Computer Modeling in
Engineering & Sciences

Tech Science Press

https://doi.org/10.32604/cmes.2026.083313

ARTICLE

Innovative Deep Learning Models for Streamflow Forecasting in High Elevation Catchments

Rana Muhammad Adnan Ikram[1], Jing-Cheng Han[1,*], Ahmed A. Ewees[2], Mo Wang[3], Ozgur Kisi[4,5,6,*], Salim Heddam[7] and Mohammad Zounemat-Kermani[8]

[1]Water Science and Environmental Research Centre, College of Chemistry and Environmental Engineering, Shenzhen University, Shenzhen, China
[2]Department of Computer, Damietta University, Damietta, Egypt
[3]College of Architecture and Urban Planning, Guangzhou University, Guangzhou, China
[4]Department of Civil Engineering, Lübeck University of Applied Science, Lübeck, Germany
[5]Department of Civil Engineering, School of Technology, Ilia State University, Tbilisi, Georgia
[6]School of Civil, Environmental and Architectural Engineering, Korea University, Seoul, Republic of Korea
[7]Faculty of Science, Agronomy Department, Hydraulics Division University, 20 Août 1955, Route El Hadaik, BP 26, Skikda, Algeria
[8]Department of Civil Engineering, Shahid Bahonar University of Kerman, Kerman, Iran

*Corresponding Authors: Jing-Cheng Han. Email: hanjc@szu.edu.cn; Ozgur Kisi. Email: ozgur.kisi@th-luebeck.de

Received: 01 April 2026; Accepted: 05 June 2026; Published: 27 July 2026

ABSTRACT: Two-phase optimized machine learning and deep learning models play a key role in enhancing the prediction accuracy of nonlinear time series modeling. This study assesses the performance of a novel two-phase optimized Long Short-Term Memory (LSTM) model with integration of Aquila Optimizer (AO) and Wild Horse Optimizer (WHO) in predicting monthly streamflow in a snow-fed catchment. The two-phase optimized LSTM-WHOAO model is compared with single-phase optimized models such as LSTM-GA (Genetic Algorithm), LSTM-GWO (Grey Wolf Optimizer), LSTM-WOA (Whale Optimization Algorithm), LSTM-AO, and LSTM-WHO. The outcomes acquired from the deep learning models were compared using four statistical measures: root-mean-square-error (RMSE), mean absolute error (MAE), Nash-Sutcliffe efficiency (NSE), and coefficient of determination (R^2). The LSTM-WHOAO model exhibited the best performance during training, with a mean RMSE of 51.930 and an R^2 of 0.851. The LSTM-WHO model also demonstrated robust performance, achieving an average RMSE of 53.900 and an R^2 value of 0.840. Other models, such as LSTM-AO and LSTM-WOA, showed average RMSE of 57.135 and 58.978, respectively, indicating improved performance over the single LSTM model. For the testing stage, the LSTM-WHOAO model remained more effective than other models, with an average RMSE of 70.413 and an R^2 of 0.755. For peak streamflow events, the LSTM-WHOAO model had the lowest absolute error (201.4%), significantly reducing prediction error compared to other models such as LSTM-GA (333.9%) and LSTM (347.1%). For peak streamflow events, the LSTM-WHOAO model had the lowest absolute error (201.4%), significantly reducing the prediction error compared to other models such as LSTM-GA (333.9%) and LSTM (347.1%). Models that incorporated snow-covered area (SCA) data, such as LSTM-WHOAO and LSTM-WHO, showed lower RMSE and higher R^2 values, underscoring the importance of considering snow cover dynamics in streamflow forecasting. The LSTM-WHOAO model proved to be the most successful, showing superior results in both the training and testing stages, as well as in peak streamflow predictions. By addressing the unique challenges of snow-fed catchments, this research offers valuable insights, especially into the application of advanced ML techniques in hydrology.

KEYWORDS: Streamflow prediction; long short-term memory; genetic algorithm; grey wolf optimizer; whale optimization algorithm; aquila optimizer; wild horse optimizer algorithm; snow covered area

1 Introduction

Long-term and mid-term accurate modeling and prediction of streamflow in high-altitude, snow-fed catchments are among the critical factors for efficient water resource management, peak flow and forecasting, and sustainable hydropower generation. According to previous research, in snow-fed catchments, snow and glaciers account for between 50% and 60% of annual river runoff [1–3]. These regions are characterized by complex hydrological processes influenced by seasonal snowmelt, variable precipitation patterns, and temperature fluctuations, posing significant challenges for streamflow prediction and downstream river flow. Hence, it is expected that accounting for the Snow-Covered Area (SCA) of a catchment (i.e., watershed) can significantly improve the accuracy of the developed mathematical or Machine Learning (ML) models. In this sense, in some cases, traditional hydrological models are not effective and robust enough to capture the dynamics of river flow or streamflow [4,5].

The application of ML techniques for predicting river flow, particularly in snow-dominated catchments and basins, has shown significant potential and improvement in the prediction accuracy of streamflow along with conceptual and empirical hydrological models. Some studies have already demonstrated the efficacy of different (i) individual ML models, such as Artificial Neural Networks (ANN) [6,7], Deep Neural Networks (DNN) [8–12], Support Vector Regression (SVR) [13], and (ii) ensemble ML models, such as Random Forest (RF) [14], and gradient boosting [15] in more accurate forecasting streamflow by incorporating snow cover data. In this regard, Abhinanda et al. [16] employed a hybrid modeling framework combining the RF model and an empirical rainfall-snowmelt-runoff module to improve streamflow forecasts in the Beas River Basin (India). They reported that the average runoff measured contributed 65% to rainfall and 35% to snowmelt. In general, the developed hybrid model performed promisingly in forecasting daily and weekly river flow. Furthermore, another similar study showed that incorporating remote sensing and atmospheric data into RF and ANN models has enhanced the simulation accuracy of snowmelt streamflow in high-altitude mountainous regions, particularly during the snowmelt period [17].

In recent years, machine learning (ML) models, particularly DNNs like Long Short-Term Memory (LSTM) networks, have shown great potential for hydrological modeling and time-series prediction tasks [18–20]. Thapa et al. [21] utilized LSTM in modeling snowmelt in the Himalayan basin. In order to evaluate the LSTM model, other individual ML models, including the nonlinear autoregressive exogenous (NARX) neural network and SVR, were also used. The models were constructed based on input data of snow coverage area (MODIS images) and meteorological variables. In summary, the LSTM model enriched with SCA inputs has outperformed the other ML models. Sarafanov et al. [13] applied a hybrid ML and snowmelt runoff physical method for flash flood modeling in the Lena River, located in Russia. They applied Ridge regression, RF, and SVR for 1- to 7-day ahead flood forecasting using climatological stations (i.e., precipitation and temperature data) and satellite images (MODIS data for snow cover fraction). The overall outcomes of the study confirm the reliability and accuracy (NSE values greater than 0.8) of the developed hybrid approach. For future studies, they suggested the application of DNN networks (e.g., LSTM) to improve forecast accuracy, which is the main focus of the current study. Similarly, Parisouj et al. [22] applied LSTM, Extreme Learning Machine (ELM), ANN, and SVR models along with a conceptual snowmelt-runoff model to predict streamflow in the Carson River (the USA) and Montmorency catchments (Canada). Four input independent variables, such as precipitation, minimum temperature, maximum temperature, and MODIS snow coverage area, were used to construct the models. Overall, the results indicated that incorporating SCA data enhanced the accuracy and robustness of the ML models.

Considering the above-mentioned studies, it can be concluded that LSTM networks have been successfully applied for streamflow prediction/forecasting based on the related data on snow coverage in snow-fed catchments. Furthermore, several studies have highlighted the capability of LSTM networks to effectively capture long-term dependencies in hydrological modeling applications [23]. Owing to this capability, LSTM models are considered highly suitable for streamflow prediction based on sequential data. However, it should be noted that the performance of standalone LSTM models can be further improved by combining them with metaheuristic optimization algorithms [24,25]. To address this, researchers have integrated metaheuristic algorithms inspired by natural processes and evolutionary principles for optimizing complex hydrological problems [26]. These algorithms can be employed to optimize the hyperparameters of deep learning (e.g., LSTM) models, enhancing their predictive accuracy. Among the various metaheuristic algorithms, the Genetic Algorithm (GA) [27], Grey Wolf Optimizer (GWO) [28], Whale Optimization Algorithm (WOA) [29], Aquila Optimizer (AO) [30], and Wild Horse Optimizer (WHO) [31] have proven their effectiveness combined with the LSTM model due to their robustness and efficiency.

It is worth noting that there are limited studies on the application of these integrated (hybrid) models (i.e., ML models + metaheuristic algorithms) specifically in snow-fed catchments. Therefore, this study aims to develop a hybrid (integrated) ML framework for monthly streamflow prediction in a snow-fed catchment in Pakistan. The specific objectives can be written as:

1. To assess the accuracy of various LSTM hybrid models, including LSTM-GA, LSTM-GWO, LSTM-WOA, LSTM-AO, LSTM-WHO, and LSTM-WHO-AO, in predicting monthly streamflow.
2. To assess the impact of integrating snow cover area (SCA) data and previous streamflow records on the predictive performance of these models.
3. To identify the most suitable hybrid model for streamflow prediction in the studied catchment.

The novelty of this study lies in the application of these advanced integrated models, particularly the hybrid model of LSTM-WHO-AO, providing a comparative analysis of their effectiveness in enhancing streamflow prediction accuracy. Furthermore, this study leverages MODIS-ERA5 merged derived SCA data to enhance the reliability and robustness of the models developed. By addressing the unique challenges of snow-fed catchments, this study contributes valuable insights into the potential of advanced machine learning approaches for hydrological forecasting applications.

2 Study Area

For this study, the Astore River is selected as a case study. The Astore River originates in Gilgit-Baltistan, a region in northern Pakistan. It passes through the Astore Valley before joining the Indus River. The river originates from the high-altitude glaciers of the Himalayas, specifically from the area surrounding the Deosai Plains and the southern slopes of the Nanga Parbat mountain range. The region's complicated geology, which was created by the Indian and Eurasian plates colliding to create magnificent mountain landscapes, shapes the course of the river. The area features a mix of sedimentary, metamorphic, and igneous rocks, with glacial deposits influencing the river's sediment load. Seasonal variations in flow patterns are the main hydrological source of the Astore River, which is fed by glacial meltwater. During the summer, water levels are influenced by precipitation and snowmelt as well as increased discharge due to glacial melt. The river's mountainous watershed contributes significantly to its sediment load, which causes severe turbidity. The dynamic and beautiful aspect of the river is enhanced by seasonal fluctuations in discharge, resulting from both glacial melt and monsoon rains. For the application of selected deep learning models, monthly streamflow data of the Astore station is collected from WAPDA for the duration of 1981 to 2015, whereas monthly snow-covered area data is obtained from merging MODIS and ERA5 snow cover products. Monthly snow-covered area (SCA) data from 2001 to 2015 were obtained from the MODIS MOD10A2 product (500 m resolution, 8-day composites, aggregated to monthly averages). For the earlier period, 1981–2000, MODIS data are not available. To extend the time series, we used ERA5 reanalysis (ERA5-Land, 0.1° resolution). ERA5 provides a continuous snow cover variable (snow cover fraction) from 1981 onward. A linear scaling correction was applied to ERA5 SCA to match the climatology of MODIS SCA during the overlapping period 2001–2015. To evaluate the prediction performance of the applied deep learning models, 20 independent runs were conducted using 70% of the data for training and 30% for testing.

