simulate and predict the streamflow of rivers in different regions. Therefore, it is necessary to determine an appropriate and reliable model for proper planning of water resources management.

Li and Cheng (2014) used SVM, ANN, and ELM for streamflow forecasting in Manwan Reservoir (in Yunnan Province of China) and Hongjiadu Reservoir (in Guizhou Province of China). They proved all three machine learning approaches had suitable performance for streamflow forecasting, and SVM simulated streamflow by correlation of 0.917 in validation phase. Also, they resulted machine learning approaches by coupling with wavelet transform can have better streamflow simulation (Li and Cheng 2014). Asefa et al. 2006 employed the SVM method to perform seasonal and hourly predictions of streamflow on several scales. The results showed a successful ability for the SVM model for modeling water management problems. The SVM's considered input was much less than the physical-based model. In addition, the seasonal streamflow forecasting had been improving by including meteorological variables as input of models (Asefa et al. 2006).

As previous studies showed, the fluctuation of the atmosphere and ocean will affect the variability of rivers. Therefore, Carrier et al. (2013) proposed a long-term traffic forecast using a data-driven kernel-based multi-class model. This study uses instruments and reconstructed waveform data in SVM. The novelty is that it improves the delay of flow prediction (Carrier et al. 2013). The SVM model can make suitable predictions for selected instruments within a lead time of 1–5 years. Compared to using a single swing, the use of a swing index helps to achieve higher predictability.

He et al. (2014) used three different types of ML-based approaches by the name of ANFIS, ANN, and SVM for streamflow modeling in a semi-arid climate. The model examines the various combinations of the lag times in streamflow time series data and selects the most suitable input variables for the modeling process via ML approaches. The result of evaluation on performance metrics showed that the SVM model was superior in comparing to the ANN and ANFIS models in predicting streamflow in semi-arid areas. Evaluation of the various documents on the SVM model led to several observations (He et al. 2014):

(i) One of the abilities of the machine learning approaches is that in addition to the mean square error of the training samples, it also minimizes the generalized error of the model. (ii) According to Mercer's hypothesis, the corresponding optimization problem is like a bulging (convex), so there is no local minimum. (iii) A large number of researchers reported that the RBF is the most suitable kernel function. The reasons are as follows: First, the adjustment parameters of RBF kernel are less than the sigmoid and polynomial kernels, which increases the complexity of model selection. Capture the situation where the relationship between class labels and attributes is non-linear rather than linear kernel. Third, the RBF kernel usually works well under the general smoothing assumption. (iv) Generally, SVM is more suitable for long-term runoff simulation than short-term runoff simulation. This shows the SVM approach potency and possibility to define hydrological time series analysis with the non-linear factors.

Conclusion

Streamflow simulation is essential for hydrological studies, irrigation management, environmental sustainability, water resources planning and management. Due to the dynamic behavior of streamflow and its interaction with other hydrological variables, streamflow modeling process needs a model that can understand these nonlinear complexities well. So, researchers have been focusing on developing the model that can overcome the complexities of the hydrological cycle (like ML approaches). Then, this study tried to analyze the applications of machine learning for runoff modeling based on literature reviews. The ML method is presented as a powerful tool to provide evidence for runoff modeling in different regions with high accuracy. This study evaluated literature reviews on the application of ANFIS, ANN, and SVM for the runoff time series forecasting on under different climate effects. Other available climatic variables (i.e., precipitation) and the lag times (delay) of runoff time series were used as the inputs of predictors models. Another purpose of this study was to consider ordinary ML models used in different climate conditions for runoff simulating, for finding possible alternatives for runoff modeling in various climates. The modeling process via ML has a huge impact on various factors that affect modeling performance. One of the most important of them is that determining effective input parameters as the key elements to achieve optimal performance of ML models. In addition, the reviewed articles also showed an overview of optimization algorithms that are combined with ML models to form hybrid models with high accuracy. This study recommends that future potential researchers use the newly developed optimization algorithms for optimizing the ability of ML models. Several examples in this study demonstrated the prediction, classification, and regression capabilities of ML related to runoff modeling problems. These examples also showed that the non-linear nature of ML should be used with caution, as this can lead to over-fitting problems. The results of the literature reviewed here indicate that ML has many uses in computational hydrology (especially in runoff forecasting). Future researchers can conduct research based on this framework to develop some new hybrid mechanisms and extend machine learning technology to overcome the complexity of hydrological predictions. The machine learning model can provide higher accuracy prediction for runoff simulation, and making ML as an efficient tool for water resources management. Future potential researchers can use hybrid-based models via hydrological and ML models for using advantages of physical-based and ML-based models for runoff simulation studies.

