Acta Geophysica
https://doi.org/10.1007/s11600-025-01570-0

RESEARCH ARTICLE - HYDROLOGY AND HYDRAULICS

Enhancing hydrological time series forecasting with a hybrid Bayesian-ConvLSTM model optimized by particle swarm optimization

Huseyin Cagan Kilinc[1] · Sina Apak[2] · Mahmut Esad Ergin[1] · Furkan Ozkan[3] · Okan Mert Katipoğlu[4] · Adem Yurtsever[5]

Received: 26 November 2024 / Accepted: 11 February 2025
© The Author(s) 2025

Abstract

Hydrological time series forecasting often relies on addressing the inherent uncertainties and complex temporal dependencies embedded in the data. This study presents an innovative hybrid framework, the Bayesian-ConvLSTM-PSO model, specifically designed to tackle these challenges. The framework synergistically combines 1D convolutional neural networks (CNNs), a convolutional Bayesian network, multi-head attention, and long short-term memory (LSTM) networks, with parameters optimized through particle swarm optimization (PSO). The fusion of the convolutional Bayesian network and 1D convolutional neural networks enhances feature robustness by capturing both probabilistic uncertainties and spatial patterns effectively. The multi-head attention model further amplifies this by focusing on the most relevant features, improving the learning process and ensuring better representation of complex temporal dependencies. The proposed model is rigorously tested on daily streamflow data from three flow measurement stations (FMS): Ahullu (D14A014), Kızıllı (D14A080), and Erenkaya (D14A127). Experimental results reveal that the Bayesian-ConvLSTM-PSO model achieves significant performance gains across various evaluation metrics, including root mean square error (RMSE), mean absolute error (MAE), determination coefficient (R^2), Kling–Gupta efficiency (KGE), and bias factor (BF). Notably, the model demonstrates exceptional accuracy with an R^2 of 0.9950, a KGE of 0.9950, and a bias factor of 0.0003, surpassing the results of PSO-1D CNN-LSTM and benchmark models, such as DNN, DNN-LSTM, and 1D ConvLSTM. These compelling findings underscore the potential of the Bayesian-ConvLSTM-PSO framework as a robust and effective tool for applications in river engineering and hydrological time series forecasting.

Keywords Multi-head attention · Particle swarm optimization · Bayesian method · Forecasting · Deep learning

Introduction

It is essential to ensure the efficient management of water in all areas where it is used, and therefore, an accurate prediction of river flow time series is crucial for understanding future water behavior (Ahmad et al. 2023; Jiang et al. 2024). Recent advances in machine learning (ML) and artificial intelligence have introduced promising techniques for improving time series forecasting accuracy. Deep learning (DL) is promising in aiding conceptual modeling linked with highly nonlinear correlations (Eren and Küçükdemiral 2024). Neural networks have been used in hydrology to predict and forecast rainfall runoff, streamflow, water quality, and precipitation using existing data (Fang et al. 2022; Hussain et al. 2020). Streamflow forecasting has long been challenged by the complexities of capturing spatialtemporal dependencies and the inherent uncertainties within hydrological data. ML has emerged as a powerful alternative to

Edited by Dr. Senlin Zhu (ASSOCIATE EDITOR) / Prof. Jochen Aberle (CO-EDITOR-IN-CHIEF).

✉ Huseyin Cagan Kilinc
huseyincagankilinc@aydin.edu.tr

[1] Department of Civil Engineering, Istanbul Aydın University, Istanbul, Turkey

[2] Department of Management Information Technology, Istanbul Aydın University, Istanbul, Turkey

[3] Department of Computer Engineering, Çukurova University, Adana, Turkey

[4] Department of Civil Engineering, Erzincan Binali Yıldırım University, Erzincan, Turkey

[5] Department of Environmental Engineering, Istanbul University-Cerrahpaşa, Istanbul, Turkey

Published online: 24 March 2025

traditional methods, leveraging data-driven approaches to reveal hidden patterns and relationships. This shift highlights ML's role in advancing hydrological forecasting by overcoming limitations faced by conventional approaches and enhancing predictive accuracy through its ability to handle complex, nonlinear dependencies. The one-dimensional convolutional neural network (1D-CNN) model addresses this by efficiently extracting spatial features from time-series data, making it possible to identify critical patterns that influence streamflow dynamics. These patterns, which are dispersed over time, are crucial for accurate forecasting (Chen et al. 2024). Among popular DL models, long short-term memory (LSTM) networks stand out for their ability to model temporal dependencies and capture nonlinear relationships. LSTM further enhances this capability by understanding complex patterns and trends in streamflow data, thereby improving the accuracy of predictions (Yu et al. 2024). Additionally, LSTM is particularly effective in managing long-term dependencies, which are essential for understanding the evolving behavior of hydrological systems under varying environmental conditions (Liu et al. 2023). For example, the study by Fang et al. (2024) demonstrates that a hybrid model combining CNN, Transformer, and LSTM models can effectively model complex, nonlinear, and non-stationary streamflow processes. In another study, integrating LSTM enables automatic adaptation to the trend of time series data (Yin et al. 2023). Experiments have shown that the LSTM hybrid model achieves higher prediction accuracy, highlighting both the feasibility and advantages of using LSTM in detail. Furthermore, probabilistic hydrological predictions are critically important for modeling uncertainties and supporting risk-based decision-making processes. In this context, a generative DL model based on a conditional variational autoencoder provided reliable predictions by considering the inherent uncertainties in hydrological variables. The model, developed using dense network layers and LSTM blocks, demonstrates that advanced DL methods can offer more reliable and comprehensive predictions by addressing uncertainties in hydrological processes (Jahangir and Quilty 2024). Similarly, Granata and Di Nunno (2024) combined a stacked multilayer perceptron–random forest model with different meta-learning algorithms to improve the generalization ability in hydrological forecasts. The findings of the study show that the selection of meta-learners plays a critical role in optimizing the performance of stacked models at different forecasting horizons. In addition, the application of innovative neural network architectures in hydrological time series forecasting is increasing (Granata et al. 2024b). While the Kolmogorov-Arnold Network offers high accuracy and interpretability in capturing complex patterns, the Transformer model stands out with its attention mechanisms, which handle long-range dependencies.

Although DL models generally perform well, configuring hyperparameters remains a significant challenge. Accurate calibration of these hyperparameters is crucial for producing reliable flow predictions. Metaheuristic algorithms play a key role in improving DL models, particularly when used in hybrid approaches. These algorithms assist in optimizing the hyperparameters necessary for achieving high performance in DL models (Narayanan and Ganesh 2024). Particle swarm optimization (PSO), a population-based heuristic search strategy, exemplifies an effective approach by focusing on cooperation and optimization within complex parameter spaces (Kilinc et al. 2024). In this context, Katipoğlu, (2023) applied PSO to optimize hyperparameters in hybrid models for hydrological drought forecasting. It has been observed that PSO has great potential for improving water supply management. Additionally, PSO has demonstrated effectiveness in reducing energy consumption while ensuring acceptable water quality throughout the optimization process (Brentan et al. 2024). Jia et al. (2024) highlighted the effectiveness of an integrated approach that combines PSO with metaheuristic algorithms and ML methods for capturing complex nonlinear relationships in streamflow prediction. Another study by Yilmaz et al. (2024) also showed that hybridization of PSO method provides higher prediction success, especially in different climate classes. It plays an important role in increasing the forecast success in data structures containing seasonal and trend components.

