



Ensemble data-driven rainfall-runoff modeling using multi-source satellite and gauge rainfall data input fusion

Vahid Nourani^{1,2} · Hüseyin Gökçekuş² · Tagesse Gichamo²

Received: 8 March 2021 / Accepted: 9 April 2021 / Published online: 29 April 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Feed Forward Neural Network (FFNN), Adaptive Neuro-fuzzy Inference System (ANFIS), and Support Vector Regression (SVR) were applied for rainfall-runoff modeling of the Gilgel Abay catchment, Blue Nile basin, Ethiopia. Daily precipitations from satellite sources and rain gauge stations and outlet discharge were used. The dominant inputs were selected by non-linear sensitivity analysis. The study was conducted in two stages. First, single models for each data source with input fusion were trained. Second, ensemble runoff modeling using rainfall data fusion from only satellite products (strategy 1) and satellite and gauge (strategy 2) was conducted by Simple Average (SA), Weighted Average (WA), and Neural Network Ensemble (NNE) methods. NNE method using input fusion of strategy 2 improved performance of the best single satellite model up to 14.5% and a single gauge model up to 8% in the validation. Strategy 2 input data fusion ensemble rainfall-runoff modeling indicated substantial improvement over satellite data-based runoff modeling. This could be due to the bias correction ability of gauge rainfall over satellite rainfall products. Overall, results showed that ensemble modeling of input fusion from multiple source satellite rainfall products is a promising option for accurate modeling of the rainfall-runoff process for ungauged or sparsely gauged catchments.

Keywords Rainfall-runoff · Satellite-rainfall · Gauge-rainfall · Artificial-intelligence · Ensemble · Gilgel-Abay

Introduction

Rainfall-runoff modeling is an important task in water resource optimization and planning activities such as flood control, river basin engineering, navigation, irrigation water management, and reservoir operation (Guimarães Santos and da Silva 2014; Noori and Kalin 2016). It also has remarkable importance in preventing and early warning of natural disasters such as drought and flood and incident mitigation in extreme cases (Shamseldin 2010).

However, rainfall-runoff process modeling is a difficult hydrologic task because of spatial and temporal dynamics of

the process with complex non-linear characteristics, chaotic disturbances, and exhibiting randomness (Singh and Sankarasubramanian 2014). A wide range of approaches such as data-driven (black-box), physically-based, and conceptual models have been already developed and applied for rainfall-runoff modeling (Shamseldin 2006). Statistical classic methods such as Autoregressive moving average (ARIMA) models are simple to use but they usually create a linear input-output relationship which may have limitations in modeling non-linear and non-stationary hydrological processes (Nourani et al. 2020). On the other hand, physically-based models such as the Soil and Water Assessment Tool (SWAT) need large size spatial and temporal hydrological data and their calibration and validation take a long time that may make them difficult to be used (Makwana and Tiwari 2014). Data-driven Artificial Intelligence (AI) models such as; Artificial Neural Network (ANN), Adaptive Neuro-fuzzy System (ANFIS), and Support Vector Regression (SVR) are black box models that can accurately model non-stationary and non-linear behavior of the hydrological processes (Gazzaz et al. 2012). Data-driven models are trained and tested for specified data and limited locations. In mountainous areas,

Communicated by: H. Babaie

✉ Vahid Nourani
vnourani@yahoo.com

¹ Faculty of Civil Engineering and Center of Excellence in Hydroinformatics, University of Tabriz, Tabriz, Iran

² Faculty of Civil and Environmental Engineering, Near East University, Near East Blv, Nicosia, Cyprus

commonly used hydrological forecasting models cannot accurately predict streamflow because of less-density rain gauge distribution. In such cases, data-driven AI techniques can accurately predict flow using cross-station or single station streamflow data (A. Danandeh Mehr et al. 2015). ANN is enthused by the studies into the biological neural networks, has a rapid, supple arrangement, self-learning, self-adaptive characteristics without the requirement for the complicated feature of fundamental progressions considered to be clearly defined in the mathematical relationships.

The application of ANN as a commonly used AI method in hydrological modeling has shown its ability to detect the complicated non-linear relation between hydrological time series, nevertheless, the model structure and parameters may not characterize the physical processes of the basin (Govindaraju 2000). ANN became popular hydrological time series forecasting tool and particularly, it has been successfully applied for runoff forecasting (e.g. see, Shamseldin 2010; Taormina and Chau 2015a). The most important strength of ANN is data handling ability, such as learning, noise tolerance, and data generalization. However, ANN application for rainfall-runoff modeling has still some limitations. For instance, (Wu et al. 2009) indicated that data noise existed in the rainfall and flow time-series could significantly affect forecasting quality. Moreover, overtraining and data quality are still problems in modeling by ANN. The aforementioned weaknesses of the ANN-based runoff modeling may be corrected by pre-processing of data, via hybrid and, ensemble approaches.

Adaptive Neuro-Inference System (ANFIS) is a hybrid and combination of the learning capability of ANN and fuzzy-logic introduced by (Jang 1993). ANFIS has proved its effectiveness in capturing the merits of both ANN and fuzzy logic methods in a particular structure (Chang et al. 2015). Numerous studies applied ANFIS to model rainfall-runoff processes (Yaseen et al. 2017) but again it shows some deficiencies in the real-world application as for ANN.

Another almost new AI model is Support Vector Regression (SVR). SVR is a non-linear regression model developed based on Support Regression Machine (SVM) with the fundamental concept of having the ability to map data with higher dimensionality using a non-linear mapping technique. SVR contemplates operational risk as the objective function to minimize the risks in place of reducing the error between measured and simulated values (Wen et al. 2015). In the last 10 years, SVR got some priorities over other AI models because of its self-learning characteristics, parallel distributed processing, avoiding over-fitting issues, and providing globally optimum results (Kaltch 2013). The main drawback of modeling via SVR is its complex computing processes for the constrained optimization issues where such disadvantages can be handled by applying a Least Square Support Vector Regression (LSSVR) algorithm that uses linear methods in

place of quadratic equations (Wang and Hu 2005). Similar to ANN and ANFIS, the SVR model has been also successfully applied for rainfall-runoff modeling (e.g. see, Ateeq-ur-Rauf et al. 2018; Kaltch 2013).

Even if such non-linear AI techniques (ANN, ANFIS, and SVR) could lead to reliable results for rainfall-runoff modeling, it is apparent that for the specific problem, different models may provide different outcomes. Therefore, combining outputs of different models by ensemble modeling would provide better efficiency of modeling by minimizing error variance compared to the individual methods (see Shamseldin and Connor 1999; Sharghi et al. 2018). Ensemble modeling captures unique features of each model and rainfall dataset thus it could improve the overall efficiency of the modeling (Homsy et al. 2020; Taormina and Chau 2015b).

The most important input in any rainfall-runoff modeling is precipitation data. Precipitation data could be derived from either densely distributed rain gauges over the basin, fairly located ground-based weather radar, and satellite sources (Prakash et al. 2018). In mountainous areas, providing accurate and reliable precipitation data is very difficult due to the less spatial coverage of rain gauge stations and orographic effects (Chen et al. 2018). Also in less developed countries, such as Ethiopia, the spatial resolution of precipitation data is usually poor because rain gauge stations are sparsely distributed and there is no ground-based weather radar due to the lack of adequate finance allocated for the meteorology research sectors. In ungagged or sparsely gauged catchments, hydrological modeling using ground-based precipitation data may not be accurate because of the unrealistic area representation of the gauge rainfall data and its associated temporal and spatial variability (Gao et al. 2017). Evenly distributed ground-based precipitation measurements for lower influence areas can best estimate precipitation data, however, some uncertainties may happen when the point rainfall is interpolated or extrapolated and applied for the large influence areas. To reduce the limitations of data acquiring from ground-based data sources, in the past few decades, precipitation values estimated from various satellite sources have been under wide range use in the regions where ground-based measurement is not available or sparsely located (Ebert et al. 2007). Recently, satellite estimated precipitation data have been widely verified as reliable, cheap, and uninterrupted data sources, particularly for areas with a lack of ground-based meteorological station accessibility (Collins et al. 2013). Moreover, the spatial coverage and temporal resolution of such data are being increased due to the advancements in radars and low orbit satellites for precipitation measurement. Spatially and temporally high-resolution satellite rainfall products are reliable inputs for hydrological modeling in areas where ground-based precipitation recording stations are unreliable or it is not periodically accessible (Gebremichael et al. 2014).

Numerous satellites have been launched for precipitation measurement, for example, Tropical Rainfall Measuring Mission (TRMM) was launched in 1997, Global Precipitation Measurement (GPM) Core Observatory was launched in 2014 which measures near-real-time precipitation and snowfall (Yong et al. 2015). The Climate Prediction Center (CPC) morphing technique (CMORPH) product was launched in 1998 and used to measure rainfall as a near-real-time rainfall product (Gebremichael et al. 2014; Joyce et al. 2004). TRMM data are available in both real-time (3B42RT) and post real-time (3B42) forms. The TRMM Multi-Satellite Precipitation (TMPA) is satellite-based precipitation from multiple satellite sources, combining relative advantages from satellites, providing more reliable and accurate gridded precipitation (Prakash et al. 2018). TRMM is ideal for tropical rainfall observation because it has suitable complementary observation devices and its orbital positioning which is positioned at a low altitude with an appropriate inclination angle that enables more frequent and more spatially comprehensive data acquisition. CMORPH retrieves higher temporal and spatial resolution rainfall data from more accurate passive microwave sensors (Ayehu et al. 2018). Even though the satellite rainfall data set is an appropriate material for hydrological modeling of ungagged catchments and each satellite source has its advantages, the spatial and temporal reliabilities of the data are highly influenced by atmospheric and topographic factors (Tang and Hossain 2012). Therefore, the fusion of rainfall data from multi-satellite sources as input ensemble may lead to a better outcome in rainfall-runoff modeling so that via the calibration step, the model would capture higher weight for the better satellite data.

This study aimed at ensemble rainfall-runoff modeling using multiple source satellite and ground gauge rainfall data sets for Gilgel-Abay, Ethiopia, using FFNN, ANFIS, and SVR models. To the best of the authors' knowledge, this is the first study that ensembles simultaneously both input data sources (gauge and satellite) and AI-based outputs to enhance the rainfall-runoff modeling and utilized input fusion strategy for bias correction of satellite rainfall products.

Gilgel-Abay, the study area of this research, is one of the important sub-catchments of the Ethiopian part of the Blue Nile river, which contributes a large proportion of the flow into the Nile River and it is very vital for hydrological, and environmental sustainability, and social and economic support of millions of peoples living in the riparian countries. In the study area, ground-based rain gauges are very sparse in space, short, and irregular in time. Moreover, the topographic and terrain condition of the area is highly variable from high to low land that may expose the data for orographic effects and cause bias and incorrect representation of the rainfall values (Gebre 2015).

Materials and methods

Proposed methodology

In this study, non-linear sensitivity analysis was applied to identify the most appropriate inputs for rainfall-runoff modeling in two steps (Fig. 1). In the first step, FFNN, ANFIS, and SVR models were trained and tested separately using satellite, gauge, and fusion of two precipitation data sets for rainfall-runoff modeling. In the second step, the outputs from the inputs fusion models were imposed into the ensemble unit to predict the future runoff values. Recently, ensemble modeling has been gaining popularity due to its significant strength to improve the accuracy of time series prediction. The main advantages of ensemble modeling as stated by (Sharghi et al. 2018) are described as follows. i) It can enable the researchers to choose an appropriate model for time series forecasting, ii) The real-world problems occasionally show both linear and non-linear features, in this circumstance, neither linear nor non-linear models are effective for time series forecasting because a small error from the linear process can be magnified via a non-linear model whereas a linear model will not be able to handle nonlinearity of a real-world process. Thus, the problem may be handled by taking advantage of all models via an ensemble of different models.

Used black box models

Artificial neural network (ANN)

ANN is an engineering conception of information in the area of AI conceptualized by inheriting human nerve functional structure (Mislan et al. 2015). ANN is a mathematical 'black-box' model containing numerous non-linear artificial neurons, which are operated side by side, that could be created as single or multiple layers. ANN is data processing methods making connections of neurons with each other to build complicated non-linear input-output interactions and it is specifically described by networking topology, testing, or training algorithms, and activation functions (Tongal and Booij 2018). ANN is a mathematical model that was able to determine a non-linear relationship within input and output parameters out of complex partial differential equation applications. The ANN models have been applied to solve very complex real-world problems such as hydrological and meteorological data preprocessing and processing. The major advantage of this model is no requirement for complex physical processes where the processes are simply described by mathematical equations (Venkata Ramana et al. 2013).