3 Methods

In this study, a standalone LSTM deep learning model, along with several hybrid approaches integrating the LSTM model with different optimization algorithms, was employed for streamflow prediction. A brief description of these models is provided below:

3.1 Long Short-Term Memory (LSTM)

LSTM is a modified kind of RNN having a high capability in handling the long-term dependencies between variables [32]. The LSTM is composed of various units organized sequentially, and in the major cases, there are four units: the forget gate, the input gate, the memory cell state, and the output gate (Fig. 1). The mathematical expression of the LSTM algorithm can be formulated as follows:

$$h_t = \sigma\left(W_{xh}x_t + W_{hh}h_{t-1} + \beta_h\right) \tag{1}$$

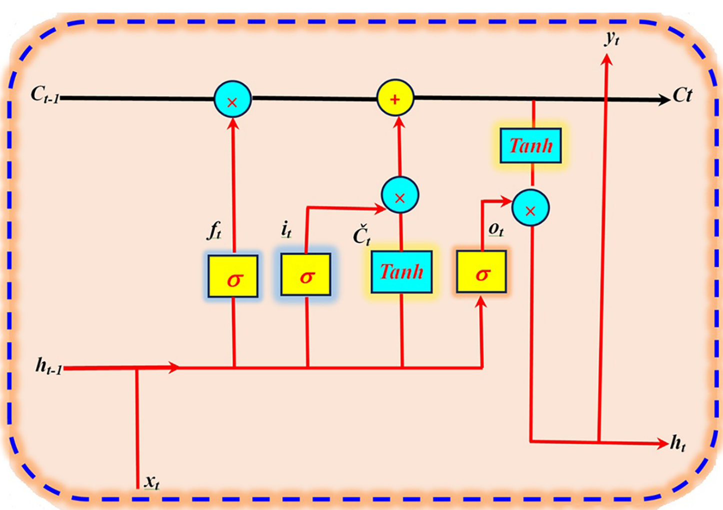

Figure 1: The long short-term memory (LSTM) architecture.

In the previous equation, h_t: $h_1, h_2, \cdots, h_t$ denotes the hidden state sequence, and x_t: $x_1, x_2, \cdots, x_t$ denotes the input time sequence, λ and β are the weights and the bias matrix, and σ is the sigmoidal activation function.

$$y_t = \sigma\left(\lambda_{hy}h_t + \beta_y\right) \tag{2}$$

The forget, input, and output gates can be expressed as follows [33]:

$$f_t = \sigma\left(\lambda_{hf} h_{t-1} + \lambda_{xf} x_t + \beta_f\right) \tag{3}$$

$$i_t = \sigma\left(\lambda_{hi} h_{t-1} + \lambda_{xi} x_t + \beta_i\right) \tag{4}$$

$$o_t = \sigma\left(\lambda_{ho} h_{t-1} + \lambda_{xo} x_t + \beta_o\right) \tag{5}$$

$$\tilde{c}_t = tanh\left(\lambda_{hc} h_{t-1} + \lambda_{xc} x_t + \beta_c\right) \tag{6}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \tag{7}$$

$$h_t = o_t \odot tanh\left(c_t\right) \tag{8}$$

In the previous equations, f_t, i_t, and o_t are the forget, input, and output gates, respectively, λ_{hf}, λ_{xf}, λ_{hi}, λ_{xi}, λ_{ho}, λ_{xo}, λ_{hc}, λ_{xc}, β_f, β_i, β_o, β_c are the weights and biases parameters to be updated during the training process. The $\tilde{c}_t$ is the candidate memory cell [33,34].

3.2 Genetic Algorithm (GA)

The GA is one of the first proposed methods for solving function optimization. The GA was developed by Holland [35], and a high number of applications can be found in the literature [36]. For GA optimization, we must identify some important components: (*i*) fitness function, (*ii*) individual, and (*iii*) population. The fitness function is the function to be optimized. For a given optimization problem, each individual is considered as a possible solution, and the population is an ensemble of individuals. The GA should be viewed as an ensemble of solutions, i.e., chromosomes implemented by three important components, i.e., selection, crossover, and mutation. The GA starts by randomly initializing the population, and during the selection, a new population is then selected based on the fitness of each individual. After that, the crossover is executed. Finally, the GA is closed by the mutation to generate a new set of probable solutions (Fig. 2), i.e., new generation, by respecting the following steps: (*i*) accounting the score of each individual as its fitness value, (*ii*) the children's are produced based on the fitness of the parents, (*iii*) a new generation has been established using the children's, and (*iv*) we stop the processes with respect to the iteration criteria [37,38].

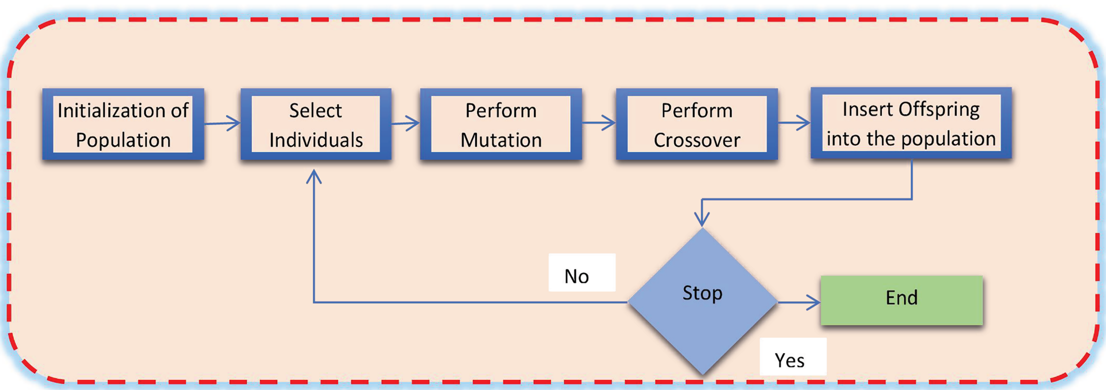

Figure 2: Flowchart of the genetic algorithm (GA) algorithm.

3.3 Grey Wolf Optimizer (GWO)

GWO is a swarm-based optimization algorithm inspired by the leadership hierarchy and hunting mechanism of grey wolves [39]. In the present study, GWO was used to optimize the hyperparameters of the LSTM model through iterative position updates guided by the three best candidate solutions (α, β, and δ wolves). The position update mechanism is expressed as follows:

$$\vec{X}(t+1) = \vec{X}_P(t) - (2\vec{a}\cdot\vec{r}_1 - \vec{a})\cdot\left|(2\cdot\vec{r}_2)\cdot\vec{X}_P(t) - \vec{X}(t)\right| \tag{9}$$

where t is the iteration number, $\vec{X}_P$ is the position vector of the prey, and $\vec{X}$ indicates the position vector of the grey wolf, $\vec{a}$ is a component that linearly decreases from 2 to 0 during the iteration process, and r_1, r_2 are random vectors in the range of [0, 1] [39].

The second step would be the hunting, and it is formulated as follows:

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1\cdot\left(\vec{D}_\alpha\right), \vec{X}_2 = \vec{X}_\beta - \vec{A}_2\cdot\left(\vec{D}_\beta\right), \vec{X}_3 = \vec{X}_\delta - \vec{A}_3\cdot\left(\vec{D}_\delta\right) \tag{10}$$

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \tag{11}$$

where $\vec{D}_\alpha$, $\vec{D}_\beta$ and $\vec{D}_\delta$ represent the distances between the search agents and the three leading wolves. $\vec{X}_\alpha$, $\vec{X}_\beta$, $\vec{X}_\delta$ are the three best solutions determined during the iteration process, and $\vec{A}$, $\vec{C}$, $\vec{D}$ are coefficient vectors. Therefore, the agents update their position in response to the prey position, and the solutions are updated accordingly (Fig. 3).

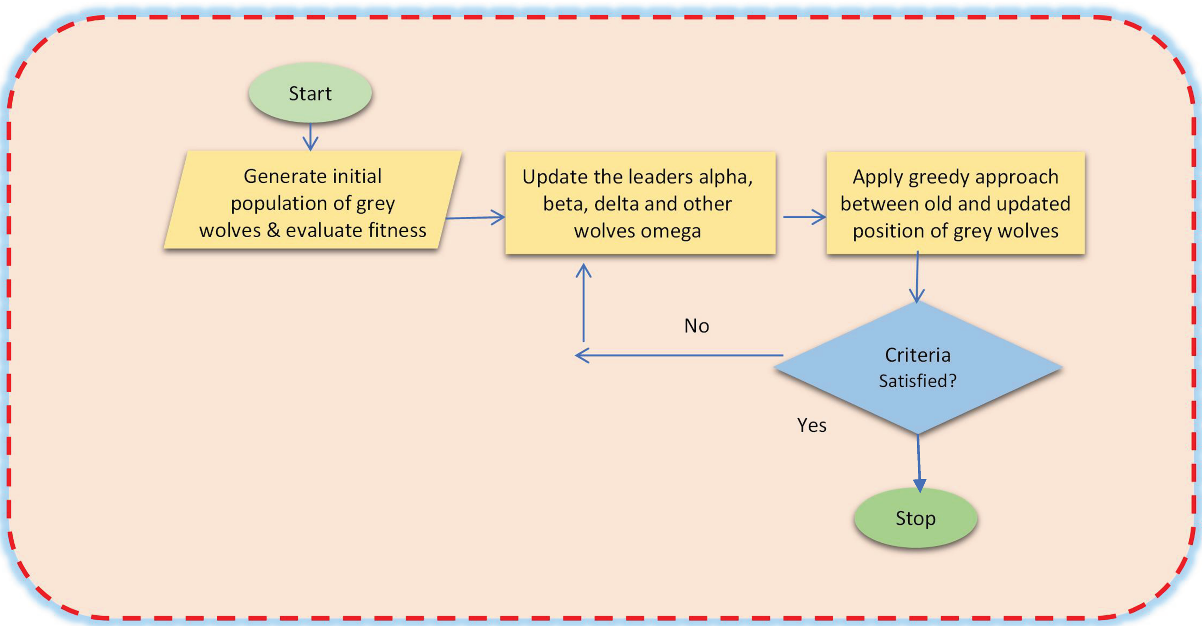

Figure 3: Flowchart of the GWO algorithm.

