Applied Water Science (2023) 13:110
https://doi.org/10.1007/s13201-023-01913-6

ORIGINAL ARTICLE

Application of novel binary optimized machine learning models for monthly streamflow prediction

Rana Muhammad Adnan[1] · Hong-Liang Dai[1] · Reham R. Mostafa[2] · Abu Reza Md. Towfiqul Islam[3,8] · Ozgur Kisi[4,5] · Ahmed Elbeltagi[6] · Mohammad Zounemat-Kermani[7]

Received: 1 August 2022 / Accepted: 20 March 2023 / Published online: 8 April 2023

Abstract

Accurate measurements of available water resources play a key role in achieving a sustainable environment of a society. Precise river flow estimation is an essential task for optimal use of hydropower generation, flood forecasting, and best utilization of water resources in river engineering. The current paper presents the development and verification of the prediction abilities of new hybrid extreme learning machine (ELM)-based models coupling with metaheuristic methods, e.g., Particle swarm optimization (PSO), Mayfly optimization algorithm (MOA), Grey wolf optimization (GWO), and simulated annealing (SA) for monthly streamflow prediction. Prediction precision of standalone ELM model was compared with two-phase optimized state-of-the-arts models, e.g., ELM–PSO, ELM–MOA, ELM–PSOGWO, and ELM–SAMOA, respectively. Hydrometeorological data acquired from Gorai and Padma Hardinge Bridge stations at Padma River Basin, northwestern Bangladesh, were utilized as inputs in this study to employ models in the form of seven different input combinations. The model's performances are appraised using Nash–Sutcliffe efficiency, root-mean-square-error (RMSE), mean absolute error, mean absolute percentage error and determination coefficient. The tested results of both stations reported that the ELM–SAMOA and ELM–PSOGWO models offered the best accuracy in the prediction of monthly streamflows compared to ELM–PSO, ELM–MOA, and ELM models. Based on the local data, the ELM–SAMOA reduced the RMSE of ELM, ELM–PSO, ELM–MOA, and ELM–PSOGWO by 31%, 27%, 19%, and 14% for the Gorai station and by 29%, 27%, 19%, and 14% for Padma Hardinge bridge station, in the testing stage, respectively. In contrast, based on external data, ELM–PSOGWO improves in RMSE of ELM, ELM–PSO, ELM–MOA, and ELM–SAMOA by 20%, 5.1%, 6.2%, and 4.6% in the testing stage, respectively. The results confirmed the superiority of two-phase optimized ELM–SAMOA and ELM–PSOGWO models over a single ELM model. The overall results suggest that ELM–SAMOA and ELM–PSOGWO models can be successfully applied in modeling monthly streamflow prediction with either local or external hydro-meteorological datasets.

Keywords Streamflow prediction · Extreme learning machine · Particle swarm optimization · Grey wolf optimization · Simulated annealing

Introduction

Accurate and reliable streamflow prediction is of great importance in designing, developing, and proper operation of water resources management, hydropower generation, and ecohydrological studies and plans. In this sense, various and vigorous modeling strategies have been introduced and implemented in the literature, such as numerical-based (Suiju and Minquan 2015), autoregressive moving average models (Liu et al. 2015), and conceptual models (Boulariah et al. 2019).

During the last decades, with the advent of high-speed processors and the development of novel Soft Computing (SC) techniques, a new era of data mining methods, e.g., Machine Learning (ML) models for modeling and simulating different phenomena in hydrology, has been emerged (Najafzadeh and Niazmardi 2021; Najafzadeh et al. 2021; Zounemat-Kermani et al. 2020a, b; Najafzadeh et al. 2021; Barzkar et al. 2022; Granata et al. 2022; Ikram et al. 2023; Mostafa et al. 2023). Accordingly, streamflow prediction

✉ Hong-Liang Dai
hldai618@gzhu.edu.cn

✉ Ozgur Kisi
ozgur.kisi@th-luebeck.de

Extended author information available on the last page of the article

using SC techniques and data-driven models has been one of the main targets for hydrologists and researchers in recent years (Adnan et al. 2020a). These models use ML techniques to map the complicated relationship between the predictors (like precipitation and/or previous observed discharge values) and the response variable (short-term or long-term streamflow) (Adnan et al. 2020b). Poul et al. (2019) challenged four different types of data-driven models, including Multi-Linear Regression (MLR), Artificial Neural Network (ANN), K-nearest neighbors, and Adaptive Neuro-Fuzzy Inference System (ANFIS) in monthly streamflow prediction based upon six input combinations. In general, the ANFIS and ANN models were slightly superior to the other methods. Fu et al. (2020) applied deep learning ML models for the prediction of daily streamflow in a tropical environment. They compared the results of the Long Short-Term Memory (LSTM), as the deep network, to the conventional backpropagation ANN. The findings of the study revealed the obvious advantages of the LSTM over the classic ANN.

It should be noted that among the successful applications of various types of ML models, the Extreme Learning Machine (ELM), which presents a simple structure yet fast learning speed single-layer ANN, has proven its accuracy and efficiency in simulating complex and nonlinear hydrological problems (Rezaie-Balf and Kisi 2018; Yaseen et al. 2020; Saberi-Movahed et al. 2020; Ikram et al. 2022).

In a study, Niu et al. (2020) utilized the ELM for the annual prediction of streamflow using data series of three hydropower reservoirs in China. It was shown that the developed ELM model outperformed several traditional methods. According to the previous researches, the majority of ML applications for streamflow prediction confirm the proper potential and accuracy of ML in comparison with the conventional statistical or conceptual models in addressing this hydrological issue (Riahi-Madvar et al. 2019). Subsequently, new ideas for upgrading and enriching ML models, especially using integrative (hybrid) ML-heuristic methods, have been being developed and tested recently (Fadaee et al. 2020).

In the field of hydrology, several researchers have reported the successful application of heuristic algorithms embedded with ML models. Sudheer et al. (2014) developed a hybrid (integrative) ML model using Support Vector Regression (SVR) and Particle Swarm Optimization (PSO) for forecasting monthly streamflow in the USA. Analyzing the forecasted results indicated that the SVR–PSO model gave a better performance than the ANN and statistical model, namely ARMA (autoregressive moving average model). Malik et al. (2020) employed several integrative models based on the SVR and six heuristic algorithms in predicting daily streamflow in India. Comparing the results showed that SVR–HHO (SVR and Harris Hawks Optimization algorithm) during calibration/validation periods had superior performance to the other developed integrative models. Jiang et al. (2020) developed an integrative (hybrid) ELM–PSO model for monthly streamflow forecasting. The findings of the study demonstrated the superiority of the ELM–PSO over standard ELM, SVR, and ANFIS models.

Considering conventional heuristic algorithms, the PSO and Simulated Annealing (SA) are among the most widely used optimization algorithms in hydrology, and numerous applications have been applied using these algorithms (Bazargan and Norouzi 2018; Hosseini et al. 2020). Recently, some new heuristic algorithms have been introduced and claimed with high capability in optimizing sophisticated problems, such as Mayfly Optimization Algorithm (MOA) (Zervoudakis and Tsafarakis 2020), Grey Wolf Optimization (GWO) (Mirjalili et al. 2014), and hybrid PSOGWO (Şenel et al. 2019). However, there has not been a report on the comprehensive assessment of the capabilities of these algorithms in dealing with hydrological issues compared to the conventional ones (here, the PSO and SA). In this regard, this study utilizes the extreme learning machine (ELM) as the main model embedded with conventional (PSO and SA) and novel (GWO, hybrid PSOGWO, MOA, and hybrid SAMOA) as predictive models to realize streamflow in subhumid region of Bangladesh.

The main contribution and innovation of this study lie in (1) developing four integrative ELM models and analyzing their performance on simulation of a complicated engineering problem, namely streamflow prediction, and (2) appraisal of all developed models based on seven different input combinations scenarios including lagged streamflow values, precipitation, and temperature. To have a clear comprehension, these new approaches include simple (ELM–MOA and ELM–GWO) and hybrid integrative (ELM–PSOGWO and ELM–SAMOA) ML models. Continuously, the results of the new mention models will be compared with the standard ELM and integrative ELM–PSO and ELM–SA to achieve an inclusive perception for their usefulness and accuracy.

Case study

For the current research, the Padma river, a large river basin, is the major downstream stretch of the Ganges River, which runs more than 2561 km^2 derived from the Gangotri glacier of the Himalayan system which was chosen as a case study. This river basin is regarded as one of the highly populated residents on the globe. The Padma River acts a crucial role in the socio-economic conditions of the country. The stretch of the Padma river flows for 108 km^2 before the confluence with Meghna River at Chandpur point. Gorai River is one of the tributaries of the Padma River. The accumulated discharge of the Padma and Brahmaputra River is 30,000 m^3/s^{-1}, and sometimes, it reaches 75,000 m^3/s^{-1} during the bank full phase (Dewan et al. 2017). Geographically, the study basin is positioned between 23° 48′ and 25° 18′ north latitudes and 88° 27′ and 89° 48′ east longitudes. Annually, 900 metric tons of sediment load passes through the river, out of which, 60% of sediment is either silt or clay, while the rest is bed load (Islam 2016). Islam et al. (2021) stated the floodplain of the river as a 'wandering' form. Padma River basin is vital for sustaining livelihood through agricultural activities, navigation, nourishment, aquaculture, and environmental sustainability perspective. For instance, the freshwater distributed by this basin is truly imperative to sustaining a riparian ecosystem of the south-western region of Bangladesh, mostly in the world's largest mangrove forest, the Sundarbans, by holding the salinity anterior downstream side into the Bay of Bengal (BOB) (Mirza 2004). Apart from this, this basin reports extreme variability of flow regime (water and sediment), triggering from monsoonal precipitation and the melting of the Himalayan ice, which causes frequent large floods of high magnitude in Bangladesh. Further, bank erosion and river shifting are common phenomena in this basin, which led to environmental degradation and population migration. This is anticipated to enhance in forthcoming years (CDMP 2014) since elevated precipitation triggered by climate change will raise the runoff into the Ganges–Brahmaputra–Meghna River systems (Moors et al. 2011). Four seasons, such as summer (March–May), monsoon (June–September), post-monsoon (October–November), and winter (December–February), have predominated in this region with significant temperature and precipitation differences (Akhter et al. 2019). Runoff in the Padma river was mostly distributed in June–October, accounting for 72.5% of the accumulated annual runoff. The dry season is from December to April, accounting for 12.1% of the accumulated annual runoff. From December to February, it was the lowest, accounting for solely 5.87% of the annual total (Islam et al. 2016). Runoff in this river basin mostly originates from heavy rainfall. Therefore, in the current study, two hydro-meteorological sites on the Padma River basin, e.g., Padma Hardinge bridge and Gorai stations were employed for predicting monthly streamflow (see Fig. 1). The data adopted in this work were collected from Bangladesh Water Development Board (BWDB).

Methods

Extreme learning machine (ELM)

ELM is a form of the single-layer feed-forwarding network (FFN) that aims to intertwine the traditional neural networks and biological learning mechanism. Due to its special structure based on the random hidden neurons mechanism—in

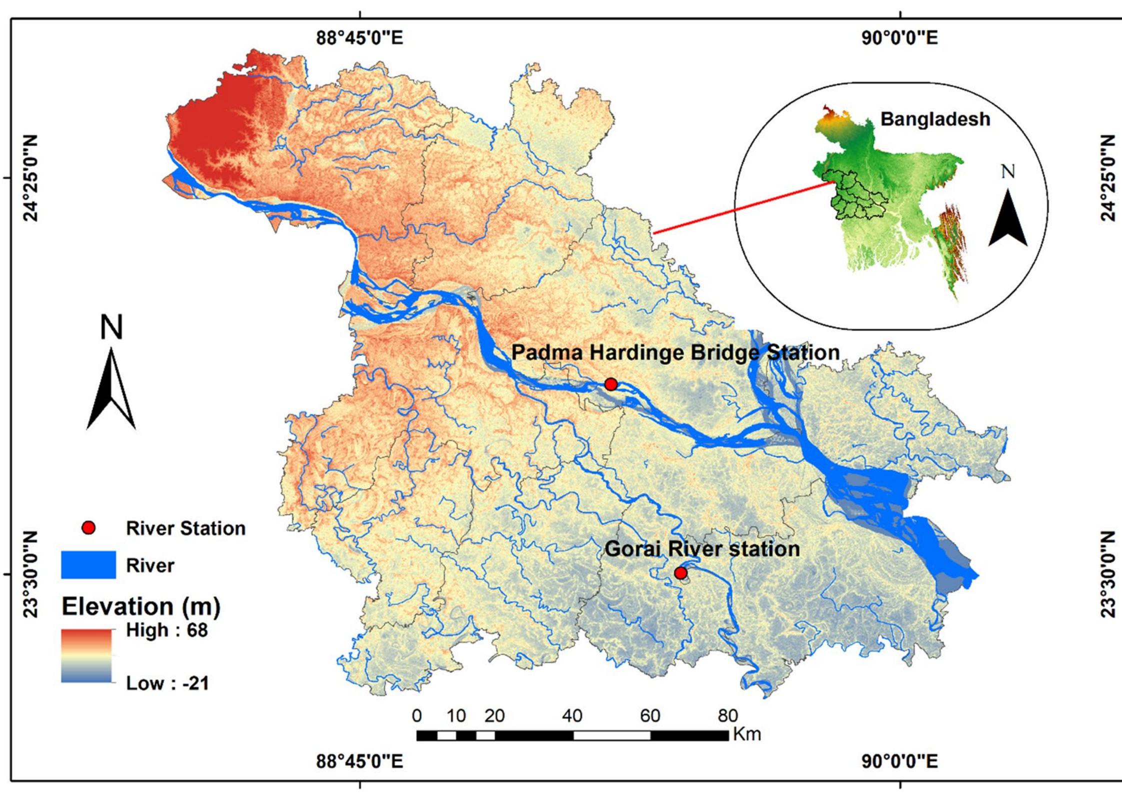

Fig. 1 Study area

which hidden neurons do not need to be tuned similar to the conventional ANNs—it can provide precise modeling results having a lower computational cost. Moreover, ELM offers other advantages such as ease of implementation, better generalization ability, and minimal human intervention. As a result of the mentioned arguments and reasons, ELM has been chosen as the core machine learning tools in this study. The main mathematical methodology of ELM is described in the following.