ANN provided the substantial methodology for managing noisy, non-linear, and non-stationary data, particularly when not fully understood the fundamental physical relationship, which makes ANN a suitable method for time series data

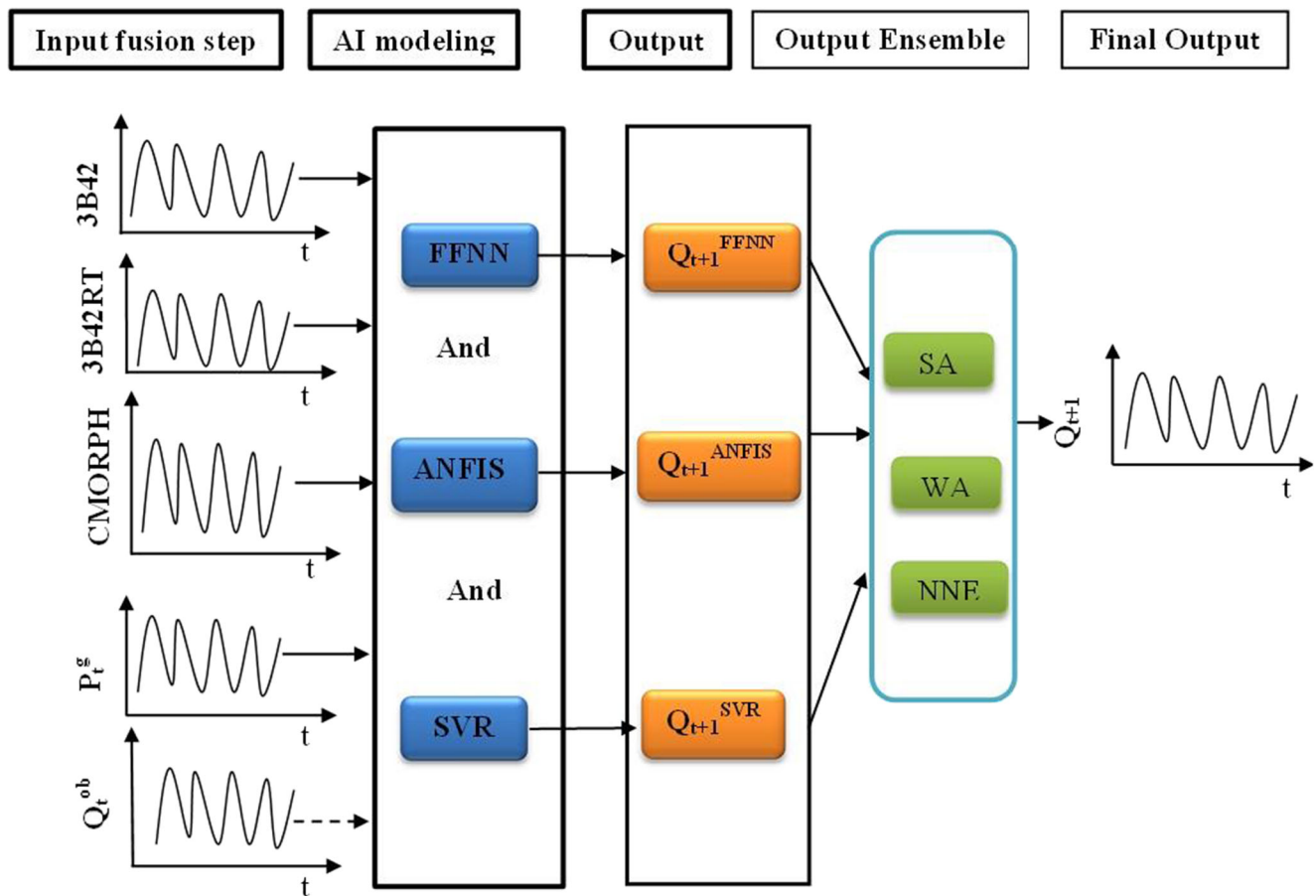


Fig. 1 Schematic of the proposed methodology (P_t^g is Thiessen-based average gauge rainfall and Q_t^{ob} is observed runoff)

forecasting. Mostly known ANN architecture in hydrological and climatological modeling is the multi-layer perceptron (MLP) trained with the backpropagation (BP) algorithm, which includes an input layer, hidden layers, and output layers. There are also extensively used ANN algorithms such as Levenberg-Marquart (LM), Conjugate gradient, Quasi-Newton and Brodyen-Fletcher-Goldfarb-Shanno are the best and efficient algorithms on fast time convergence.

The FFNN trained with Back Propagation (BP) algorithm is the most extensive applied ANN architecture in forecasting several hydrological time series problems and it is also applied in the current study. The FFNN architecture comprises input layers, hidden layers, and output layers, and weights and activation functions (Fig. 2.). The inputs are transformed into output by the following equations.

$$y_1 = f_1 \left[\sum_{k=1}^K w_{1k} f_2 \left(\sum_{j=1}^J w_{kj} x_j + b_k \right) + b_l \right] \quad (1)$$

$$f_2(p) = \frac{2}{1 + e^{-2p}} - 1 \quad (2)$$

where W_{kj} represents weight which connects input and output layers, W_{lk} symbolizes the joining weight between the hidden neuron and output neuron, b_k and b_l stands for the bias of the

corresponding hidden and output layer neurons $f_1(\cdot)$ stands for the linear activation function and $f_2(\cdot)$ denotes the tansigmoidal activation function of the model.

Adaptive Neuro-fuzzy inference system (ANFIS)

Fuzzy Logic (FL) describes computational methods of thought and problem-solving increases the reasoning ability and decision-making ability of human minds (Chandwani et al. 2015). Fuzzy logic has a strong capability of connecting diverse inputs to single output without complex computations, such as normalization, linearization, and homogenization like traditional statistical techniques. The assumption of FL is different from classical models. Classical models assume that the variables have exact numerical values which are related by mathematical functions and output is crisp numbers but in FL, values of variables are linguistically defined, related by If-Then rules and the outputs can be fuzzy subsets then defuzzified to crisp numbers. Modeling by FL takes account of the fuzzification of sets, specifying basic rules, choosing inference techniques, and defuzzification to obtain prediction results.

The adaptive neuro-fuzzy inference system (ANFIS) was first presented by (Jang 1993) to resolve various real-world

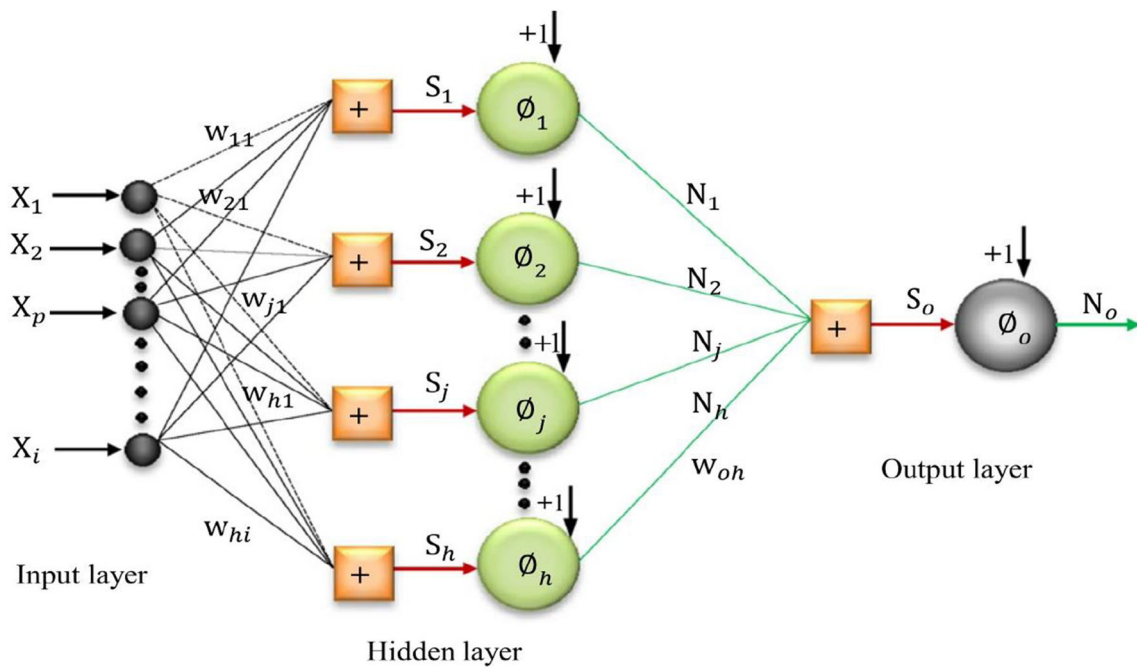


Fig. 2 Typical architecture of three-layered FFNN

problems. ANFIS uses backpropagation gradient descent and least square algorithms that are created by the hybrid-learning algorithm and that can adjust fuzzy membership function parameters by iterative tuning. The main aim behind ANFIS training is to rule the resulting components and optimum premise by training the fuzzy-inference system (FIS) with ANFIS to adjust the membership function parameter to balance with the training database on error selected criterion. ANFIS having the training and testing data, the least square data model is designated which is the parameter linked to the FIS model. ANFIS combination gives a hybrid intelligent system that synergizes fuzzy logic and artificial neural network by conjoining human cognitive ability with neural network and fuzzy logic (Talei et al. 2010) to handle the limitations of the ANN and FIS. ANFIS is a powerful toolbox to model a problem with uncertain and doubtful input data (Moghaddammia et al. 2009) that can handle complexity and noise such as streamflow forecasting and rainfall-runoff modeling. ANFIS is often known as a tool that can universally approximator and which have the capability of approximating any real-world continuous data sets to an acceptable accuracy range. The ANFIS structure is combined of five layers similar to multiple layer FFNN and named based on their functional operation as presented in (Fig. 3). Calibration of ANFIS needs a determination of fuzzy language rules unlike to neural network which tuned weights. The ANFIS membership function calibration is achieved by applying backpropagation and/or least mean square but Takagi Sugeno fuzzy model is calibrated by the conventional least square method. Considering FIS with two inputs and one output as x , y , and f , the Sugeno first-order

fuzzy model used in this study has ideal rule sets which are if-then rules and are specified by:

Rule 1: If $\mu(x)$ is A_1 and $\mu(y)$ is B_1 ; then

$$f_1 = p_1x + q_1y + r_1 \tag{3}$$

Rule 2: If $\mu(x)$ is A_2 and $\mu(y)$ is B_2 ; then

$$f_2 = p_2x + q_2y + r_2 \tag{4}$$

Where A_1 and A_2 are x inputs membership functions, B_1 and B_2 are y input membership functions while the output function parameters are p_1 , q_1 , r_1 and p_2 , q_2 and r_2 a five-layer ANFIS architecture is described as:

Layer 1: Each node i is an adaptive node in this layer with a node function of:

$$Q_i^1 = \mu_{A_i}(x) \text{ for } i = 1, 2 \text{ or } Q_i^1 = \mu_{B_i}(y) \text{ for } i = 3, 4 \tag{5}$$

where Q_i^1 is input and x or y is membership grades.

Layer 2: T-norm operator connecting each rule in this layer between inputs ‘AND’ operator as:

$$Q_i^2 = w_i = \mu_{A_i}(x) \cdot \mu_{B_i}(y) \text{ for } i = 1, 2 \tag{6}$$

Layer 3: “Normalized firing strength” is the output in this layer

$$Q_i^3 = w = \frac{w_i}{w_1 + w_2}, i = 1, 2 \tag{7}$$

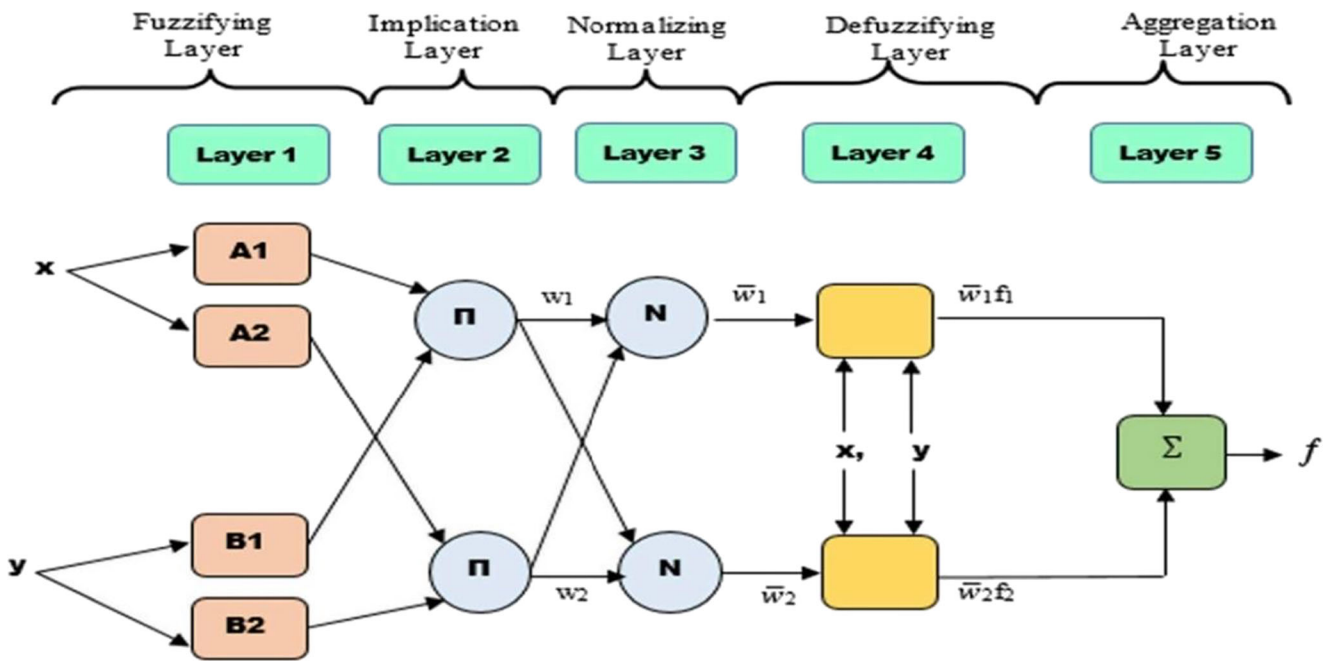


Fig. 3 First-order Sugeno ANFIS and FIS model architecture (Jang 1993)

Layer 4: Each node i in this layer is an adaptive node and achieves the resulting of the rules as:

$$Q_i^4 = w(p_i x + q_i y + r_i) = w f_i \tag{8}$$

w represent the output of layer 3 p_i, q_i and r_i are consequents of parameters.

Layer 5: the overall output of all incoming signals is calculated in this layer:

$$Q_i^5 = w(p_i x + q_i y + r_i) = \sum w_i f_i = \frac{\sum w_i f_i}{\sum w_i} \tag{9}$$

Support vector regression (SVR)

SVR was created based on the Support Vector Machine (SVM) conception, which is used for non-linear regression and classification of the problems (Nourani et al. 2020). In contrary to many other black box predicting approaches, SVR reduces operational risks as an objective function rather than minimizing the error between the actual and predicted parameters. SVR is the type of AI model that is based on a supervised-learning technique with two-layered networks. In the first layers of SVR, weights are non-linear and it is linear in the second layer. In SVR, first, linear regression is created on the data and then the results go through a non-linear kernel to handle the non-linear characteristic of the input data (W. C. Wang et al. 2013). SVR can solve regression problems by applying an alternative loss function, which is modified including distance measure, and the architecture of SVM is given in (Fig. 4.).