3.4 Whale Optimization Algorithm (WOA)

The WOA was developed by Mirjalili and Lewis [40] and was inspired by the hunting behavior of humpback whales' social behavior (Fig. 4). The overall WOA optimization process can be classified into three steps: (*i*) encircling prey, (*ii*) bubble-net attack strategy, and (*iii*) random search mechanism [41]. According to Mirjalili and Lewis [40], the optimization process of the WOA can be formulated as follows. For the encircling prey:

$$\vec{X}(t+1) = \vec{X}^*(t) - (2\vec{a}\cdot\vec{r} - \vec{a})\cdot\left|(2\cdot\vec{r})\cdot\vec{X}^*(t) - \vec{X}(t)\right| \tag{12}$$

where, $\vec{X}^*$ is the optimal solution at the iteration (t), $\vec{X}$ is the actual solution, $\left|(2\cdot\vec{r})\cdot\vec{X}^*(t) - \vec{X}(t)\right|$ correspond to the absolute value, $\vec{a}$ is linearly decreased from 2 to 0, $\vec{r}$ is a random vector.

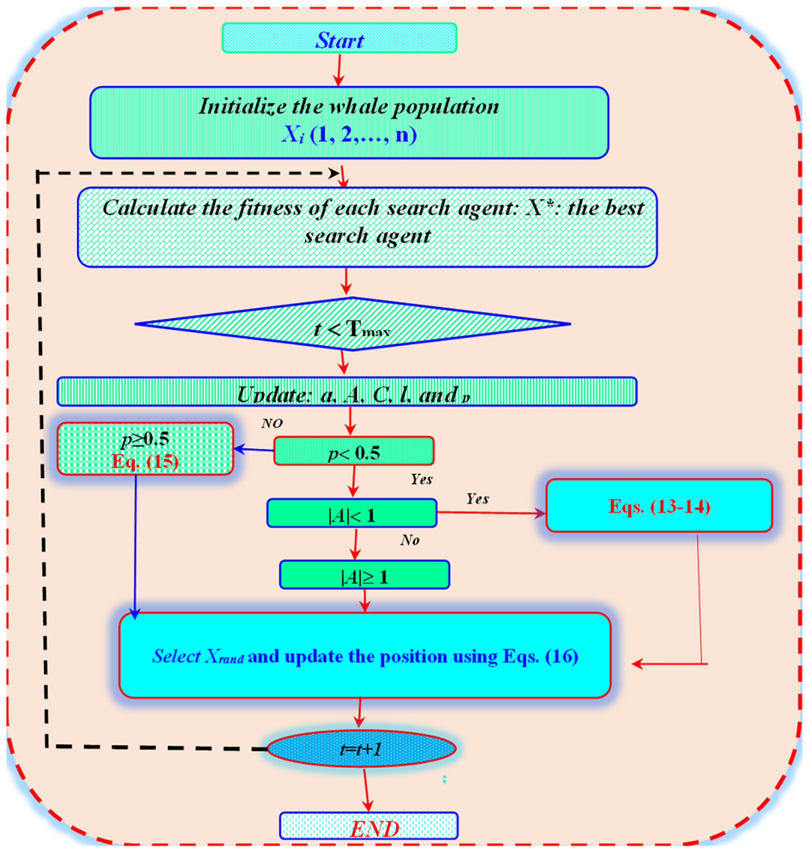

Figure 4: Flowchart of whale optimization algorithm (WOA).

The bubble-net attack strategy, i.e., the exploitation phase, is achieved in two phases, namely, the Shrinking encircling mechanism and the Spiral updating position, and it can be formulated as follows:

$$\vec{X}(t+1) = \begin{cases} \vec{X}^*(t) - (2\vec{a}\cdot\vec{r} - \vec{a})\cdot\left|(2\cdot\vec{r})\cdot\vec{X}^*(t) - \vec{X}(t)\right| & if\ p < 0.5 \\ \vec{D}'\cdot e^{bl}\cdot\cos(2\pi l) + \vec{X}^*(t) & if\ p \geq 0.5 \end{cases} \tag{13}$$

where p is a random number in [0, 1].

3.5 *Aquila Optimizer (AO)*

AO is an optimization method developed by Abualigah et al. [42] and mainly inspired by the hunting behavior of Aquila. Generally speaking, an ensemble of candidates (N) with a dimension (D) leads to the creation of a matrix (X) which reflects the number of solutions $(N \times D)$. The algorithm starts with an initialization stage as follows (Fig. 5):

$$X_{ij} = \beta \times \left(UB_j - LB_j\right) + LB_j, i = 1, 2, \ldots.., N, j = 1, 2, \ldots.., Dim \tag{14}$$

where X_{ij} the probable candidate solution, N denotes the total number of candidate solutions, and dim refers to the size of the dimensionality of the optimization problem, β is a random number varying from 0 to 1, *UB* and *LB* are the lower and upper bounds. According to Abualigah et al. [42], the AO algorithm can be achieved in four steps and they are denoted as X_1, X_2, X_3 and X_4, and they are denoted as: (*i*) the expanded exploration stage (i.e., high soar and vertical stoop), (*ii*) the narrowed exploration (i.e., the contour flight with short glide), (*iii*) the expanded exploitation (i.e., low flight with short decent), and (*iv*) the narrowed exploitation (i.e., the walking and grabbing prey) [43]. The first stage (i.e., the expanded exploration) can be mathematically formulated as follows (i.e., the X_1):

$$X_1(t+1) = X_{best}(t) \times \left(1 - \frac{t}{T}\right) + \left(X_M(t) - X_{best}(t) * \beta\right) \tag{15}$$

where $\left(1 - \frac{t}{T}\right)$ is used for controlling the iteration process by considering that t and T are the actual and maximum number of iterations, X_{best} the optimal solution obtained at the *i*th iteration, X_M corresponds to the mean values of all solutions, and $X_1(t+1)$ corresponds to the solution at the next iteration.

The third step (X_3), which is the expanded exploitation, can be expressed as follows:

$$X_3(t+1) = \left(X_{best}(t) - X_M(t)\right) \times \alpha - \beta + \left((UB - LB) \times \beta + LB\right) \times \delta \tag{16}$$

where $X_3(t+1)$ corresponds to the solution at the next iteration t, X_{best} is the position of the prey, i.e., the best solution, UB and LB are the upper and lower bounds, α and β are adjustment parameters equal to 0.1. More details about the AO can be found in [42].

3.6 Wild Horse Optimizer Algorithm (WHO)

Naruei and Keynia [44] developed the wild horse optimizer algorithm (WHO). The WHO algorithm can be achieved in five steps as follow: (*i*) an initial population is randomly generated and a new a system of hours is established, (*ii*) grazing and mating of horses, (*iii*) a leader called stallion is designated for leading the group, (*iv*) exchange and selection of leaders, and (*v*) save the optimal solution [44]. In the first step, a new initial population is randomly generated as follows:

$$(\vec{x}) = \{\vec{x}_1, \vec{x}_2, \vec{x}_3, \vec{x}_4, \cdots\cdots\cdots\cdots\cdots, \vec{x}_n\} \tag{17}$$

In the second step, the grazing behavior is simulated as follows:

$$\overline{X}_{i,G}^{j} = 2Z\cos(2\pi RZ) \times \left(ST^j - X_{i,G}^j\right) + ST^j \tag{18}$$

where $\overline{X}_{i,G}^{j}$ is the updated place of the group member at the end of the iteration t, and the $X_{i,G}^{j}$ is the actual position of the group member, ST^j is the position of the leader called the stallion, Z is an adaptive mechanism, R is a uniform random number in the range [−2, 2], and π is equal to 3.14 [44].

The group leader can be formulated as follows:

$$\overline{ST}_{G_i} = \begin{Bmatrix} 2Z\cos(2\pi RZ) \times (\mathrm{WH} - ST_{G_i}) + \mathrm{WH} & if\ R_3 > 0.5 \\ 2Z\cos(2\pi RZ) \times (\mathrm{WH} - ST_{G_i}) - \mathrm{WH} & if\ R_3 \le 0.5 \end{Bmatrix} \tag{19}$$

where the next location of the leader is $\overline{ST}_{G_i}$ of the group i, the location of the water hole is WH, and the present location of the leader is ST_{G_i} of the group i, and R is a uniform random number in the range $[-2, 2]$ [44].

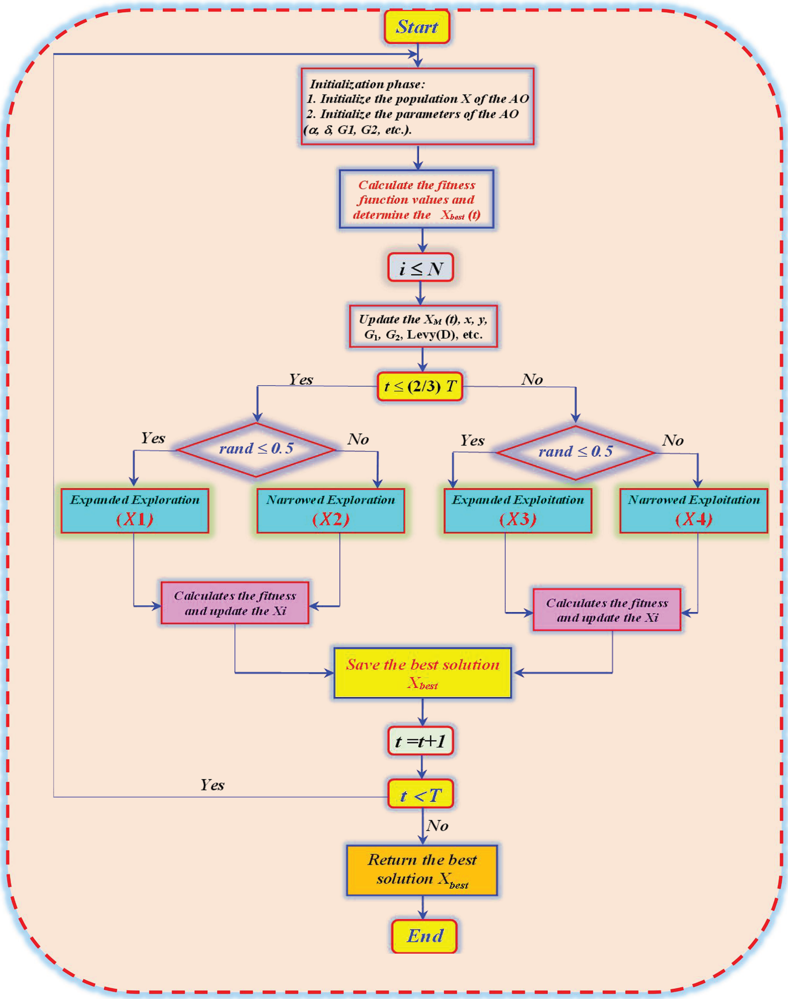

Figure 5: Flowchart of the aquila optimizer (AO) algorithm.