Recently, multi-factor models have been developed to increase the accuracy of ML models used in streamflow estimation (Huang and Wu 2024; Wan et al. 2024). The multi-head attention mechanism significantly boosts prediction accuracy by allowing the model to simultaneously attend to various temporal and spatial features of the data, thereby capturing complex patterns and interactions that are crucial for accurate forecasting of floods, water level changes, and long-term water trends (Wang et al. 2023b). Likewise, Lee and Ahn (2024) developed a new model for stream forecasting. The multi-head attention mechanism in the basic structure provides a great advantage especially in flood forecasting by increasing the ability of the model to focus on different data subsets. It has been stated also in the study of Demiray et al. (2024) that the multi-head attention mechanism has a decisive effect on the model's ability to better capture long-term dependencies in time series. By incorporating the multi-head attention mechanism, the model's ability to capture intricate relationships within the input sequence is enhanced, making it particularly valuable for hydrological forecasting. In the literature, studies on the effectiveness of traditional and data-driven methods in forecasting hydrological processes are increasing. For instance, a study on short- and medium-term river flow rate forecasting based on the additive regression of radial basis function neural networks model achieved high accuracy by

optimizing lagged values and hyperparameters using Bayesian optimization (BO) (Granata et al. 2024a). Moreover, the Bayesian approach has become increasingly popular in water resources and hydrological engineering by integrating prior knowledge of model parameters and streamflow data (Bates and Dowdy 2024; Ramkumar and Jothiprakash 2024). It allows the model to not only predict streamflow with high accuracy but also to express the confidence of these predictions (Wei et al. 2024). This dual capability is essential in scenarios where decision-making relies on understanding the range of possible outcomes, especially under conditions of data scarcity or environmental change (Ahmed et al. 2024; Jeon et al. 2024).

This study introduces a novel approach for daily time series prediction of river flow in the Yeşilırmak basin in Turkey by leveraging the complementary strengths of multiple advanced techniques. Our model integrates 1D-CNN, which excel at extracting local patterns from sequential data, with a convolutional Bayesian network to handle uncertainty and provide probabilistic predictions, crucial for robust decision-making in hydrological forecasting. Additionally, we incorporate multi-head attention mechanisms, which enhance the model's ability to focus on different parts of the input sequence, capturing long-range dependencies and temporal relationships that are often missed by conventional methods. Our contributions are twofold: (1) the development of a hybrid model that synergistically combines 1D convolutional layers, a convolutional Bayesian network, multi-head attention, and LSTM networks, all optimized through PSO. This integration enables our model to effectively capture complex patterns and dependencies in hydrological data, significantly enhancing prediction accuracy; and (2) the application of this innovative approach to the challenging domain of hydrological time series forecasting, where it demonstrates significant improvements over both benchmark and state-of-the-art models. These results not only underscore the resilience and effectiveness of our approach but also establish a new benchmark for forecasting accuracy in this field, offering a powerful tool for advancing water resource management.

Materials and methods

Characterization of the study area

Yeşilırmak Basin is located at the coordinates of 40° 38′ 54" North Latitude and 35° 49′ 52" East Longitude which was determined as the research area in this study (Gazioğlu et al. 2019). Yeşilırmak Basin covers the area located in the northern part of Anatolia and discharges its waters to the Black Sea through the Yeşilırmak River. In the Yeşilırmak Basin, floods that spread over a wide area and occur over time as a result of long-term rainfall and affect large basins are generally observed. In addition, since high sloped lands generally occupy a significant area in the Yeşilırmak Basin, there are problems such as erosion at various levels and stoniness and shallow soil profile. The Yeşilırmak Basin, which is approximately 5% of Turkey's surface area (39,595 km^2) and 519 km long, ranks third among 25 river basins in terms of surface area in the country. In this basin, where agriculture is an important source of income, floods, erosion, domestic and industrial wastes caused by the irregularity of the flow regime of the Yeşilırmak River create important problems in this basin and this situation reveals that the region needs hydrological rehabilitation. In addition, in terms of the quality and quantity of Turkey's water resources, protection action plans have been prepared by scientific and technological research council of Turkey (TUBITAK) and the Ministry of Environment and Forestry in some basins, including the Yeşilırmak basin, by taking into account the pollution status, pressure and effects, drinking water and protected areas in 25 main water basins. The basin has been addressed by both the government and a numerous of researchers. Pinarlik and Selek (2024) conducted a study aimed at evaluating the operational performance of selected reservoirs in Turkey's Yeşilırmak Basin. In another study, the Yeşilırmak Basin was chosen as the focus area to forecast the likelihood of a one-month hydrological drought (Katipoğlu 2023).

Datasets and preprocessing

Flow data of D14A014-Ahullu, D14A080-Kızıllı, D14A127-Erenkaya flow measurement stations (FMS) in Yeşilırmak Basin were obtained from hydraulic state works (DSI). The positions of the relevant stations in the basin are shown in Fig. 1. Flow data of stations at various end points within the basin were used. The catchment areas of Ahullu, Kızıllı, and Erenkaya are 259 km^2, 338.6 km^2, and 89.42 km^2, respectively, with corresponding elevations of 140 m, 1193 m, and 1497 m. The mentioned stations were selected due to their ability to capture seasonal transitions at each station, their location along different branches of the river, and the presence of more complete daily flow data, as determined by the analysis of data obtained from DSI. The requirement for at least 10 years of measurements for analysis was effective in selecting the stations. In the research, the analysis timespan between 2005 and 2014 was determined. In this study, various percentages were tried and the optimum result being that 80% of the data set was used in the training phase, and the remaining 20% was used in the testing phase. The characteristic of the Yeşilırmak Basin is the high number of existing dam lakes. With these effects, dams provide a mitigating function in extraordinary peak flow situations. Previously observed flow values were selected as input variables dataset to produce multi-step forecasts for up to five days for the