Considering the problem of approximation, the set of data $(x_1, y_1), \dots, (x_1, y_1), x \in R^N, y \in R$ with a linear function.

$$f(x, a) = (w * x) + b \tag{10}$$

The ideal regression equation is obtained by minimizing the empirical risk

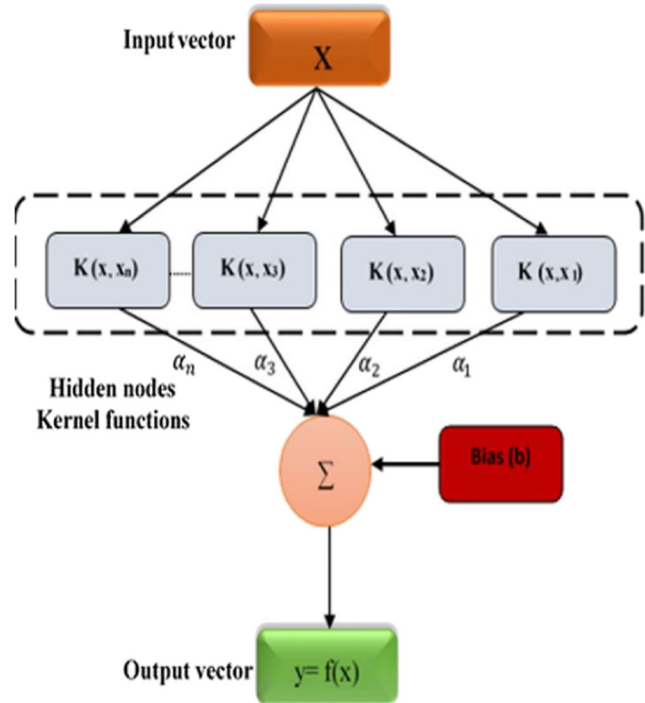


Fig. 4 Structure of the SVM model

$$R_{emp}(w, b) = \frac{1}{l} \sum |y_i - f(x_i, \alpha)|_\varepsilon \tag{11}$$

The most general loss function with ε -insensitive zone explained as

$$|y - f(x, \alpha)| = \begin{cases} \varepsilon, & \text{if } |y - f(x, \alpha)| \leq \varepsilon \\ |y - f(x, \alpha)|, & \text{otherwise} \end{cases} \tag{12}$$

otherwise, the objective is to found a function $f(x, \alpha)$ which has at most ε deviation from the actual observed targets y_i for all the training data and simultaneously as flat as possible. This is equivalent to minimizing functional

$$\varphi(w, \xi^*, \xi) = \|w\|^2/2 + C(\sum \xi_i^* + \sum \xi_i) \tag{13}$$

where C is a pre-defined value and ξ^*, ξ are slack variables representative of upper and lower constraints on the outputs of the system represented in the following equations:

$$\begin{aligned} y_i - (wx_i) + b &\leq \varepsilon \xi_i, i = 1, 2, \dots, l \\ ((wx_i) + b) &\leq \varepsilon + \xi_i^*, i = 1, 2, \dots, l \\ \xi_i^* &\geq 0 \text{ and } \xi_i \geq 0, i = 1, 2, \dots, l \end{aligned} \tag{14}$$

Lagrange function would be formulated from objective function and corresponding constraint by applying a dual set of variables as the following equation:

$$\begin{aligned} L = & \|w\|^2/2 + C \sum (\xi_i + \xi_i^*) \\ & - \sum \alpha_i [\varepsilon + \xi_i - y_i + (wx_i) + b] \\ & - \sum \alpha_i^* [\varepsilon + \xi_i^* - y_i + (wx_i) + b] \\ & \sum (\eta_i \xi_i + \eta_i^* \xi_i^*) \end{aligned} \tag{15}$$

From the saddle point situation, the partial derivatives of L with respect to main variables (w, b, ξ_i^*, ξ_i) have to vanish for ideality. Replacing the result of derivation into the eq. (15) produces dual optimization.

$$W(\alpha^*, \alpha) = -\varepsilon \sum (\alpha_i^* + \alpha_i) + \sum y_i (\alpha_i^* - \alpha_i) - (1/2) \times \sum \sum (\alpha_i^* - \alpha_i) (\alpha_j^* - \alpha_j) (x_i x_j) \tag{16}$$

which has to be maximized subject to constraints

$$\begin{aligned} \sum \alpha_i^* &= \sum \alpha_i, 0 \leq C, \text{ and} \\ 0 &\leq \alpha_i \leq C \text{ for } i = 1, 2, \dots, l \end{aligned} \tag{17}$$

After the coefficients α_i^* and α_i are found from eq. (17) the required vectors can now be determined as:

$$\begin{aligned} w_0 &= \sum (\alpha_i^* - \alpha_i) x_i \text{ and} \\ \sum (\alpha_i^* - \alpha_i) (x_i x) &+ b_0 \end{aligned} \tag{18}$$

For the non-linear SVR model, a non-linear mapping kernel could be applied to map the data into larger dimensional

characteristics place where linear regression is fitted. The quadratic equation to be maximized can be re-written as:

$$\begin{aligned} w(\alpha^*, \alpha) = & -\varepsilon \sum (\alpha_i^* + \alpha_i) + \sum y_i (\alpha_i^* - \alpha_i) - (1/2) \\ & \times \sum \sum (\alpha_i^* - \alpha_i) (\alpha_j^* - \alpha_j) K(x_i, x_j) \end{aligned} \tag{19}$$

and the regression function is given by:

$$f(x) = w_0 + b_0 \tag{20}$$

where

$$\begin{aligned} w_0 x &= \sum (\alpha_i^0 - \alpha_i^{0*}) k(x_i, x) \text{ and} \\ b &= -(1/2) \sum (\alpha_i^0 - \alpha_i^{0*}) [k(x_r, x_i) + k(x_s, x_j)] \end{aligned} \tag{21}$$

Ensemble unit

For similar sets of data, obviously, one AI model may outperform others and when various sets of data are used, the results of different models would be entirely different. To use the benefits of each model without missing the general nature of data, the ensemble technique was developed which uses individual model’s output as input with a definite importance level allocated to each with the assistance of an arbitrator to offer the output (Kiran and Ravi 2008). The accuracy of the combination of outputs from different individual models usually will be better than the accuracy of the best single model (Asaad Y Shamseldin and Connor 1999). The importance of ensemble modeling is that each output from an individual model may be considered as representative of the source of data that may be separate from the other models and combining all information from different sources may enable to optimize all input information to the model. For boosting prediction results, several methods of an ensemble such as neural network, random forest regression, simple average, least square, weighted average, and Bates-Granger has been employed (Elkiran et al. 2019; Homsy et al. 2020; Ribeiro et al. 2020; Shamshirband et al. 2019; Shiru and Park 2020). Seasonal rainfall was successfully predicted by ensemble techniques by using different genetic programming models (Danandeh Mehr 2020; Danandeh Mehr 2021) which could be used as pre-processing of rainfall for different hydrological modeling. This study applied three ensemble techniques namely; simple average, weighted average and, neural network ensemble methods to improve the performance of AI-based individual rainfall-runoff modeling. The selected ensemble methods consume less time for modeling and more efficient as reported in the previous studies.

Simple average ensemble (SA)

In the SA ensemble technique, FFNN, ANFIS, and SVR are modeled individually and the SA output is produced by taking the average of the outputs of the individual models as:

$$Q_{o-} = \frac{1}{n} \sum_{i=1}^n Q_{oi} \quad (22)$$

Where Q_{o-} is average discharge from the simple ensemble model, Q_{oi} is discharged from the i th single model, and n is the number of individual models (here, $n = 3$).

Weighted average ensemble (WA)

Weighted average ensemble applies different weights on the outputs of models outputs based on the relative importance of the results as:

$$Q_{o-} = \frac{1}{n} \sum_{i=1}^n w_i Q_{oi} \quad (23)$$

where w_i is the applied weight on the output of an i th model that can be computed based on the model performance as:

$$w_i = \frac{DC_i}{\sum_{i=1}^n DC_i} \quad (24)$$

DC_i is the performance measure (e.g., coefficient of determination) of the i th single model.

Non-linear neural network ensemble method (NNE)

In a non-linear neural network ensemble technique, the results of individual models are taken as inputs of the neural ensemble; each is assigned to one neuron of the input layer. The modeling steps of the neural ensemble modeling are similar to FFNN where the best topology and iteration number of the neural ensemble combination should be attained using the trial-error process and the sigmoid may be considered as hidden and output activation functions.

Sensitivity analysis

The performance of any model is influenced by the relevance and quality of inputs concerning the output. A large number of inputs could cause the complexity of modeling and overfitting that leads to unrealistic results, especially in AI-based modeling. On the other hand, an insufficient number of inputs can reduce modeling accuracy. Several statistical and data-driven sensitivity analysis methods such as cobweb plots, Sobol' indices, linear regressions, neural network, and partial derivative (PaD) were widely applied to understand the impacts of

inputs on the outputs (Tunkiel et al. 2020). Statistical methods such as correlation coefficient, cobweb plots, Sobol' indices, linear regressions might not be suitable high-dimensional data sets and could not well capture the non-linear hydrologic processes. The neural network is affirmed as a powerful tool to analyze sensitivity on output imposed by input parameters since the neural network can well handle the non-linearity of hydro-meteorological data and handles the large-dimensionality of inputs (Nourani and Sayyah Fard 2012). To determine the most influential and relevant input parameters on the runoff, the ANN-based sensitivity analysis was applied in this study to detect the sensitivity of the inputs such as discharge, rainfall, and temperature with different lag times on the output. The hydro-meteorological parameters with different time lags were considered as potential inputs to predict runoff via a FFNN. The performance of each parameter in terms of DC was used to rank the influence extent of each input to the output and only significantly important parameters were used as inputs for the AI-based modeling of the rainfall-runoff process. Accordingly, different time lags of discharge, precipitation, and temperature were used as inputs and single-ahead discharge was considered as the target for the FFNN based sensitivity analysis.

Performance evaluation

There are several techniques applied for evaluating the predicting efficiency of models such as coefficient of determination (DC), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). According to some studies, (e.g., Legates and McCabe 1999) to have an effective comparison, the model efficiency performance should include at least one goodness-of-fit (e.g., DC) and at least one absolute error measure (e.g., RMSE). The performance of the proposed models could be evaluated using the standard evaluation criteria such as coefficient of determination (Eq. 25) and root mean square error (Eq. 26).

$$DC = 1 - \frac{\sum_{i=1}^N (Q_{obsi} - Q_{comi})^2}{\sum_{i=1}^N (Q_{obsi} - Q_{obs})^2} \quad (25)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Q_o(t) - Q_s(t))^2} \quad (26)$$

where DC is the determination coefficient (or Nash-Sutcliffe criterion) (Nourani et al. 2019), RMSE is the root mean square error, Q_o is observed discharge, N is the number of observations, Q is the average of the observed discharge and Q_s is the predicted discharge at time t .

Case study and used data

Study area

The Gilgel Abay watershed is situated in the north-western part of Ethiopia in the latitude of 10°56' to 11°51' N and longitudes 36°44' to 37°23'E and has an area of 1635km² (see Fig. 5). The basin is among the sub-basins of Lake Tana that can contribute more than 60% of runoff to the Lake Tana basin (Wale et al. 2009). Most of the catchment is characterized by mountains topography where the elevation varies from 1805 m to 3518 m above mean sea level and the land slopes vary 0% to 6%. The area is characterized by a cool semi-humid climate with an annual temperature of 17-20 °C, the wet season occurs from June–September and dry season occurs from October–May and the annual mean rainfall is 1416 mm. The watershed has one stream gauging station located at the basin outlet. The textural class of the soil is proportionally distributed among the basin (33.3% clay, 33.7% clay loam, and 33% silt loam). The dominant soil type of the basin is Haplic Luvisols and 74% of the catchment is covered by rain-fed cropland, 15% grassland, and 11% woodlands and forested at higher altitudes.

Used data sets

The data set used for this study includes 5 years (2014–2018) daily rainfall, streamflow, and average temperature.

Precipitation and temperature were collected from Ethiopian National Meteorological Agency for five stations (Sekela, Wetet Abay, Adit, Dangila, and Gundil,), the first two are located inside the basin and the rest located around the basin (see Fig. 5). Streamflow data recorded at the outlet gauging station of the main river, obtained from the Ethiopian Ministry of Water Irrigation and Energy. The first 3.5 years of the data were used for training and the rest 1.5 years of data were utilized for verification of the models.