3.7 Hybrid of Wild Horse Optimizer Algorithm & Aquila Optimizer (WHOAO)

Among the disadvantages of the WHO, we can report its poor convergence ability; therefore, in this study, we propose a new method aimed at improving the performance of the WHO algorithm. The Improved WHO (WHOAO) is based on the AO algorithm. The major advantage of the combination of the WHO and the AO algorithms is the improvement of the WHO by balancing global and local exploitation. Furthermore, the new optimization method, WHOAO, can help in avoiding the issue of population stabilization and, more precisely, overcoming the issue of local optima. The proposed framework was developed to address one of the main limitations of the original WHO algorithm, namely its weak convergence behavior. This limitation motivated the integration of the AO strategy into the WHO search mechanism. In the first optimization phase, the expanded exploration behavior of AO is selectively incorporated into the WHO position-update process to replace the grazing behavior phase under a randomized condition. According to the implemented strategy, when the generated random value exceeds 0.5, the AO exploration equation is activated, allowing the search agents to move toward broader regions of the solution space. This selective integration improves population diversity, strengthens the exploration capability of WHOAO, and provides a better balance between exploration and exploitation. Once promising candidate solutions are identified, the second optimization phase starts by training the LSTM's configuration, including the number of hidden units, training epochs, mini-batch size, and learning rate. The first phase is mainly responsible for improving the global search behavior of WHOAO and reducing premature convergence, whereas the second phase concentrates on refining the LSTM configuration through MSE-based evaluation to improve prediction accuracy and model stability.

The interaction between AO and WHO follows a sequential hybrid optimization process within the WHOAO framework.

1. The AO exploration equation is conditionally integrated into the WHO search process to replace the grazing behavior when the random value exceeds 0.5, which improves exploration capability and reduces premature convergence.
2. The WHOAO algorithm then searches for suitable LSTM training settings, including hidden units, learning rate, mini-batch size, and training epochs, using MSE as the fitness criterion.
3. The LSTM network is finally constructed and trained using the optimized settings to process the input data and generate the prediction results.

This structure allows WHOAO to improve convergence behavior and search efficiency, while the LSTM model independently learns the internal network parameters.

3.8 LSTM-WHOAO

This study presents a new method aimed at improving the accuracy of the WHO algorithm. The goal is to enhance its exploration capability by integrating the refined exploitation strategy from the Aquila Optimizer (AO) algorithm. By assimilating AO's localized search mechanism, the modified WHO, termed Improved WHO (WHOAO), demonstrates superior adaptability in swiftly exploiting optimal solutions compared to its original counterpart. Consequently, the proposed WHOAO method is applied to optimize the hyperparameters of the LSTM technique, resulting in heightened efficacy of LSTM applications. This novel approach effectively mitigates the constraints of the original WHO algorithm while concurrently offering a more resilient and effective optimization framework for LSTM methodologies.

To boost the exploration phase of the WHO, the expanded exploration phase of AO (Eq. (20)) is selectively employed in lieu of the WHO's grazing behavior phase based on a randomized criterion. If the random value exceeds 0.5, the AO phase is adopted.

$$x_{n+1} = x_{best}(t) \times (1 - \frac{t}{T}) + mean(x(t)) - x_{best}(t) \times rand \tag{20}$$

Here, x_{best} refers to the best-selected solution. t and T represent the current and total iteration numbers. $x(t)$ signifies the solution at iteration t. The variable rand is a randomly generated value ranging from 0 to 1.

Furthermore, the application of the WHOAO algorithm extends to the optimization of weights and bias parameters within the LSTM framework. The WHOAO-LSTM framework initiates by defining the requisite parameters and acquiring experimental data, which is subsequently split into training-validation-testing sets. The primary objective of the WHOAO algorithm revolves around exploring optimal weights and biases for the LSTM method to enhance its accuracy. To evaluate the efficacy of candidate parameters, the MSE is utilized, as depicted in Eq. (21):

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(Q_{o\,i} - Q_{c\,i}\right)^2 \tag{21}$$

In the above equation, $Q_{o\,i}$ denotes the sample size. The candidate parameters yielding the minimum error are deemed optimal solutions. The WHOAO algorithm iteratively explores various parameter combinations until the maximum number of fitness function evaluations is attained. Upon determining the optimal parameters, they are utilized to optimize the LSTM model, marking the initiation of the testing phase.

4 Performance Indicators and Model Setup

The performance of LSTM, LSTM-GA, LSTM-GWO, LSTM-WOA, LSTM-AO, LSTM-WHO, and LSTM-WHOAO models was investigated for monthly streamflow prediction in the high-altitude snow-fed Astore River catchment in Pakistan using antecedent streamflow values selected through autocorrelation analysis. The comparison models were selected to represent both established and recently successful optimization paradigms. Specifically, LSTM-GA (Genetic Algorithm) serves as a baseline representing a classic, widely used evolutionary algorithm in hydrological modeling. LSTM-GWO (Grey Wolf Optimizer) and LSTM-WOA (Whale Optimization Algorithm) were included because they have emerged in the past decade as highly successful, state-of-the-art metaheuristic algorithms across various engineering and hydrological applications. This selection allows a fair assessment of the proposed LSTM-WHOAO against both a traditional method (GA) and contemporary competitive algorithms (GWO, WOA). All LSTM-based models were configured using the proposed WHOAO optimization framework. The optimization process considered several training parameters, including the number of hidden units (32–256), learning rate (0.001–0.05), mini-batch size (16–64), and training epochs (50–200). The search ranges were selected according to preliminary experiments and model stability considerations. The AO exploration mechanism was conditionally integrated into the WHO search process to improve exploration capability and reduce premature convergence toward local optima. The WHOAO algorithm iteratively searched for the optimal parameter combination using MSE as the fitness criterion. After obtaining the optimized parameter settings, the LSTM model was constructed and trained using the selected configuration to generate the prediction results. SCA (snow-covered area) data generated by merging MODIS with ERA5 and MN (month number) were also used in the inputs to improve the models' accuracy. The outcomes acquired from the deep learning models were compared using the following criteria in the current study:

$$RMSE\text{: } Root\ Mean\ Square\ Error = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[(Q_0)_i - (Q_C)_i\right]^2} \tag{22}$$

$$MAE\text{: } Mean\ Absolute\ Error = \frac{1}{N}\sum_{i=1}^{N}\left|(Q_0)_i - (Q_C)_i\right| \tag{23}$$

$$NSE\text{: } Nash - Suutcliffe = 1 - \frac{\sum_{i=1}^{N}\left[(Q_0)_i - (Q_C)_i\right]^2}{\sum_{i=1}^{N}\left[(Q_0)_i - \overline{Q}_0\right]^2}, -\infty < NSE \le 1 \tag{24}$$

where $Q_c, Q_o, \overline{Q}_o$ are calculated, observed, and average streamflow, respectively, and N is the total number of data.

5 Results and Discussion

5.1 Results

The training and testing outcomes of the LSTM models in streamflow prediction (Table 1) offer valuable insights into the model's performance and its generalization ability. The training stage metrics show a clear improvement as more input features are added to the model. For instance, the RMSE and MAE decrease significantly when moving from a simple model with only previous streamflow data (Qt-1) to more complex models that include additional time-lagged streamflow values (Qt-11, Qt-12), snow-covered area (SCA), and month number (MN). The high R^2 and NSE values in the training set indicate that the model is capturing the underlying patterns in the data well, explaining a substantial portion of the variability in streamflow. The models with more input features perform relatively well on the testing data, indicating that including features like month number (MN) and snow-covered area (SCA) does improve generalization to some extent. The inclusion of the month number (MN) is crucial for capturing seasonal patterns in streamflow, which are common due to variations in precipitation, snowmelt, and other seasonal factors. The improved metrics in models that include MN suggest that the model is better able to predict streamflow by accounting for these seasonal variations. The models that include a combination of lagged streamflow values, SCA, and MN perform better overall, suggesting that these additional inputs provide valuable information that improves model robustness. The relatively good performance on the test set, indicates that the model has learned useful patterns, but further refinement is needed to enhance its generalization ability.

Table 1: Training and test statistics of the models for Streamflow prediction—LSTM.

Model Inputs	Training Period				Test Period			
	RMSE	MAE	R^2	NSE	RMSE	MAE	R^2	NSE
Qt-1	91.85	63.56	0.568	0.559	95.68	68.93	0.557	0.549
Qt-1, Qt-11	61.28	37.92	0.811	0.805	82.18	51.26	0.676	0.667
Qt-1, Qt-11, Qt-12	59.36	36.13	0.826	0.821	81.73	49.56	0.682	0.673
Qt-1, Qt-11, Qt-12, SCA	56.46	35.71	0.839	0.831	77.63	47.46	0.702	0.691
Qt-1, Qt-11, Qt-12, MN	53.81	33.62	0.846	0.839	75.26	46.82	0.705	0.701
Qt-1, Qt-11, Qt-12, SCA, MN	51.19	31.46	0.853	0.847	73.68	43.56	0.711	0.706
Mean	62.325	39.733	0.791	0.784	81.027	51.265	0.672	0.665

The LSTM models optimized with the Genetic Algorithm (LSTM-GA) provide insights into the model's performance during both the training and testing stages (Table 2). The models demonstrate consistent improvement in performance metrics with more inputs. In the training stage, RMSE drops from 49.62 to 71.26 in testing, and R^2 increases from 0.865 in training to 0.739 in testing. Despite the typical increase in RMSE (by approximately 18 points) and MAE (by around 11 points) from training to testing, the model generalizes well. The inclusion of lagged streamflow values, SCA, and MN significantly improves accuracy, with NSE values reaching 0.858 in training and 0.677 in testing. The LSTM models optimized using Grey Wolf Optimization (LSTM-GWO) models show a consistent improvement in performance metrics in the training stage as more inputs are incorporated (Table 3). The models show consistent improvement in RMSE, which decreases from 48.85 in training to 70.72 in testing with all inputs. R^2 improves from 0.876 in training to 0.757 in testing, indicating good generalization. Although the R^2 drops from 0.810 in training to 0.697 in testing, the model still performs effectively. The inclusion of lagged values, SCA, and MN enhances the model's ability to predict streamflow, capturing crucial seasonal and temporal patterns.

Table 2: Training and test statistics of the models for Streamflow prediction—LSTM-GA.

Model Inputs	Training Period				Test Period			
	RMSE	MAE	R^2	NSE	RMSE	MAE	R^2	NSE
Qt-1	91.28	62.92	0.573	0.564	93.82	66.25	0.561	0.552
Qt-1, Qt-11	60.84	36.83	0.826	0.815	80.56	50.34	0.684	0.673
Qt-1, Qt-11, Qt-12	58.62	35.84	0.837	0.829	79.48	48.39	0.695	0.681
Qt-1, Qt-11, Qt-12, SCA	55.83	34.62	0.842	0.832	76.17	46.81	0.708	0.701
Qt-1, Qt-11, Qt-12, MN	52.18	32.99	0.851	0.844	74.82	45.62	0.729	0.721
Qt-1, Qt-11, Qt-12, SCA, MN	49.62	30.61	0.865	0.858	71.26	43.53	0.739	0.732
Mean	61.395	38.968	0.799	0.790	79.352	50.157	0.686	0.677

Table 3: Training and test statistics of the models for Streamflow prediction—LSTM-GWO.