Funding Open access funding provided by Lund University.

References

Abbot J, Marohasy J (2012) Application of artificial neural networks to rainfall forecasting in Queensland, Australia. Adv Atmos Sci. https://doi.org/10.1007/s00376-012-1259-9

Abdollahi S, Raeisi J, Khalilianpour M et al (2017) Daily mean streamflow prediction in perennial and non-perennial rivers using four data driven techniques. Water Resour Manag. https://doi.org/10.1007/s11269-017-1782-7

Abdulelah Al-Sudani Z, Salih SQ, Sharafati A, Yaseen ZM (2019) Development of multivariate adaptive regression spline integrated with differential evolution model for streamflow simulation. J Hydrol. https://doi.org/10.1016/j.jhydrol.2019.03.004

Abudu S, Cui CL, King JP, Abudukadeer K (2010) Comparison of performance of statistical models in forecasting monthly streamflow of Kizil River, China. Water Sci Eng. https://doi.org/10.3882/j.issn.1674-2370.2010.03.003

Adamowski J, Chan HF, Prasher SO, Sharda VN (2012) Comparison of multivariate adaptive regression splines with coupled wavelet transform artificial neural networks for runoff forecasting in Himalayan micro-watersheds with limited data. J Hydroinform. https://doi.org/10.2166/hydro.2011.044

Adnan RM, Petroselli A, Heddam S et al (2021) Comparison of different methodologies for rainfall–runoff modeling: machine learning vs conceptual approach. Nat Hazards. https://doi.org/10.1007/s11069-020-04438-2

Alizadeh A, Rajabi A, Shabanlou S et al (2021) Modeling long-term rainfall-runoff time series through wavelet-weighted regularization extreme learning machine. Earth Sci Inform. https://doi.org/10.1007/s12145-021-00603-8

Asefa T, Kemblowski M, McKee M, Khalil A (2006) Multi-time scale stream flow predictions: the support vector machines approach. J Hydrol. https://doi.org/10.1016/j.jhydrol.2005.06.001

Bray M, Han D (2004) Identification of support vector machines for runoff modelling. J Hydroinform 6:265–280

Carrier C, Kalra A, Ahmad S (2013) Using paleo reconstructions to improve streamflow forecast lead time in the western United States. J Am Water Resour Assoc. https://doi.org/10.1111/jawr.12088

Chadalawada J, Herath HMVV, Babovic V (2020) Hydrologically informed machine learning for rainfall-runoff modeling: a genetic programming-based toolkit for automatic model induction. Water Resour Res. https://doi.org/10.1029/2019WR026933

Chang FJ, Chen YC (2001) A counterpropagation fuzzy-neural network modeling approach to real time streamflow prediction. J Hydrol. https://doi.org/10.1016/S0022-1694(01)00350-X

Chiang YM, Chang LC, Chang FJ (2004) Comparison of static-feedforward and dynamic-feedback neural networks for rainfall-runoff modeling. J Hydrol. https://doi.org/10.1016/j.jhydrol.2003.12.033

Cigizoglu HK (2005) Application of generalized regression neural networks to intermittent flow forecasting and estimation. J Hydrol Eng. https://doi.org/10.1061/(asce)1084-0699(2005)10:4(336)

Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn. https://doi.org/10.1023/A:1022627411411

Dibike YB, Velickov S, Solomatine D, Abbott MB (2001) Model induction with support vector machines: introduction and applications. J Comput Civ Eng. https://doi.org/10.1061/(asce)0887-3801(2001)15:3(208)

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. https://doi.org/10.1007/s11269-006-9027-1

Greco R (2012) A fuzzy-autoregressive model of daily river flows. Comput Geosci. https://doi.org/10.1016/j.cageo.2012.02.031