As can be seen in Fig. 2, a single hidden layer resides in the ELM network. The input weights between the input and hidden layers are only initialized once and do not need to be conditioned (Zhu et al. 2019; Prasad et al. 2018; Yadav et al. 2017; Zhang et al. 2020). Iterative testing is used to prepare the outcome weights between the un-seen and outcome layers. The training and process time of an ELM is much quicker than that of the comparable single-layer FFN since the input weights remain in their initial state and only the output weights are trained. Huang et al. (2006) suggested the ELM, a basic three-layer structure algorithm, to address the shortcomings of conventional soft computing techniques. The input weight and the bias values are generated at random in the ELM structure. ELM uses a basic simplified inverse operation of the hidden layer output matrix to measure the output weight matrix between hidden and output layers analytically (Fan et al. 2018; Zhu et al. 2020). The ELM is a promising time series prediction method because of its interpolation and uniform approximation capabilities. ELM can be represented mathematically as a function with L hidden nodes and N training data, as seen in Huang et al., (2006):

$$\sum_{i=1}^{L} w_i g\left(W_{\mathrm{in}(i)}, b_i, x_j\right) = \sum_{i=1}^{L} w_i g\left(W_{\mathrm{in}(i)}, x_j + b_i\right) = Y_j,\ j = 1, \ldots, N. \tag{1}$$

where x_j represent the input vector, $W_{\mathrm{in}(i)}$ denotes the weight vtor of the input, $W_{\mathrm{in}(i)}.x_j$ corresponds the inner product of $W_{\mathrm{in}(i)}$ and x_j, b_i denotes the bias of the ith hidden node, $g(\bullet)$ denotes the sigmoid function, w_i refers to the weight matrix of the output, and y_j is the modeled output of the ELM. At the start of the ELM algorithm, the input weight and bias are selected at random.

Particle swarm optimization (PSO)

The PSO algorithm is a type of optimization algorithm to simulate the behavior of swarm intelligence (Qi et al. 2018; Xi et al. 2021). It is an efficient globally optimized search algorithm through the system intelligence guidelines that comes from the collaboration and rivalry between parts of the group. In the research space of N-dimensional, it can specifically be represented randomly as a particle group (the quantity is m) and each implemented a special structure of ANN (Chen et al. 2020; Wang et al. 2018). The primary architectural theory of PSO is closely related to two studies: the first is the evolutionary algorithm; the PSO swarm mode where the mostly optimized objective solution space can be searched simultaneously. The other is synthetic life and in particular the examination of artificial systems with life-characteristics. The efficiency of the particles location was measured by the statistical error in the training phase. More specifically, at generating the lowest values of MSE, the ANN structure is defined by a particle with a superlative performance (Elsheikh and Abd Elaziz 2019). The next swarm was produced based on updating the particles' locations, which takes into account the best position of swarm and particle. Particle swarms gradually shifted to the optimal location up to the maximum number of iterations. The velocity and direction of the particle can be modified as follows:

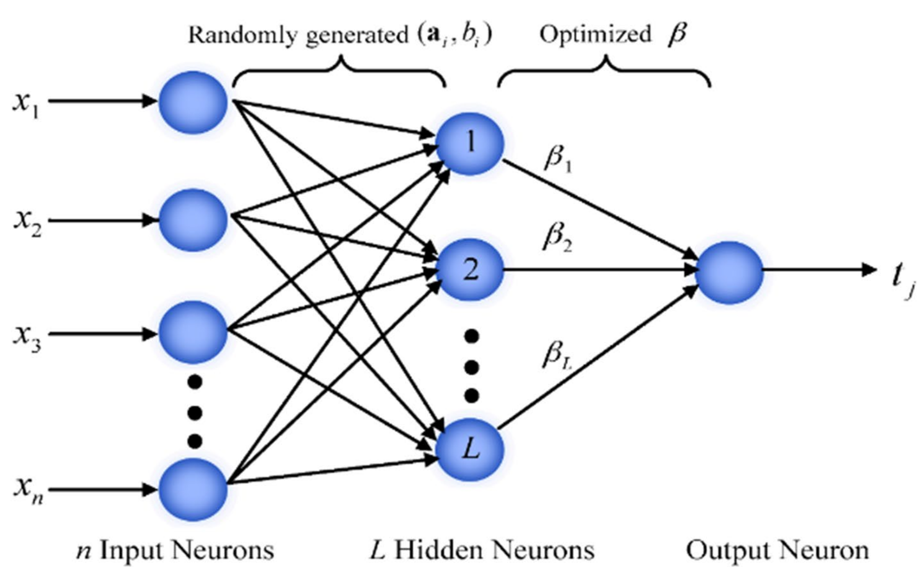

Fig. 2 ELM structure

$$V_i^{k+1} = \omega \times V_i^k + c_1 \times \text{rand}() \times \left(P_i^k - X_i^k\right) + c_2 \times \text{rand}() \times \left(G_i^k - X_i^k\right) \quad (2)$$

$$X_i^{k+1} = X_i^k + V_i^{k+1} \quad (3)$$

where V_i^{k+1} and V_i^k denote the particle (i) velocity at iteration (k) and $(k+1)$, rand () represents a random value between 0 and 1. c_1, c_2 are called a learning or acceleration factor and equaled 2 for both in this work. ω refers to the coefficient of inertia. P and G represent the best place for a particle and a swarm, respectively. In the present work, number of iterations, population, and runs were 100, 25 and 7, respectively. ω ranged from 0.2 to 0.9.

Mayfly optimization algorithm (MOA)

The suggested process of optimization considers as a PSO adjustment and the combination of main PSO, genetic and firefly algorithms benefits (Gao et al. 2020; Zervoudakis and Tsafarakis 2020). In reality, it offers for researchers who seek to advance the efficiency of the PSO algorithm with certain techniques and local search, a powerful hybrid algorithm, dependent upon the mayflies' actions, as PSO has proved that modifications are required to ensure optimum performance in high-dimensional areas. The algorithm functions like this. Initially, the random generation of two groups of mayflies is male and female, respectively (Mansouri et al. 2019). The individual fly can be put on a random basis as a d-dimensional vector solution $x=(x_1,\ldots,x_d)$ in problem space, and its output is assessed on the previous definition of objective function $f(x)$. The velocity of the mayfly $v=(v_1,\ldots,v_d)$ is known as the shift of its path, and each mayfly's flying path is a complex relationship between individual and social experience. In specific, they will change their direction to their best (P_{best}) location and the best position achieved by any swarm flying (G_{best}) (Haddad et al. 2006). In Swarms, male's mayflies will continue to investigate or manipulate during iterations. The speed will be modified to their present fitness data and historical optimal values of fitness (Chen and Shi 2019; Mansouri et al. 2019; Zhou et al. 2018). The females will update their speeds in a particular way. In biological terms, the female only live with wings in one day to at most seven days, so the female mayflies need to search the male mayflies to intermarriage and spread. They will then update their speeds on the basis of their masculine flies. The top best male and female mayflies are regarded as the first mate and the second best female/male mayflies as the second mates and so on. In the present study, the parameters used for this algorithm are presented in Table 1.

Simulated annealing algorithm (SA)

The SA is an algorithm that belongs to meta-heuristic methods and is known also as Monte Carlo Annealing (Shao and Zuo 2020; Tufano et al. 2020; Turhan and Bilgen 2020). The SA algorithm is a physical annealing method based on simulation that has successfully extended to multiple dynamic problem optimization. The concept behind the SA stems from thermodynamics and illustrates how to mirror the solidifying mechanism of the fluid into a crystalline solid (Abdel-Basset et al. 2021; Redi et al. 2020; Silva et al. 2020; Tang et al. 2020). Annealing is a physical mechanism utilized to tend to harden metals beginning at maximum temperature and cool off slowly. At first, SA parameters including the initial temperature T_0, final temperature T_{final} and cooling rate are initialized. The original temperature is the maximum and will be gradually refreshed using the rate of cooling before the end temperature is reached (Ben Messaoud 2020; Meng et al. 2020). A sharp decline in temperature causes the molecules to select the right location, normally not the best possible case, therefore quench the object. The algorithm starts with a randomly created solution. The approach is based on progressive improvement of the present solution. During iterations process, a new neighboring solution for the current solution is chosen. If the latest neighboring solution is stronger, the existing solution is modified. In addition, when the adjacent solution is stronger, the optimal solution is refreshed. When the end temperature is hit, the model stops (Liu et al. 2020; Zhao et al. 2020). For each iteration (), the actual temperature T is modified by:

$$T = T * \tau,\ 0 < \tau < 1 \quad (4)$$

Table 1 Parameters setting of each optimization algorithms

PSO	Cognitive component (c_1)	2
	Social component (c_2)	2
	inertia weight	0.2–0.9
MOA	positive attraction constants (a_1)	3
	positive attraction constants (a_2)	3.5
	Visibility coefficient (β)	0.1
	gravitational coefficient (g)	0.98
	Initial nuptial dance coefficient d_0	3
	Initial random walk coefficient fl_0	3
	Random value crossover r_{of}	0.95
SA	Initial temperature (T_0)	100
	Final temperature (T_f)	0.01
	Cooling rate τ	0.7
GWO	a	Decreased from 2 to 0
PSOGWO	As in both PSO and GWO	
SAMOA	As in both SA and MOA	
All algorithms	Population	25
	Number of iterations	100
	Number of runs for each Algorithm	7

Boltzmann distribution gives the probability according to mathematical thermodynamics that a molecule is at a certain energy degree as follows:

$$P(\varepsilon_i) = \frac{\exp^{-\varepsilon_{i/kT}}}{\sum_{j=1}^{m} \exp^{-\varepsilon_{j/kT}}} \tag{5}$$

where ε_i is the energy at the state (i), k represents the constant of Boltzmann, T equals the thermodynamic temperature and m denotes the overall states number. In this study, the initial and final temperatures were 100 and 0.01, respectively, and the rate of cooling was 0.7 (Table 1).

Grey wolf optimizer (GWO)

GWO is an advanced successful introduced swarm intelligence algorithm suggested by Mirjalili et al. (2014) in last decade. The kind of algorithms is based on the imitation of the social order and hunting activities of the Grey Wolf Herd. GW is purely hierarchical social species comprising $\propto$, β, δ, and ω. In this respect, $\propto$ is the leader who distributes several assignments to individuals at various levels to achieve global optimization. Due to its simple structure, insignificant parameter change required and high accuracy, the GWO algorithm has always been used for functional optimization (Dehghani et al. 2019; Himanshu et al. 2020; Wang et al. 2018). The location of the ith wolf is defined as $X_i = X^I{}_1, X^I{}_2\ldots, X^I{}_d$. $X^I{}_d$ stands for the position of the ith wolf is d-dimensional space, for a population containing N grey wolves ($X = X_1, X_2,\ldots, X_N$). The specific role of hunting is as follows:

$$D = \left|C * X_i(t) - X(t)\right| \tag{6}$$

$$X_i(t+1) = X(t) - A * D \tag{7}$$

where A and C stand for the coefficient vectors, t is number of iterations, $X(t)$ is the position vector of GW, $X_i\,(t)$ is target position vector of the GW, D corresponds to the distance between the prey and the grey wolf.

Defining the coefficient vector as follows:

$$A = 2 * a * r_1 - a \tag{8}$$

$$C = 2 * r_2 \tag{9}$$

$$a = 2 - i * \left(\frac{2}{Maximum\ iteration}\right) \tag{10}$$

where r_1 and r_2 correspond to the random vectors with values (0, 1) and a represents the iteration factor. In this study, the population, number of iterations, and number of runs for all algorithms were 25, 100, and 7, respectively, as demonstrated in Table 1. In addition, a decreased from 2 to 0. GW has strong food quest ability. $\propto$ is the boss who will serve in all activities and sometimes β and δ can participate. β and δ can also provide $\propto$ with successful goal information in the GWO as an ideal solution (Mohammadi et al. 2020; Tikhamarine et al. 2020, 2019; Yu and Lu 2018). Therefore, α, β, and δ are the three ideal alternatives in fact and their adjusted positions:

$$D_{\propto} = \left|C_1 * X_{\propto}(t) - X(t)\right| \tag{11}$$

$$D_{\beta} = \left|C_2 * X_{\beta}(t) - X(t)\right| \tag{12}$$

$$D_{\delta} = \left|C_3 * X_{\delta}(t) - X(t)\right| \tag{13}$$

$$X_1 = X_{\propto} - A_1 * D_{\propto} \tag{14}$$

$$X_2 = X_{\beta} - A_2 * D_{\beta} \tag{15}$$

$$X_3 = X_{\delta} - A_3 * D_{\delta} \tag{16}$$

$$X(t+1) = \frac{X_1 + X_2 + X_3}{3} \tag{17}$$

where X_α, X_β, and X_δ correspond to the current positions of the three optimal solutions α, β, and δ, respectively; $X(t)$ stands for the target position; D_α, D_β, and D_δ denote the distances from the prey to the three solutions, respectively; $X(t+1)$ is the position vector with updated searching factor; C and A represent the random vectors.

Hybrid PSO–GWO algorithm

Without modifying the overall operation of PSO and GWO algorithms, our hybrid PSO–GWO algorithm has been built. In nearly all real-world issues, the PSO can produce good results. Nevertheless, the PSO algorithm has to be reduced to minimal in the local solution. The GWO algorithm in our approach suggested follows the PSO algorithm in order to minimize the risk of slipping through a minimum local level (Şenel et al. 2019). The PSO algorithm leads those particles into the random spot, as discussed in PSO description section, with no potential to escape local minima. These directions can be risky for going away from the global minimum. The GWO algorithm's scanning ability can be used to avoid these threats by directing those particles into locations which are partly enhanced by the GWO algorithm rather than randomly guided. The runtime therefore is expanded, since besides the PSO algorithm, the GWO algorithm is still used. The longer time will be considered tolerable, based on the optimization problem solved, when the results are effective and the additional time required is considered. The improved success will accommodate extra time in the sector like leather, in which losses are much more important as long as the solution is accomplished in a proper time. The flowchart of this algorithm is presented in Fig. 3.