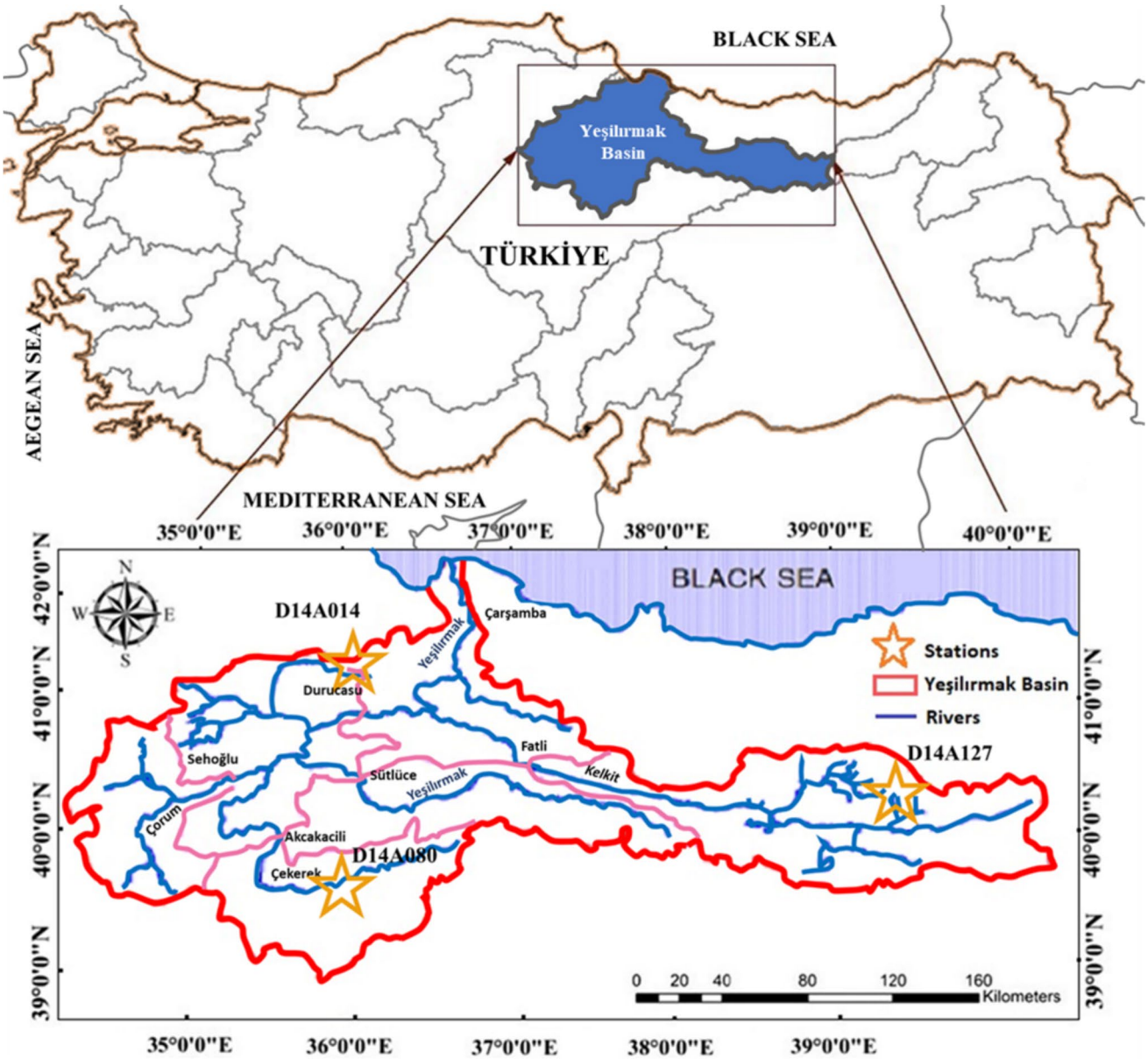


Fig. 1 Yeşilırmak Basin in Turkey map and location of FMSs in basin

flow at the three flow measurement stations included in the study. After evaluating different lag groups, it was determined that optimal results were achieved with a short-term prediction at up to five days lag, which was subsequently applied to all models in the study.

The data set consists of daily flow rate values. Daily flow potentials of FMSs are presented in Figure S1. Examining of the flow data, it is observed that the flow rate is at the highest in March, April and May for all stations and decreases to the lowest level in July, August and September. The reason for this is the climatic conditions prevailing in the basin. Incorporating precipitation into water during winter increases flow levels, whereas in summer, the onset of intense evaporation and the lack of precipitation lead to a decrease in water levels.

Particle swarm optimization

Kennedy and Eberhart (1995) created PSO, a well-known metaheuristic algorithm. PSO consists of a randomly selected set of possible solutions. At each iteration, the PSO algorithm tries to find the best of the swarm by updating its particles. All parameters of each result set are applied separately to the system to be optimized, and

Gbest, the best of the herd, and *Pbest*, the best of the particle, are calculated. The particle swarm repeats the process which is depicted and presented in Figure S2 until an ideal solution is found.

PSO updates the position and velocity of each particle after it starts working. A fitness function determines the fitness value of each particle in the population throughout the iteration process. The velocity value is also utilized to determine the direction and distance of the particle's travel. The position and velocity information of each particle is specified in Eq. (1) and Eq. (2) (Gogineni and Chintalacheruvu 2024).

$$X_i(k+1) = X_i(k) + V_i(k+1), \tag{1}$$

$$V_i(k+1) = w.V_i(k) + c_1.r_1\left[P_{best} - X_i(k)\right] + c_2.r_2\left[G_{best} - X_i(k)\right]. \tag{2}$$

In the equations above: i represents the swarm's particle count, the iteration count is denoted by k, the inertia weight is represented by w, r_1 and r_2 are randomly distributed variables in the scale *[0,1],* and c_1 and c_2 are the acceleration coefficients. *Pbest* is the particle's own best position, *Gbest* is the best position of the particles in the swarm, X_i. position of the particle, V_i. represents the speed of the particle (Istiqomah et al. 2024).

A Bayesian convolutional neural network

A Bayesian convolutional neural network integrates Bayesian inference with the structure of CNNs to effectively model uncertainty in deep learning tasks. In a Bayesian convolutional neural network, the weights W are treated as random variables with a prior distribution $P(W)$ often chosen as a Gaussian prior $(W;o,\sigma^2 I)$. Given observed data$D = \{(X_i, Y_i)\}_{i=1}^{N}$, the posterior distribution over the weights is defined by Bayes' theorem:

$$P(W|D) = (P(D|W)p(W))/P(D). \tag{3}$$

Since this posterior is typically intractable due to the complex, high-dimensional nature of CNNs, approximation techniques like variational inference are employed. Variational inference approximates the posterior with a simpler distribution $q(W|\theta)$ by minimizing the Kullback–Leibler (KL) divergence between the variational approximation and the true posterior, effectively optimizing the evidence lower bound (ELBO):

$$\mathcal{L}(\theta) = \mathbb{E}_q(W|\theta)[\log p(D|W)] - KL(q(W|\theta)||p(W)). \tag{4}$$

To make predictions for a new input x^*, the Bayesian convolutional neural network calculates the predictive distribution by marginalizing over the posterior distribution of the weights:

$$p(y^*|x^*, D) = \int p(y^*|x^*, W)p(w|D)dw. \tag{5}$$

This integral is approximated using Monte Carlo sampling, where multiple samples $w_1, w_2, w_3, \ldots, w_T$ are drawn from the variational distribution $(W|\theta)$, or by employing techniques like Monte Carlo Dropout. This methodology provides a means to quantify both aleatoric and epistemic uncertainty, making Bayesian convolutional neural network's particularly valuable in applications that require reliable uncertainty estimation like our study, more details are presented in (Shridhar et al. 2019).

One-dimensional convolutional neural networks

The CNN model, which is among the successful deep learning models in current time series predictions, is a deep learning technique that uses an efficient neural network, which focuses on processing sequence data, to find patterns in the data (Seunghwan and Chung 2022). In addition, a 1D-CNN model has been developed to model one-dimensional data with a convolutional neural network. The input layer, convolution layer, max-pooling layer, and flattening layer are components of the CNN model. CNN may be employed for data feature extraction and data prediction analysis utilizing one-dimensional time series data (Chen et al. 2017). CNNs offer the benefit of simplified training through minimal parameter requirements and streamlined data preprocessing. The following Eq. (3) describes the output of a CNN model with one-dimensional input data.

$$s(t) = (x * w)(t) = \sum x(a)w(t-a), \tag{6}$$

where $s(t)$ is the feature map, which is the output layer, w is the kernel map, and x is the input data.