Hadi SJ, Tombul M (2018) Monthly streamflow forecasting using continuous wavelet and multi-gene genetic programming combination. J Hydrol. https://doi.org/10.1016/j.jhydrol.2018.04.036

Haykin S (2004) A comprehensive foundation. Neural Netw 2(2004):41

He Z, Wen X, Liu H, Du J (2014) A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. J Hydrol. https://doi.org/10.1016/j.jhydrol.2013.11.054

Hornik K, Stinchcombe M, White H (1989) Multilayer feedforward networks are universal approximators. Neural Netw. https://doi.org/10.1016/0893-6080(89)90020-8

Hu TS, Lam KC, Ng ST (2005) A modified neural network for improving river flow prediction. Hydrol Sci J. https://doi.org/10.1623/hysj.50.2.299.60649

Humphrey GB, Gibbs MS, Dandy GC, Maier HR (2016) A hybrid approach to monthly streamflow forecasting: integrating hydrological model outputs into a Bayesian artificial neural network. J Hydrol. https://doi.org/10.1016/j.jhydrol.2016.06.026

Jain A, Kumar AM (2007) Hybrid neural network models for hydrologic time series forecasting. Appl Soft Comput J. https://doi.org/10.1016/j.asoc.2006.03.002

Jang JSR (1993) ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans Syst Man Cybern 23:665–685. https://doi.org/10.1109/21.256541

Jothiprakash V, Magar RB (2012) Multi-time-step ahead daily and hourly intermittent reservoir inflow prediction by artificial intelligent techniques using lumped and distributed data. J Hydrol. https://doi.org/10.1016/j.jhydrol.2012.04.045

Kagoda PA, Ndiritu J, Ntuli C, Mwaka B (2010) Application of radial basis function neural networks to short-term streamflow forecasting. Phys Chem Earth. https://doi.org/10.1016/j.pce.2010.07.021

Katambara Z, Ndiritu JG (2010) A hybrid conceptual-fuzzy inference streamflow modelling for the Letaba River system in South Africa. Phys Chem Earth. https://doi.org/10.1016/j.pce.2010.07.032

Kentel E (2009) Estimation of river flow by artificial neural networks and identification of input vectors susceptible to producing unreliable flow estimates. J Hydrol. https://doi.org/10.1016/j.jhydrol.2009.06.051

Kişi Ö (2007) Streamflow forecasting using different artificial neural network algorithms. J Hydrol Eng. https://doi.org/10.1061/(asce)1084-0699(2007)12:5(532)

Kisi O, Nia AM, Gosheh MG et al (2012) Intermittent streamflow forecasting by using several data driven techniques. Water Resour Manag. https://doi.org/10.1007/s11269-011-9926-7

Kreinovich V, Nguyen HT, Yam Y (2000) Fuzzy systems are universal approximators for a smooth function and its derivatives. Int J Intell Syst. https://doi.org/10.1002/(SICI)1098-111X(200006)15:6%3c565::AID-INT6%3e3.0.CO;2-0

Li BJ, Cheng CT (2014) Monthly discharge forecasting using wavelet neural networks with extreme learning machine. Sci China Technol Sci. https://doi.org/10.1007/s11431-014-5712-0

Liu Z, Zhou P, Chen G, Guo L (2014) Evaluating a coupled discrete wavelet transform and support vector regression for daily and monthly streamflow forecasting. J Hydrol. https://doi.org/10.1016/j.jhydrol.2014.06.050

Mohammadi B, Mehdizadeh S (2020) Modeling daily reference evapotranspiration via a novel approach based on support vector regression coupled with whale optimization algorithm. Agric Water Manag. https://doi.org/10.1016/j.agwat.2020.106145

Mohammadi B, Ahmadi F, Mehdizadeh S et al (2020a) Developing novel robust models to improve the accuracy of daily streamflow modeling. Water Resour Manag. https://doi.org/10.1007/s11269-020-02619-z

Mohammadi B, Guan Y, Aghelpour P et al (2020b) Simulation of Titicaca lake water level fluctuations using hybrid machine learning technique integrated with grey wolf optimizer algorithm. Water. https://doi.org/10.3390/w12113015