Proposed hybrid MOA–SA algorithm

For on-going optimization issues, the MOA algorithm was designed above. The new hybrid algorithm called MOA–SA is for fitness function problems of each particle in the current population and selecting the best particle and swam. Our hybrid MOA–SA algorithm was created without altering the overall operation of MOA and SA algorithms. Each MOA solution vector shall be transformed, i.e., only 0 s and 1 s and evaluated in binary form. The S-shaped transition mechanism is used for this conversion. The option of selecting a specific feature in a solution vector is given by this function. The initial temperature is the highest and will be steadily refreshed using the cooling rate to reach the final temperature, as discussed in the SA algorithm description. The algorithm begins with a solution generated at random. The method is focused on gradual changes to the existing approach. A new neighboring solution for the current solution is selected during the iteration process. The final steps

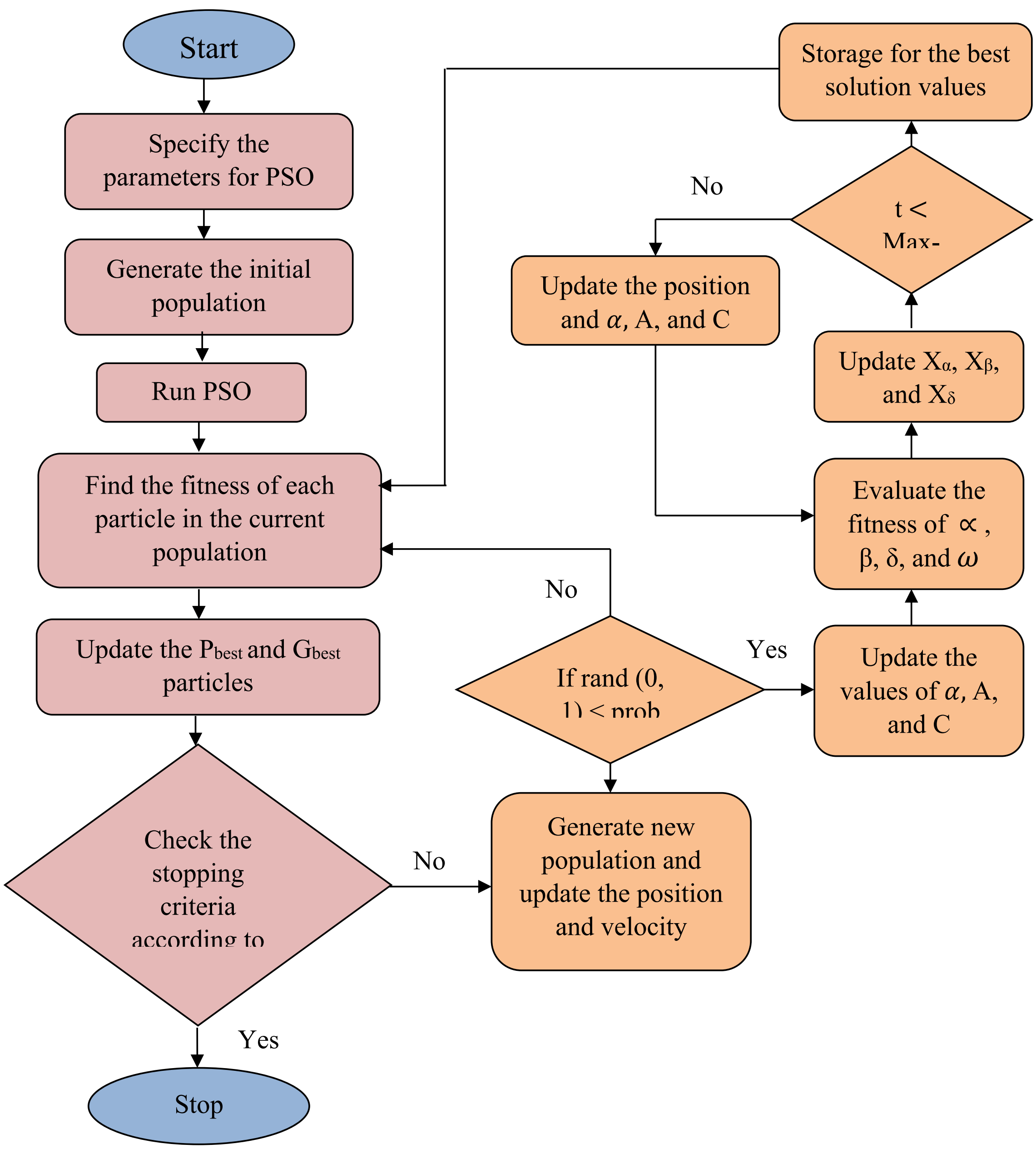

Fig. 3 Flowchart of the PSO–GWO Algorithm

are checking the iteration process and terminate the computational process and store the optimal solution. The flowchart for the new hybrid algorithm (MOA–SA) is shown in Fig. 4.

Application and results

The ability of two-phase optimized ELM models, ELM–PSOGWO and ELM–SAMOA, is investigated in monthly streamflow prediction. Various combinations of streamflow, precipitation and temperature data obtained from two stations, Pakistan, are used as model inputs. The outcomes acquired from the ELM–PSOGWO and ELM–SAMOA are compared with two single-phase optimized models, ELM–PSO, ELM–MOA, and standalone ELM. The following criteria are utilized in assessment of the employed models:

RMSE : Root Mean Square Error

$$= \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[(S_o)_i - (S_c)_i\right]^2} \tag{18}$$

$$\text{MAE : Mean Absolute Error } = \frac{1}{N}\sum_{i=1}^{N}\left|(S_o)_i - (S_c)_i\right| \tag{19}$$

MAPE : Mean Abs. Percentage Error

$$= \frac{1}{N}\sum_{i=1}^{N}\frac{|(S_o)_i - (S_c)_i| * 100}{(S_c)_i} \tag{20}$$

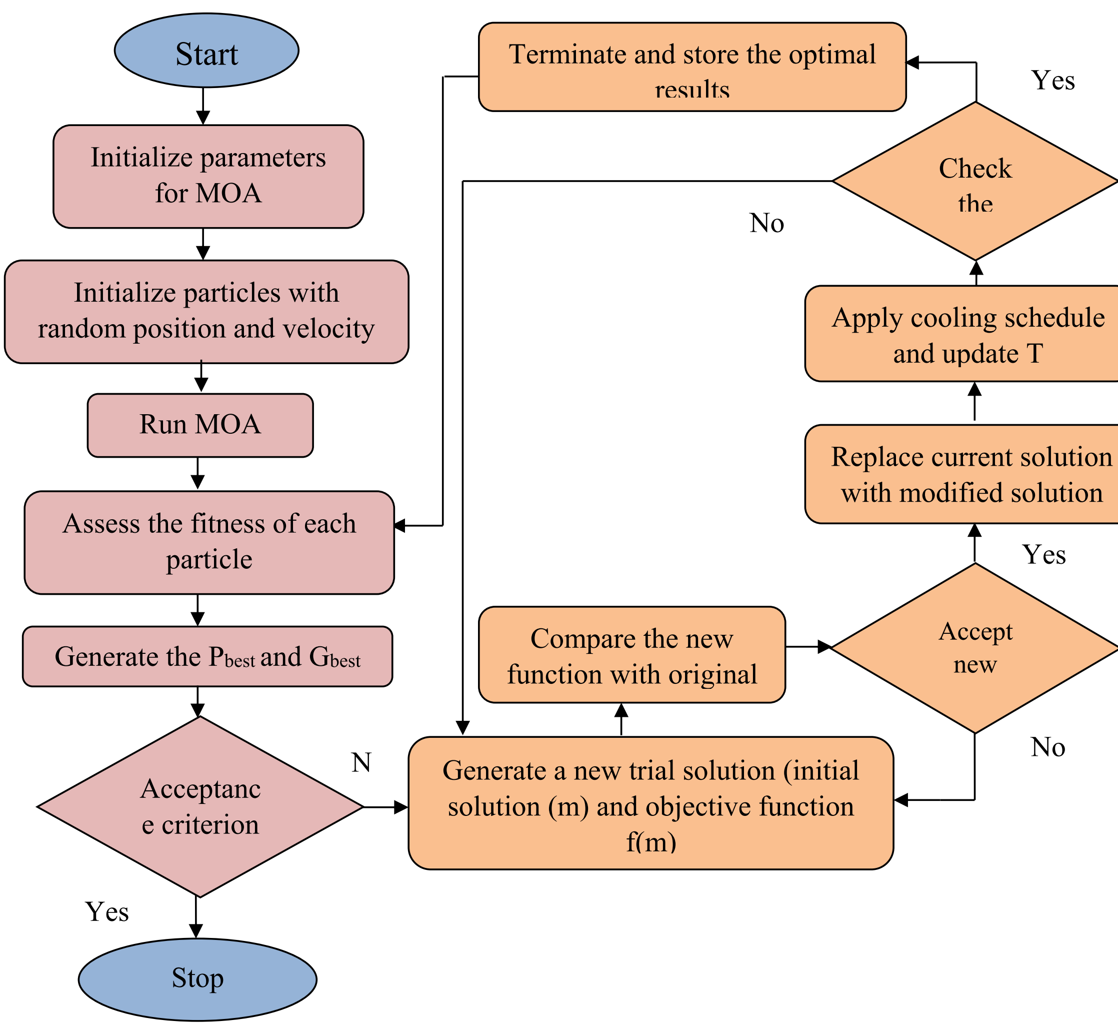

Fig. 4 Flowchart of the MOA–SA Algorithm

$$\text{NSE:Nash} - \text{Sutcliffe efficiency} = 1 - \frac{\sum_{i=1}^{N}\left[(S_o)_i - (S_c)_i\right]^2}{\sum_{i=1}^{N}\left[(S_o)_i - \overline{S}_o\right]^2}, -\infty < \text{NSE} \leq 1 \quad (21)$$

where $S_c, S_o, \overline{S}_o$ are calculated, observed and mean of the observed streamflows, respectively, and N is the quantity of the data. Several control parameter values were considered in model development phase, and the optimal ones were decided with respect to lowest square error (Najafzadeh and Niazmardi 2021; Najafzadeh et al. 2021). These values are listed in Table 1 for each algorithm. The table also shows the population and iteration numbers and the number of runs which are necessary to get more robust results from the meta-heuristic algorithms. Input combinations listed in Table 2 were decided taking into account the correlation analysis (autocorrelation and/or partial auto-correlation functions). As seen from Table 2, first three involve lagged streamflow values, in the inputs iv and v, precipitation data are imported and after temperature data are included in the input combinations.

Table 3 represents training and testing statistics of single-phase optimized, two-phase optimized and standalone ELM models in predicting streamflows of Padma Station. As it is clear from the table, two-phase optimized models are superior to the other models, while the standalone model has the worst results in streamflow prediction. It is also clear that the two-phase optimized ELM–SAMOA has the lowest RMSE, MAE and MAPE and the highest NSE and R^2 in both training (RMSE = 3448 m^3/s, MAE = 2282 m^3/s, NSE = 0.928, R^2 = 0.928, MAPE = 10.78%) and testing (RMSE = 4042 m^3/s, MAE = 2413 m^3/s, NSE = 0.896, R^2 = 0.897, MAPE = 15.24%) stages; an increase in RMSE of ELM, ELM–PSO, ELM–MOA, ELM–PSOGWO is by 31%, 27%, 19% and 14% applying the ELM-SAMOA in the test stage, respectively. All models' outcomes reveal that importing precipitation information deteriorates the accuracy while lagged temperature inputs improve the efficiency of the single-phase optimized ELM–MOA and the two-phase optimized ELM–SAMOA models (compare the inputs iii and iv/v). The RMSE and MAE of the ELM–SAMOA decrease from 4488 and 2419 m^3/s to 4042 and 2413 m^3/s by 11% and 0.2%; the NSE and R^2 increase from 0.872 and 0.878 to 0.896 and 0.897 by 3% and 2%.

Training and testing results of the ELM-based models in predicting streamflows of Gorai Station are compared in Table 4. Here, also the superiority of the two-phase optimized ELM models over other models and the standalone model has the last rank in accuracy. The ELM-SAMOA with third input combination produced lower RMSE (413.1 m^3/s), MAE (240.4 m^3/s), MAPE (18.94%) and higher NSE (0.855) and R^2 (0.857) than those of the best ELM with input iii (RMSE = 531.3 m^3/s, MAE = 289.6 m^3/s, NSE = 0.760, R^2 = 0.762, MAPE = 27.29%), ELM–PSO with input vii (RMSE = 523.38 m^3/s, MAE = 281.17 m^3/s, NSE = 0.767, R^2 = 0.788, MAPE = 27.15%), ELM–MOA with input iv (RMSE = 492.8 m^3/s, MAE = 276.2 m^3/s, NSE = 0.793, R^2 = 0.810, MAPE = 23.68%) and ELM–PSOGWO with input iv (RMSE = 470.5 m^3/s, MAE = 275.1 m^3/s, NSE = 0.811, R^2 = 0.819, MAPE = 23.17%) in the testing stage. By applying the two-phase optimized ELM–SAMOA, the RMSE was improved by 29%, 27%, 19% and 14% compared to ELM, ELM–PSO, ELM–MOA and ELM–PSOGWO, respectively. On the contrary to Padma Station, here adding precipitation information into the inputs improves the accuracy of the ELM–MOA (RMSE from 521.5 to 482.8 m^3/s, MAE from 281.17 to 276.2 m^3/s, NSE from 0.767 to 0.793 and R^2 from 0.788 to 0.810) and ELM–PSOGWO (RMSE from 600.6 to 470.5 m^3/s, MAE from 324.7 to 275.1 m^3/s, NSE from 0.693 to 0.811 and R^2 from 0.812 to 0.819). Importing temperature input slightly improves efficiency of ELM–PSO; from input combination iii to vii, the RMSE, MAE, NSE and R^2 were improved by 0.6%, 6.9%, 0.4% and 3% in the testing stage, respectively.