Model Inputs	Training Period				Test Period			
	RMSE	MAE	R^2	NSE	RMSE	MAE	R^2	NSE
Qt-1	90.78	62.31	0.577	0.566	93.28	65.89	0.564	0.556
Qt-1, Qt-11	57.34	34.58	0.839	0.827	78.49	49.16	0.704	0.691
Qt-1, Qt-11, Qt-12	56.48	34.03	0.846	0.839	77.81	47.92	0.707	0.697
Qt-1, Qt-11, Qt-12, SCA	54.91	33.84	0.852	0.847	74.76	45.96	0.715	0.708
Qt-1, Qt-11, Qt-12, MN	52.68	30.72	0.869	0.865	73.92	44.52	0.736	0.729
Qt-1, Qt-11, Qt-12, SCA, MN	48.85	29.39	0.876	0.864	70.72	41.36	0.757	0.748
Mean	60.173	37.478	0.810	0.801	78.163	49.135	0.697	0.688

LSTM-WOA shows a marked improvement in performance metrics as more inputs are incorporated (Table 4). The models show significant improvements as more inputs are added. RMSE decreases from 90.25 to 46.75 in training and from 92.93 to 68.26 in testing, while R^2 increases from 0.579 to 0.882 in training and from 0.568 to 0.774 in testing. Despite a slight decline in performance from training to testing, with mean RMSE increasing from 58.978 to 76.607, the models maintain strong generalization. The inclusion of lagged streamflow values, SCA, and MN enhances model accuracy, particularly in capturing seasonal dynamics. The LSTM models optimized using the Aquila Optimizer (LSTM-AO) significant improvement as more inputs are included (Table 5). The models exhibited considerable improvement with the inclusion of additional input variables. Specifically, RMSE decreased from 89.93 to 43.91 during the training phase and from 92.37 to 66.52 during testing, while the R^2 values increased from 0.582 to 0.894 in training and from 0.570 to 0.790 in testing. Although RMSE increases from 57.135 in training to 74.945 in testing, the models maintain good generalization. Including lagged streamflow values, SCA, and MN improves the models' ability to predict streamflow accurately. The LSTM models optimized using the Wild Horse Optimizer (LSTM-WHO) models show substantial improvement as more inputs are incorporated (Table 6). The models demonstrated notable performance improvements as the number of input variables increased. RMSE values decreased from 89.17 to 40.36 during training and from 91.82 to 63.72 during testing, whereas the R^2 values improved from 0.586 to 0.914 in the training phase and from 0.575 to 0.814 in the testing phase. Although testing metrics are higher, with mean RMSE increasing from 53.900 to 72.735, the models still generalize well. The inclusion of lagged streamflow values, SCA, and MN is crucial for capturing temporal and seasonal dynamics, improving prediction accuracy. The LSTM models optimized using the combination of Wild Horse Optimizer and Aquila Optimizer (LSTM-WHOAO) models show significant improvement with more inputs in the training stage (Table 7). The models show substantial improvement with more inputs. The LSTM models optimized using the combination of Wild Horse and Aquila Optimizers (LSTM-WHOAO) show substantial improvement with more inputs. RMSE decreases from 88.86 to 38.91 in training and from 90.84 to 58.31 in testing, while R^2 increases from 0.598 to 0.932 in training and from 0.582 to 0.852 in testing. Although there is an increase in RMSE from 51.930 in training to 70.413 in testing, the models maintain strong explanatory power and predictive accuracy. Including lagged streamflow values, SCA, and MN further improves model performance, demonstrating effective generalization and capturing the underlying dynamics of streamflow.

Table 4: Training and test statistics of the models for Streamflow prediction—LSTM-WOA.

Model Inputs	Training Period				Test Period			
	RMSE	MAE	R^2	NSE	RMSE	MAE	R^2	NSE
Qt-1	90.25	61.85	0.579	0.569	92.93	65.27	0.568	0.559
Qt-1, Qt-11	56.8	33.17	0.847	0.841	77.68	48.76	0.713	0.702
Qt-1, Qt-11, Qt-12	55.67	32.84	0.851	0.84	76.68	47.18	0.718	0.709
Qt-1, Qt-11, Qt-12, SCA	52.84	32.76	0.871	0.862	72.48	44.72	0.734	0.725
Qt-1, Qt-11, Qt-12, MN	51.56	28.49	0.878	0.873	71.61	43.28	0.756	0.733
Qt-1, Qt-11, Qt-12, SCA, MN	46.75	28.07	0.882	0.873	68.26	40.35	0.774	0.765
Mean	58.978	36.197	0.818	0.810	76.607	48.260	0.711	0.699

Table 5: Training and test statistics of the models for Streamflow prediction—LSTM-AO.

Model Inputs	Training Period				Test Period			
	RMSE	MAE	R^2	NSE	RMSE	MAE	R^2	NSE
Qt-1	89.93	61.28	0.582	0.573	92.37	64.95	0.57	0.561
Qt-1, Qt-11	54.82	31.72	0.856	0.849	75.82	47.58	0.728	0.719
Qt-1, Qt-11, Qt-12	53.72	30.81	0.863	0.857	74.83	46.58	0.732	0.724
Qt-1, Qt-11, Qt-12, SCA	50.62	31.82	0.881	0.876	70.41	43.62	0.751	0.734
Qt-1, Qt-11, Qt-12, MN	49.81	26.82	0.893	0.885	69.72	41.06	0.766	0.754
Qt-1, Qt-11, Qt-12, SCA, MN	43.91	27.72	0.894	0.885	66.52	38.12	0.790	0.783
Mean	57.135	35.028	0.828	0.821	74.945	46.985	0.723	0.713

The performance of different LSTM methods provides valuable insights into their effectiveness in predicting streamflow. It is evident from the average statistics of the LSTM models in the training stage (Tables 1–7) that all models show progressive improvement as they are optimized with more advanced algorithms. The combined Wild Horse Optimizer and Aquila Optimizer (LSTM-WHOAO) model achieves the best performance in the training stage, with the lowest mean RMSE (51.930) and highest mean R^2 (0.851), indicating a superior fit to the training data. The models optimized with GA, GWO, WOA, AO, and WHO all perform better than the standard LSTM, with incremental gains in RMSE and R^2. Similar trends are observed in the testing stage, where the LSTM-WHOAO model also performs the best, with the lowest RMSE (70.413) and the highest R^2 (0.755), showing good generalization to unseen data. The LSTM-GWO and LSTM-WOA models also show significant improvements over the standard LSTM, indicating that these optimization techniques effectively enhance model generalization. The models optimized with WHOAO show the least performance drop from training to testing, suggesting the best generalization ability. This is crucial in practical applications where the model needs to perform well on unseen data. Models like LSTM-WOA and LSTM-AO also show better generalization compared to the standard LSTM, with higher R^2 and NSE values in the testing stage. The standard LSTM model exhibits a more significant performance drop, indicating potential overfitting. The advanced optimization techniques help mitigate this by finding a balance between model complexity and generalization. The combination of Wild Horse Optimizer and Aquila Optimizer (WHOAO) proves to be the most effective, improving both training accuracy and testing performance. This suggests that the hybrid approach leverages the strengths of both optimizers to fine-tune the model parameters more effectively. Each optimization technique provides incremental improvements over the basic LSTM model, with WHO, AO, and WOA showing particularly strong results. The comparative analysis shows that LSTM models optimized with advanced algorithms significantly outperform the standard LSTM model in streamflow prediction. Among these, the LSTM-WHOAO model demonstrates the best overall performance in both accuracy and generalization. These results suggest that employing sophisticated optimization techniques can substantially enhance the predictive capabilities of LSTM models in hydrological applications. For real-world scenarios, where accuracy and reliability are critical, the use of models like LSTM-WHOAO is recommended.

Table 6: Training and test statistics of the models for Streamflow prediction—LSTM-WHO.

Model Inputs	Training Period				Test Period			
	RMSE	MAE	R^2	NSE	RMSE	MAE	R^2	NSE
Qt-1	89.17	60.86	0.586	0.577	91.82	63.07	0.575	0.566
Qt-1, Qt-11	51.89	29.68	0.869	0.862	74.91	46.43	0.733	0.726
Qt-1, Qt-11, Qt-12	49.54	27.82	0.876	0.868	71.06	44.13	0.741	0.731
Qt-1, Qt-11, Qt-12, SCA	46.81	26.13	0.892	0.885	68.72	42.72	0.762	0.753
Qt-1, Qt-11, Qt-12, MN	45.63	25.32	0.902	0.893	66.18	39.76	0.786	0.777
Qt-1, Qt-11, Qt-12, SCA, MN	40.36	25.28	0.914	0.901	63.72	36.81	0.814	0.805
Mean	53.900	32.515	0.840	0.831	72.735	45.487	0.735	0.726

Table 7: Training and test statistics of the models for Streamflow prediction—LSTM-WHOAO.

Model Inputs	Training Period				Test Period			
	RMSE	MAE	R^2	NSE	RMSE	MAE	R^2	NSE
Qt-1	88.86	60.21	0.598	0.592	90.84	62.76	0.582	0.574
Qt-1, Qt-11	49.67	28.47	0.876	0.869	73.46	45.38	0.748	0.739
Qt-1, Qt-11, Qt-12	47.17	25.72	0.882	0.874	69.58	43.75	0.756	0.739
Qt-1, Qt-11, Qt-12, SCA	45.61	24.46	0.899	0.892	66.48	41.86	0.778	0.772
Qt-1, Qt-11, Qt-12, MN	41.36	22.48	0.916	0.917	63.81	37.74	0.811	0.804
Qt-1, Qt-11, Qt-12, SCA, MN	38.91	20.99	0.932	0.927	58.31	34.06	0.852	0.845
Mean	51.930	30.388	0.851	0.845	70.413	44.258	0.755	0.746

Scatter plots, Taylor diagrams, and violin plots are utilized to compare the performances of these deep learning models graphically. LSTM-WHOAO provided better predictions close to the observed values according to scatter, Taylor, and violin plots (Figs. 6–8). It is clear from the Taylor diagram that the LSTM-WHOAO has the highest correlation (the least angle represents the least correlation) with the observed data and the lowest standard deviation (the least radial distance from the origin). The violin plot in Fig. 8 shows the distribution shape for each method, with the width representing the density of data at different residual levels. The horizontal lines show the median, and the markers indicate the mean for each method. It is clear from the figure that the distribution of the LSTM-WHOAO model predictions are closer to the observed one compared to other models.