Remotely sensed rainfall data from the satellite may give better spatial and temporal resolutions of data in a case where ground-based rain-gauge stations are sparse. Several satellite-based rainfall products namely: Global Climatology Project Multi-satellites (GPCP-MS), Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN), National Oceanographic and Atmospheric Administration Climate Prediction Center (NOAA-CPC) Merged Analysis (CMAP), and TRMM products (Dinku et al. 2007). High-resolution satellite rainfall algorithms combine rainfall information from remotely sensed, more accurate, and infrequent microwave and more frequent and less accurate infrared algorithms. For this study, TRMM (Tropical Rain Measuring Mission) 3B42RT v7, which provides real-time data, and TRMM 3B42 v7 post-real-time data, and CMORPH data, were downloaded and used at daily temporal and 0.25°x0.25° spatial resolutions for the study period. Nine satellite rainfall grids cover the whole watershed as shown in Fig. 5. These satellite rainfall products were selected

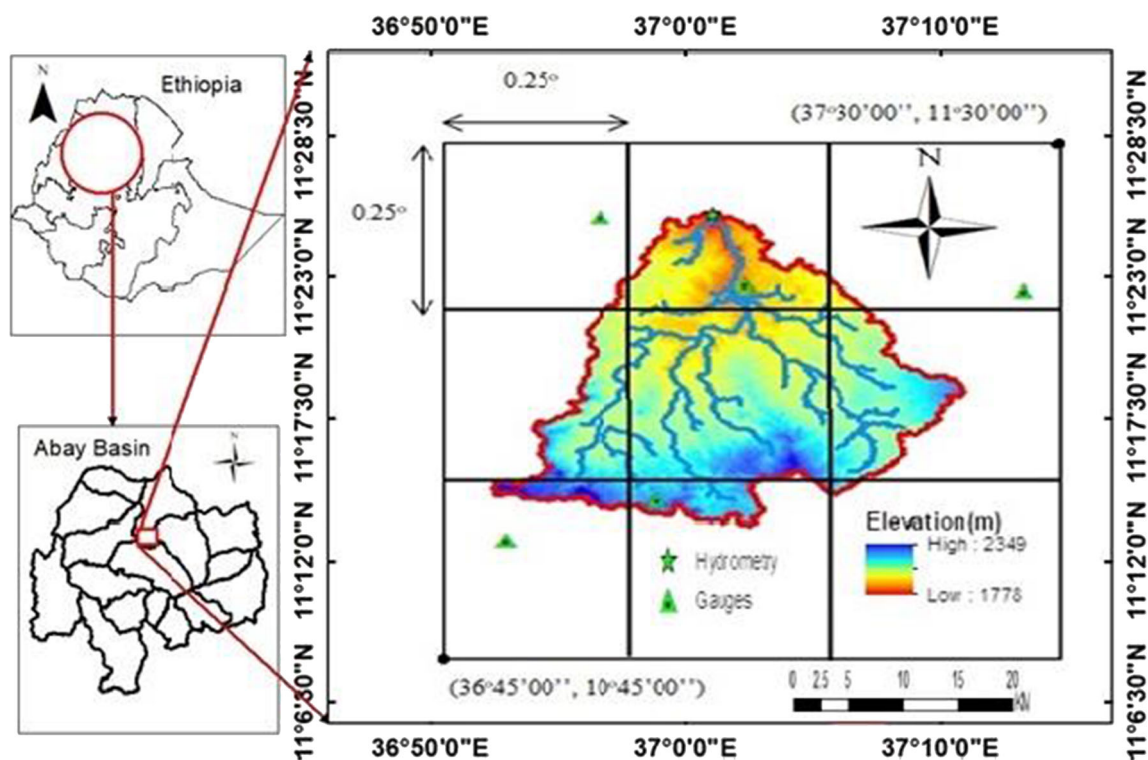


Fig. 5 Map of the study area and the locations of meteorological and hygrometry stations and satellite grids

because they already led to good performance for the study area in previous studies (Menberu M. Bitew et al. 2012; Menberu M. Bitew and Gebremichael 2011) and the temporal resolution of the products are available daily which is suitable for AI modeling.

TRMM precipitation sensor spacecraft contains precipitation radar, microwave imager, and infrared and visible ray scanner. TRMM estimates rainfall based on three steps; i) received raw products are calibrated and geo-located, ii) the products are derived by geographical and physical features for the same location and resolution for the raw data, iii) the time-averaged product is mapped into uniform space and time grids. The 3B42RT is a near-real-time version (about 9 h later real-time) which covers latitude from 60°N to 60°S and the 3B42 is the post-real-time product (10 to 15 days later the end of every month) that covers latitude from 50°N to 50°S (Li et al. 2018). Both products are version 7 and have 0.25° by 0.25° spatial resolution and 1-day temporal resolution. The 3B42RT uses TRMM Combined Instrument (TCI) dataset, which contains TRMM Precipitation Radar (PR) and TRMM Microwave Imager (TMI), to calibrate rainfall estimations acquired from low orbit microwave satellites (Ochoa et al. 2014). The 3B42RT combines all of the estimates at a given time interval and data gaps are filled from analysis of geostationary earth orbit infrared information that is locally calibrated based on merged microwave products. The 3B42 uses gauge data analysis such as Global Precipitation Climatology Center (GPCC) 1° by 1°, monthly rain gauge analysis, and Climate Assessment and Monitoring System (CAMS) 0.5° by 0.5° rain gauge analysis (Rudolf et al. 1994).

The Climate Prediction Center (CPC) morphing technique (CMORPH) product is a near-real-time rainfall product (Gebremichael et al. 2014). It is usually available after 18 h of observation developed by the United States National Oceanic and Atmospheric Administration (NOAA). To estimate precipitation data, the CMORPH algorithm uses passive microwave information from near-orbit satellite radiometers and infrared information from geostationary satellites. CMORPH algorithm is not merging passive microwave and infrared precipitation estimates but it uses rainfall estimates derived from passive microwave observations and transmits this information in space using motion vectors derived from geostationary infrared data (Dinku et al. 2007). In the first step, the time sequence of features motion is governed from infrared ray information, and then these data are used to provide the displacement motion for morphing from one instantaneous microwave estimate to the next. In this process, CMORPH combines the higher retrieval accuracy of passive microwave and the superior spatial and temporal resolution of infrared ray information. The statistics of daily discharge at the outlet and rainfall of stations and satellite sources are given in Table 1.

All input data were normalized to keep the range of data in the specific range that means between 0 and 1 via:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (27)$$

Where, X_{norm} is normalized value, X is observed value, X_{min} is the minimum observed value and X_{max} is the maximum observed value.

Rain gauge and satellite rainfall datasets

Hydrological modeling requires accurate rainfall data over the whole basin, however; rainfall is highly variable in both space and time. Rain gauges in developing countries are installed very sparsely and that cannot give accurate rainfall to exactly represent the catchment. For hydrological modeling, average rainfall over the catchment is required and it can be determined by several techniques. Among the techniques, the Thiessen polygon method is the most popular for practical problems. This method divides the catchment into smaller areas with different geometric shapes, assigns a weight for each polygon, and assumes that the rainfall at any point in the watershed is similar to the nearest rain gauge. For the study area, the polygons around rain-gauge stations and their respective weights are shown in Fig. 6. Using this method, average gauge time series rainfall over the watershed was computed and used in the modeling. According to the Thiessen polygon results, rainfall in the watershed is mainly influenced by Wetet-Abay and Sekela gauges because those stations are located inside the watershed.

Daily Thiessen polygon average rainfall and satellite datasets (3B42RT, 3B42, and CMORPH) for (2014–2018) are plotted in Fig. 7a,b, and c where satellite rainfall products averaged over 0.25° by 0.25° are compared with average gauge rainfall.

CMORPH satellite rainfall dataset was relatively close to ground station rainfall especially since high rainfall during the summer season is much closer to ground station rainfall record (Fig. 7a.). But it overestimates low rainfall for all seasons. The 3B42 satellite rainfall product is weak to capture both high and low rainfall values. It has a tendency to underestimate the majority of peak rainfall values and overestimate low-intensity rainfall (see Fig. 7c.) and it produces some false spikes during the dry season. 3B42RT satellite rainfall products are fairly good to capture low-intensity rainfalls but they underestimate peak rainfall events (Fig. 7b.). Overall, usually, CMORPH performed better than 3B42RT and 3B42 in capturing seasonal and diurnal cycles of the rainfall over the study area.

Table 1 Statistics of the used daily data sets

Rainfall data (mm)				
Data source	Maximum	Minimum	Mean	Standard deviation
Adet (gauge)	70.2	0	3.49	7.32
Dangila (gauge)	61	0	4.41	8.52
Gundil (gauge)	90.2	0	6.73	11.73
Sekela (gauge)	72.3	0	6.04	9.91
Wetet-Abay (gauge)	70.5	0	5.32	9.90
3B42 (satellite, S ¹)	48.34	0	3.51	6.05
3B42RT (satellite, S ²)	54.18	0	3.62	6.42
CMORPH (satellite, S ³)	71.15	0	5.72	9.34
Flow at gauging station (m ³ /day)	355.46	4.69	58.40	60.71

Results and discussion

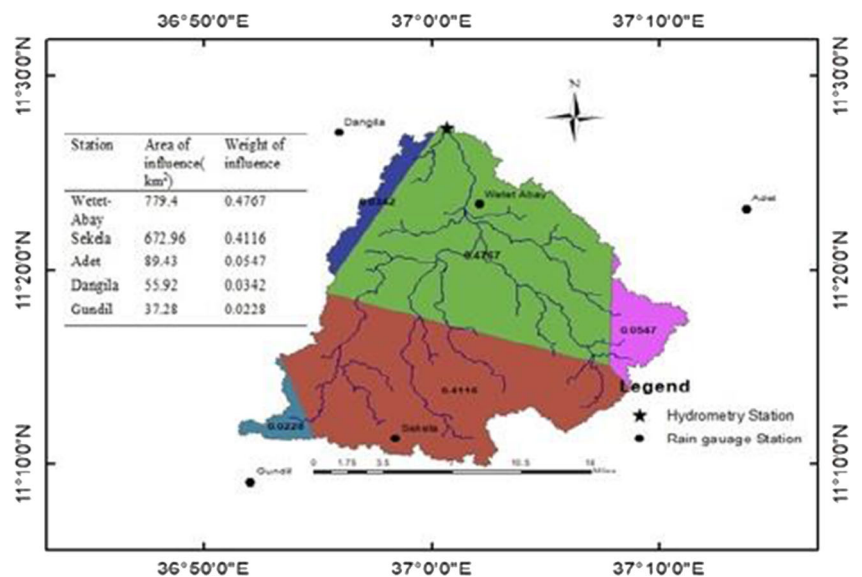
Results of sensitivity analysis and dominant inputs selection

Relevant and dominant input selection is the most important step in any black-box modeling because the quality and relevance of input data could significantly affect the output. In this study, the effect and sensitivity of each input data i.e., discharge ($Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3},$ and Q_{t-4}), rainfall ($P_t, P_{t-1}, P_{t-2}, P_{t-3},$ and P_{t-4}), and temperature ($T_t, T_{t-1}, T_{t-2}, T_{t-3},$ and T_{t-4}) to the output i.e., single-ahead discharge (Q_{t+1}) were analyzed via neural network modeling and ranked based on the mean DC value of each parameter obtained in the calibration and validation phases of FFNN modeling (Table 2). Accordingly, the parameters which are significantly important for rainfall-runoff modeling were selected by the t-student test and used as the inputs of the models. Based on the sensitivity analysis

result (Table 2), the parameters found to be relevant for this study are $Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, P_t, P_{t-1}, P_{t-2}, P_{t-3},$ and P_{t-4} , however, $T, T_{t-1}, T_{t-2}, T_{t-3},$ and T_{t-4} were irrelevant since their contribution to runoff is very low, therefore, these parameters were not considered as inputs for the modeling. Hence, the inputs selected for this study were $Q_t, Q_{t-1}, Q_{t-2}, P_t,$ and P_{t-1} to get reliable modeling outputs.

The proposed modeling in this study comprises two steps, (1) separate modeling of rainfall-runoff using satellite, gauge, and input fusion rainfall data by different non-linear models were created and the modeling performance of each input data source for the respective model was evaluated; (2) The ensemble was conducted using two linear ensemble methods (weighted average and simple average) and one non-linear (Neural network) to appraise the efficiency of single modeling. In this way, the outputs of the inputs fusion models were used as inputs for the ensemble techniques. The results are presented in the following sub-sections.

Fig. 6 Thiessen polygons and their weights of influence



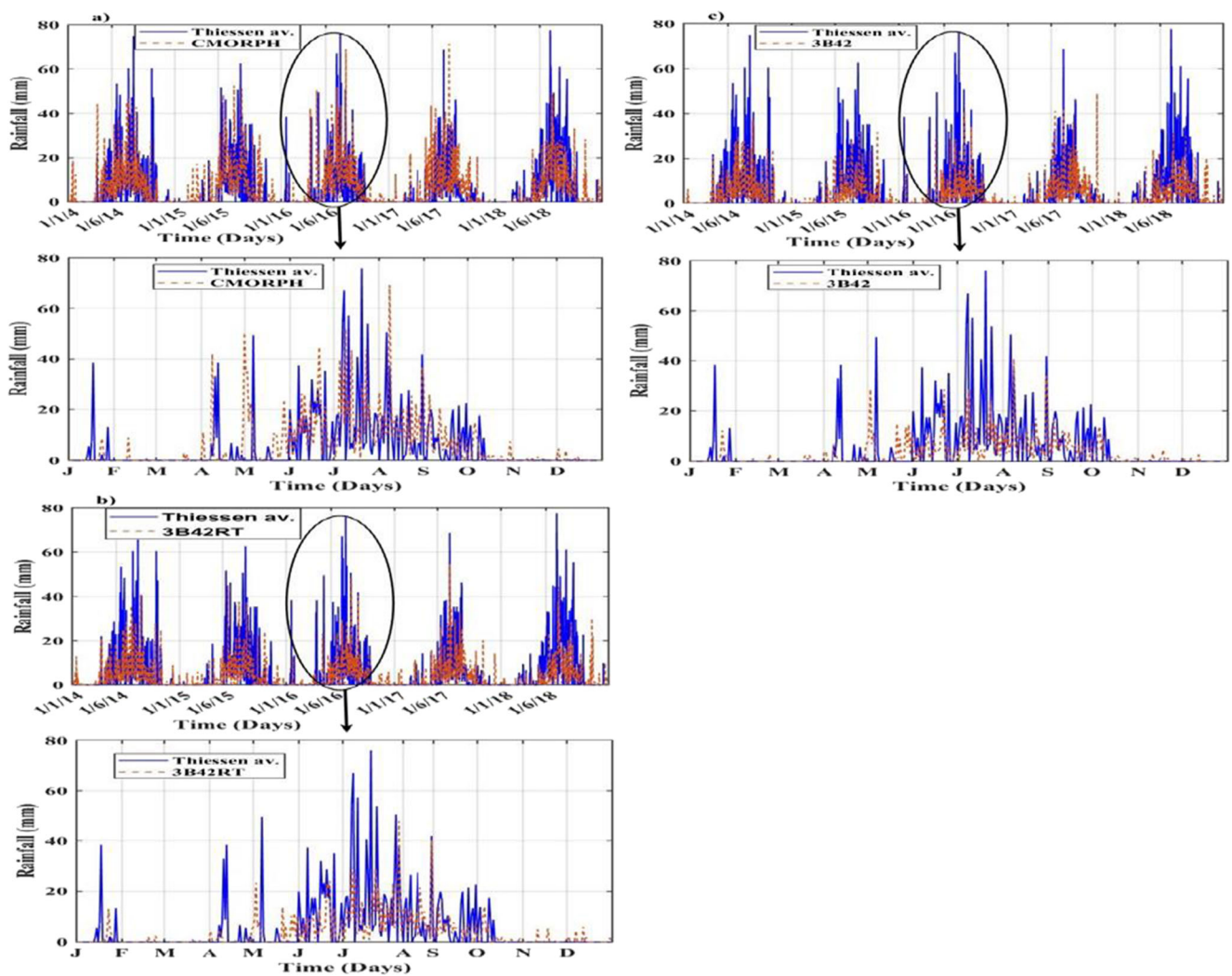


Fig. 7 Rainfall time series of Thiessen average and CMORPH for 2014–2018 and a detail for 2016 (a), Thiessen average and 3B42RT for 2014–2018 and a detail for 2016 (b), and Thiessen average and 3B42 for 2014–2018 and detail for 2016 (c)

Table 2 Results of Sensitivity analysis

Inputs Parameters	Mean DC	Rank
Q_t	0.8438	1
Q_{t-1}	0.8317	2
Q_{t-2}	0.8242	3
P_t	0.7826	4
P_{t-1}	0.7724	5
Q_{t-3}	0.6041	6
Q_{t-4}	0.5698	7
P_{t-2}	0.5549	8
P_{t-3}	0.5421	9
P_{t-4}	0.5392	10
T	0.1877	11
T_{t-1}	0.1855	12
T_{t-2}	0.1824	13
T_{t-3}	0.1786	14
T_{t-4}	0.1755	15

Results of individual rainfall-runoff models

Using satellite and gauge data sets, AI-based rainfall-runoff models were created using the single non-linear models of FFNN, ANFIS, and SVR.