Figures 5 and 6 illustrate the time variation diagrams of the observed and predicted streamflows by the best ELM-based models for the Padma and Gorai stations, respectively. From the figures, the two-phase optimized ELM-SAMOA model appears to be better simulate the streamflows compared to other models while the standalone ELM cannot adequately catch the observed values in both stations. The scatter diagrams of the streamflow predictions are compared in Figs. 7 and 8 for two stations. From these graphs, it is also apparent that the ELM-SAMOA offers better accuracy with less scattered estimates compared to other alternatives. The implemented models are further compared on Taylor and violin diagrams in Figs. 9 and 10 for the Padma and Gorai stations. It is apparent from the Taylor graphs that the ELM–SAMOA has the closest standard deviation to observed one with the highest correlation and the lowest square error and it is followed by the other two-phase optimized ELM–PSOGW, single-phase optimized ELM–MOA,

Table 2 The input combinations used for model development

Input combinations	Variables
(i)	Qt-1
(ii)	Qt-1,Qt-11
(iii)	Qt-1,Qt-11,Qt-12
(iv)	Qt-1,Qt-11,Qt-12,Pt10
(v)	Qt-1,Qt-11,Qt-12,Pt10, Pt-11
(vi)	Qt-1,Qt-11,Qt-12,Pt10, Pt-11,Tt-9
(vii)	Qt-1,Qt-11,Qt-12,Pt10, Pt-11,Tt-9,Tt-10

Table 3 The results of the single-phase, two-phase optimized and standalone ELM models in prediction streamflows of Padma Station

Models	Input Combinations	Training					Testing				
		RMSE	MAE	NSE	R^2	MAPE	RMSE	MAE	NSE	R^2	MAPE
ELM	I	9559	6482	0.443	0.461	58.06	10,388	7105	0.313	0.336	71.04
	II	5933	3753	0.785	0.786	29.34	6215	3816	0.754	0.758	28.56
	III	**4947**	**2834**	**0.851**	**0.851**	**18.26**	**5295**	**3032**	**0.822**	**0.823**	**19.75**
	IV	4979	3093	0.849	0.852	18.17	5427	3195	0.813	0.814	21.38
	V	4805	3024	0.859	0.859	17.84	5647	3393	0.797	0.798	25.47
	VI	4957	3029	0.850	0.850	18.49	5405	3131	0.814	0.816	24.59
	VII	4549	2862	0.874	0.874	17.16	5402	3368	0.812	0.814	24.48
ELM-PSO	I	8916	6232	0.515	0.516	54.18	9412	6379	0.436	0.439	60.25
	II	5558	3620	0.812	0.813	23.59	5367	2952	0.817	0.823	22.37
	III	**4668**	**2760**	**0.867**	**0.867**	**17.48**	**5122**	**2885**	**0.833**	**0.836**	**19.08**
	IV	4738	2967	0.863	0.864	17.69	5390	3266	0.815	0.817	21.07
	V	4339	2670	0.885	0.887	16.27	5549	3198	0.804	0.805	24.29
	VI	4752	2874	0.862	0.862	17.16	5177	2987	0.829	0.834	23.97
	VII	4158	2624	0.895	0.899	16.05	5208	3183	0.827	0.828	24.05
ELM-MOA	I	8876	6291	0.520	0.520	53.89	9291	6628	0.450	0.460	53.18
	II	5473	3379	0.817	0.817	23.14	5303	3233	0.821	0.821	21.56
	III	4367	2566	0.884	0.885	16.25	4867	2823	0.849	0.851	18.49
	IV	4357	2615	0.884	0.886	16.20	5106	2794	0.834	0.848	19.67
	V	4139	2695	0.896	0.897	16.13	5474	3111	0.809	0.810	20.24
	VI	**4104**	**2577**	**0.897**	**0.899**	**16.02**	**4789**	**2733**	**0.854**	**0.855**	**18.05**
	VII	3874	2607	0.909	0.909	15.95	5011	2992	0.840	0.841	18.37
ELM-PSOGWO	I	8825	6235	0.525	0.525	52.76	8860	6199	0.500	0.501	52.49
	II	4963	2777	0.850	0.851	20.75	4855	3055	0.850	0.853	20.92
	III	**3918**	**2313**	**0.906**	**0.907**	**12.76**	**4622**	**2491**	**0.858**	**0.869**	**17.68**
	IV	4256	2708	0.890	0.891	14.68	4861	2777	0.850	0.861	17.85
	V	3742	2341	0.915	0.915	13.80	5395	3198	0.815	0.816	20.08
	VI	4031	2620	0.901	0.901	13.98	4719	2810	0.864	0.866	17.46
	VII	3515	2320	0.925	0.925	13.50	4991	3061	0.841	0.847	19.72
ELM-SAMOA	I	8740	6117	0.534	0.534	51.68	8837	6238	0.503	0.503	52.18
	II	4953	3091	0.850	0.852	19.49	4622	2503	0.864	0.872	17.49
	III	3997	2418	0.903	0.903	13.68	4488	2419	0.872	0.878	17.24
	IV	4087	2542	0.898	0.898	14.51	4794	2960	0.854	0.857	18.28
	V	3681	2286	0.917	0.918	12.24	5080	3023	0.836	0.837	19.37
	VI	3929	2617	0.906	0.906	12.85	4484	2586	0.872	0.872	17.56
	VII	**3448**	**2282**	**0.928**	**0.928**	**10.78**	**4042**	**2413**	**0.896**	**0.897**	**15.24**

Bold values defines the best accuracy (meaning the slowest RMSE and MAE and highest NSE and R2)

ELM–PSO and standard ELM models, respectively. It is clear from the violin diagrams that the distribution of the ELM–SAMOA predictions is closer to the observed one while the standard ELM model has the most different distribution. All the graphs justify the testing statistics provided in Tables 3 and 4 that the ELM–SAMOA acted better than the other models in prediction of monthly streamflows.

Table 5 compares the single-phase, two-phase optimized and standalone ELM models with respect to *t*-test for both stations. In the table, the statistics were calculated according to the significance level of 5% (two-tailed test). Higher *t*-statistics (*t*-stat) than the critical one shows that there is no significant difference in mean between the computed and observed data. The model having higher *t*-stat indicates better robustness. It is clearly seen from Table 5 that the two-phase optimized ELM–SAMOA has higher statistics compared to the other models in Padma and Gorai stations.

Table 4 The results of the single-phase, two-phase optimized and standalone ELM models in prediction streamflows of Gorai Station

Models	Input Combinations	Training					Testing				
		RMSE	MAE	NSE	R^2	MAPE	RMSE	MAE	NSE	R^2	MAPE
ELM	I	1063.7	777.3	0.520	0.523	50.28	868.5	565.0	0.358	0.424	59.34
	II	770.8	433.2	0.748	0.748	28.64	607.3	364.6	0.686	0.698	34.62
	III	**727.8**	**352.5**	**0.775**	**0.775**	**27.86**	**531.3**	**289.6**	**0.760**	**0.762**	**27.29**
	IV	658.4	359.4	0.816	0.817	22.62	566.8	330.7	0.726	0.737	30.38
	V	452.0	276.2	0.913	0.914	14.08	560.6	338.0	0.732	0.734	30.17
	VI	412.9	215.3	0.928	0.928	13.85	580.3	338.1	0.713	0.722	30.76
	VII	379.0	213.3	0.939	0.939	13.68	580.2	311.4	0.713	0.756	28.37
ELM-PSO	I	1046.2	718.6	0.535	0.535	49.78	843.00	553.14	0.395	0.452	57.82
	II	714.2	352.7	0.784	0.784	25.49	592.05	342.13	0.701	0.703	33.29
	III	705.2	338.2	0.789	0.789	25.36	526.56	300.63	0.764	0.764	27.08
	IV	531.7	296.7	0.880	0.880	18.27	539.94	311.35	0.752	0.756	29.84
	V	420.4	220.2	0.925	0.925	14.18	553.30	304.25	0.739	0.785	27.48
	VI	405.4	221.3	0.930	0.930	13.72	577.69	342.24	0.716	0.724	30.67
	VII	**378.8**	**213.2**	**0.939**	**0.939**	**13.65**	**523.38**	**281.17**	**0.767**	**0.788**	**27.15**
ELM-MOA	I	1028.0	704.5	0.551	0.552	48.27	826.9	564.3	0.418	0.459	56.37
	II	705.7	361.8	0.789	0.789	25.16	563.8	332.7	0.729	0.741	29.38
	III	579.1	262.4	0.858	0.858	20.38	521.5	326.4	0.768	0.772	26.84
	IV	**481.4**	**285.2**	**0.902**	**0.903**	**15.74**	**492.8**	**276.2**	**0.793**	**0.810**	**23.68**
	V	378.5	200.1	0.939	0.939	13.29	546.5	291.0	0.746	0.763	27.17
	VI	389.0	212.3	0.936	0.936	13.38	574.1	369.2	0.719	0.720	30.54
	VII	360.2	190.3	0.945	0.945	13.08	522.9	301.2	0.767	0.789	27.02
ELM-PSOGWO	I	1000.4	659.3	0.575	0.575	47.24	813.5	568.1	0.436	0.474	55.34
	II	655.1	304.5	0.818	0.818	23.65	538.7	297.3	0.753	0.765	27.61
	III	530.2	300.1	0.881	0.882	17.95	600.6	324.7	0.693	0.812	23.78
	IV	**434.2**	**222.3**	**0.920**	**0.920**	**13.91**	**470.5**	**275.1**	**0.811**	**0.819**	**23.17**
	V	365.9	192.4	0.943	0.944	13.12	535.0	295.8	0.756	0.780	26.19
	VI	388.0	210.8	0.936	0.936	13.34	567.0	305.7	0.726	0.796	25.87
	VII	306.8	168.9	0.960	0.960	12.85	483.8	287.1	0.801	0.811	24.38
ELM-SAMOA	I	998.4	656.5	0.577	0.577	46.39	807.1	565.3	0.445	0.484	54.82
	II	630.0	301.2	0.832	0.832	22.87	515.9	285.3	0.773	0.774	26.34
	III	**513.9**	**273.5**	**0.888**	**0.888**	**17.25**	**413.1**	**240.4**	**0.855**	**0.857**	**18.94**
	IV	410.8	224.8	0.928	0.928	13.82	442.1	259.0	0.834	0.836	17.96
	V	349.7	191.2	0.948	0.948	13.08	518.1	287.1	0.771	0.789	25.43
	VI	333.7	182.4	0.953	0.953	12.92	557.0	300.7	0.736	0.793	24.19
	VII	306.6	168.6	0.960	0.960	12.81	472.5	278.8	0.810	0.817	22.38

Bold values defines the best accuracy (meaning the slowest RMSE and MAE and highest NSE and R2)

Two-phase optimized ELM models were also compared with the single-phase optimized and standalone ELM models in estimating streamflows of Gorai Station (downstream) using data of Padma Station (upstream). Same input combinations were taken into account, and the model outcomes are listed in Table 6 with respect to RMSE, MAE, NSE, R^2 and MAPE. In this application, also two-phase optimized ELM models perform superior to the single-phase optimized and standalone ELM models; however, the difference between the ELMSAMOA and ELM–PSO is marginal. The best two-phase optimized ELM–PSOGWO model with input ii has lower RMSE (511.9 m^3/s) and higher NSE (0.777) than those of the best ELM with input iv (RMSE = 612.1 m^3/s, NSE = 0.681), ELM–PSO with input v (RMSE = 538 m^3/s, NSE = 0.753),

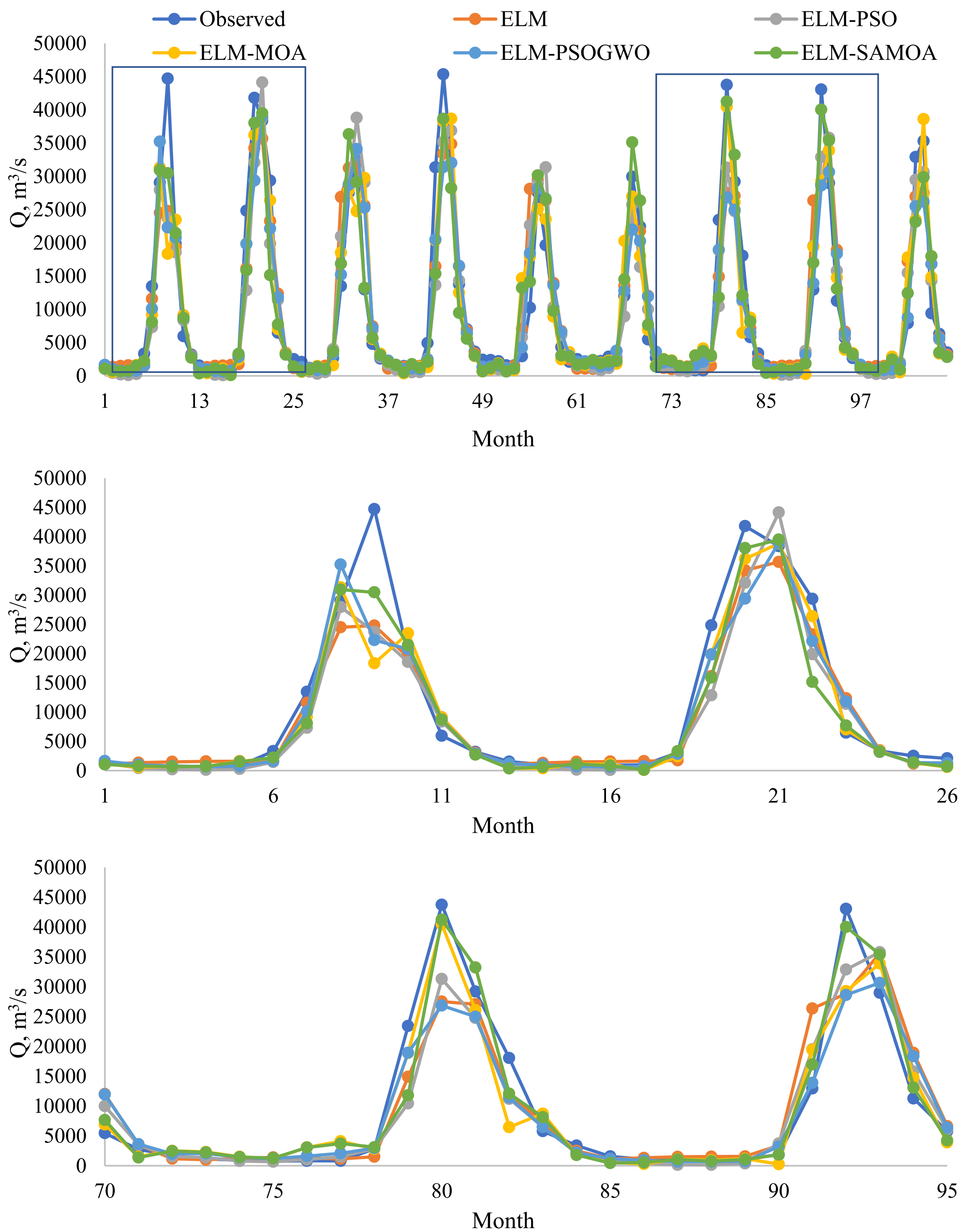

Fig. 5 Time variation graphs of the observed and predicted streamflows by different ELM-based models in the test period of Padma River Station