Mohammadi B, Linh NTT, Pham QB et al (2020c) Adaptive neuro-fuzzy inference system coupled with shuffled frog leaping algorithm for predicting river streamflow time series. Hydrol Sci J. https://doi.org/10.1080/02626667.2020.1758703

Mohammadi B, Guan Y, Moazenzadeh R, Safari MJS (2021a) Implementation of hybrid particle swarm optimization-differential evolution algorithms coupled with multi-layer perceptron for suspended sediment load estimation. CATENA. https://doi.org/10.1016/j.catena.2020.105024

Mohammadi B, Moazenzadeh R, Christian K, Duan Z (2021b) Improving streamflow simulation by combining hydrological process-driven and artificial intelligence-based models. Environ Sci Pollut Res. https://doi.org/10.1007/s11356-021-15563-1

Mutlu E, Chaubey I, Hexmoor H, Bajwa SG (2008) Comparison of artificial neural network models for hydrologic predictions at multiple gauging stations in an agricultural watershed. Hydrol Process. https://doi.org/10.1002/hyp.7136

Nayak PC, Sudheer KP, Jain SK (2007) Rainfall-runoff modeling through hybrid intelligent system. Water Resour Res. https://doi.org/10.1029/2006WR004930

Niu WJ, Feng ZK, Zeng M et al (2019) Forecasting reservoir monthly runoff via ensemble empirical mode decomposition and extreme learning machine optimized by an improved gravitational search algorithm. Appl Soft Comput J. https://doi.org/10.1016/j.asoc.2019.105589

Nourani V, Komasi M, Mano A (2009) A multivariate ANN-wavelet approach for rainfall-runoff modeling. Water Resour Manag. https://doi.org/10.1007/s11269-009-9414-5

Nourani V, Kisi Ö, Komasi M (2011) Two hybrid artificial intelligence approaches for modeling rainfall-runoff process. J Hydrol. https://doi.org/10.1016/j.jhydrol.2011.03.002

Nourani V, Gökçekuş H, Gichamo T (2021) Ensemble data-driven rainfall-runoff modeling using multi-source satellite and gauge rainfall data input fusion. Earth Sci Inform. https://doi.org/10.1007/s12145-021-00615-4

Okkan U, Ersoy ZB, Ali Kumanlioglu A, Fistikoglu O (2021) Embedding machine learning techniques into a conceptual model to improve monthly runoff simulation: a nested hybrid rainfall-runoff modeling. J Hydrol. https://doi.org/10.1016/j.jhydrol.2021.126433

Oppel H, Schumann AH (2020) Machine learning based identification of dominant controls on runoff dynamics. Hydrol Process. https://doi.org/10.1002/hyp.13740

Özger M (2009) Comparison of fuzzy inference systems for streamflow prediction. Hydrol Sci J. https://doi.org/10.1623/hysj.54.2.261

Parisouj P, Mohebzadeh H, Lee T (2020) Employing machine learning algorithms for streamflow prediction: a case study of four river basins with different climatic zones in the United States. Water Resour Manag. https://doi.org/10.1007/s11269-020-02659-5

Parvinizadeh S, Zakermoshfegh M, Shakiba M (2021) A simple and efficient rainfall–runoff model based on supervised brain emotional learning. Neural Comput Appl. https://doi.org/10.1007/s00521-021-06475-9

Pramanik N, Panda RK (2009) Application of neural network and adaptive neuro-fuzzy inference systems for river flow prediction. Hydrol Sci J. https://doi.org/10.1623/hysj.54.2.247

Qu J, Ren K, Shi X (2021) Binary grey wolf optimization-regularized extreme learning machine wrapper coupled with the boruta algorithm for monthly streamflow forecasting. Water Resour Manag. https://doi.org/10.1007/s11269-021-02770-1

Safari MJS, RahimzadehArashloo S, DanandehMehr A (2020) Rainfall-runoff modeling through regression in the reproducing kernel Hilbert space algorithm. J Hydrol. https://doi.org/10.1016/j.jhydrol.2020.125014

Sang YF (2013) A review on the applications of wavelet transform in hydrology time series analysis. Atmos Res 122:8