Results of the FFNN model

FFNN trained by BP and the LM algorithm was applied in this study with one hidden layer and variable hidden neurons because of its fast convergence ability and popularity. The optimum number of hidden neurons was determined by the trial and error method for each data source. Hence, the range of hidden neurons applied in this study varies from 9 to 21 for the prediction of runoff (see Table 3). Among the satellite rainfall products, CMORPH rainfall data with FFNN structure of (5–17-1) with 17 hidden neurons gave the best result with DC of 0.8597 and 0.7744 in calibration and validation phases,

Table 3 Rainfall-runoff results of FFNN, ANFIS, and SVR models for satellite and average gauge rainfall data (inputs of all models are Q_t , Q_{t-1} , Q_{t-2} , P_t , and P_{t-1})

Model	RF data source	Structure	DC		RMSE ^a	
			Training	Validation	Training	Validation
FFNN	3B42	5–21-1	0.8603	0.7724	0.06423	0.0740
	3B42RT	5–9-1	0.8625	0.7726	0.0641	0.0742
	CMORPH	5–17-1	0.8597	0.7744	0.0633	0.0728
	Gauge	5–11-1	0.9141	0.8368	0.0612	0.7007
ANFIS	3B42	Gaussian	0.8684	0.7803	0.0633	0.0717
	3B42RT	Gaussian	0.8571	0.7928	0.0636	0.0690
	CMORPH	Triangular	0.8677	0.7986	0.0624	0.0687
	Gauge	Gaussian	0.9205	0.8452	0.05818	0.0667
SVR	3B42	RBF	0.8571	0.7678	0.0652	0.0737
	3B42RT	RBF	0.8558	0.7688	0.0653	0.0736
	CMORPH	RBF	0.8578	0.7732	0.0658	0.0732
	Gauge	RBF	0.9082	0.8342	0.0590	0.0702

RMSE^a: RMSE is unitless because the data are normalized

respectively (see Table 3). Comparing the performance of the three satellite data sets in FFNN modeling, it is practical to select datasets to perform better with fewer hidden neurons. In this case, the 3B42RT dataset is superior over CMORPH and 3B42 datasets. Therefore, FFNN using the 3B42RT rainfall dataset could forecast runoff at a short time and low cost for the study area. For average gauge rainfall data, the FFNN model achieved DC of 0.9141 and 0.8368 in training and validation steps, respectively. Ground station rainfall data indicated superiority over satellite rainfall products in runoff prediction and this could be due to variation of capturing ability of satellite spacecraft resulting biases for different rainfall magnitudes. For instance, 3B42RT and 3B42 underestimate the majority of peak rainfalls (see Fig. 7b and c) that could reduce predicted runoff from observed runoff. Runoff time series predicted by FFNN for best models using two different rainfall data sources (ground station and CMORPH satellite) are plotted versus observed runoff (Fig. 8a.). As it is shown in Fig. 8a, FFNN accurately modeled low flows at dry seasons however; it led to less accurate results in capturing the high flows in wet seasons for both data sets.

Results of the ANFIS model

Sugeno type ANFIS was applied in this study and the membership functions (MPs) were calibrated by input-output parameters through a hybrid optimization algorithm. Various MPs were deployed for ANFIS modeling and the best ANFIS structures were characterized by MF and different iterations of epochs. For rainfall-runoff simulation, Gaussian, Trapezoidal, and Triangular shaped MFs were applied for all (satellite and ground station) data sets, and the MFs that gave the best results at optimum epoch are presented in Table 3.

From satellite rainfall products, CMORPH rainfall products with Triangular MF performed well with DC of 0.8677 and 0.7986 at training and validation stages, respectively (see Table 3). All satellite rainfall products performed fairly well for rainfall-runoff simulation by ANFIS at optimum epoch iteration and appropriate MFs. Using the ground rain gauge station data sets, ANFIS achieved the best results with DC of 0.9205 and 0.8452 at training and validation stages, respectively (Table 3). ANFIS structure constructed with Gaussian MF gave the best result. Runoff time series predicted by ANFIS using two different rainfall data sources (ground station and satellite) are plotted versus observed runoff in Fig. 8b. ANFIS accurately predicted peak runoff in summer seasons since the watershed receives high rainfall at this season, however; it slightly overestimated low flow in the dry season for both CMORPH and average gauge rainfall data sets (see Fig. 8b).

Results of the SVR model

In SVR modeling, Radial Base Function (RBF) kernel was used to create the models for ground-based and satellite data sets. RBF was selected over the sigmoid and polynomial kernels because it uses fewer tuning parameters and has been already confirmed that RBF outperforms the other kernels (Sharghi et al. 2018). The results of the SVR model for satellite data sets are presented in Table 3 and it is shown that all three satellite rainfall products gave fairly good results but CMORPH surpassed the 3B42 and 3B42RT having DC values of 0.8578 and 0.7732 in the training and validation stage, respectively. The rainfall-runoff result of SVR showed a good performance for average rain gauge data sets with DC of 0.9082 and 0.8342 in training and validation stages,

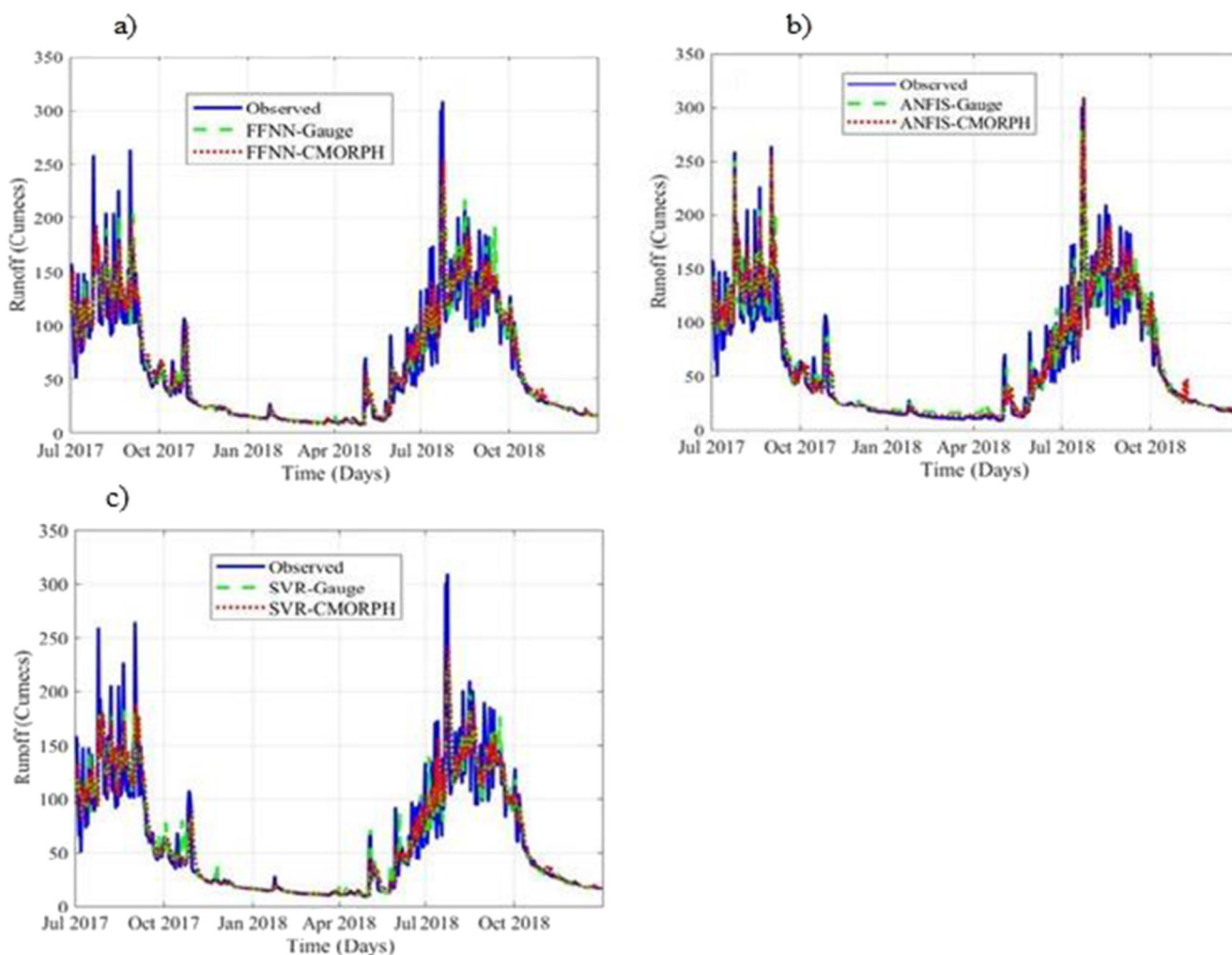


Fig. 8 Observed vs. predicted runoff (average gauge and CMORPH rainfall) via **a** FFNN, **b** ANFIS, **c**SVR, in the validation phase

respectively (see Table 3). SVR could accurately model low flow in the dry season and normal flow in the wet season, however, it underestimated peak flows in wet seasons using both CMORPH and rain gauge data sets (see Fig. 8c).

Overall, the runoff prediction performance of ANFIS surpassed the FFNN and SVR models for both average gauge and satellite data sets (see Fig. 9).

Results of modeling by input fusion

In input fusion modeling, two input fusion strategies were deployed. Strategy 1) Only satellite rainfall products were combined as inputs (without average gauge rainfall) to predict runoff (Table 4). Strategy 2) Rainfall from all three satellite rainfall data products and the average gauge were combined and imposed into the input layer of the models to predict runoff (see Table 5), to see the combined effects of data from different sources on runoff prediction. In this section, all inputs from both data sources were combined and modeled by all AI models then ensemble modeling was performed using

WA, SA, and NNE techniques. Usually prior to use satellite data in hydrological modeling, data should be bias-corrected according to the ground-based gauge data (Menberu M. Bitew et al. 2012). Satellite rainfall is “bias-corrected” by following two steps. First, the bias on the satellite rainfall products is determined by dividing the daily average satellite rainfall products on the pixel that comprises the rain gauge to the corresponding gauge rainfall value. Second, the original daily satellite rainfall product is multiplied by bias to remove the bias in satellite rainfall data. However, in this study, gauge rainfall data were imposed directly into the models along with satellite data that this can act as a bias correction method of the satellite data.