ELM–MOA with input ii (RMSE = 543.4 m^3/s, NSE = 0.749) and ELM-SAMOA with input iv (RMSE = 535.7 m^3/s, NSE = 0.756) in the testing stage. Implementing the ELM–PSOGWO improved the RMSE of the best ELM, ELM–PSO, ELM–MOA and ELM–SAMOA by 20%, 5.1%, 6.2% and 4.6% in the testing stage, respectively. By including precipitation inputs, the accuracy of standalone ELM (input iv), (ELM–PSO (input v) and ELM–SAMOA (input iv) was improved, whereas the temperature data did not increase the efficiency of the

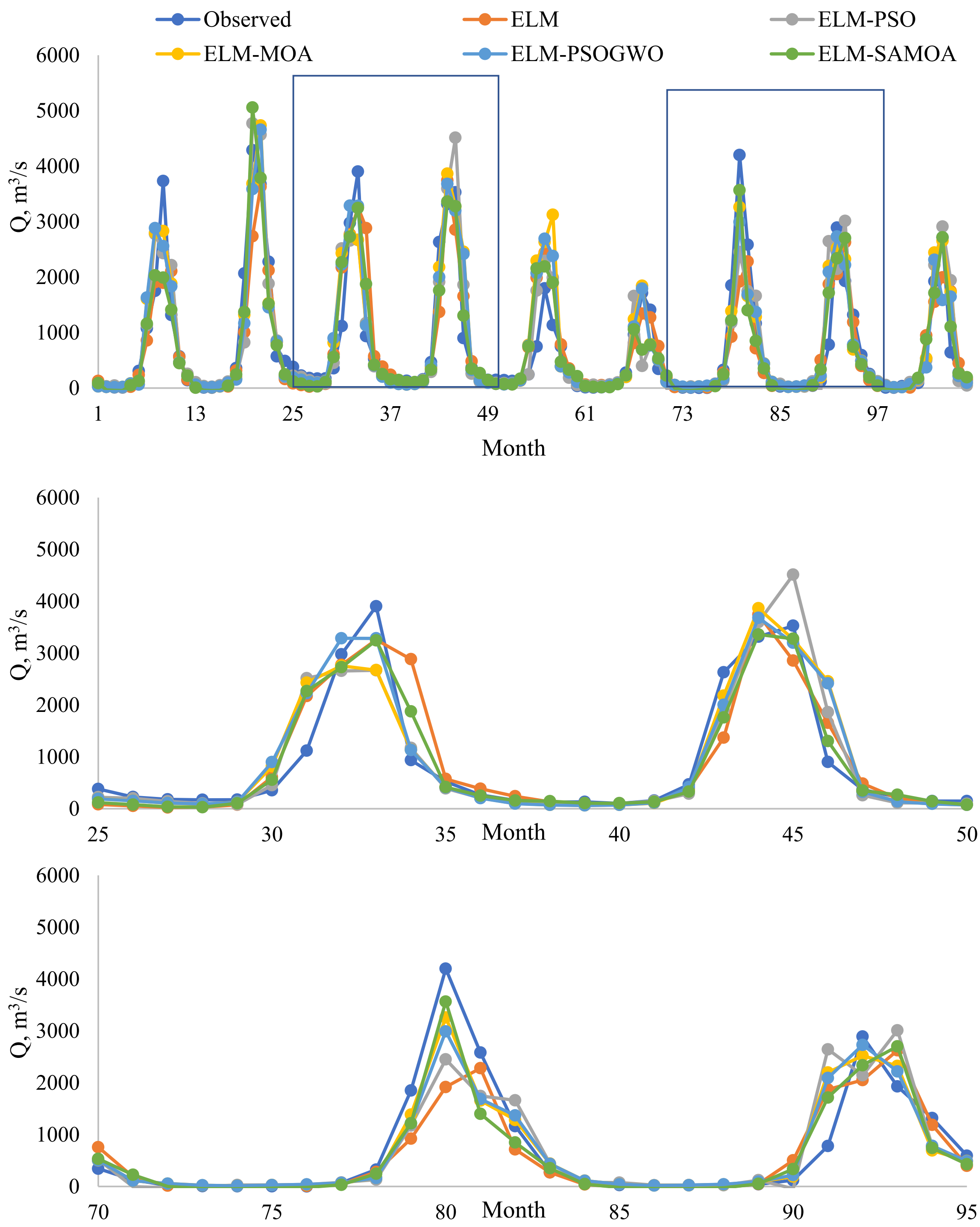

Fig. 6 Time variation graphs of the observed and predicted streamflows by different ELM-based models in the test period of Gorai Station

models in estimating Gorai's streamflows utilizing data of upstream station. Time variation and scatter diagrams of the estimated streamflows by different ELM-based models are illustrated in Figs. 11 and 12. It is clear from the visual comparisons; two-phase optimized ELM models have closer streamflow estimates to the observed values, and their scatters are less compared to single-phase optimized and standalone ELM models. This part of study is very useful especially for the basins having missing streamflow

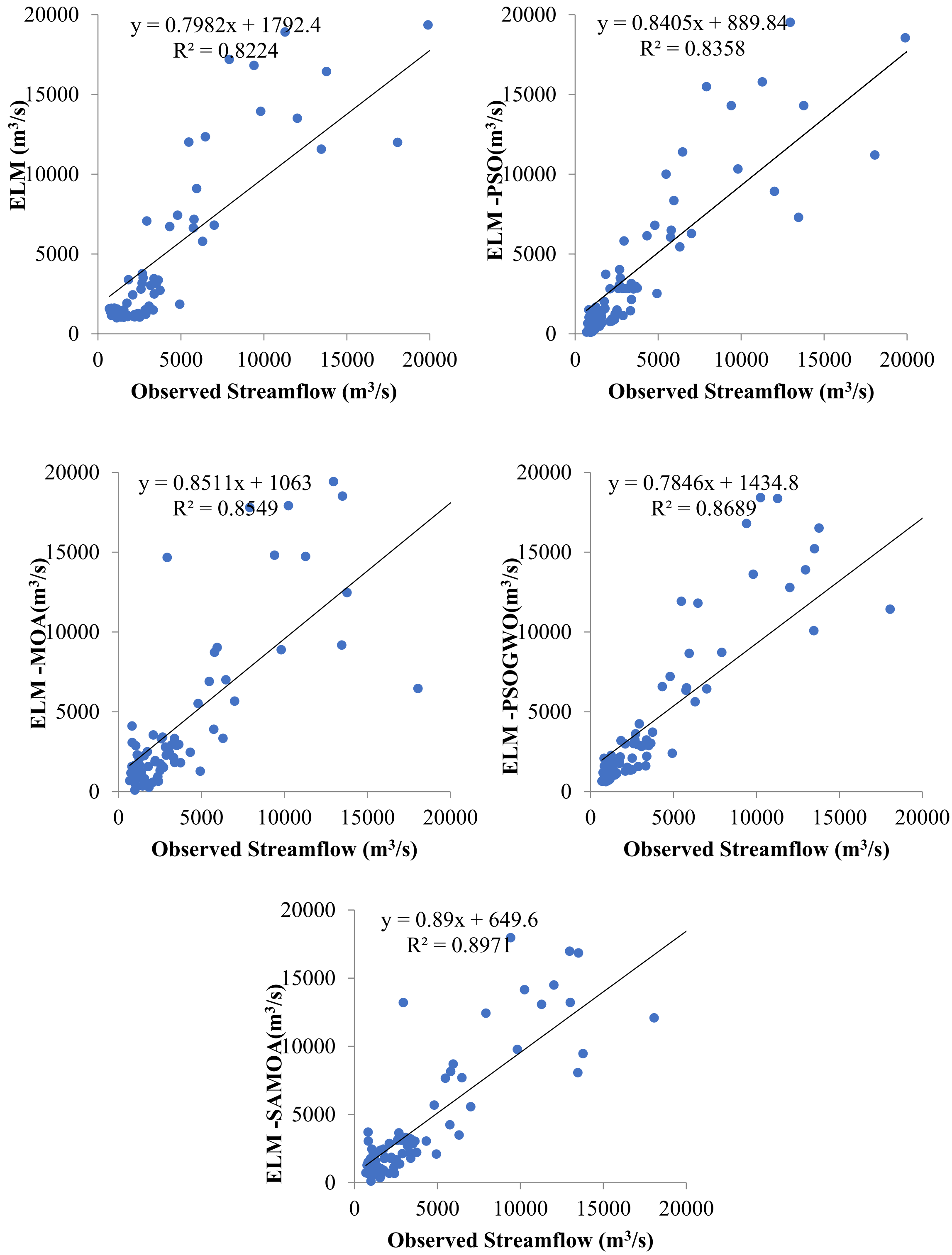

Fig. 7 Scatterplots of the observed and predicted streamflows different ELM-based models in the test period of Padma

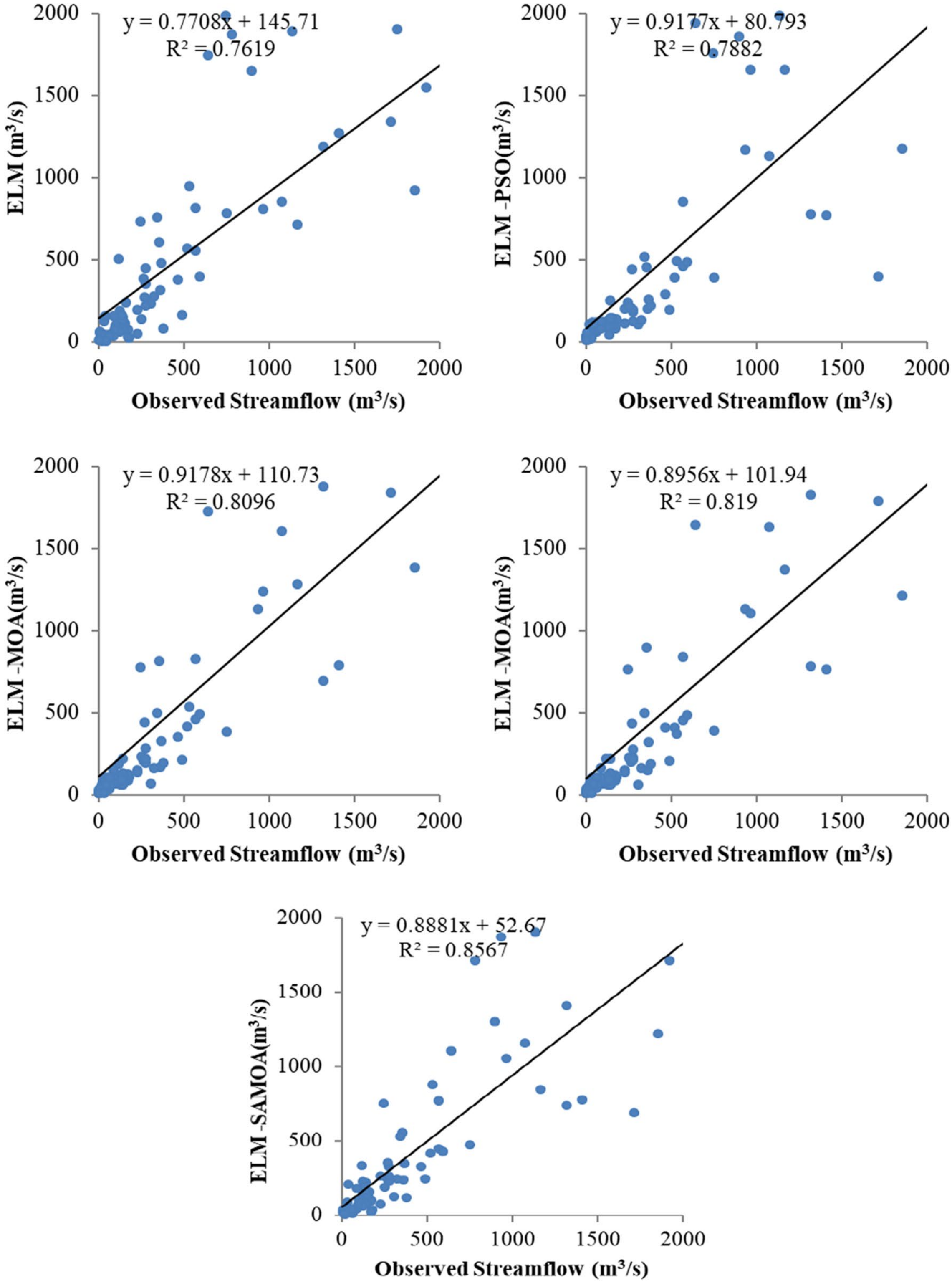

Fig. 8 Scatterplots of the observed and predicted streamflows different ELM-based models in the test period of Gorai