Long short-term memory networks

LSTM is a version of recurrent neural networks (RNN) designed to solve the gradient vanishing challenge. LSTM has memory cells that can preserve long-term dependencies more effectively. In the LSTM model, a few different mathematical operations are performed compared to RNN, and the parameters are calculated and stored in its memory. By the memory advantage of the LSTM, the model can access the input and output information from previous steps when needed (Senel 2023). LSTM, an artificial neural network model widely used in deep learning, classifies, processes, and predicts time series data by utilizing a cell state to retain essential information and incorporating memory cells with gating units within its recurrent hidden layers. These gates play pivotal roles in the processing flow. The computations conducted within the segments of the LSTM block structure

are illustrated through the equations presented below (Kilinc et al. 2023).

$$i_t = \sigma\left(W_{i,x}xX_t + W_{i,h}xh_{t-1} + b_i\right)(\text{input gate}), \tag{7}$$

$$f_t = \sigma\left(W_{f,x}xX_t + W_{f,h}xh_{t-1} + b_f\right)(\text{gate to forget}), \tag{8}$$

$$\hat{C}_t = \tanh\left(W_{c,x}xX_t + W_{c,h}xh_{t-1} + b_c\right)(\text{intermediate cell state}), \tag{9}$$

$$o_t = \sigma(W_{o,x}xX_t + W_{o,h}xh_{t-1} + b_o(\text{output gate}), \tag{10}$$

$$c_t = c_{t-1}xf_t + i_txC_t(\text{cell state}), \tag{11}$$

$$ht = o_t + tanh\left(c_t\right)(\text{new cell state}). \tag{12}$$

The input vector x_t, it is fed into the input gate (Eq. (7)), f_t represents the forget gate (Eq. (8)), C denotes the intermediate cell state (Eq. (9)), o_t signifies the output gate (Eq. (10)), c_t stands for the cell state (Eq. (11)), h_t denotes the output vector (Eq. (12)). The sigmoid activation function σ and the hyperbolic tangent activation function tanh are employed accordingly throughout the LSTM computations.

Proposed model

The proposed model architecture is designed to leverage the strengths of multiple advanced techniques to enhance the accuracy and robustness of time series data prediction. The model begins by feeding streaming data into a 1D convolutional (Conv) layer, which is particularly adept at capturing local patterns and spatial hierarchies within the data. The 1D Conv layer operates by applying convolutional filters across the input sequence, identifying significant features such as boundaries or trends that may be crucial for understanding the underlying structure of the data. This initial feature extraction is essential for identifying key characteristics that serve as the foundation for deeper, more abstract representations in subsequent layers. Following this, the extracted features are processed in parallel by two distinct layers: another 1D Conv layer and a convolutional Bayesian model. The additional 1D Conv layer further refines the local feature representations by applying more complex filters, thereby capturing deeper hierarchical structures within the data. This progression allows the model to recognize increasingly abstract features, enhancing its ability to distinguish subtle differences in the data.

Meanwhile, the convolutional Bayesian model introduces a probabilistic framework that enables the model to incorporate uncertainty directly into the feature representation. This is particularly important in real-world data, where noise and variability can obscure the true signal. By modeling this uncertainty, the Bayesian component helps the model make more robust predictions, especially when dealing with ambiguous or noisy data. This dual-pathway approach, combining deterministic feature extraction with probabilistic reasoning, ensures that both concrete and uncertain aspects of the data are thoroughly explored, enhancing the model's ability to generalize across different data scenarios. The outputs of these parallel layers are then integrated into a multi-head attention mechanism. The multi-head attention model is crucial for its ability to focus on different parts of the input sequence simultaneously, thereby capturing long-range dependencies and complex interactions within the data. Each "head" in the multi-head attention mechanism can learn to pay attention to different aspects of the sequence, which allows the model to prioritize the most relevant features for the task at hand. This attention-driven feature selection process ensures that the model does not overlook critical information, enhancing its ability to make accurate classifications even in complex scenarios. After the multi-head attention mechanism has refined and weighted the feature representations, the output is fed into a LSTM layer. The LSTM is particularly well-suited for handling sequential data because it is designed to capture temporal dependencies and patterns over time. Unlike traditional neural networks, LSTMs have a memory cell that allows them to retain information from earlier in the sequence and use it to inform current decisions. This makes them highly effective in contexts where the order and timing of inputs are important, such as in time-series analysis or natural language processing. In the context of this model, the LSTM ensures that the temporal dynamics of the data are fully captured, maintaining a coherent understanding of how features evolve over time and contributing to more accurate predictions.

Finally, the entire model is optimized using PSO. In this model, PSO adjusts the parameters of the entire network to maximize prediction accuracy. By simulating a swarm of particles (each representing a potential solution), PSO iteratively explores the solution space, with particles adjusting their positions based on their own experience and that of their neighbors. This global optimization strategy allows the model to escape local minima and find more globally optimal parameter settings, ultimately leading to better performance in classification tasks. The pseudocode of the proposed approach and main architecture are presented in Table 1 and Fig. 2 below.

Results

Optimized hyperparameters

In this section, it is implemented PSO to fine-tune hyperparameters for a deep learning model. PSO is an optimization

Table 1 Pseudocode of PSO method for hyperparameter optimization of proposed model

Swarm size N, Maximum number of iterations Max Iterations
Best position (solution) found by the swarm.
Initialization:

- Initialize swarm.
- Set the best-known positions of all particles to their initial positions.

For each particle i in the swarm: Initialize position P_i randomly in the search space for 1D CNN-LSTM hyperparameters.

1. Initialize velocity V_i randomly.
2. For each particle i in the swarm:
3. Set $P_{best_i} = P_i$
4. Evaluate fitness F_i of the particles using the 1D CNN, Bayesian Network, Multi-Head Attention, and LSTM hyperparameters.
5. Set the global best position based on the particle with the best fitness:
6. $G_{best_i} = P_{best_i}$, where i is the particle with the best fitness in the swarm
7. For iteration = 1 to Max iteration do
8. For each particle i in the swarm do
9. Update velocity and position:
10. $V_i = W \times V_i + c_1 \times (P_{best_i} - P_i) + c_2 \times (G_{best_i})$
11. $p_i = p_i + v_i$
12. Evaluate fitness F_i of the particle:
13. Feed streaming data into the initial 1D Conv layer.
14. Extracted features are processed in parallel:
15. a. Through another 1D Conv layer to refine local features.
16. b. Through a Convolutional Bayesian network to model uncertainty.
17. The outputs from these parallel layers are then integrated into a Multi-Head Attention mechanism.
18. Feed the output of the Multi-Head Attention mechanism into an LSTM layer to capture temporal dependencies.
19. Update personal best:
20. If $F_i < F(P_{best_i})$ then
21. Set $P_{best_i} = P_i$
22. Update global best:
23. $G_{best} = P_{best_i}$, where i is the particle with the best fitness in the swarm using deep model
24. end

algorithm that simulates the social behavior of birds flocking or fish schooling to find the best solution in a search space. In our case, the objective function encapsulates the model training process, which involves creating a model with specified filters (1 × 1, 1 × 3, and 1 × 5), kernel size (1, 2, and 3), learning rate (0.001, 0.01, and 0.1), and LSTM units (32, 64, and 128). The PSO algorithm explores the hyperparameter space defined by lower and upper bounds to minimize the MSE on a validation set. With a swarm size of 5 and a maximum of 5 iterations, the optimal hyperparameters are determined. The results of the PSO optimization process are summarized in Table 2, which provides a clear view of the selected hyperparameters for each experiment (Ahullu, Kizilli, and Erenkaya).