The strategy 1 input fusion results gave promising improvements over runoff predicting using individual satellite rainfall data sources however it indicated slightly lower performance as compared with average gauge-based runoff modeling (see Table 4). Inputs fusion of satellite with an average gauge as strategy 2, significantly improved the runoff prediction accuracy over both average gauge and satellite-

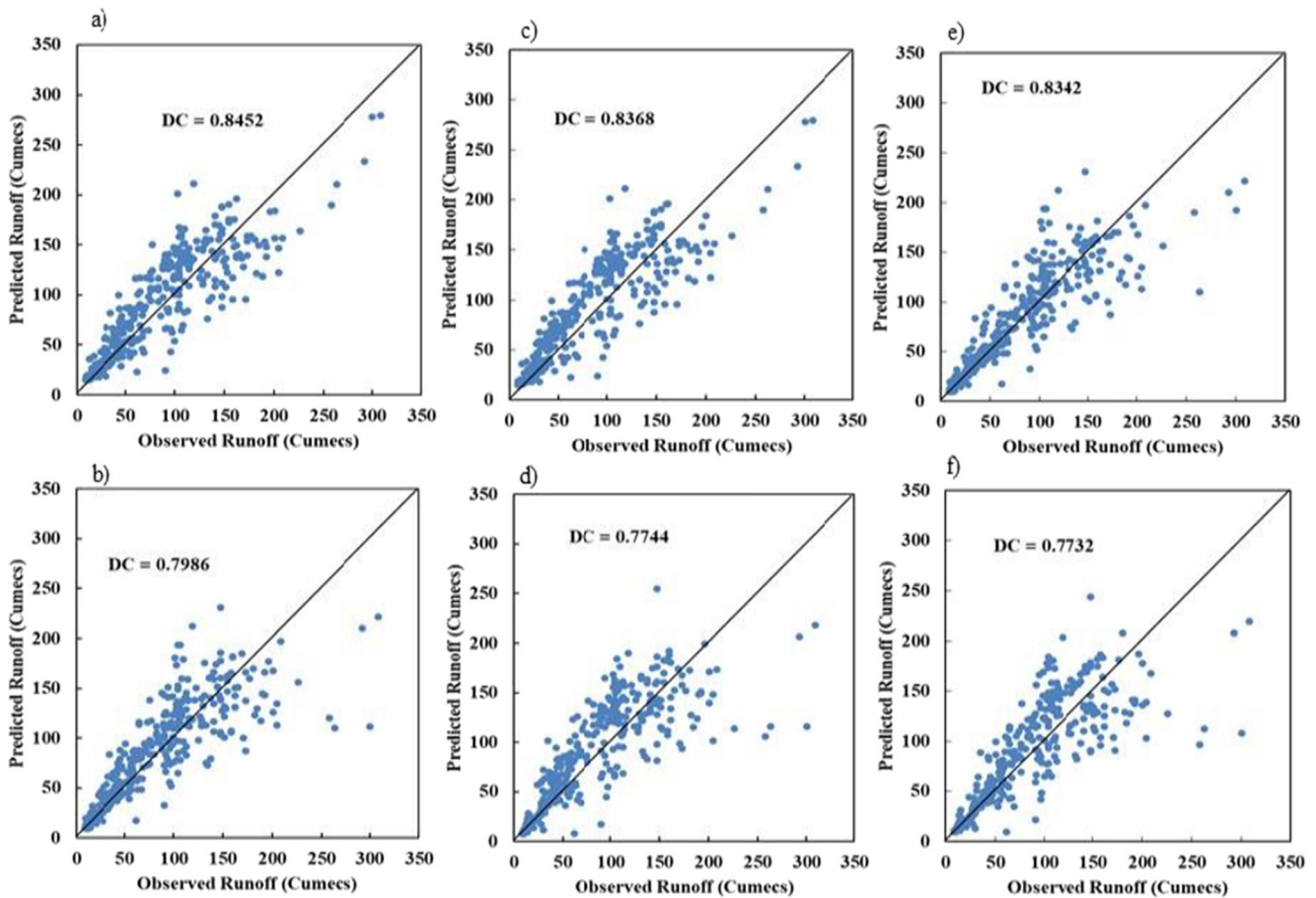


Fig. 9 Scatter plot of observed and predicted results of **a** ANFIS-Gauge average, **b** ANFIS-CMORPH, **c** FFNN-Gauge average, **d** FFNN-CMORPH, **e** SVR-Gauge average, **f** SVR-CMORPH at validation phase

based rainfall for all AI models (Table 5). Particularly, for satellite-based rainfall, the reason for the improvement of the runoff prediction accuracy could be related to the bias correction capacity when gauge data are imposed to the models as well as satellite data. It is fact that gauge rainfall is more accurate than satellite rainfall products since the quality of satellite information depends on cloud conditions, revisit time of satellites, and their orbital positioning which results in bias to rainfall estimations. In strategy 2 input fusions, both sources of inputs combined to predict rainfall-runoff thus boosted the

models’ performance as compared with strategy 1. It is worth mentioning that gauge rainfall corrected the bias of satellite-based rainfall products and improved the rainfall-runoff prediction efficiency of the models.

When the modeling results for gauge and satellite rainfall data sources are compared, the gauge-based data indicated superiority in both single modeling and input fusion stages. The reason is related to the fact that the gauge rainfall can capture the real physical relationships between rainfall and runoff at the watershed level. The satellite-based rainfall

Table 4 Results of rainfall-runoff modeling by input data fusion using only satellite data sets (strategy 1) (inputs for all models are Q_t , Q_{t-1} , Q_{t-2} , P_t CMORP, P_t 3B42RT, and P_t 3B42)

Model	Model structure	DC		RMSE ^a	
		Training	Validation	Training	Validation
FFNN	6–9–1	0.9086	0.8086	0.0642	0.0726
ANFIS	Gaussian	0.9187	0.8147	0.0608	0.0692
SVR	RBF	0.8908	0.8043	0.0646	0.0732

RMSE^a: RMSE is unit-less because the data are normalized

Table 5 Results of rainfall-runoff modeling for input fusion of both gauge and satellite data sets (strategy 2) (inputs for all models are Q_t , Q_{t-1} , Q_{t-2} , P_t Gauge, P_t CMORP, P_t 3B42RT, and P_t 3B42)

Model	Model structure	DC		RMSE ^a	
		Training	Validation	Training	Validation
FFNN	7–10–1	0.9205	0.8492	0.05634	0.06514
ANFIS	Gaussian	0.9360	0.8644	0.0518	0.0625
SVR	RBF	0.9175	0.8428	0.0588	0.0661

RMSE^a: RMSE is unitless because the data are normalized

showed weak performance in single modeling and satellite rainfall input fusion (strategy 1 modeling) and this could be due to the bias of satellite data. The satellite-based rainfall could lead to reasonable results however, over-estimate or under-estimate of runoff was noticed based on seasonal and temporal variations as compared to observed runoff. This could be because of topographic variations of the watershed and the accuracy of the sensor of the satellites to retrieve the information. The study by (Gebremichael et al. 2014) on CMORPH, 3B42RT, and 3B42 satellite rainfall products indicated that they may overestimate daily rainfall at lowlands and underestimate at mountainous areas, as the study watershed has high topographic variations that vary between 1778 and 2349 m above sea level (see Fig. 5), hence, it could be more vulnerable to the topographic effects. The accuracy of rainfall measurement by satellite also depends on the algorithms they utilized. The study by (Bitew and Gebremichael 2010) indicated that satellite which uses microwave algorithms performs better than that which uses infrared waves. As the result indicated that, the CMORPH that uses the microwave algorithm surpasses 3B42RT that uses a combination of microwave and infrared algorithms, and 3B42, which uses the infrared algorithm (see Table 3).

All AI models used in this study led to promising results (see Fig. 10a) for both satellite and gauge data sets (see Table 3); however, all models could not equally perform and capture temporal variations of the runoff. To have a better visualization of predicted runoff by each model, Fig. 10b indicates the plots for the wet season (July – October 2017) and Fig. 10c shows the plots for the dry season (December – March 2017/2018). The results reveal that ANFIS could precisely predict peak runoff in the wet season (Fig. 10b) but it overestimated the dry season low flow regime (see Fig. 10c). SVR and FFNN models were good for predicting low flow during the dry season (Fig. 10c); however, they indicated less accuracy in simulating peak flows in the wet season (see Fig. 10b).

To further investigate modeling performances at different time spans, for each season, 2 different intervals were picked and the predicted runoff by each model was compared with the observed runoff values. For the wet season, points 1 and 2 on 18 July, and on 1 September 2017, respectively, were considered. At point 1, observed = 140.92 m³/s, FFNN = 106.8792 m³/s, ANFIS = 136.0592 m³/s, SVR = 106.5292 m³/s and at point 2, observed = 263.92 m³/s, FFNN = 101.0892 m³/s, ANFIS = 250.6792 m³/s, and SVR = 109.9292 m³/s, these indicate that ANFIS gave more close predictions to the observed runoff with regard to FFNN and SVR models in both points as it is shown in Fig. 10b. For the dry season, points 3 and 4 on 17 December 2017, and 26 January 2018, respectively, were considered. At point 3, observed = 17.4292 m³/s, FFNN = 17.4492 m³/s, ANFIS = 19.2492 m³/s, and SVR = 16.8192 m³/s, and at point 4,

observed = 18.4292 m³/s, FFNN = 17.492 m³/s, ANFIS = 22.7292 m³/s, and SVR = 16.9792 m³/s, which show FFNN and SVR results are more close to the observed values than ANFIS model as shown in Fig. 10c. The results at these selected points indicate that different models at different time spans could deduce different data aspects. Therefore, the combination of models via ensemble techniques could improve the performance of the modeling and may lead to a better accuracy level of modeling. To this end, outputs ensemble techniques by two linear (SA, WA) and one nonlinear (NNE) approach were applied to improve the overall efficiency of the modeling.

Results of input fusion-ensemble modeling

Ensemble modeling can boost the overall runoff prediction capacity of individual models (FFNN, ANFIS, and SVR). The outputs from the single models obtained by input fusion were used as inputs of the ensemble unit using three proposed ensemble techniques. The runoff already simulated by FFNN, ANFIS, and SVR models using average gauge rainfall, and satellite products were combined by SA, WA, and NNE ensemble techniques. In this study, two strategies of ensembles were applied, 1) ensemble of runoff outputs for only satellite input fusion, 2) ensemble of runoff outputs for input fusion of both satellite and average gauge data sources.

To obtain the weights of the WA ensemble, DCs at the validation stage were used according to Eq. 24. Similar to FFNN modeling, the NNE ensemble was developed by FFNN with BP, using the Levenberg Marquardt algorithm for training with one hidden layer and variable hidden neurons. The best epoch and hidden neuron numbers of the ensemble structure were determined by trial and error. NNE was selected among other non-linear ensemble methods because of its higher performance but other neural methods (e.g. ANFIS and SVR) can be also similarly used.

The outputs of FFNN, ANFIS, and SVR from input fusion modeling (both strategies) were applied as inputs to ensemble techniques of SA, WA, and NNE. The results obtained from the ensemble techniques are presented in Tables 6 and 7 with input-output structures for SA where *a*, *b*, *c* indicate weights generated by FFNN, ANFIS, and SVR applied for the WA ensemble. In terms of DC and RMSE, strategy 2 input fusion ensemble modeling certainly improved modeling accuracy over individual models on separate and inputs fusion modeling (see Tables 3, 4, and 5). The ensemble result of runoff for input fusion of strategy 2 has indicated the superiority over ensemble runoff for input fusion of only satellite rainfall products (strategy 1). This could be because of the bias correction capability of gauge rainfall on satellite rainfall products. Hence in strategy 1 of the ensemble, the used inputs were raw satellite rainfall products that were bias uncorrected whereas gauge rainfall corrected bias of satellite rainfall in

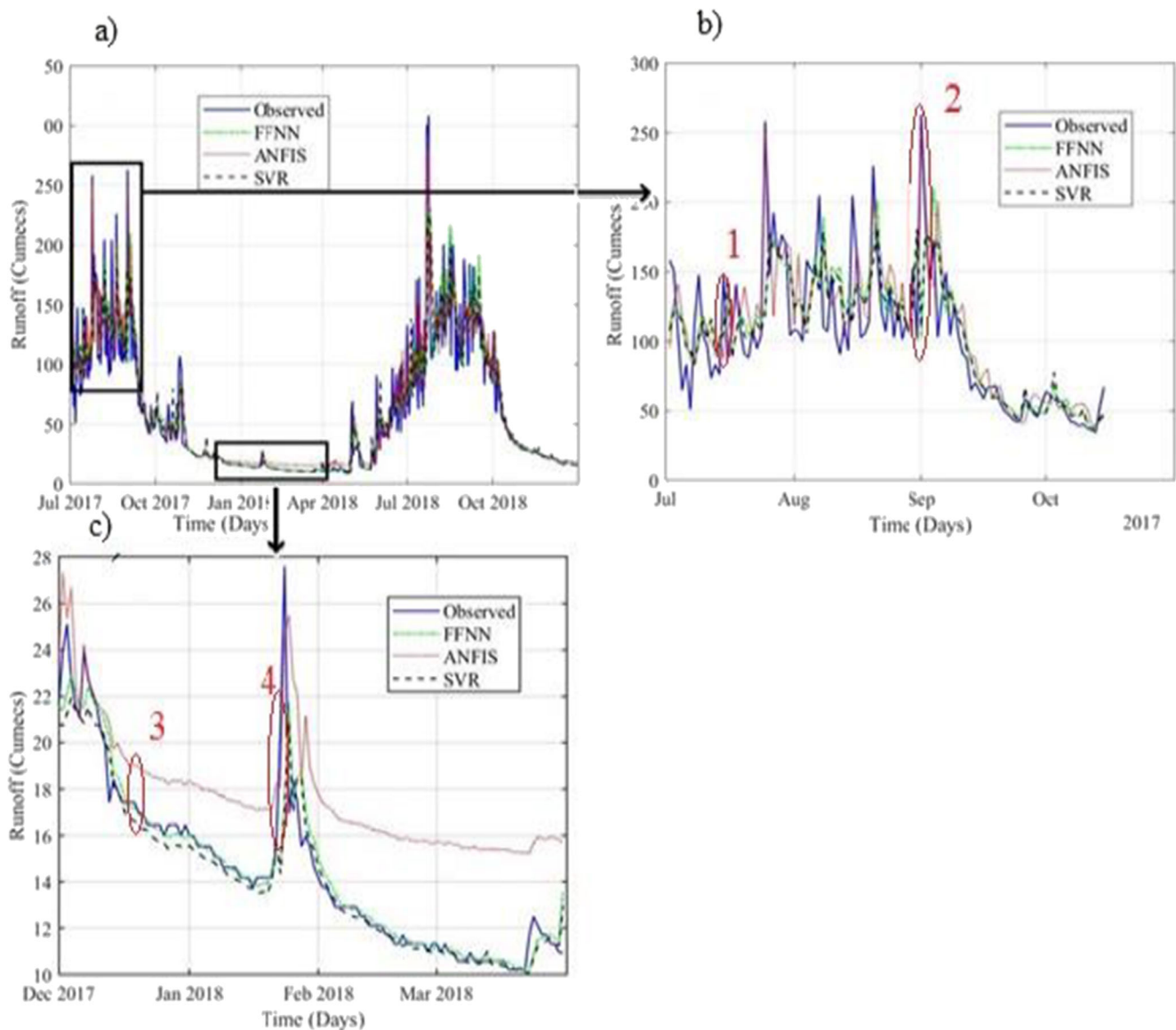


Fig. 10 Observed and predicted runoff time series in validation phase computed via FFNN, ANFIS, and SVR (for gauge data set), **a** the whole verification period, **b** detail for a wet season, **c** detail for a dry season

strategy 2. Anyway, the result of strategy 1 showed that it can be a good option of inputs for rainfall-runoff modeling in ungagged and sparsely gauged catchments (see Table 5).