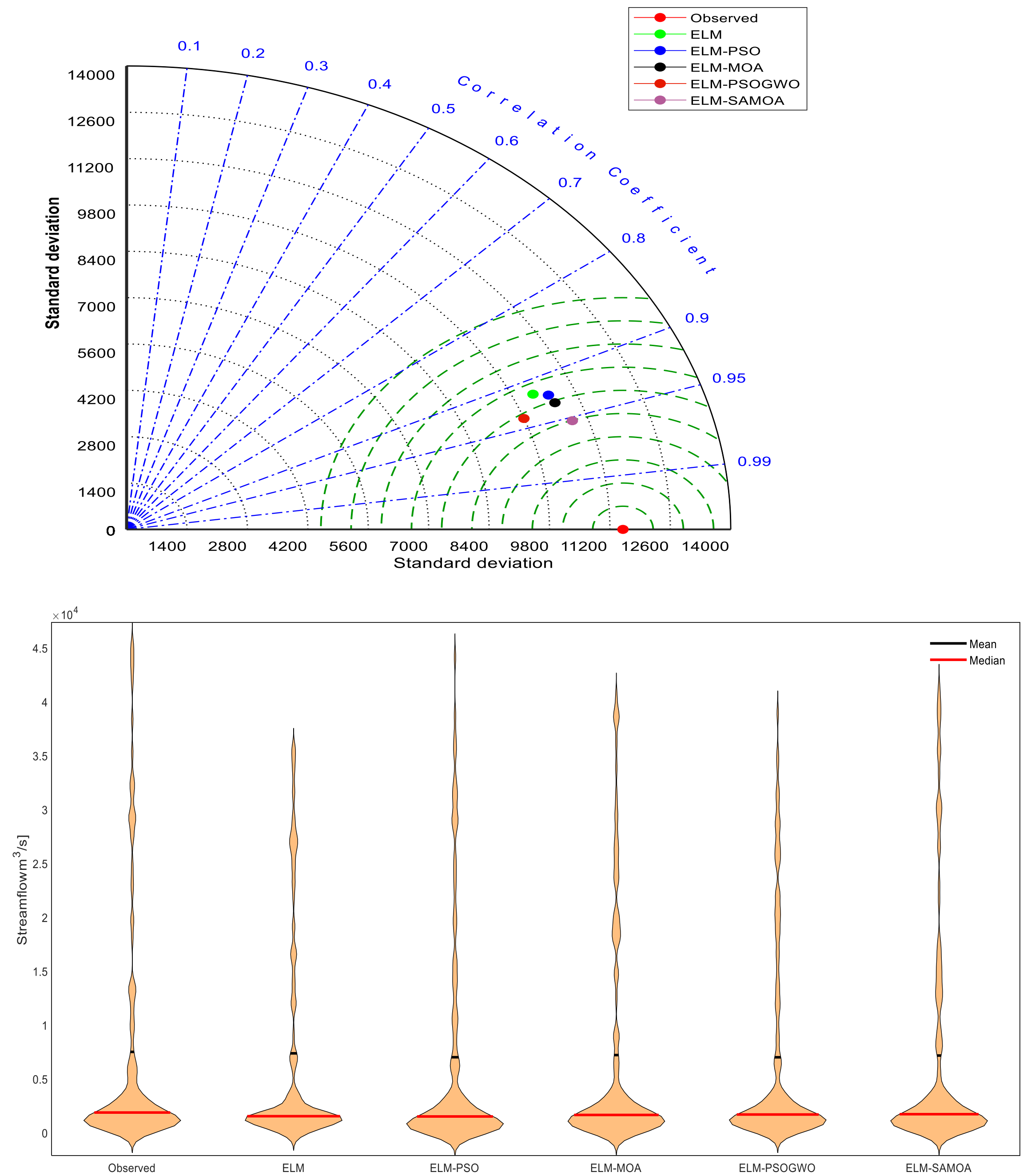

Fig. 9 Taylor and violin diagrams of different ELM-based models in the test period of Padma Station

data. In such basins, streamflow can be easily estimated using external (e.g., upstream) data.

Overall, two-phase optimized ELM models, ELM–PSOGWO and ELM–SAMOA models, perform superior to the single-phase and standalone ELM models in streamflow prediction. Among the two-phase optimized ELM models, the ELM–SAMOA acted better than the other. In the second application, the ELM–PSOGWO offered better efficiency compared to other ELM-based models. The main advantages of the two-phase optimization approaches are improvement in exploration and exploitation abilities of single-phase metaheuristic algorithms. Because for a robust and generalizable optimization algorithm is to balance the ability of exploitation and exploration efficiently in order to find best solution/parameters of a machine learning model. Results of this study also endorsed the effectiveness of PSOGWO- and SAMOA-based ELM models by improving exploration and exploitation capabilities.

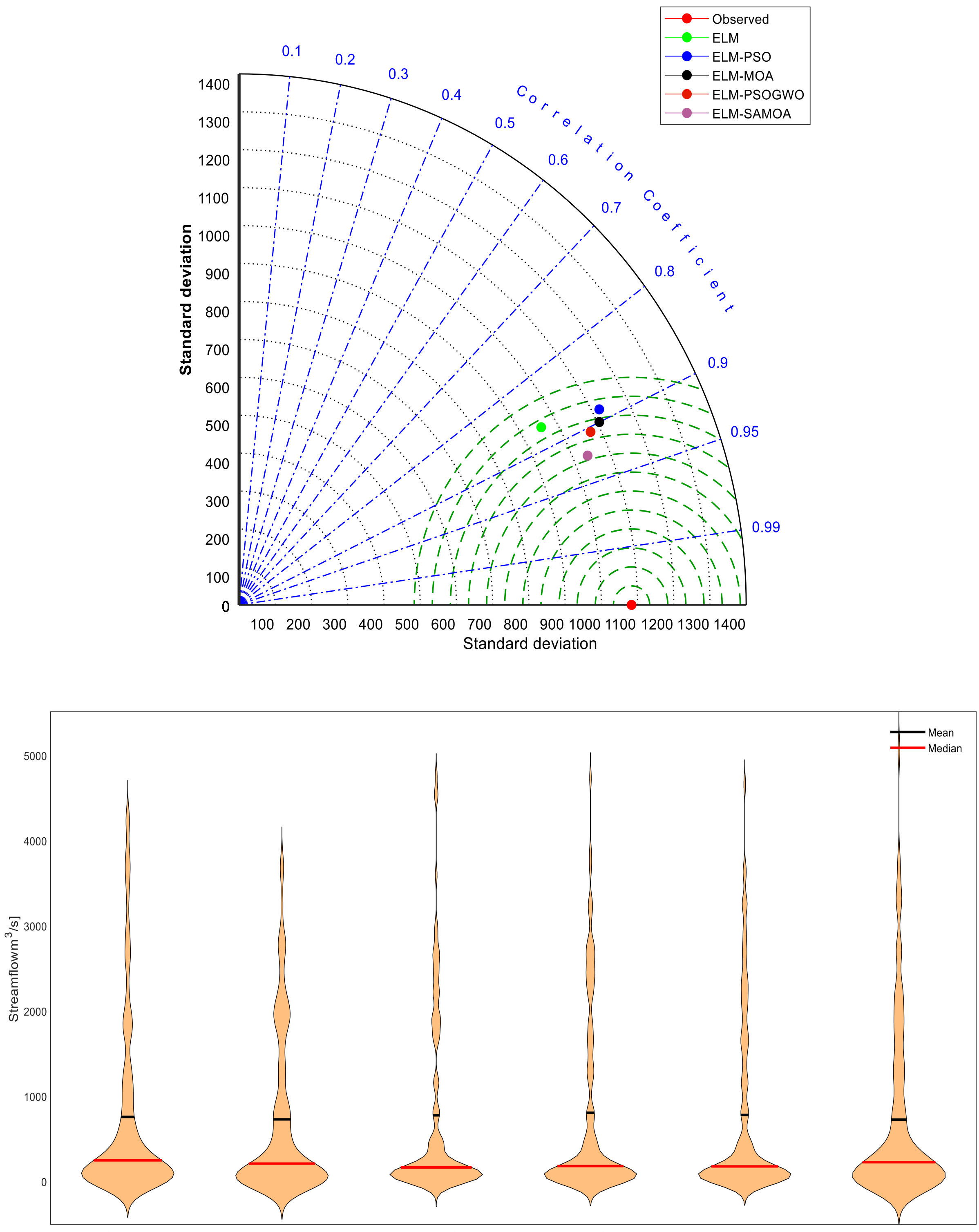

Fig. 10 Taylor and violin diagrams of different ELM-based models in the test period of Gorai Station

Concluding remarks

The ability of two-phase optimized ELM models was investigated in monthly streamflow prediction using lagged streamflow, precipitation and temperature data as inputs. The outcomes were compared with the single-phase optimized and standalone ELM models. In the first application, each station's streamflows were predicted using local data while

Table 5 t-test of single-phase, two-phase optimized and standalone ELM models for both stations using best input combination results

	Models				
	ELM	ELM-PSO	ELM-MOA	ELM-PSOGWO	ELM-SAMOA
Padma station					
t-stat	1.364	1.114	0.971	1.338	1.424
p value	0.328	0.267	0.333	0.183	0.157
t-critical	1.982	1.982	1.982	1.982	1.982
Gorai station					
t-stat	– 1.526	-0.316	0.104	– 0.583	1.086
p value	0.129	0.752	0.916	0.560	0.279
t-critical	1.982	1.982	1.982	1.982	1.982

Table 6 The results of the single-phase, two-phase optimized and standalone ELM models in estimation of Gorai's streamflow using data of Padma Station (Upstream)

Models	Input Combinations	Training					Testing				
		RMSE	MAE	NSE	R^2	MAPE	RMSE	MAE	NSE	R^2	MAPE
ELM	I	1145.5	804.1	0.443	0.444	59.34	892.9	635.9	0.321	0.455	58.19
	II	858.2	430.1	0.687	0.690	34.18	691.8	393.4	0.592	0.760	28.17
	III	766.4	354.0	0.751	0.751	29.47	622.9	371.3	0.670	0.769	28.02
	IV	**743.5**	**339.3**	**0.765**	**0.766**	**28.82**	**612.1**	**352.3**	**0.681**	**0.777**	**27.64**
	V	747.3	347.1	0.763	0.764	28.75	686.0	383.5	0.599	0.772	27.85
	VI	694.0	312.9	0.796	0.796	24.62	672.6	378.0	0.615	0.775	27.72
	VII	724.7	403.2	0.777	0.778	25.94	760.7	471.4	0.507	0.755	28.58
ELM-PSO	I	1135.0	795.7	0.453	0.453	58.19	870.8	621.1	0.354	0.479	56.48
	II	818.0	425.0	0.716	0.717	33.26	573.0	370.8	0.720	0.767	27.34
	III	720.6	328.7	0.780	0.780	25.09	608.6	333.8	0.685	0.786	26.08
	IV	730.7	352.2	0.773	0.774	25.82	640.1	361.3	0.651	0.795	25.54
	V	**687.5**	**324.5**	**0.799**	**0.799**	**24.29**	**538.0**	**311.6**	**0.753**	**0.802**	**24.76**
	VI	684.6	359.2	0.801	0.802	24.08	659.0	430.2	0.630	0.784	26.17
	VII	666.5	352.1	0.811	0.812	23.47	651.4	418.7	0.639	0.752	27.95
ELM-MOA	I	1132.8	791.9	0.455	0.455	57.29	868.7	629.5	0.357	0.472	55.19
	II	**788.4**	**410.0**	**0.736**	**0.739**	**30.73**	**543.4**	**334.4**	**0.749**	**0.822**	**22.38**
	III	713.5	339.0	0.784	0.786	26.81	595.0	353.9	0.698	0.786	25.49
	IV	726.6	343.4	0.776	0.776	27.08	601.5	357.8	0.699	0.807	24.37
	V	691.3	341.3	0.797	0.803	24.36	595.4	371.3	0.698	0.792	25.81
	VI	669.6	326.7	0.810	0.810	22.18	693.0	439.4	0.591	0.794	25.64
	VII	642.3	363.3	0.825	0.825	21.76	685.3	456.5	0.600	0.761	24.82
ELM-PSOGWO	I	1114.3	778.0	0.473	0.473	55.72	864.3	630.5	0.364	0.473	52.42
	II	**747.3**	**371.3**	**0.763**	**0.764**	**28.64**	**511.9**	**338.2**	**0.777**	**0.829**	**21.71**
	III	714.5	317.2	0.783	0.785	27.42	592.1	326.1	0.701	0.784	24.95
	IV	709.4	320.3	0.786	0.787	27.02	537.7	312.1	0.754	0.815	23.43
	V	652.2	332.2	0.819	0.820	22.02	598.6	358.9	0.695	0.801	24.94
	VI	648.1	318.8	0.822	0.822	21.76	574.6	383.3	0.719	0.785	23.64
	VII	569.4	310.1	0.862	0.863	20.83	638.8	428.6	0.652	0.790	23.07
ELM-SAMOA	I	1100.5	768.7	0.486	0.486	53.27	843.2	610.2	0.394	0.493	51.46
	II	724.3	354.1	0.777	0.777	26.49	554.5	365.5	0.738	0.835	20.87
	III	711.2	306.4	0.785	0.785	26.07	585.4	346.1	0.708	0.811	23.62
	IV	**534.3**	**299.3**	**0.879**	**0.879**	**17.08**	**535.7**	**307.4**	**0.756**	**0.844**	**19.07**
	V	640.4	343.1	0.826	0.827	22.49	576.8	346.8	0.717	0.805	23.34
	VI	589.5	306.4	0.853	0.853	19.07	598.3	360.1	0.695	0.802	23.08
	VII	479.9	286.0	0.902	0.902	16.45	736.6	415.8	0.538	0.799	23.24