Performance analysis of proposed model

Experimental results obtained from the proposed LSTM-based model for short-term flow prediction are discussed in this section. The model was trained and evaluated on a comprehensive dataset comprising hydrological time series data. The experiments conducted in this study aimed to

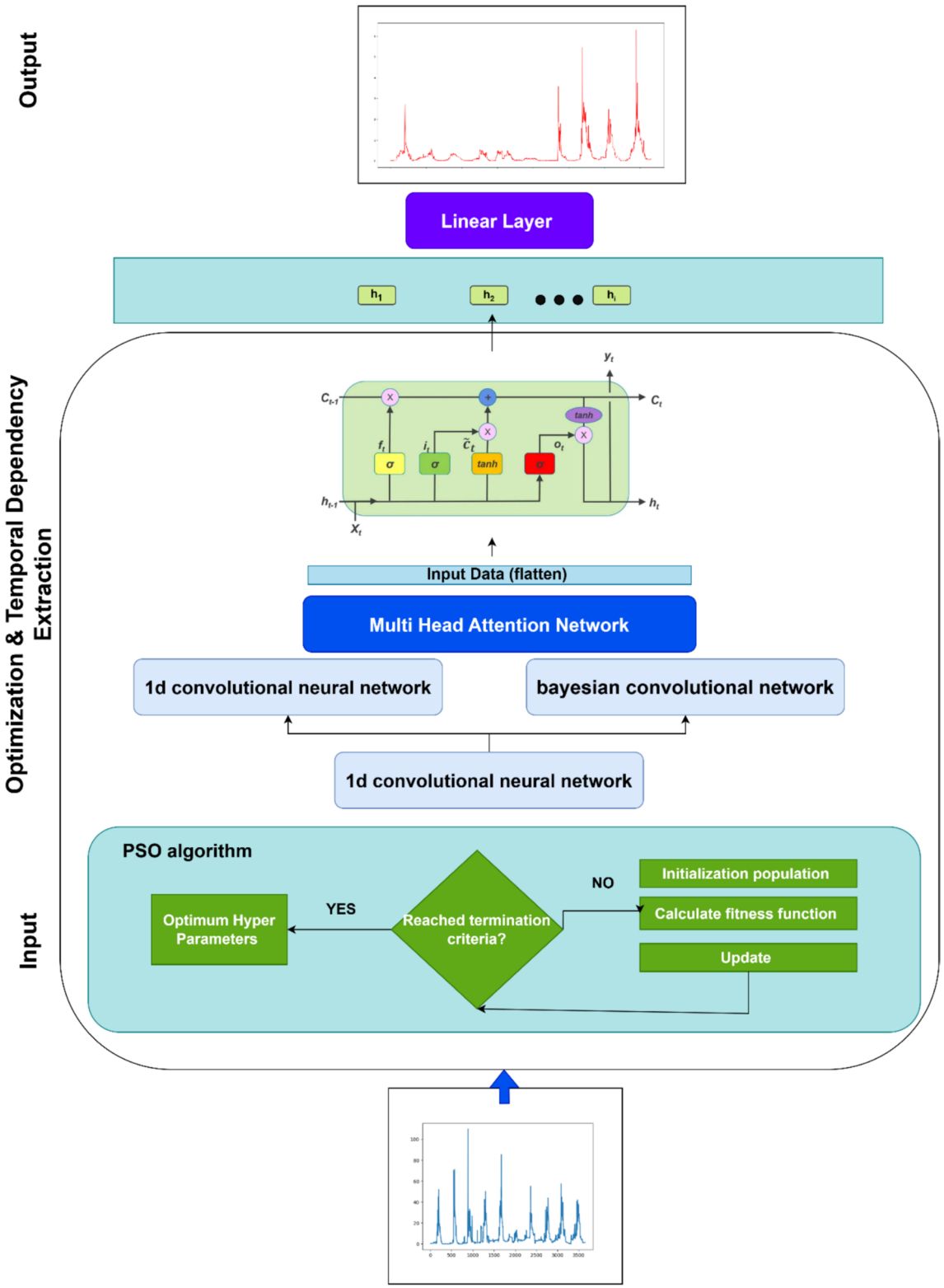


Fig. 2 Architecture of the proposed Bayesian-ConvLSTM-PSO Model

Table 2 Optimized hyperparameters and their search ranges for Ahullu, Kizilli, and Erenkaya datasets

Hyperparameter	Search range	Ahullu	Kizilli	Erenkaya
Filters	(1 × 1, 1 × 3, and 1 × 5)	(1 × 3)	(1 × 3)	(1 × 3)
First Kernel size	[32–64–128]	64	64	64
Second Kernel size	[32–64–128]	32	64	128
Third Kernel size	[32–64–128]	128	64	32
Learning rate	[0.001, 0.01, 0.1]	0.001	0.001	0.001

evaluate the effectiveness of hyperparameter optimization in improving the predictive performance of the LSTM network. We conducted a series of experiments using different hyperparameter combinations, including varying LSTM units, learning rates, and batch sizes, to determine the optimal configuration. The model's performance was estimated via mean squared error (MSE) on a separate validation set. Results demonstrate the impact of hyperparameter tuning on the model's ability to capture complex temporal patterns in streamflow data. Additionally, the optimized LSTM model is compared with the baseline configurations to showcase the improvements achieved through the hyperparameter optimization strategy. To further demonstrate the improvement of the proposed model, it was compared with four additional models, including three benchmark models and one hybrid model. For a fair comparison, these benchmark models were selected and designed to closely match the proposed model in terms of the number of trainable parameters. DNN models are limited because they cannot capture time dependencies in sequential data and are prone to over-learning. However, DNN-LSTM models are successful in capturing sequential dependencies, but they have difficulties such as high computational cost and large data requirements due to their complex structure. Although 1D ConvLSTM models can be used to capture local features and long-term dependencies, their performance may decrease in long sequences and the model requires precise hyperparameter tuning. These benchmark models can become too complex due to excessive parameterization and have a risk of error in the results obtained. For example, the performance of a 1D-CNN-based model depends significantly on the correct selection and application of hyperparameters, such as filter size, number of layers and learning rate (Kilichev and Kim 2023). However, thanks to PSO's convergence property, determining optimal hyperparameters can be made easier with fewer evaluations than other methods. In a study, it was shown that it can be obtained optimal hyperparameter values by making fewer function evaluations using PSO optimization, and therefore can be considered a more effective method for hyperparameter optimization in certain situations (Agrawal and Chakraborty 2021). Accordingly, in this study, the convergence property of PSO was utilized to overcome the limitations of the benchmark models.

It systematically evaluated the impact of different algorithmic components on the accuracy of classification in our proposed Bayesian-ConvLSTM-PSO model. The study compares this novel approach with several comparison models, including PSO-1D CNN-LSTM, DNN, DNN-LSTM, and 1D Conv-LSTM. By isolating and analyzing the contributions of each section within the proposed model, it is aimed to quantify the influence of individual components, such as the integration of Bayesian optimization, ConvLSTM, and PSO, on the overall performance. The comparison across multiple stations Ahullu, Kizilli, and Erenkaya highlights the performance of the proposed Bayesian-ConvLSTM-PSO model in time series prediction tasks. Based on the results presented in Table 3, for the Ahullu station, the proposed Bayesian-ConvLSTM-PSO method outperforms all other models. It achieves the lowest root mean square error (RMSE) of 0.0245 and MAE of 0.0085, indicating superior prediction accuracy. The R^2 value of 0.9750 signifies a very strong correlation between the predicted and observed values, demonstrating the model's effectiveness in capturing the underlying patterns in the data. Additionally, the KGE value of 0.7500 reflects a balanced performance in terms of correlation, variability, and bias, while the BF of 0.0006 indicates minimal bias in the predictions. In comparison, the baseline models, such as PSO-1D CNN-LSTM and DNN-LSTM, have higher RMSE and MAE values and significantly lower R^2 values, suggesting less accurate and less reliable predictions. At the Kizilli station, the proposed model also leads the performance metrics, with an RMSE of 0.0195 and MAE of 0.0065, both the lowest among the models evaluated. The method's R^2 value of 0.9950 and KGE of 0.9950 underscore its robustness and accuracy in predicting the hydrological time series at this station. The BF of 0.0003 further confirms the low bias in its predictions. The baseline models, including PSO-1D CNN-LSTM and 1D conv-LSTM, while still competitive, show slightly higher RMSE and MAE values, and their R^2 values, though close, do not match the precision of the proposed model. For the Erenkaya station, the hybrid Bayesian-ConvLSTM-PSO method again demonstrates its superiority with an RMSE of 0.0180 and MAE of 0.0068, outperforming all other models. The R^2 value of 0.9755 indicates a strong predictive capability, while the KGE of 0.9550 confirms the model's ability to capture the variability in the data effectively. Although the BF is slightly higher at 0.5000 compared to other stations, it remains within acceptable limits, indicating good calibration of the predictions. The baseline models, such as DNN and DNN-LSTM, show much higher RMSE and MAE values, along with significantly lower R^2 values, indicating poorer performance compared to the proposed model. Overall, the hybrid Bayesian-ConvLSTM—PSO method consistently