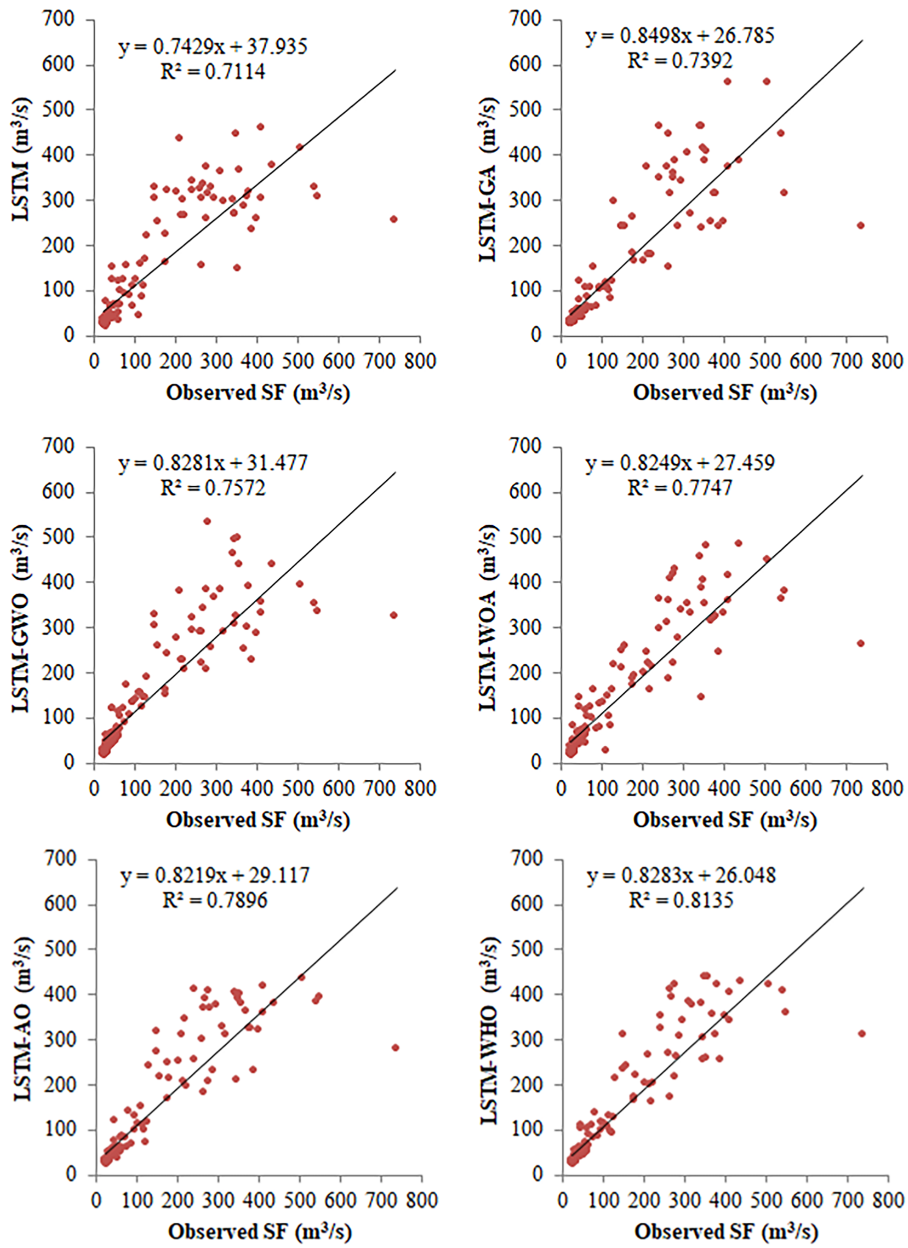

Figure 6: (Continued)

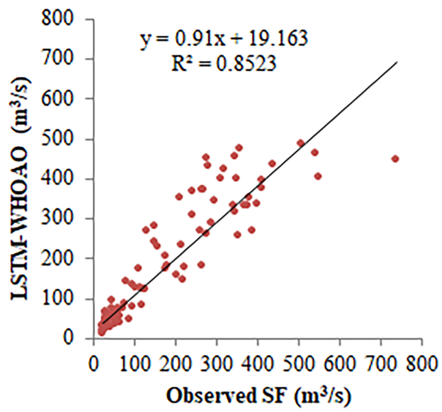

Figure 6: Scatterplots of the observed and predicted streamflow by different models in the test period using the best input combination.

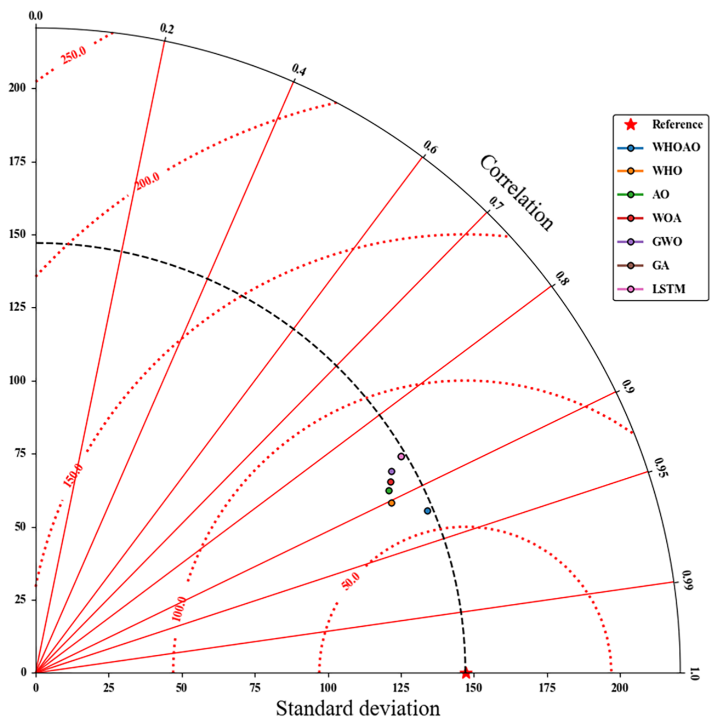

Figure 7: Taylor diagrams of the predicted streamflow by different models in the test period using the best input combination.

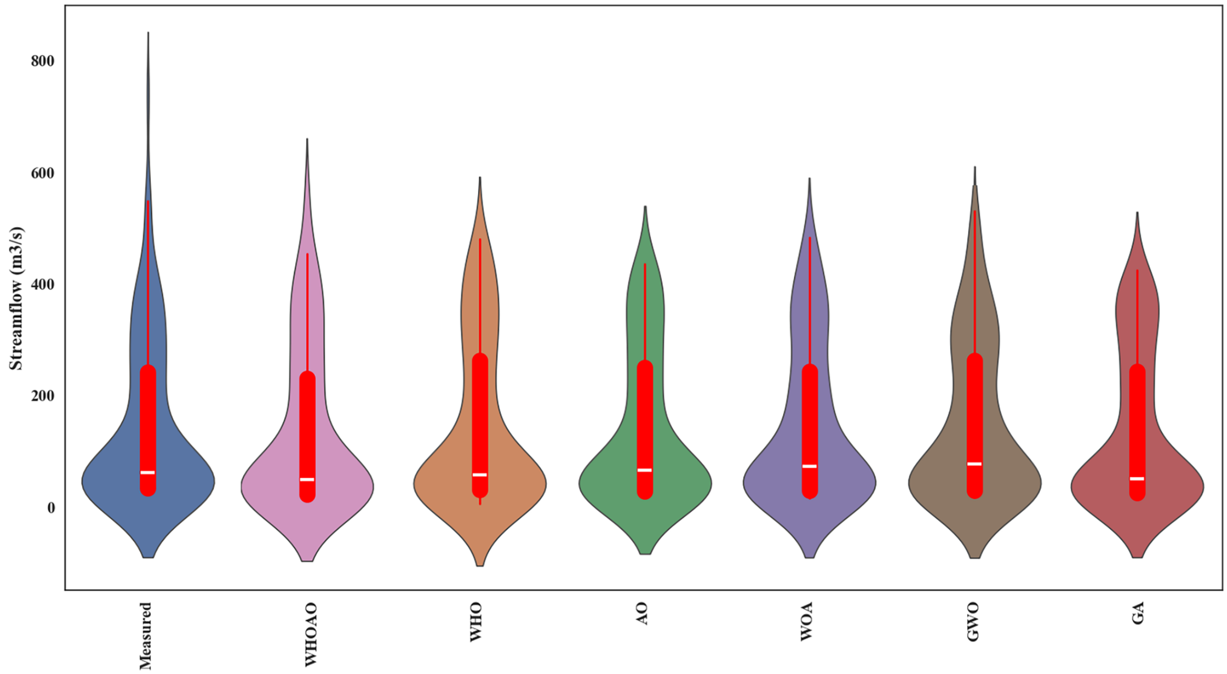

Figure 8: Violin Charts of the predicted streamflow by different models in the test period using the best input combination.

Peak streamflow events are critical for effective flood management, water resource planning, and ensuring the safety of hydropower infrastructure. Table 8 compares the performance of different LSTM-based models—LSTM, LSTM-GA, LSTM-GWO, LSTM-WOA, LSTM-AO, LSTM-WHO, and LSTM-WHOAO—in predicting peak streamflow events exceeding 360 cubic meters per second (m^3/s). The LSTM-WHOAO achieved the lowest absolute error (201.4%), indicating that it consistently provided more accurate predictions of peak streamflow compared to the other models. This suggests that the combination of Wild Horse Optimizer and Aquila Optimizer effectively enhances the model's ability to capture extreme flow events. LSTM-WHO also performed well, with a relatively low absolute error (233.7%) and showed better accuracy than the standalone LSTM and other hybrid models. LSTM-AO (249.9%) and LSTM-WOA (261.7%) models showed moderate performance improvements over LSTM, LSTM-GA, and LSTM-GWO, but were outperformed by LSTM-WHO and LSTM-WHOAO. LSTM-GWO and LSTM-GA exhibited higher errors, indicating that these optimizers might be less effective in capturing peak flow dynamics compared to WHO and AO. For certain peak flow events, like in June 2009 (412.2 m^3/s), the LSTM-GA and LSTM-WHOAO models showed the lowest relative prediction error, highlighting their potential effectiveness in predicting specific extreme events. Conversely, for the peak in August 2010 (738.4 m^3/s), the LSTM-WHOAO showed a significant improvement in prediction accuracy (39.6%) compared to other models, where errors were substantially higher. The LSTM-WHOAO displayed more consistent performance across different peaks, reflecting its robustness in predicting a wide range of peak flow events. The comparative analysis indicates that hybrid models, particularly LSTM-WHOAO, provide the most reliable predictions for peak streamflow events in snow-fed catchments. These models significantly reduce the relative prediction error compared to standalone LSTM models and other hybrid configurations, making them more suitable for forecasting extreme flow events, which are critical for flood risk management and water resource planning. The LSTM-WHOAO model's consistent performance across various peak events highlights its potential for practical application in hydrology, especially in regions prone to high variability in streamflow.