Bold values defines the best accuracy (meaning the slowest RMSE and MAE and highest NSE and R2)

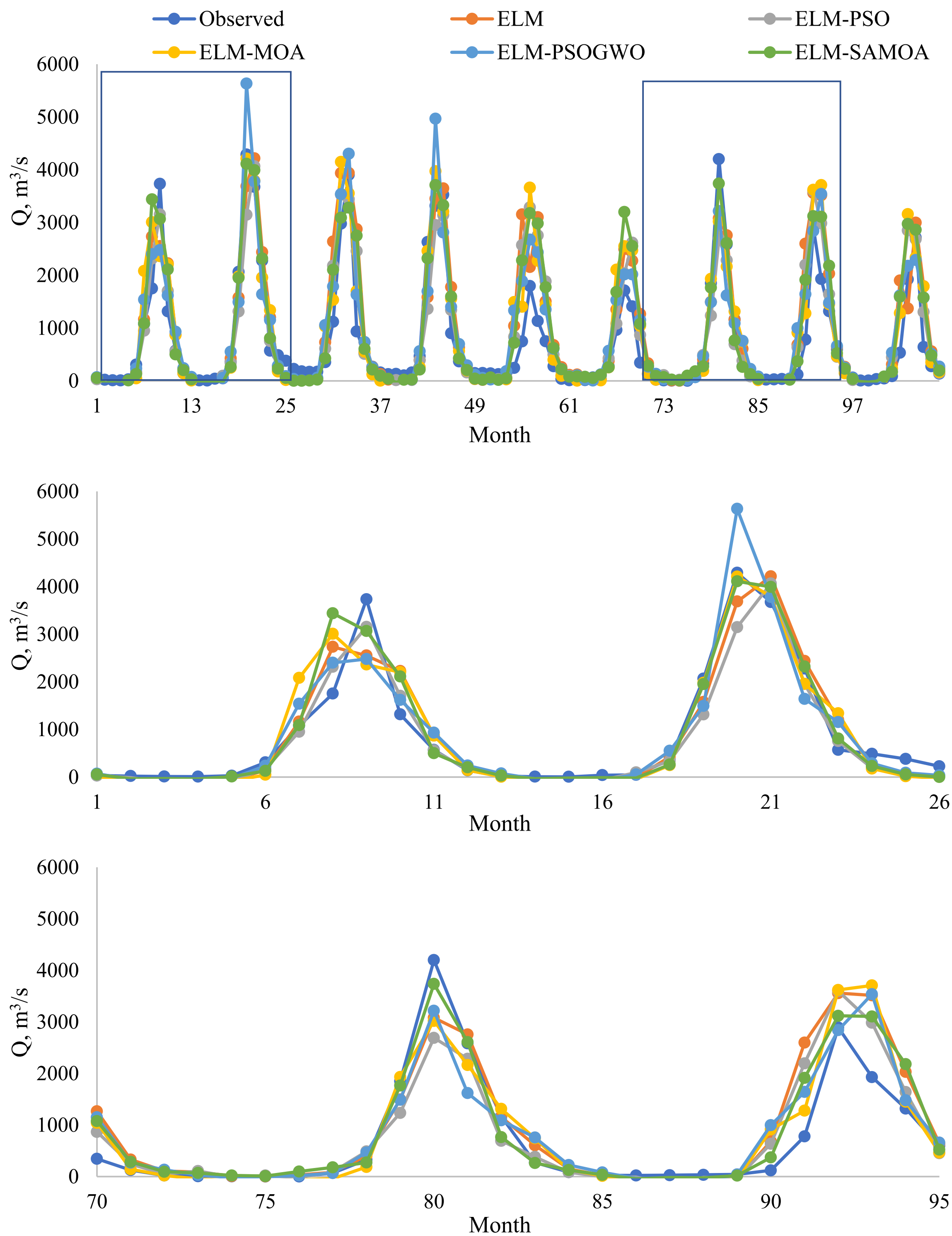

Fig. 11 Time variation graphs of the observed and predicted streamflows by different ELM-based models in the test period of Gorai station (d/s) using Padma station data (u/s)

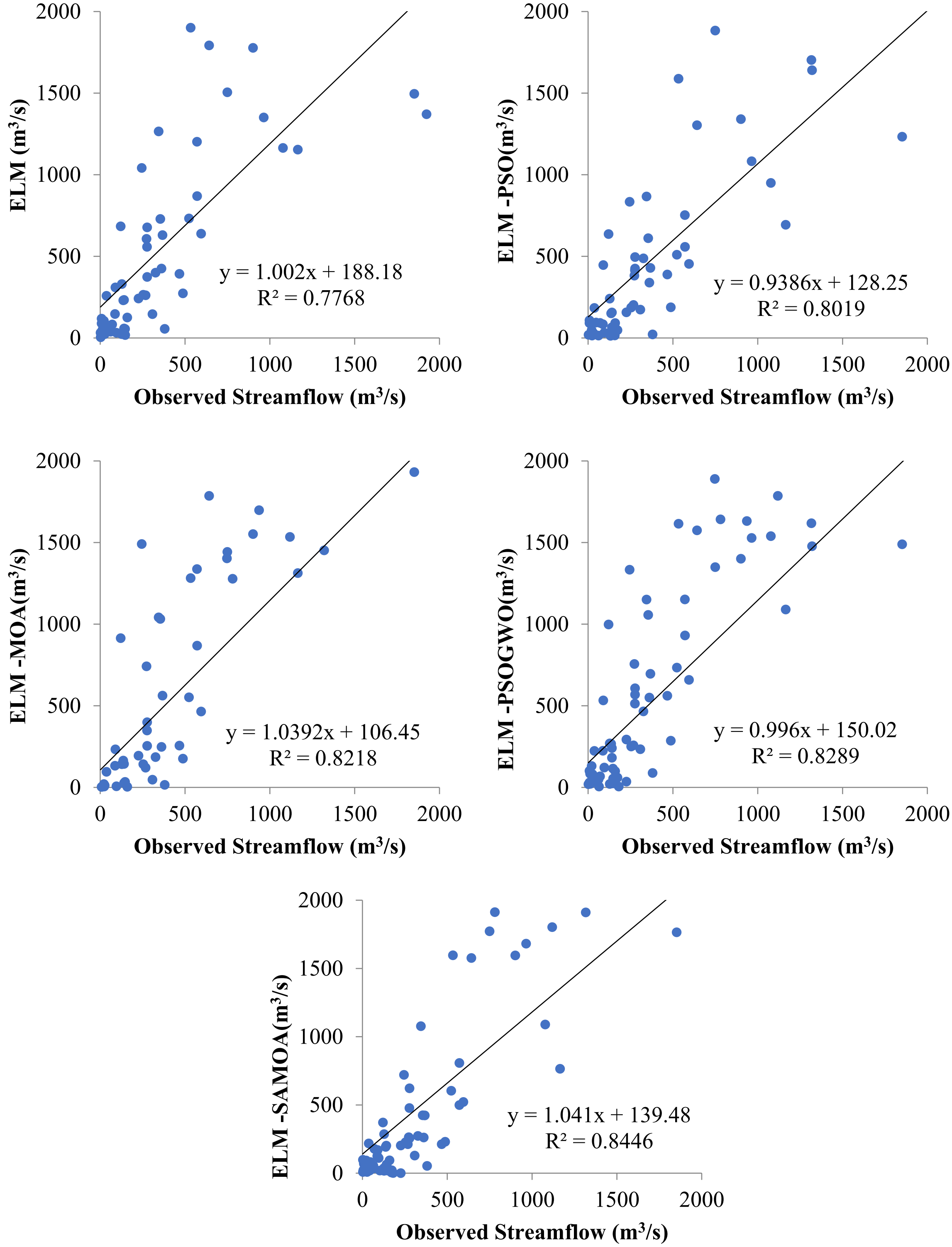

Fig. 12 Scatterplots of the observed and predicted streamflows by different ANFIS-based models in the test period of Gorai Station (d/s) using Padma station (u/s) data

in the second application, streamflows of one station were estimated using other station data. The following conclusions were reached from the benchmark outcomes.

- Based on the RMSE, MAE, NSE, R^2 and MAPE criteria and graphical methods (e.g., time variation, scatter, Taylor and violin diagrams), the two-phase optimized ELM–SAMOA offered the best accuracy in prediction of monthly streamflows using local data; improvement in RMSE of ELM, ELM–PSO, ELM–MOA and ELM–PSOGWO is by 31%, 27%, 19% and 14% for one station and 29%, 27%, 19% and 14% for other station, in the testing stage, respectively.
- In the second application, the two-phase optimized ELM–PSOGWO acted as the best model in streamflow estimation with external data; improvement in RMSE of ELM, ELM–PSO, ELM–MOA and ELM–SAMOA is by 20%, 5.1%, 6.2% and 4.6% in the testing stage, respectively.
- The outcomes suggested the use of two-phase optimization compared to single-phase one while the standalone ELM model provided the worst efficiency.

Author contributions The authors have equally made contributions. All authors read and approved the original manuscript.

Funding The authors would also like to express their sincere appreciation to the associate editor and the anonymous reviewers for their comments and suggestions. This work was supported by the National Social Science Foundation of China (18BTJ029), Key Projects of National Statistical Science Research Projects (2020LZ10), and Tertiary Education Scientific Research Project of Guangzhou Municipal Education Bureau (202235324). The APC was funded by the Postdoctoral Start-up Research Fund of Guangzhou University.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Abdel-Basset M, Ding W, El-Shahat D (2021) A hybrid Harris Hawks optimization algorithm with simulated annealing for feature selection. Artif Intell Rev. https://doi.org/10.1007/s10462-020-09860-3

Adnan RM, Liang Z, Heddam S, Zounemat-Kermani M, Kisi O, Li B (2020a) Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydro-meteorological data as inputs. J Hydrol 586:124371

Adnan RM, Zounemat-Kermani M, Kuriqi A, Kisi O (2020) Machine learning method in prediction streamflow considering periodicity component. Intelligent data analytics for decision-support systems in hazard mitigation. Springer, Sipngapore, pp 383–403

Akhter S, Eibek KU, Islam S, Islam ARMT, Shen S, Chu R (2019) Predicting spatiotemporal changes of channel morphology in the reach of Teesta River, Bangladesh using GIS and ARIMA modeling. Quat Int 513:80–94. https://doi.org/10.1016/j.quaint.2019.01.022

Barzkar A, Najafzadeh M, Homaei F (2022) Evaluation of drought events in various climatic conditions using data-driven models and a reliability-based probabilistic model. Nat Hazards 110(3):1931–1952

Bazargan J, Norouzi H (2018) Investigation the effect of using variable values for the parameters of the linear Muskingum method using the particle swarm algorithm (PSO). Water Resour Manag 32:4763–4777

Ben Messaoud R (2020) Extraction of uncertain parameters of single-diode model of a photovoltaic panel using simulated annealing optimization. Energy Rep 6:350–357. https://doi.org/10.1016/j.egyr.2020.01.016

Boulariah O, Meddi M, Longobardi A (2019) Assessment of prediction performances of stochastic and conceptual hydrological models: monthly stream flow prediction in northwestern Algeria. Arab J Geosci 12:1–14

CDMP (Comprehensive Disaster Management Programme) (2014) Trend and impact analysis of internal displacement due to the impacts of disaster and climate change. Study Report, Ministry of Disaster Management and Relief, Dhaka

Chen J, Shi J (2019) A multi-compartment vehicle routing problem with time windows for urban distribution—a comparison study on particle swarm optimization algorithms. Comput Ind Eng 133:95–106. https://doi.org/10.1016/j.cie.2019.05.008

Chen H, Fan DL, Fang L, Huang W, Huang J, Cao C, Yang L, He Y, Zeng L (2020) Particle swarm optimization algorithm with mutation operator for particle filter noise reduction in mechanical fault diagnosis. Int J Pattern Recognit Artif Intell. https://doi.org/10.1142/S0218001420580124

Dehghani M, Riahi-Madvar H, Hooshyaripor F, Mosavi A, Shamshirband S, Zavadskas EK, Chau K-w (2019) Prediction of hydropower generation using Grey wolf optimization adaptive neuro-fuzzy inference system. Energies 12:1–20. https://doi.org/10.3390/en12020289

Dewan A, Corner R, Saleem A, Rahman MM, Haider MR, Rahman MM, Sarker MH (2017) Assessing channel changes of the Ganges–Padma River system in Bangladesh using Landsat and hydrological data. Geomorphology 276:257–279. https://doi.org/10.1016/j.geomorph.2016.10.017

Elsheikh AH, Abd Elaziz M (2019) Review on applications of particle swarm optimization in solar energy systems. Int J Environ Sci Technol 16:1159–1170. https://doi.org/10.1007/s13762-018-1970-x