Sanikhani H, Kisi O (2012) River flow estimation and forecasting by using two different adaptive neuro-fuzzy approaches. Water Resour Manag. https://doi.org/10.1007/s11269-012-9982-7

Sedighi F, Vafakhah M, Javadi MR (2016) Rainfall-runoff modeling using support vector machine in snow-affected watershed. Arab J Sci Eng. https://doi.org/10.1007/s13369-016-2095-5

Sharafati A, Khazaei MR, Nashwan MS et al (2020) Assessing the uncertainty associated with flood features due to variability of rainfall and hydrological parameters. Adv Civ Eng. https://doi.org/10.1155/2020/7948902

Siddiqi TA, Ashraf S, Khan SA, Iqbal MJ (2021) Estimation of data-driven streamflow predicting models using machine learning methods. Arab J Geosci. https://doi.org/10.1007/s12517-021-07446-z

Siqueira H, Boccato L, Luna I et al (2018) Performance analysis of unorganized machines in streamflow forecasting of Brazilian plants. Appl Soft Comput J. https://doi.org/10.1016/j.asoc.2018.04.007

Terzi Ö, Ergin G (2014) Forecasting of monthly river flow with autoregressive modeling and data-driven techniques. Neural Comput Appl. https://doi.org/10.1007/s00521-013-1469-9

Tikhamarine Y, Souag-Gamane D, Ahmed AN et al (2020a) Rainfall-runoff modelling using improved machine learning methods: Harris hawks optimizer vs particle swarm optimization. J Hydrol. https://doi.org/10.1016/j.jhydrol.2020.125133

Tikhamarine Y, Souag-Gamane D, Najah Ahmed A et al (2020b) Improving artificial intelligence models accuracy for monthly streamflow forecasting using grey Wolf optimization (GWO) algorithm. J Hydrol. https://doi.org/10.1016/j.jhydrol.2019.124435

Tongal H, Booij MJ (2018) Simulation and forecasting of streamflows using machine learning models coupled with base flow separation. J Hydrol. https://doi.org/10.1016/j.jhydrol.2018.07.004

Toth E, Brath A (2007) Multistep ahead streamflow forecasting: role of calibration data in conceptual and neural network modeling. Water Resour Res. https://doi.org/10.1029/2006WR005383

Tripathi S, Srinivas VV, Nanjundiah RS (2006) Downscaling of precipitation for climate change scenarios: a support vector machine approach. J Hydrol. https://doi.org/10.1016/j.jhydrol.2006.04.030

Tripathy A, Schwefel H-P (1982) Numerical optimization of computer models. J Oper Res Soc. https://doi.org/10.2307/2581158

Uysal G, Şorman AA, Şensoy A (2016) Streamflow forecasting using different neural network models with satellite data for a snow dominated region in Turkey. Procedia Eng 154:1185

Valipour M, Montazar AA (2012a) Optimize of all effective infiltration parameters in furrow irrigation using visual basic and genetic algorithm programming. Aust J Basic Appl Sci 6:132

Valipour M, Montazar AA (2012b) Sensitive analysis of optimized infiltration parameters in SWDC model. Adv Environ Biol 6:2574

Wedding DK (1997) Fuzzy sets and fuzzy logic: theory and applications. Neurocomputing. https://doi.org/10.1016/s0925-2312(97)88327-0

Wen X, Feng Q, Deo RC et al (2019) Two-phase extreme learning machines integrated with the complete ensemble empirical mode decomposition with adaptive noise algorithm for multi-scale runoff prediction problems. J Hydrol. https://doi.org/10.1016/j.jhydrol.2018.12.060

Wu JS, Han J, Annambhotla S, Bryant S (2005) Artificial neural networks for forecasting watershed runoff and stream flows. J Hydrol Eng. https://doi.org/10.1061/(asce)1084-0699(2005)10:3(216)

Zadeh LA (1965) Fuzzy sets. Inf Control 8:338–353. https://doi.org/10.1016/S0019-9958(65)90241-X

Zhou Y, Guo S, Chang FJ (2019) Explore an evolutionary recurrent ANFIS for modelling multi-step-ahead flood forecasts. J Hydrol. https://doi.org/10.1016/j.jhydrol.2018.12.040