Table 3 Accuracy metrics of each model based on specificized station dataset

Station	Model	RMSE	MAE	R^2	KGE	BF
Ahullu	Bayesian-ConvLSTM-PSO	0.0245	0.0085	0.9750	0.7500	0.0006
	PSO-1D CNN-LSTM	0.0285	0.0093	0.9457	0.6836	0.0008
	DNN	0.0534	0.0324	0.2431	0.1550	0.0028
	DNN-LSTM	0.0284	0.0089	0.6481	0.7078	0.0008
	1D conv-LSTM	0.0284	0.0105	0.6265	0.7186	0.0008
Kizilli	Bayesian-ConvLSTM-PSO	0.0195	0.0065	0.9950	0.9950	0.0003
	PSO-1D CNN-LSTM	0.0208	0.0076	0.9579	0.9505	0.0004
	DNN	0.1153	0.0660	0.2876	0.3171	0.0131
	DNN-LSTM	0.0229	0.0092	0.9493	0.8903	0.0005
	1D conv-LSTM	0.0205	0.0079	0.9594	0.9434	0.0004
Erenkaya	Bayesian-ConvLSTM-PSO	0.0180	0.0068	0.9755	0.9550	0.5000
	PSO-1D CNN-LSTM	0.0226	0.0104	0.9472	0.8513	0.0005
	DNN	0.1128	0.0547	0.3076	0.3991	0.0127
	DNN-LSTM	0.0279	0.0116	0.9200	0.8232	0.0007
	1D conv-LSTM	0.0252	0.0138	0.9346	0.8527	0.0006

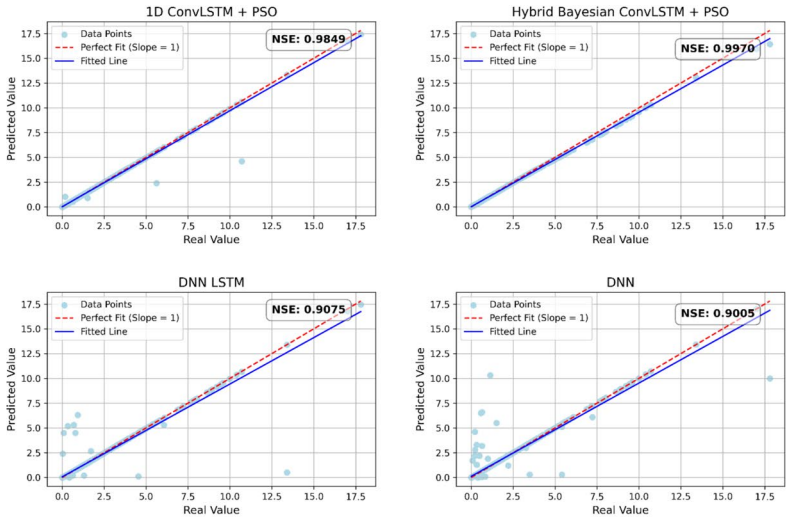
demonstrates superior accuracy, reliability, and bias minimization across all three stations when compared to the comparison models.

The proposed model consistently achieved a Nash–Sutcliffe Efficiency (NSE) score of 0.99 across all three datasets, demonstrating its exceptional accuracy and robustness in predicting outcomes (Fig. 3). In comparison, the 1DConv LSTM combined with PSO also performed admirably, matching the proposed model's performance on the Ahullu and Erenkaya datasets with an NSE of 0.99, and achieving a slightly lower NSE of 0.98 on the Kizilli dataset. The DDN LSTM model exhibited a more varied performance, with strong results on the Erenkaya dataset (NSE 0.99) but lower accuracy on the Kizilli (NSE 0.90) and Ahullu (NSE 0.93) datasets. Similarly, the DNN model showed high accuracy on the Erenkaya dataset (NSE 0.99) but underperformed on the Kizilli (NSE 0.90) and Ahullu (NSE 0.88) datasets. These results underscore the consistency and superiority of the proposed model, while also highlighting the effectiveness of the 1D ConvLSTM-PSO model, particularly on the Ahullu and Erenkaya datasets. The variability in performance of the DDN LSTM and DNN models across different datasets further emphasizes the robustness of the proposed approach.

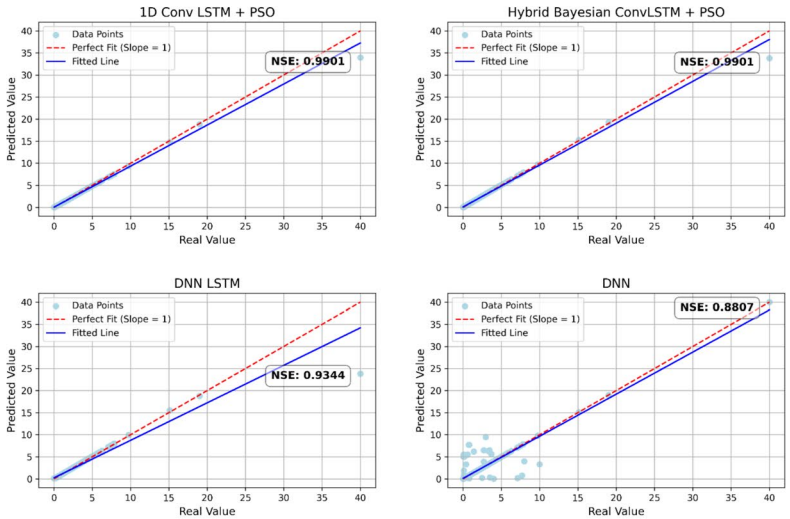
Comparison with state-of- the-art models

After demonstrating that our proposed system, which integrates a PSO algorithm with a Bayesian convolutional framework and a multi-head attention mechanism, significantly improves accuracy, it is proceeded to compare it with these state-of-the-art models. In this study, it has been selected three state-of-the-art models: (i) GRU (Wang et al. 2024), (ii) attention-based long short-term memory (A-LSTM) (Muhammad et al. 2024), and (iii) long short-term memory recurrent neural network (LSTM-RNN) (Sahoo et al. 2019) which are closely related to our proposed approach and recently proposed, particularly in their use of deep learning parameters. These methods primarily utilize LSTM networks and attention mechanisms. For a fair comparison, all parameters of the state-of-the-art models were set based on their respective foundational papers.