Fadaee M, Mahdavi-Meymand A, Zounemat-Kermani M (2020) Suspended sediment prediction using integrative soft computing

models: on the analogy between the butterfly optimization and genetic algorithms. Geocarto Int 37:961–977
Fan J, Yue W, Wu L, Zhang F, Cai H, Wang X, Lu X, Xiang Y (2018) Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. Agric for Meteorol 263:225–241. https://doi.org/10.1016/j.agrformet.2018.08.019
Fu M, Fan T, Za Ding, Salih SQ, Al-Ansari N, Yaseen ZM (2020) Deep learning data-intelligence model based on adjusted forecasting window scale: application in daily streamflow simulation. IEEE Access 8:32632–32651
Gao ZM, Zhao J, Li SR, Hu YR (2020) The improved mayfly optimization algorithm. J Phys Conf Ser. https://doi.org/10.1088/1742-6596/1684/1/012077
Granata F, Di Nunno F, Najafzadeh M, Demir I (2022) A stacked machine learning algorithm for multi-step ahead prediction of soil moisture. Hydrology 10(1):1
Haddad OB, Afshar A, Mariño MA (2006) Honey-bees mating optimization (HBMO) algorithm: a new heuristic approach for water resources optimization. Water Resour Manag 20:661–680. https://doi.org/10.1007/s11269-005-9001-3
Himanshu N, Kumar V, Burman A, Maity D, Gordan B (2020) Grey wolf optimization approach for searching critical failure surface in soil slopes. Eng Comput. https://doi.org/10.1007/s00366-019-00927-6
Hosseini FS, Choubin B, Mosavi A, Nabipour N, Shamshirband S, Darabi H, Haghighi AT (2020) Flash-flood hazard assessment using ensembles and Bayesian-based machine learning models: application of the simulated annealing feature selection method. Sci Tot Environ 711:135161
Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: theory and applications. Neurocomputing 70:489–501. https://doi.org/10.1016/j.neucom.2005.12.126
Ikram RMA, Mostafa RR, Chen Z, Parmar KS, Kisi O, Zounemat-Kermani M (2023) Water temperature prediction using improved deep learning methods through reptile search algorithm and weighted mean of vectors optimizer. J Mar Sci Eng 11(2):259
Ikram RMA, Hazarika BB, Gupta D, Heddam S, Kisi O (2022) Streamflow prediction in mountainous region using new machine learning and data preprocessing methods: a case study. Neural Comput Appl 1–18
Islam ARMT (2016) Assessment of fluvial channel dynamics of padma river in northwestern Bangladesh. Univ J Geosci 4:41–49. https://doi.org/10.13189/ujg.2016.040204
Islam ARMT, Sein ZMM, Ongoma V et al (2016) Geomorphological and land use mapping: A case study of Ishwardi under Pabna district, Bangladesh. Adv Res 4(6):378–387. https://doi.org/10.9734/AIR/2015/14149
Islam ARMT, Talukdar S, Mahato S et al (2021) Machine learning algorithm-based risk assessment of riparian wetlands in Padma River Basin of Northwest Bangladesh. Environ Sci Pollut Res. https://doi.org/10.1007/s11356-021-12806-z
Jiang Y, Bao X, Hao S, Zhao H, Li X, Wu X (2020) Monthly streamflow forecasting using ELM-IPSO based on phase space reconstruction. Water Resour Manag 34:3515–3531
Liu Y et al (2015) Analyzing effects of climate change on streamflow in a glacier mountain catchment using an ARMA model. Quatern Int 358:137–145
Liu B, Wang R, Zhao G, Guo X, Wang Y, Li J, Wang S (2020) Prediction of rock mass parameters in the TBM tunnel based on BP neural network integrated simulated annealing algorithm. Tunn Undergr Space Technol 95:1–12. https://doi.org/10.1016/j.tust.2019.103103
Malik A, Tikhamarine Y, Souag-Gamane D, Kisi O, Pham QB (2020) Support vector regression optimized by meta-heuristic algorithms for daily streamflow prediction. Stoch Env Res Risk Assess 34:1755–1773
Mansouri N, Mohammad HZB, Javidi MM (2019) Hybrid task scheduling strategy for cloud computing by modified particle swarm optimization and fuzzy theory. Comput Ind Eng 130:597–633. https://doi.org/10.1016/j.cie.2019.03.006
Meng X, Fu Y, Yuan J (2020) Estimating solubilities of ternary water-salt systems using simulated annealing algorithm based generalized regression neural network. Fluid Phase Equilib 505:112357. https://doi.org/10.1016/j.fluid.2019.112357
Mirjalili S, Mirjalili SM, Lewis A (2014) Grey wolf optimizer. Adv Eng Softw 69:46–61
Mirza MMQ (2004) The Ganges water diversion: environmental effects andimplications-an introduction. In: Mirza MMQ (ed) The Ganges waterdiversion: environmental effects and implications. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 1–12
Mohammadi B, Guan Y, Aghelpour P, Emamgholizadeh S, Zolá RP, Zhang D (2020) Simulation of titicaca lake water level fluctuations using hybrid machine learning technique integrated with grey wolf optimizer algorithm. Water (Switzerland) 12:1–18. https://doi.org/10.3390/w12113015
Moors EJ, Groot A, Biemans H et al (2011) Adaptation to changing water resources in the Ganges basin, northern India. Environ Sci Policy 14:758–769
Mostafa RR, Kisi O, Adnan RM, Sadeghifar T, Kuriqi A (2023) Modeling potential evapotranspiration by improved machine learning methods using limited climatic data. Water 15(3):486
Najafzadeh M, Niazmardi S (2021) A novel multiple-kernel support vector regression algorithm for estimation of water quality parameters. Nat Resour Res 30(5):3761–3775
Najafzadeh M, Homaei F, Farhadi H (2021) Reliability assessment of water quality index based on guidelines of national sanitation foundation in natural streams: Integration of remote sensing and data-driven models. Artif Intell Rev 54(6):4619–4651
Niu W-j, Feng Z-k, Chen Y-b, Zhang H-r, Cheng C-t (2020) Annual streamflow time series prediction using extreme learning machine based on gravitational search algorithm and variational mode decomposition. J Hydrol Eng 25:04020008
Poul AK, Shourian M, Ebrahimi H (2019) A comparative study of MLR, KNN, ANN and ANFIS models with wavelet transform in monthly stream flow prediction. Water Resour Manag 33:2907–2923
Prasad R, Deo RC, Li Y, Maraseni T (2018) Soil moisture forecasting by a hybrid machine learning technique: ELM integrated with ensemble empirical mode decomposition. Geoderma 330:136–161. https://doi.org/10.1016/j.geoderma.2018.05.035
Qi C, Fourie A, Chen Q (2018) Neural network and particle swarm optimization for predicting the unconfined compressive strength of cemented paste backfill. Constr Build Mater 159:473–478. https://doi.org/10.1016/j.conbuildmat.2017.11.006
Redi AANP, Jewpanya P, Kurniawan AC, Persada SF, Nadlifatin R, Dewi OAC (2020) A simulated annealing algorithm for solving two-echelon vehicle routing problem with locker facilities. Algorithms 13:1–14. https://doi.org/10.3390/a13090218
Rezaie-Balf M, Kisi O (2018) New formulation for forecasting streamflow: evolutionary polynomial regression vs. extreme learning machine. Hydrol Res 49:939–953
Riahi-Madvar H, Dehghani M, Seifi A, Salwana E, Shamshirband S, Mosavi A, Chau K-w (2019) Comparative analysis of soft computing techniques RBF MLP, and ANFIS with MLR and MNLR for predicting grade-control scour hole geometry. Eng Appl Comput Fluid Mech 13:529–550
Saberi-Movahed F, Najafzadeh M, Mehrpooya A (2020) Receiving more accurate predictions for longitudinal dispersion coefficients

in water pipelines: training group method of data handling using extreme learning machine conceptions. Water Resour Manage 34(2):529–561

Şenel FA, Gökçe F, Yüksel AS, Yiğit T (2019) A novel hybrid PSO–GWO algorithm for optimization problems. Eng Comput 35:1359–1373. https://doi.org/10.1007/s00366-018-0668-5

Shao W, Zuo Y (2020) Computing the halfspace depth with multiple try algorithm and simulated annealing algorithm. Comput Stat 35:203–226. https://doi.org/10.1007/s00180-019-00906-x

Silva BH, Machado IM, Pereira FM, Pagot PR, França FHR (2020) Application of the simulated annealing algorithm to the correlated WMP radiation model for flames. Inverse Probl Sci Eng 28:1345–1360. https://doi.org/10.1080/17415977.2020.1732956

Sudheer C, Maheswaran R, Panigrahi BK, Mathur S (2014) A hybrid SVM–PSO model for forecasting monthly streamflow. Neural Comput Appl 24:1381–1389

Suiju L, Minquan F (2015) Three-dimensional numerical simulation of flow in Daliushu reach of the Yellow River. Int J Heat Technol 33:107–114

Tang S, Peng M, Xia G, Wang G, Zhou C (2020) Optimization design for supercritical carbon dioxide compressor based on simulated annealing algorithm. Ann Nucl Energy 140:107107. https://doi.org/10.1016/j.anucene.2019.107107

Tikhamarine Y, Souag-Gamane D, Kisi O (2019) A new intelligent method for monthly streamflow prediction: hybrid wavelet support vector regression based on grey wolf optimizer (WSVR–GWO). Arab J Geosci. https://doi.org/10.1007/s12517-019-4697-1

Tikhamarine Y, Souag-Gamane D, Najah Ahmed A, Kisi O, El-Shafie A (2020) Improving artificial intelligence models accuracy for monthly streamflow forecasting using grey Wolf optimization (GWO) algorithm. J Hydrol 582:124435. https://doi.org/10.1016/j.jhydrol.2019.124435

Tufano A, Accorsi R, Manzini R (2020) A simulated annealing algorithm for the allocation of production resources in the food catering industry. Br Food J 122:2139–2158. https://doi.org/10.1108/BFJ-08-2019-0642

Turhan AM, Bilgen B (2020) A hybrid fix-and-optimize and simulated annealing approaches for nurse rostering problem. Comput Ind Eng 145:106531. https://doi.org/10.1016/j.cie.2020.106531

Wang D, Tan D, Liu L (2018) Particle swarm optimization algorithm: an overview. Soft Comput 22:387–408. https://doi.org/10.1007/s00500-016-2474-6

Xi H, Liao P, Wu X (2021) Simultaneous parametric optimization for design and operation of solvent-based post-combustion carbon capture using particle swarm optimization. Appl Therm Eng 184:116287. https://doi.org/10.1016/j.applthermaleng.2020.116287

Yadav B, Ch S, Mathur S, Adamowski J (2017) Assessing the suitability of extreme learning machines (ELM) for groundwater level prediction. J Water l Dev 32:103–112. https://doi.org/10.1515/jwld-2017-0012

Yaseen ZM, Faris H, Al-Ansari N (2020) Hybridized extreme learning machine model with salp swarm algorithm: a novel predictive model for hydrological application. Complexity 2020:1–14

Yu S, Lu H (2018) An integrated model of water resources optimization allocation based on projection pursuit model—Grey wolf optimization method in a transboundary river basin. J Hydrol 559:156–165. https://doi.org/10.1016/j.jhydrol.2018.02.033

Zervoudakis K, Tsafarakis S (2020) A mayfly optimization algorithm. Comput Ind Eng. https://doi.org/10.1016/j.cie.2020.106559

Zhang L, Chen X, Zhang Y, Wu F, Chen F, Wang W, Guo F (2020) Application of GWO–ELM model to prediction of caojiatuo landslide displacement in the three gorge reservoir area. Water (Switz). https://doi.org/10.3390/w12071860

Zhao S, Dong J, Monte C, Sun X, Zhang W (2020) New phase function development and complete spectral radiative properties measurements of aerogel infused fibrous blanket based on simulated annealing algorithm. Int J Therm Sci 154:106407. https://doi.org/10.1016/j.ijthermalsci.2020.106407

Zhou H, Pang J, Chen PK, Chou FD (2018) A modified particle swarm optimization algorithm for a batch-processing machine scheduling problem with arbitrary release times and non-identical job sizes. Comput Ind Eng 123:67–81. https://doi.org/10.1016/j.cie.2018.06.018

Zhu S, Heddam S, Wu S, Dai J, Jia B (2019) Extreme learning machine-based prediction of daily water temperature for rivers. Environ Earth Sci 78:1–17. https://doi.org/10.1007/s12665-019-8202-7

Zhu B, Feng Y, Gong D, Jiang S, Zhao L, Cui N (2020) Hybrid particle swarm optimization with extreme learning machine for daily reference evapotranspiration prediction from limited climatic data. Comput Electron Agric 173:105430. https://doi.org/10.1016/j.compag.2020.105430

Zounemat-Kermani M, Matta E, Cominola A, Xia X, Zhang Q, Liang Q, Hinkelmann R (2020) Neurocomputing in surface water hydrology and hydraulics: a review of two decades retrospective, current status and future prospects. J Hydrol 588:125085

Zounemat-Kermani M, Alizamir M, Fadaee M, Sankaran Namboothiri A, Shiri J (2020) Online sequential extreme learning machine in river water quality (turbidity) prediction: a comparative study on different data mining approaches. Water Environ J 35:355

Authors and Affiliations

Rana Muhammad Adnan[1] · Hong-Liang Dai[1] · Reham R. Mostafa[2] · Abu Reza Md. Towfiqul Islam[3,8] · Ozgur Kisi[4,5] · Ahmed Elbeltagi[6] · Mohammad Zounemat-Kermani[7]

Rana Muhammad Adnan
rana@gzhu.edu.cn

Reham R. Mostafa
reham_2006@mans.edu.eg

Abu Reza Md. Towfiqul Islam
towfiq_dm@brur.ac.bd

Ahmed Elbeltagi
ahmedelbeltagy81@mans.edu.eg

Mohammad Zounemat-Kermani
zounemat@uk.ac.ir

1 School of Economics and Statistics, Guangzhou University, Guangzhou 510006, China

2 Information Systems Department, Faculty of Computers and Information Sciences, Mansoura University, Mansoura 35516, Egypt

[3] Department of Disaster Management, Begum Rokeya University, Rangpur 5400, Bangladesh

[4] Department of Civil Engineering, Lübeck University of Applied Science, 23562 Lübeck, Germany

[5] Civil Engineering Department, Ilia State University, 0162 Tbilisi, Georgia

[6] Agricultural Engineering Department, Faculty of Agriculture, Mansoura University, Mansoura 35516, Egypt

[7] Department of Water Engineering, Shahid Bahonar University of Kerman, Kerman, Iran

[8] Department of Development Studies, Daffodil International University, Dhaka 1216, Bangladesh