The ensemble runoff results obtained using input fusion of strategy 2 are depicted in Table 7 which the best ensemble

model i.e., NNE increased the prediction performance of best satellite (CMORPH) rainfall-based single models up to 14.4%, 12%, and 14.5% for FFNN, ANFIS, and SVR respectively in the validation stage (see Tables 3 and 7). It also improved the performance of runoff prediction of input fusion

Table 6 Results of ensemble rainfall-runoff using input fusion data, strategy 1 (inputs for all techniques are Q_{FFNN} , Q_{ANFIS} , Q_{SVR})

Ensemble techniques	Model structure	DC		RMSE ^a	
		Training	Validation	Training	Validation
SA	3–1	0.9145	0.8412	0.056663	0.064447
WA	0.3367,0.3307, 0.3368	0.9235	0.8469	0.05377	0.062343
NNE	3–8–1	0.9278	0.8696	0.04823	0.057123

RMSE^a: RMSE is unit-less because the data are normalized

Table 7 Results of ensemble rainfall-runoff modeling using input fusion data, strategy 2 (inputs of all techniques are Q_{FFNN} , Q_{ANFIS} , and Q_{SVR})

Ensemble techniques	Model structure	DC		RMSE ^a	
		Training	Test	Training	Test
SA	3–1	0.9324	0.8644	0.0556	0.0635
WA	0.3367,0.3307, 0.3368	0.9382	0.8678	0.0529	0.0622
NNE	3–8-1	0.9626	0.9045	0.0445	0.0545

RMSE^a: RMSE is unitless because the data are normalized

of only satellite rainfall models (strategy 1) up to 10.6%, 9.5%, and 11% for FFNN, ANFIS, and SVR models respectively, in the verification stage (see Tables 4 and 7). In strategy 2 of the input fusion ensemble, NNE improved the single model runoff prediction of gauge-based rainfall data by 7.5%, 8%, and 6.6% for FFNN, SVR, and ANFIS models respectively (see Tables 3 and 7). Moreover, the NNE ensemble modeling increased the performance of the input fusion of gauge and satellite rainfall models (strategy 2) up to 6%, 4.5%, and 7% for FFNN, ANFIS, and SVR models respectively, in the validation stage (see Tables 5 and 7).

From the obtained results, it is logical to conclude that the ensemble of input fusion from different data sources could improve the modeling reliability of separate single models and input fusions.

Figure 11 indicates the scatter plots of runoff predicted by single best models for inputs of gauge rainfall, satellite rainfall, strategy 1, and strategy 2 ensemble models versus observed runoff values. As is mentioned earlier, NNE performed better than the other linear ensemble methods because i) NNE uses a non-linear relationship to simulate runoff; hence, it well catches the non-linear behavior of the rainfall-runoff process, unlike linear ensemble methods. ii) The results from individual models could affect the ensembles by SA and WA methods, indicating that less performing models may lead to poor ensemble results by SA and WA since the single models and ensemble models (SA, WA) directly connect the linear relationships. iii) The inaccuracies resulted from single models might be propagated and combined via WA and SA ensemble techniques due to the direct amalgamation of single models. The ensemble models did not lead to significant improvement of DCs in the training phase for some models, however; a remarkable improvement attained in the verification phase for all models and that was the major focus of the methodology used in this study.

The efficiency of the ensemble techniques (SA, WA, and NNE) are presented in Fig. 12 by two-dimensional graphic transparency (Taylor diagram) that can vividly display the predicted and observed values for accurate comparisons. In this diagram, standard deviation (SD) and DC are combined in a metrics form to build multi-performance metrics in a single combination and it can describe the statistical similarity between observed and predicted runoff values. The goal of the

Taylor diagram is to summarize the multi-performances in a single combination, which measures the level of agreement between observed and predicted runoff values. The Taylor diagrams for the input fusion of strategy 1 and input fusion of strategy 2 ensembles are presented in Fig. 12 for SA, WA, and NNE ensemble techniques. In this diagram, the computed values closer to the observed runoff values belong to the model that performed well. Hence, NNE performance surpassed the SA and WA ensemble techniques for both models because the observed and predicted points are close to each other in NNE than the others. The input fusion of strategy 2 ensemble could improve the runoff prediction for the satellite rainfall products than gauge rainfall, which indicates the satellite rainfall products have a limitation of accuracy to capture rainfall values more accurately. From this result, it is worth mentioning that the utilization of input fusion of multiple sources of satellite and gauge rainfall products then ensemble modeling can improve rainfall-runoff modeling more precisely. The result also indicated that the gauge rainfall in strategy 2 input fusion significantly corrected the bias of rainfall of satellites that could occur due to the aforementioned reasons.

Conclusions

This study focused on rainfall-runoff modeling to predict single-step-ahead runoff discharge of Gilgel Abay catchment using 5 years (2014–2018) daily data of three satellite data sets (CMORPH, 3B42RT, and 3B42) and Thiessen polygon averaged rainfall of five stations (Gundil, Dangila, Adet, Wetet-Abay and Sekela) as inputs into different AI models. The most relevant and dominant inputs were selected by FFNN based non-linear sensitivity analysis method. Firstly, rainfall-runoff modeling using data from each of satellite and gauge as well as input fusion of only satellite data sets (strategy 1), and all satellites and gauge (strategy 2) was conducted by each of the AI models, separately. Secondly, runoff values obtained by input fusion models were combined by employing SA, WA, and NNE ensemble techniques to improve the accuracy of runoff predictions. Among the satellite data sets, modeling by CMORPH satellite data performed better via all models; however, it tended to overestimate low flows. The models using 3B42 and 3B42RT products underestimated high-

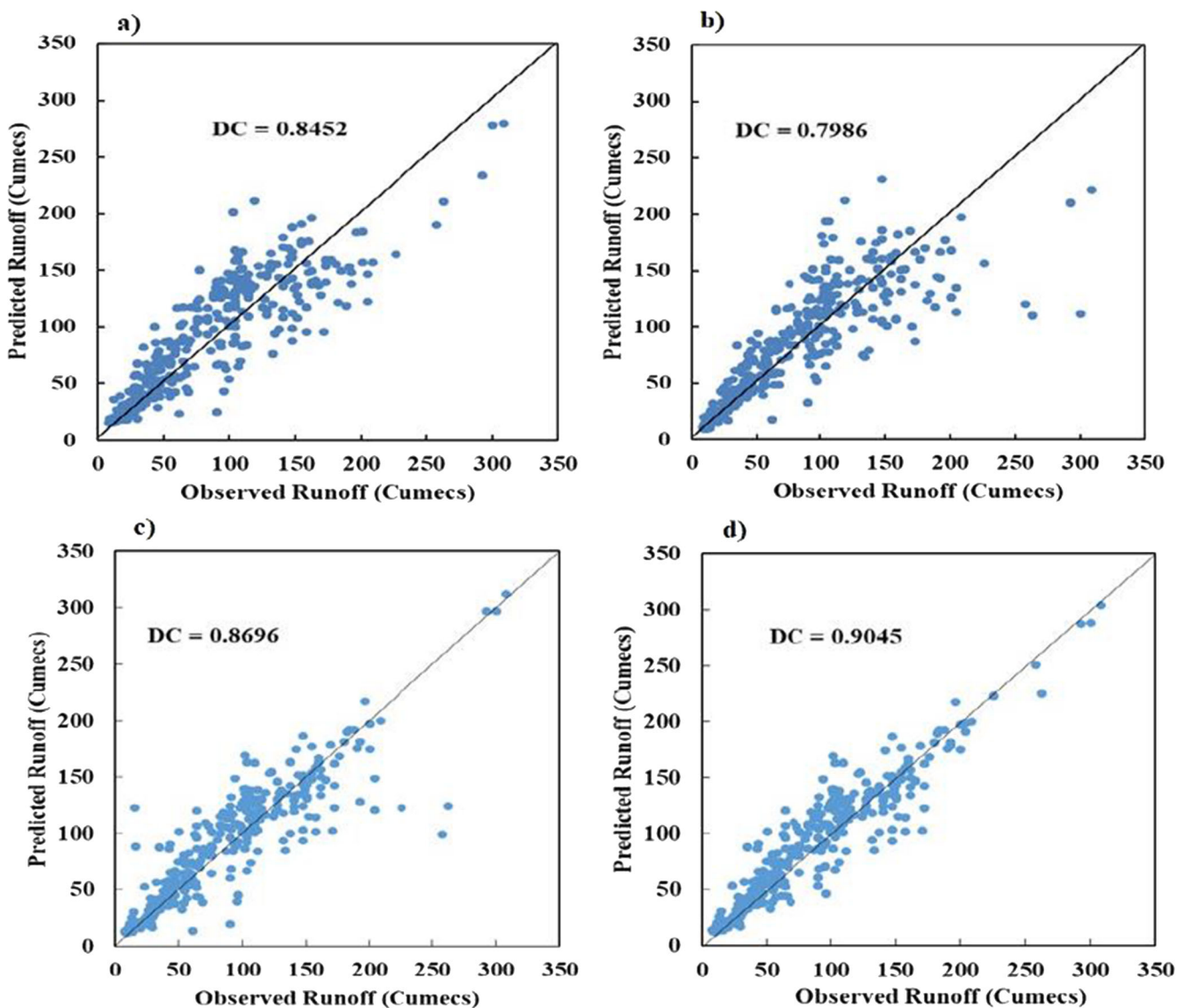


Fig. 11 Scatter plots for **a** Gauge-ANFIS **b** CMORPH-ANFIS **c** Strategy 1-NNE ensemble **d** Strategy 2-NNE ensemble in the validation stage

runoff and 3B42 produced random false spikes in the dry season. From the AI models applied, ANFIS revealed the best performance in average gauge and all satellite rainfall products that could be because the model combines the learning ability of neural network and fuzzy logic in a single framework. Input data fusion from two data sources showed substantial improvements over the outputs of satellite data sources but indicated slight improvement as compared with modeling by gauge rainfall data. This could be due to gauge rainfall bias correction capacity for erratic satellite rainfall products. To improve the prediction performance of the single models, ensemble-modeling SA, WA (linear), and NNE (non-linear) techniques were applied for the input fusion of strategies 1 and 2. In the input fusion of strategy 2, the ensemble of runoff from different satellite rainfall products and the average gauge was conducted to enhance prediction performance. In

this stage, NNE led best results and improved the performance of best satellite (CMORPH) rainfall-based single models up to 14.4%, 12%, and 14.5% for FFNN, ANFIS, and SVR, respectively in the validation stage. It also improved the single modeling runoff prediction using gauge-based rainfall data by 7.5%, 8%, and 6.6% for FFNN, SVR, and ANFIS models, respectively. Among the ensemble techniques, NNE was a robust and precise ensemble technique for accurate rainfall-runoff modeling because of its ability to handle the non-linear nature of the process. Overall, the output of this study contributes a promising suggestion about utilizing a fusion of multiple sources of satellite rainfall products for ungauged and sparsely gauged catchments especially in developing countries where it could increase the accuracy of rainfall-runoff modeling through the provision of reliable input data. Moreover, it is recommended that future studies should focus

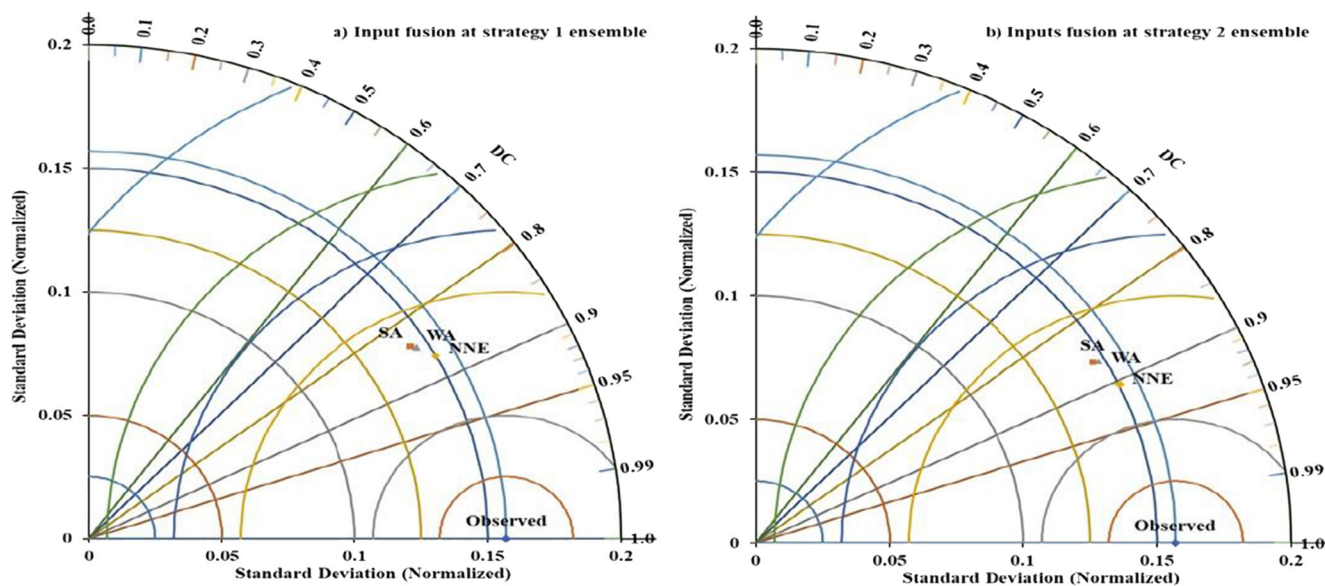


Fig. 12 Taylor diagrams indicating the ensemble performance of SA, WA, and NNE **a** input fusion in strategy 1 ensemble **b** input fusion in strategy 2 ensemble in the validation phase

on the ensemble modeling of AI and physically based models for simulation of rainfall-runoff using rainfall data from multiple satellite sources.