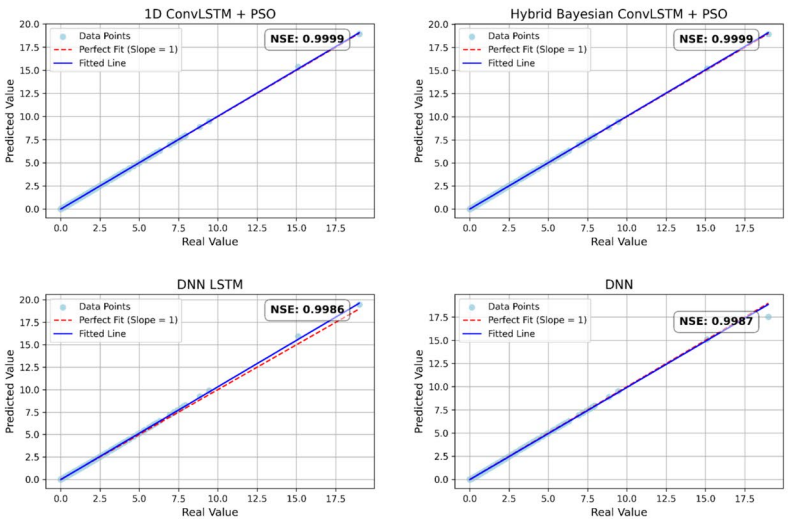
Based on the results demonstrated in Table 4, for the Ahullu station, the proposed model demonstrates superior performance with a reduced RMSE of 0.0245 and MAE of 0.0085 compared to the other models, indicating higher prediction accuracy. The R^2 value of 0.9750 suggests a much stronger correlation between the predicted and observed values than that of the other models. Additionally, the KGE value of 0.7500 reflects a better balance between the correlation, variability, and bias of the predictions. The BF of 0.0006 also indicates that the proposed model exhibits the least bias in predictions. At the Kizilli station, the proposed model outperforms the other models, achieving the lowest RMSE (0.0195) and MAE (0.0065), which further confirms its predictive accuracy. The R^2 value of 0.9950, coupled with a KGE of 0.9950, underscores the robustness of the proposed model in capturing the underlying data patterns. The BF of 0.0003 highlights minimal bias, ensuring reliable predictions. For the Erenkaya station, the proposed model again shows enhanced performance with an RMSE of 0.0180 and MAE of 0.0068, both lower than those of the comparative models. The R^2 value of 0.9755 indicates a strong predictive capability, while the KGE of 0.9550 confirms the method's ability to capture the variability in the data effectively. The BF of 0.5000, being lower than that of the other models, indicates better calibration of predictions.



I. KIZILLI dataset



II. AHULLU dataset



III. ERENKAYA dataset

Fig. 3 NSE score comparison for Kizilli, Ahullu, and Erenkaya datasets

Table 4 Accuracy metrics of the proposed model compared with state-of-the-art models

Station	Model	RMSE	MAE	R^2	KGE	BF
AHULLU	Proposed model	0.0245	0.0085	0.9750	0.7500	0.0006
	GRU (Wang et al. 2024)	0.0289	0.0093	0.9457	0.6836	0.0008
	A-LSTM (Muhammad et al. 2024)	0.0508	0.0219	0.8991	0.9726	1.6348
	LSTM-RNN (Sahoo et al. 2019)	0.0258	0.9989	0.8698	0.9815	0.0258
KIZILLI	Proposed model	0.0195	0.0065	0.9950	0.9950	0.0003
	GRU (Wang et al. 2024)	0.0224	0.0099	0.9898	0.9972	0.2751
	A-LSTM (Muhammad et al. 2024)	0.0281	0.0075	0.9799	0.9975	0.2012
	LSTM-RNN (Sahoo et al. 2019)	0.0239	0.0013	0.9991	0.9925	0.6881
ERENKAYA	Proposed model	0.0180	0.0068	0.9755	0.9550	0.5000
	GRU (Wang et al. 2024)	0.4460	0.0598	0.9638	0.9015	1.6362
	A-LSTM (Muhammad et al. 2024)	0.1187	0.0410	0.9474	0.9753	2.3852
	LSTM-RNN (Sahoo et al. 2019)	0.5064	0.0334	0.9533	0.8915	1.0269

In this study, it is conducted an error distribution analysis to evaluate the performance of the proposed system against state-of-the-art models across the datasets in Fig. 4. In the Ahullu dataset, all models, including the proposed system, exhibited error distributions within a narrow range from 0 to 8. The close alignment of these error ranges indicates that all models performed similarly, with minimal variance in prediction accuracy. For the Erenkaya dataset, a notable divergence in performance was observed. The proposed system, along with the A-LSTM (Muhammad et al. 2024) and LSTM-RNN (Sahoo et al. 2019) models, achieved errors close to zero, indicating high accuracy and consistency in their predictions. However, the GRU (Brentan et al. 2024) model displayed a broader error range, spanning from −7.5 to 10, suggesting less stable and reliable predictions. In the Kizilli dataset, the proposed model maintained a small

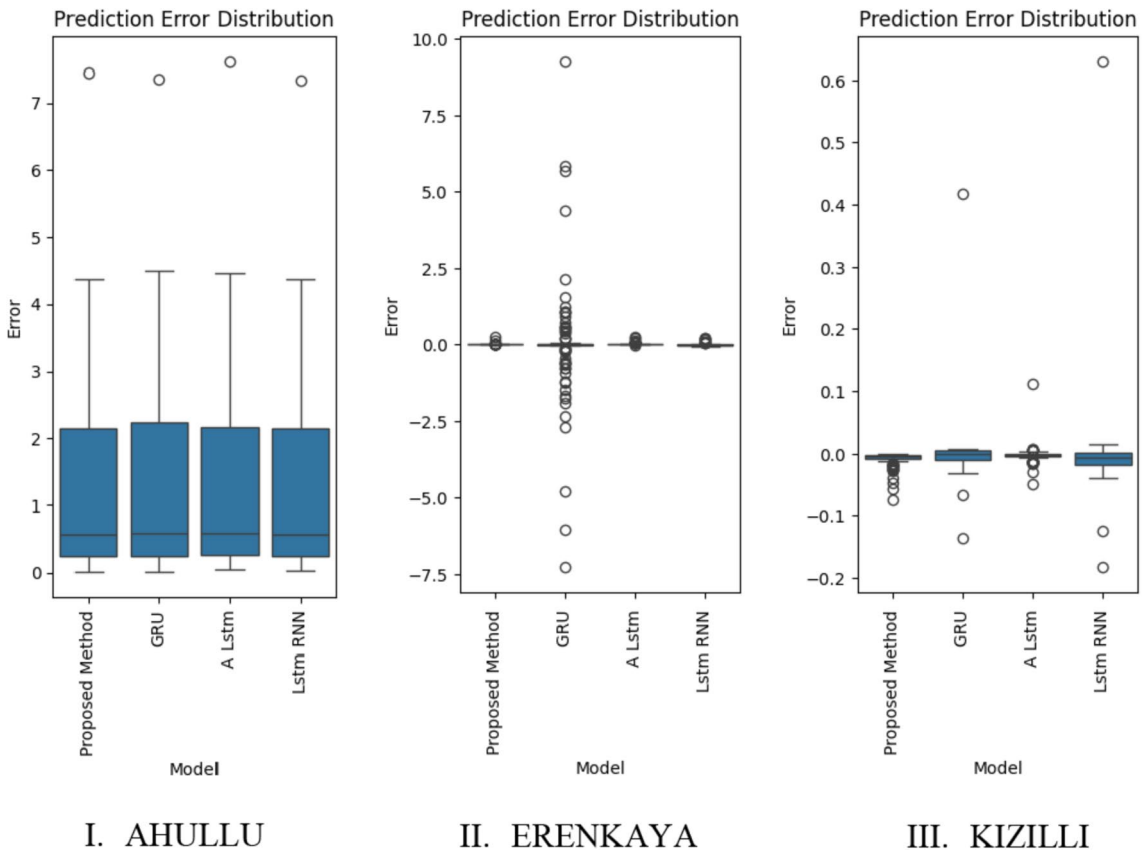


Fig. 4 Prediction error distribution of the proposed system compared to state-of-the-art models