Table 8: The comparison of different models in peak streamflow prediction for the test period.

Date	Observed Values	Relative Prediction Error Percentage						
	Peaks > 360	LSTM %	LSTM-GA %	LSTM-GWO %	LSTM-WOA %	LSTM-AO %	LSTM-WHO %	LSTM-WHOAO %
May-06	389.4	39.3	37.8	41.3	37.1	40.7	34.6	31.3
Jun-08	399.0	34.9	37.1	27.7	16.4	19.2	11.9	40.9
Jun-09	412.2	25.9	9.5	19.8	13.0	12.5	16.6	9.3
Jun-10	413.5	−11.1	−35.5	14.0	−0.5	−1.2	2.2	4.0
Jul-10	549.6	44.1	42.7	38.9	30.8	28.4	34.6	26.5
Aug-10	738.4	65.2	67.2	56.1	64.5	61.8	57.9	39.6
Jun-12	369.0	22.0	32.0	31.7	14.5	1.3	3.4	10.5
Jul-12	540.4	39.1	17.6	34.9	33.1	29.0	24.7	14.7
Aug-12	379.5	15.9	17.0	−3.3	14.8	14.6	−11.3	7.7
Jul-13	437.1	13.5	11.1	−0.9	−10.8	12.8	2.3	0.3
Jul-13	509.0	18.4	−10.1	22.9	11.3	14.7	17.1	4.8
Jul-15	376.3	17.8	16.3	20.3	14.9	13.6	17.2	11.8
Absolute Error		347.1	333.9	311.7	261.7	249.9	233.7	201.4

To assess whether the performance improvements of LSTM-WHOAO over the baseline models are statistically meaningful, paired t-tests were conducted on the absolute prediction errors over the testing period. Table 9 summarizes the results. All comparisons yield negative t-statistics, indicating that LSTM-WHOAO consistently produces lower absolute errors than each baseline. Furthermore, all p-values are below 0.05, confirming that the differences are statistically significant. Therefore, the superior performance of the proposed LSTM-WHOAO model is unlikely to be due to chance. Notably, the smallest absolute t-statistic ($|t|$ = 1.32) occurs against LSTM-WHO, suggesting that the Wild Horse Optimizer alone is a strong competitor, while the largest improvement is observed against standard LSTM ($|t|$ = 3.53). In terms of prediction speed, the proposed LSTM-WHOAO model demonstrated competitive computational efficiency during inference. Based on the average runtime across different input configurations, LSTM-WHOAO required approximately 0.858 s per prediction, compared with 0.883 s for LSTM-WHO and 0.957 s for LSTM-AO. The reduced computational time of the proposed model may be attributed to the improved parameter optimization achieved through the WHOAO strategy.

Table 9: Paired t-test results comparing LSTM-WHOAO against baseline models over the testing period.

Baseline Model	t-Statistic	p-Value	Significant (p < 0.05)
LSTM	−3.5309	0.00064	Yes
LSTM-GA	−3.2825	0.00418	Yes
LSTM-GWO	−2.9681	0.00266	Yes
LSTM-WOA	−2.7316	0.00135	Yes
LSTM-AO	−1.8134	0.04132	Yes
LSTM-WHO	−1.3247	0.01628	Yes

5.2 Discussion

The LSTM models optimized with AO and WHO outperformed traditional LSTM models and other hybrid models in the study, indicating their superior ability to capture the complex dynamics of streamflow in snow-fed catchments. This aligns with recent findings by Qiao et al. [45] and Shabani et al. [46], who demonstrated that these optimizers can significantly enhance the performance of deep learning models in hydrological applications. The inclusion of SCA data in the model inputs was shown to significantly enhance the predictive accuracy of the models in this study, which is consistent with the findings of studies like Thapa et al. [21] and Parisouj et al. [22]. These studies emphasized the importance of incorporating SCA data, derived from remote sensing sources like MODIS, to improve the reliability of streamflow forecasts, particularly in snow-fed catchments. The study's results indicate that while all models showed some degree of overfitting, the optimized models (especially LSTM-WHOAO) were better at generalizing to unseen data compared to traditional LSTM models. This finding is in line with the literature, where hybrid models have been shown to provide a more balanced approach, reducing overfitting while maintaining high predictive accuracy [20,47]. Uncertainties arising from input data and model predictions are not explicitly quantified in this study. Future work will adopt Monte Carlo dropout and ensemble methods to assess predictive uncertainty, as well as propagate input uncertainties through the modeling chain.

6 Conclusions

This study evaluated the performance of various LSTM-based hybrid models, including LSTM-GA, LSTM-GWO, LSTM-WOA, LSTM-AO, LSTM-WHO, and LSTM-WHOAO, in predicting streamflow in a snow-fed catchment. The research aimed to enhance predictive accuracy by integrating these models with metaheuristic optimization algorithms and incorporating Snow-Covered Area (SCA) data. The LSTM-WHOAO model exhibited the best performance during the training stage, with an average RMSE of 51.930 and an R^2 of 0.851. This model demonstrated its ability to fit the training data closely, reducing prediction errors and effectively capturing the streamflow dynamics in the snow-fed catchment. The LSTM-WHO model also performed strongly, with an average RMSE of 53.900 and an R^2 of 0.840. This indicates that the Wild Horse Optimizer, when combined with LSTM, can significantly enhance model accuracy. Other models, such as LSTM-AO and LSTM-WOA, showed average RMSE values of 57.135 and 58.978, respectively, demonstrating improved performance over the standalone LSTM model, which had a higher RMSE. In the testing stage, the LSTM-WHOAO model continued to outperform other models, with an average RMSE of 70.413 and an R^2 of 0.755. This highlights its strong generalization ability, maintaining accuracy even when applied to unseen data. The LSTM-WHO and LSTM-AO models also showed strong testing results, with RMSE values of 72.735 and 74.945, respectively, and R^2 values above 0.720. These models demonstrated better generalization compared to the simpler LSTM model. The standalone LSTM model had the highest RMSE (81.027) and the lowest R^2 (0.672), indicating its limited ability to accurately predict streamflow without optimization algorithms.

For peak streamflow events, the LSTM-WHOAO model had the lowest absolute error (201.4%), significantly reducing prediction error compared to other models such as LSTM-GA (333.9%) and LSTM (347.1%). This performance indicates the model's robustness in handling extreme hydrological events, which are crucial for flood management and water resource planning. Across all models, the inclusion of SCA data consistently improved prediction accuracy. Models that incorporated SCA, such as LSTM-WHOAO and LSTM-WHO, showed lower RMSE and higher R^2 values, underscoring the importance of considering snow cover dynamics in streamflow forecasting. Overall, the study demonstrated that integrating LSTM models with advanced metaheuristic optimizers significantly enhances predictive accuracy and generalization. The LSTM-WHOAO model, in particular, proved the most effective, showing superior performance across both training and testing stages and in peak streamflow predictions.

The accuracy and reliability of the models are highly dependent on the quality and availability of input data, particularly SCA data. Limited data or low-resolution datasets could lead to less accurate predictions. The hybrid models, especially those combining multiple optimization algorithms, are computationally intensive and complex. This complexity may limit their practicality for real-time applications or for users with limited computational resources. While the models were effective in the studied snow-fed catchment, their generalizability to other regions with different hydrological conditions remains uncertain. The models may require retraining and adjustment to perform well in different geographic contexts. Despite the improved accuracy, there is a risk of overfitting, particularly in models with many parameters. This could limit the model's performance when applied to new or different datasets.

Future research should focus on improving the resolution and accuracy of input data, particularly SCA data from remote sensing technologies. Utilizing multi-source data integration could further enhance model performance. Efforts should be made to simplify the hybrid models, balancing between complexity and performance. This could involve pruning unnecessary model parameters or optimizing the computational efficiency of the algorithms. To validate the robustness of these models, it is recommended to test them across different catchments with varying hydrological characteristics. This will help assess their generalizability and adaptability to different environmental conditions. The developed models, especially LSTM-WHOAO, should be integrated into real-time streamflow forecasting systems. This would involve testing their operational performance and making necessary adjustments for practical use in flood forecasting and water management. A limitation of this study is the application to a single catchment. Future work will evaluate the LSTM WHOAO model across additional basins with contrasting hydrological regimes to assess its transferability. Future research can focus on employing SHAP to quantify input variable contributions and enhance model interpretability, coupling LSTM-WHOAO with other climatic variables to assess climate change impacts on snowmelt-driven flows, with other deep learning models, and exploring physics-informed neural networks (PINNs) that embed the shallow water equations or the snowmelt runoff model as physical constraints. For wider adoption, creating user-friendly software or tools that encapsulate these advanced models would facilitate their use by hydrologists and water resource managers, thereby aiding decision-making.

Acknowledgement: The authors would like to thank the China National Key R&D Program (No. 2024YFC3013300), National Natural Science Foundation of China (No. 52479019), Major Basic Research Development Program of the Science and Technology, Qinghai Province (2025-HZ-805), and Shenzhen Talent Research Start-up Fund.

Funding Statement: This work was financially supported by the China National Key R&D Program (No. 2024YFC3013300), National Natural Science Foundation of China (No. 52479019), Major Basic Research Development Program of the Science and Technology, Qinghai Province (2025-HZ-805), and Shenzhen Talent Research Start-up Fund.

Author Contributions: Writing—original draft, Rana Muhammad Adnan Ikram; Validation, Ahmed A. Ewees; Data curation, Mo Wang; Conceptualization, Jing-Cheng Han; Methodology, Salim Heddam and Mohammad Zounemat-Kermani; Writing—review & editing, Ozgur Kisi and Mohammad Zounemat-Kermani; Supervision, Jing-Cheng Han. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the Corresponding Author upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process: AI-assisted tools were used to support language editing and improve the manuscript's clarity and readability. All scientific content, data analysis, interpretation of results, and conclusions were developed and verified solely by the authors, who take full responsibility for the publication's content.