References

- Ateeq-ur-Rauf, Ghumman AR, Ahmad S, Hashmi HN (2018) Performance assessment of artificial neural networks and support vector regression models for stream flow predictions. *Environ Monit Assess* 190(12):704. <https://doi.org/10.1007/s10661-018-7012-9>
- Ayehu GT, Tadesse T, Gessesse B, Dinku T (2018) Validation of new satellite rainfall products over the Upper Blue Nile Basin, Ethiopia. *Atmos Meas Tech* 11(4):1921–1936. <https://doi.org/10.5194/amt-11-1921-2018>
- Bitew MM, Gebremichael M (2010) Assessment of high-resolution satellite rainfall for streamflow simulation in medium watersheds of the East African highlands. *Hydrol Earth Syst Sci Discuss* 7(5):8213–8232. <https://doi.org/10.5194/hessd-7-8213-2010>
- Bitew MM, Gebremichael M (2011) Evaluation of satellite rainfall products through hydrologic simulation in a fully distributed hydrologic model. *Water Resour Res* 47(6):1–11. <https://doi.org/10.1029/2010WR009917>
- Bitew MM, Gebremichael M, Ghebremichael LT, Bayissa YA (2012) Evaluation of high-resolution satellite rainfall products through streamflow simulation in a hydrological modeling of a small mountainous watershed in Ethiopia. *J Hydrometeorol* 13(1):338–350. <https://doi.org/10.1175/2011JHM1292.1>
- Chandwani V, Vyas SK, Agrawal V, Sharma G (2015) Soft Computing Approach for Rainfall-runoff Modelling: A Review. *Aquat Procedia* 4(Icwrcoe), 1054–1061. <https://doi.org/10.1016/j.aqpro.2015.02.133>
- Chang FJ, Chiang YM, Ho YH (2015) Multistep-ahead flood forecasts by neuro-fuzzy networks with effective rainfall-run-off patterns. *J Flood Risk Manag* 8(3):224–236. <https://doi.org/10.1111/jfr3.12089>
- Chen Y, Huang J, Sheng S, Mansaray LR (2018) A new downscaling-integration framework for high-resolution monthly precipitation estimates: Combining rain gauge observations, satellite-derived precipitation data and geographical ancillary data. *Remote Sens Environ* 214:154–172. <https://doi.org/10.1016/j.rse.2018.05.021>
- Collins M, Achuta Rao K, Ashok K, Bhandari S, Mitra AK, Prakash S, Srivastava R, Turner A (2013) Observational challenges in evaluating climate models. *Nat Clim Chang* 3(11):940–941
- Danandeh Mehr A (2020) An ensemble genetic programming model for seasonal precipitation forecasting. *SN Appl Sci* 2(11). <https://doi.org/10.1007/s42452-020-03625-x>
- Danandeh Mehr A (2021) Seasonal rainfall hindcasting using ensemble multi-stage genetic programming. *Theor Appl Climatol* 143(1–2): 461–472. <https://doi.org/10.1007/s00704-020-03438-3>
- Danandeh Mehr A, Kahya E, Şahin A, Nazemosadat MJ (2015) Successive-station monthly streamflow prediction using different artificial neural network algorithms. *Int J Environ Sci Technol* 12(7):2191–2200. <https://doi.org/10.1007/s13762-014-0613-0>
- Dinku T, Ceccato P, Grover-Kopec E, Lemma M, Connor SJ, Ropelewski CF (2007) Validation of satellite rainfall products over East Africa's complex topography. *Int J Remote Sens* 28(7):1503–1526. <https://doi.org/10.1080/01431160600954688>
- Ebert EE, John EJ, Kidd C (2007) Comparison of near-real-time precipitation estimates from satellite observations and numerical models. 88(1): 47–64
- Elkiran G, Nourani V, Abba SI (2019) Multi-step ahead modelling of river water quality parameters using ensemble artificial intelligence-based approach. *J Hydrol* 577:123962. <https://doi.org/10.1016/j.jhydrol.2019.123962>
- Gao Z, Long D, Tang G, Zeng C, Huang J, Hong Y (2017) Assessing the potential of satellite-based precipitation estimates for flood frequency analysis in ungauged or poorly gauged tributaries of China's Yangtze River basin. *J Hydrol* 550:478–496. <https://doi.org/10.1016/j.jhydrol.2017.05.025>
- Gazzaz NM, Aris AZ, Juahir H, Ramli NF, Yusoff MK (2012) Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. *Mar Pollut Bull* 64(11):2409–2420

- Gebre SL (2015) Application of the HEC-HMS Model for Runoff Simulation of Upper Blue Nile River Basin. *J Waste Water Treat Anal* 06(02). <https://doi.org/10.4172/2157-7587.1000199>
- Gebremichael M, Bitew MM, Hirpa FA, Tesfay GN (2014) Accuracy of satellite rainfall estimates in the Blue Nile Basin: Lowland plain versus highland mountain. *Water Resour Res* 50:8775–8790. <https://doi.org/10.1002/2013WR014500>. Received
- Govindaraju RS (2000) Artificial neural networks in hydrology. II: Hydrological applications. *J Hydrol Eng* 5:124–137
- Guimarães Santos CA, da Silva GBL (2014) Daily streamflow forecasting using a wavelet transform and artificial neural network hybrid models. *Hydrol Sci J* 59(2):312–324. <https://doi.org/10.1080/02626667.2013.800944>
- Homsri R, Shiru MS, Shahid S, Ismail T, Harun SB, Al-Ansari N, Chau KW, Yaseen ZM (2020) Precipitation projection using a CMIP5 GCM ensemble model: a regional investigation of Syria. *Eng Appl Comput Fluid Mech* 14(1):90–106. <https://doi.org/10.1080/19942060.2019.1683076>
- Jang JSR (1993) ANFIS: Adaptive-Neural-Network-Based Fuzzy Inference System. *IEEE Trans Syst Man Cybern* 23(3):665–685. <https://doi.org/10.1109/21.256541>
- Joyce RJ, Janowiak JE, Arkin PA, Xie P (2004) CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J Hydrometeorol* 5(3):487–503. [https://doi.org/10.1175/1525-7541\(2004\)005<0487:CAMTPG>2.0.CO;2](https://doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2)
- Kalteh AM (2013) Monthly river flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform. *Comput Geosci* 54:1–8
- Kiran NR, Ravi V (2008) Software reliability prediction by soft computing techniques. *J Syst Softw* 81(4):576–583
- Legates DR, McCabe GJ (1999) Evaluating the use of ‘goodness-of-fit’ measures in hydrologic and hydroclimatic model validation. *Water Resour Res* 35(1):233–241
- Li D, Christakos G, Ding X, Wu J (2018) Adequacy of TRMM satellite rainfall data in driving the SWAT modeling of Tiaoxi catchment (Taihu lake basin, China). *J Hydrol* 556:1139–1152. <https://doi.org/10.1016/j.jhydrol.2017.01.006>
- Makwana JJ, Tiwari MK (2014) Intermittent Streamflow Forecasting and Extreme Event Modelling using Wavelet based Artificial Neural Networks. 4857–4873. <https://doi.org/10.1007/s11269-014-0781-1>
- Mislan H, Hardwinarto S, Sumaryono, Aipassa M (2015) Rainfall monthly prediction based on artificial neural network: a case study in tenggarong station, East Kalimantan - Indonesia. *Procedia Comput Sci* 59:142–151. <https://doi.org/10.1016/j.procs.2015.07.528>
- Moghaddamnia A, Ghafari Gousheh M, Piri J, Amin S, Han D (2009) Evaporation estimation using artificial neural networks and adaptive neuro-fuzzy inference system technique. *Adv Water Resour* 32(1): 88–97
- Noori N, Kalin L (2016) Coupling SWAT and ANN models for enhanced daily streamflow prediction. *J Hydrol* 533:141–151
- Nourani V, Sayyah Fard M (2012) Sensitivity analysis of the artificial neural network outputs in simulation of the evaporation process at different climatologic regimes. *Adv Eng Softw* 47(1):127–146. <https://doi.org/10.1016/j.advengsoft.2011.12.014>
- Nourani V, Elkiran G, Abdullahi J (2019) Multi-station artificial intelligence based ensemble modeling of reference evapotranspiration using pan evaporation measurements. *J Hydrol* 577(March): 123958. <https://doi.org/10.1016/j.jhydrol.2019.123958>
- Nourani V, Gökçekuş H, Umar IK (2020) Artificial intelligence based ensemble model for prediction of vehicular traffic noise. *Environ Res* 180:108852. <https://doi.org/10.1016/j.envres.2019.108852>
- Ochoa A, Pineda L, Crespo P, Willems P (2014) Evaluation of TRMM 3B42 precipitation estimates and WRF retrospective precipitation simulation over the Pacific-Andean region of Ecuador and Peru. *Hydrol Earth Syst Sci* 18:3179–3193. <https://doi.org/10.5194/hess-18-3179-2014>
- Prakash S, Mitra AK, Aghakouchak A, Liu Z, Norouzi H, Pai DS (2018) A preliminary assessment of GPM-based multi-satellite precipitation estimates over a monsoon dominated region. *J Hydrol* 556: 865–876. <https://doi.org/10.1016/j.jhydrol.2016.01.029>
- Ribeiro VHA, Reynoso-Meza G, Siqueira HV (2020) Multi-objective ensembles of echo state networks and extreme learning machines for streamflow series forecasting. *Eng Appl Artif Intell* 95:103910. <https://doi.org/10.1016/j.engappai.2020.103910>
- Rudolf B, Hauschild H, Rueth W, Schneider U (1994) Terrestrial Precipitation Analysis: Operational Method and Required Density of Point Measurements. In: *Global Precipitations and Climate Change*. Springer. https://doi.org/10.1007/978-3-642-79268-7_10
- Shamseldin AY (2006) Topics related to rainfall-runoff models. In *River Basin Modelling for Flood Risk Mitigation*. (D. W. K. & A. Y. Shamseldin (ed.)). Taylor and Francis
- Shamseldin AY (2010) Artificial neural network model for river flow forecasting in a developing country. 22–35. <https://doi.org/10.2166/hydro.2010.027>
- Shamseldin AY, Connor KMO (1999) A real-time combination method for the outputs of different rainfall-runoff models. *Hydrol Sci J* 44(6):895–912. <https://doi.org/10.1080/02626669909492288>
- Shamshirband S, Jafari Nodoushan E, Adolf JE, Abdul Manaf A, Mosavi A, Chau KW (2019) Ensemble models with uncertainty analysis for multi-day ahead forecasting of chlorophyll a concentration in coastal waters. *Eng Appl Comput Fluid Mech* 13(1):91–101. <https://doi.org/10.1080/19942060.2018.1553742>
- Sharghi E, Nourani V, Behfar N (2018) Earthfill dam seepage analysis using ensemble artificial intelligence based modeling. *J Hydroinf* jh2018151. <https://doi.org/10.2166/hydro.2018.151>
- Shiru MS, Park I (2020) Comparison of ensembles projections of rainfall from four bias correction methods over nigeria. *Water (Switzerland)* 12(11):1–16. <https://doi.org/10.3390/w12113044>
- Singh H, Sankarasubramanian A (2014) Systematic uncertainty reduction strategies for developing streamflow forecasts utilizing multiple climate models and hydrologic models. *Water Resour Res* 50(2): 1288–1307. <https://doi.org/10.1002/2013WR013855>
- Talei A, Hock L, Chua C, Quek C (2010) A novel application of a neuro-fuzzy computational technique in event-based rainfall – runoff modeling. *Expert Syst Appl* 37(12):7456–7468. <https://doi.org/10.1016/j.eswa.2010.04.015>
- Tang L, Hossain F (2012) Investigating the similarity of satellite rainfall error metrics as a function of Koppen climate classification. *Atmos Res* 104(105):182–192
- Taormina R, Chau K (2015a) Neural network river forecasting with multi-objective fully informed particle swarm optimization. *J Hydroinf* 17:99–113
- Taormina R, Chau KW (2015b) Data-driven input variable selection for rainfall-runoff modeling using binary-coded particle swarm optimization and Extreme Learning Machines. *J Hydrol* 529:1617–1632. <https://doi.org/10.1016/j.jhydrol.2015.08.022>
- Tongal H, Booij MJ (2018) Simulation and forecasting of streamflows using machine learning models coupled with base flow separation. *J Hydrol* 564:266–282. <https://doi.org/10.1016/j.jhydrol.2018.07.004>
- Tunkiel AT, Sui D, Wiktorski T (2020) Data-driven sensitivity analysis of complex machine learning models: A case study of directional drilling. *J Pet Sci Eng* 195:107630. <https://doi.org/10.1016/j.petrol.2020.107630>
- Venkata Ramana R, Krishna B, Kumar SR, Pandey NG (2013) Monthly Rainfall Prediction Using Wavelet Neural Network Analysis. *Water Resour Manag* 27(10):3697–3711. <https://doi.org/10.1007/s11269-013-0374-4>
- Wale A, Rientjes THM, Gieske ASM, Getachew HA (2009) Ungauged catchment contributions to Lake Tana’s water balance. *Hydrol Process* 23(26):3682–3693

- Wang H, Hu D (2005) Comparison of SVM and LS-SVM for regression. *Neural Netw Brain* 1:2079–2283
- Wang WC, Xu DM, Chau KW, Chen S (2013) Improved annual rainfall-runoff forecasting using PSO–SVM model based on EEMD. *J Hydroinf* 15(4):1377–1390
- Wen X, Si J, He Z, Shao H (2015) Support-Vector-Machine-Based Models for Modeling Daily Reference Evapotranspiration With Limited Climatic Data in Extreme Arid Regions. *Water Resour Manag* 29(July):3195–3209. <https://doi.org/10.1007/s11269-015-0990-2>
- Wu CL, Chau KW, Li YS (2009) Methods to improve neural network performance in daily flows prediction. *J Hydrol* 372:80–93
- Yaseen ZM, Ebtehaj I, Bonakdari H, Deo RC, Danandeh Mehr A, Mohtar WHMW, Diop L, El-shafie A, Singh VP (2017) Novel approach for streamflow forecasting using a hybrid ANFIS-FFA model. *J Hydrol* 554:263–276. <https://doi.org/10.1016/j.jhydrol.2017.09.007>
- Yong B, Liu D, Gourley JJ, Tian Y, Huffman GJ, Ren L, Hong Y (2015) Global view of real-time TRMM multisatellite precipitation analysis: implications for its successor global precipitation measurement mission. *Bull Am Meteorol Soc* 96(283–296):283–296

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