References

1. Singh P, Jain SK. Snow and glacier melt in the Satluj River at Bhakra Dam in the western Himalayan region. Hydrol Sci J. 2002;47(1):93–106. doi:10.1080/02626660209492910.
2. Singh P, Jain SK, Kumar N. Estimation of snow and glacier-melt contribution to the Chenab River, western Himalaya. Mt Res Dev. 1997;17(1):49. doi:10.2307/3673913.
3. Mangukiya NK, Sharma A. Deep learning-based approach for enhancing streamflow prediction in watersheds with aggregated and intermittent observations. Water Resour Res. 2025;61(1):e2024WR037331. doi:10.1029/2024wr037331.
4. Sofi MS, Rautela KS, Muslim M, Bhat SU, Rashid I, Kuniyal JC. Modeling the hydrological response of a snow-fed river in the Kashmir Himalayas through SWAT and Artificial Neural Network. Int J Environ Sci Technol. 2024;21(3):3115–28. doi:10.1007/s13762-023-05170-7.
5. Shahariar S, Zobeyer H, Jahan N, Ali M. Enhanced streamflow prediction using a hybrid CNN-LSTM model: integrating climate-driven spatio-temporal features in a deep learning framework. Water Resour Manag. 2026;40(4):134.
6. Shikhar KC, Bhattarai KP, Shan TD, Mishra S, Joshi I, Singh AK. Comprehensive performance analysis of training functions in flow prediction model using artificial neural network. Water SA. 2024;52(2):190–200. doi:10.17159/wsa/2024.v50.i2.4099.
7. Wang G, Hao X, Yao X, Wang J, Li H, Chen R, et al. Simulations of snowmelt runoff in a high-altitude mountainous area based on big data and machine learning models: taking the xiying river basin as an example. Remote Sens. 2023;15(4):1118. doi:10.3390/rs15041118.
8. Zhou Y, Duan Y, Yao H, Li X, Li S. Incorporating hydrological constraints with deep learning for streamflow prediction. Expert Syst Appl. 2025;259(1):125379. doi:10.1016/j.eswa.2024.125379.
9. Workneh HA, Jha MK. Utilizing deep learning models to predict streamflow. Water. 2025;17(5):756. doi:10.3390/w17050756.
10. Tang X, Qin G, Wu X, Zhao Y, Li H. Causality-Guided Deep learning for streamflow predicting in a mountainous region. J Hydrol. 2025;653:132719. doi:10.1016/j.jhydrol.2025.132719.
11. Frank C, Rußwurm M, Fluixa-Sanmartin J, Tuia D. Short-term runoff forecasting in an Alpine catchment with a long short-term memory neural network. Front Water. 2023;5:1126310. doi:10.3389/frwa.2023.1126310.
12. Ren W, Li X, Zheng D, Zeng R, Su J, Mu T, et al. Enhancing flood simulation in data-limited glacial river basins through hybrid modeling and multi-source remote sensing data. Remote Sens. 2023;15(18):4527. doi:10.3390/rs15184527.
13. Sarafanov M, Borisova Y, Maslyaev M, Revin I, Maximov G, Nikitin NO. Short-term river flood forecasting using composite models and automated machine learning: the case study of *Lena* River. Water. 2021;13(24):3482. doi:10.3390/w13243482.
14. Gogineni A, Chintalacheruvu MR, Kale RV. Modelling of snow and glacier melt dynamics in a mountainous river basin using integrated SWAT and machine learning approaches. Earth Sci Inform. 2024;17(5):4315–37. doi:10.1007/s12145-024-01397-1.
15. Chandran D, Chithra NR. Predictive performance of ensemble learning boosting techniques in daily streamflow simulation. Water Resour Manag. 2025;39(3):1235–59. doi:10.1007/s11269-024-04029-x.
16. Abhinanda R, Kasiviswanathan KS, Patidar S, Adeloye AJ, Soundharajan BS, Ojha CSP. A novel physics-aware machine learning-based dynamic error correction model for improving streamflow forecast accuracy. Water Resour Res. 2023;59(2):e2022WR033318. doi:10.1029/2022wr033318.
17. Shrestha RR, Dibike YB, Bonsal BR. Snowpack driven streamflow predictability under future climate: contrasting changes across two western Canadian river basins. J Hydrometeorol. 2022;23(7):1113–29. doi:10.1175/jhm-d-21-0214.1.
18. Arsenault R, Martel JL, Brunet F, Brissette F, Mai J. Continuous streamflow prediction in ungauged basins: long short-term memory neural networks clearly outperform traditional hydrological models. Hydrol Earth Syst Sci. 2023;27(1):139–57. doi:10.5194/hess-27-139-2023.

19. Li S, Qin H, Xia F, Xu X, Cai Y. Application of LSTM coupled models in runoff simulation and prediction: a review. Appl Water Sci. 2026;16(3):89. doi:10.1007/s13201-025-02747-0.
20. Le XH, Nguyen DH, Jung S, Yeon M, Lee G. Comparison of deep learning techniques for river streamflow forecasting. IEEE Access. 2021;9:71805–20. doi:10.1109/access.2021.3077703.
21. Thapa S, Li H, Li B, Fu D, Shi X, Yabo S, et al. Impact of climate change on snowmelt runoff in a Himalayan basin. Nepal Environ Monit Assess. 2021;193(7):393. doi:10.1007/s10661-021-09197-6.
22. Parisouj P, Mohammadzadeh Khani H, Islam MF, Jun C, Bateni SM, Kim D. AI-based runoff simulation based on remote sensing observations: a case study of two river basins in the United States and Canada. JAWRA J Am Water Resour Assoc. 2023;59(2):299–316.
23. Kheimi M, Almadani M, Zounemat-Kermani M. Stochastic (S [ARIMA]), shallow (NARnet, NAR-GMDH, OS-ELM), and deep learning (LSTM, Stacked-LSTM, CNN-GRU) models, application to river flow forecasting. Acta Geophys. 2024;72(4):2679–93. doi:10.1007/s11600-023-01157-7.
24. Mahsa M, Lee T. Comparison of optimization algorithms in deep learning-based neural networks for hydrological forecasting: case study of Nam river daily runoff. J Korean Soc Hazard Mitig. 2018;18(6):377–84. doi:10.9798/kosham.2018.18.6.377.
25. Jia F, Zhu Z, Dai W, Le VV. Short-term forecasting of streamflow by integrating machine learning methods combined with metaheuristic algorithms. Expert Syst Appl. 2024;245(1):123076. doi:10.1016/j.eswa.2023.123076.
26. Mahdavi-Meymand A, Sulisz W, Zounemat-Kermani M. Hybrid and integrative evolutionary machine learning in hydrology: a systematic review and meta-analysis. Arch Comput Meth Eng. 2024;31(3):1297–340. doi:10.1007/s11831-023-10017-y.
27. Choudhary K, Jha GK, Jaiswal R, Kumar RR. A genetic algorithm optimized hybrid model for agricultural price forecasting based on VMD and LSTM network. Sci Rep. 2025;15(1):9932. doi:10.1038/s41598-025-94173-0.
28. Liang J, Du Y, Xu Y, Xie B, Li W, Lu Z, et al. Using Adaptive Chaotic Grey Wolf Optimization for the daily streamflow prediction. Expert Syst Appl. 2024;237(3):121113. doi:10.1016/j.eswa.2023.121113.
29. Wang WC, Wang B, Chau KW, Zhao YW, Zang HF, Xu DM. Monthly runoff prediction using gated recurrent unit neural network based on variational modal decomposition and optimized by whale optimization algorithm. Environ Earth Sci. 2024;83(2):72. doi:10.21203/rs.3.rs-2680780/v1.
30. Chahal A, Gulia P, Gill NS, Sultana N. Aquila optimizer-based hybrid predictive model for traffic congestion in an IoT-enabled smart city. Int J Intell Syst. 2024;2024(1):5577278. doi:10.1155/2024/5577278.
31. Motwakel A, Alabdulkreem E, Gaddah A, Marzouk R, Salem NM, Zamani AS, et al. Wild horse optimization with deep learning-driven short-term load forecasting scheme for smart grids. Sustainability. 2023;15(2):1524. doi:10.3390/su15021524.
32. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–80. doi:10.1162/neco.1997.9.8.1735.
33. Hu Y, Li H, Zhang C, Wang T, Chu W, Li R. Investigate the rainfall-runoff relationship and hydrological concepts inside LSTM. Environ Model Softw. 2025;192(10):106527. doi:10.1016/j.envsoft.2025.106527.
34. Cai K, He J, Li Q, Shangguan W, Li L, Hu H. Meta-LSTM in hydrology: advancing runoff predictions through model-agnostic meta-learning. J Hydrol. 2024;639(10):131521. doi:10.1016/j.jhydrol.2024.131521.
35. Holland JH. Genetic algorithms. Sci Am. 1992;267(1):66–72.
36. Golberg DE. Genetic algorithms in search, optimization, and machine learning. Boston, MA, USA: Addison Wesley; 1989. 36 p.
37. Guo J, Huang Y, Li Z, Li J, Jiang C, Chen Y. Performance prediction and optimization of lateral exhaust hood based on back propagation neural network and genetic algorithm. Sustain Cities Soc. 2024;113(4):105696. doi:10.1016/j.scs.2024.105696.
38. Jhong YD, Chen CS, Jhong BC, Tsai CH, Yang SY. Optimization of LSTM parameters for flash flood forecasting using genetic algorithm. Water Resour Manag. 2024;38(3):1141–64. doi:10.1007/s11269-023-03713-8.
39. Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. Adv Eng Softw. 2014;69:46–61. doi:10.1016/j.advengsoft.2013.12.007.

40. Mirjalili S, Lewis A. The whale optimization algorithm. Adv Eng Softw. 2016;95(12):51–67. doi:10.1016/j.advengsoft.2016.01.008.
41. Xue Z, Yi X, Feng W, Kong L, Wu M. Prediction and mapping of soil thickness in Alpine canyon regions based on whale optimization algorithm optimized random forest: a case study of Baihetan Reservoir area in China. Comput Geosci. 2024;191(1):105667. doi:10.1016/j.cageo.2024.105667.
42. Abualigah L, Yousri D, Abd Elaziz M, Ewees AA, Al-qaness MAA, Gandomi AH. *Aquila* optimizer: a novel meta-heuristic optimization algorithm. Comput Ind Eng. 2021;157(11):107250. doi:10.1016/j.cie.2021.107250.
43. Bai L, Pei Z, Wang J, Zhou Y. Multi-strategy improved aquila optimizer algorithm and its application in railway freight volume prediction. Electronics. 2025;14(8):1621. doi:10.3390/electronics14081621.
44. Naruei I, Keynia F. Wild horse optimizer: a new meta-heuristic algorithm for solving engineering optimization problems. Eng Comput. 2022;38(4):3025–56. doi:10.1007/s00366-021-01438-z.
45. Qiao X, Peng T, Sun N, Zhang C, Liu Q, Zhang Y, et al. Metaheuristic evolutionary deep learning model based on temporal convolutional network, improved *Aquila* optimizer and random forest for rainfall-runoff simulation and multi-step runoff prediction. Expert Syst Appl. 2023;229(12):120616. doi:10.1016/j.eswa.2023.120616.
46. Shabani M, Asadi MA, Fathian H. Improving the daily pan evaporation estimation of long short-term memory and support vector regression models by using the Wild Horse Optimizer algorithm. Water Supply. 2024;24(4):1315–34. doi:10.2166/ws.2024.063.
47. Sabzipour B, Arsenault R, Troin M, Martel JL, Brissette F, Brunet F, et al. Comparing a long short-term memory (LSTM) neural network with a physically-based hydrological model for streamflow forecasting over a Canadian catchment. J Hydrol. 2023;627(6):130380. doi:10.1016/j.jhydrol.2023.130380.