error range, fluctuating slightly above zero to − 1. The GRU (Brentan et al. 2024) model showed a somewhat larger error spread, ranging from 0 to − 2. The A-LSTM (Muhammad et al. 2024) model's errors were marginally below zero and extended up to 1, while the LSTM-RNN (Sahoo et al. 2019) model demonstrated the widest error distribution, ranging from − 2 to 6. These results underscore that the methods employing attention mechanisms, such as the proposed system and A-LSTM (Muhammad et al. 2024), consistently achieve more stable and accurate results across different datasets when compared to other models. The attention mechanisms likely contribute to the models' ability to focus on relevant input features, thereby reducing prediction errors and enhancing overall performance.

To provide a more detailed comparison, it was employed a Taylor diagram, which visually represents the correlation, standard deviation, and root-mean-square error of each method across three different datasets. The results, as depicted in the Taylor diagram (Fig. 5), indicate that all the methods achieved significant results, with values close to one. This outcome demonstrates that our proposed algorithm, similar to other state-of-the-art models, achieved

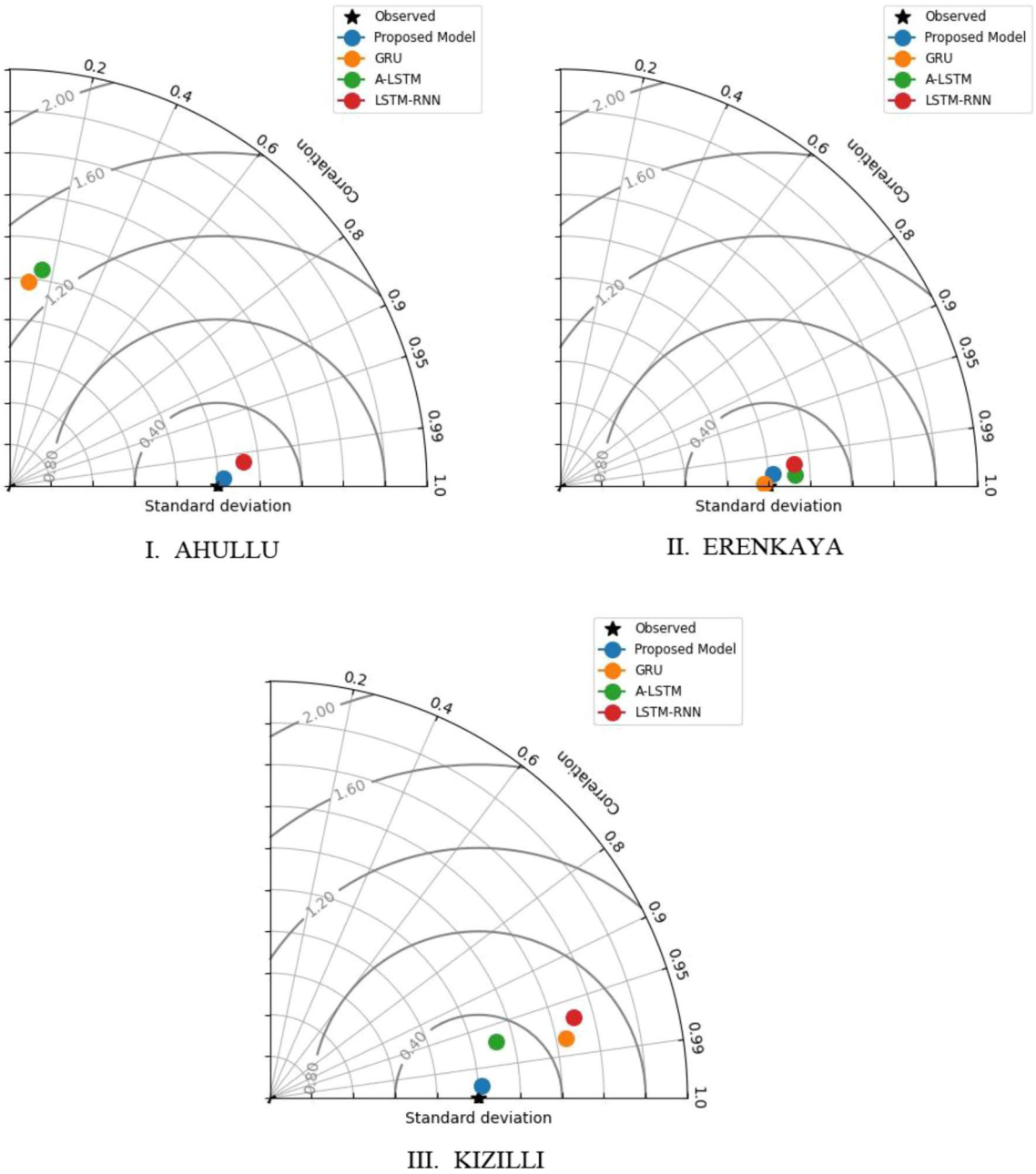


Fig. 5 Performance comparison of proposed algorithms using Taylor diagram

results that are compatible and highly effective across the datasets, reinforcing its robustness and reliability in classification tasks.

Discussion

This study focused on optimizing a hybrid model that integrates the Bayesian network with PSO, and the results demonstrated high accuracy rates. In the ablation study, the proposed Bayesian-ConvLSTM-PSO model achieved exceptionally high R^2 values, such as 0.9950. Moreover, while the PSO-1D CNN-LSTM model yielded an R^2 value of 0.9457 on the Ahullu dataset, the hybrid model improved this to 0.9750. Significant increases were observed in RMSE and MAE values, while notable improvements were achieved in R^2 and KGE values. The results indicate that combining the Bayesian method with LSTM networks plays a crucial role in enhancing accuracy and reducing uncertainty. BO is particularly useful when calculating objective functions is computationally costly. It is not impacted by the parameter issues that commonly affect traditional population-based algorithms like genetic algorithms and PSO. This makes it especially valuable when the objective function is costly to compute (Hutter, Frank; Kotthoff,Lars; Vanschoren, 2019).

In this study, rather than employing Bayesian Optimization, a novel hybrid model was developed by combining convolutional Bayesian network, LSTM, multi-head attention, and 1D convolutional layers. Studies show that the attention mechanism increases the prediction accuracy by facilitating the extraction of dynamic features. Zhou et al. (2022) highlighted the superiority of an attention mechanism and a CNN-LSTM-based hybrid model for daily urban water demand forecasting. This study emphasizes the relevant features in multivariate time series and eliminates unnecessary data using the attention mechanism. It aligns with the advantages of the current study through its ability to improve accuracy and manage uncertainty in data forecasting. Similarly, Wang et al. (2023a) used the independent spatiotemporal attention mechanism in water level estimation to improve LSTM. STA-LSTM consistently exhibited low prediction errors, with MAPE and RMSE values of 3.3 and 0.75, respectively. Recently, Zhang et al. (2024) addressed the complexities of long-term dependencies and environmental data patterns by integrating spatiotemporal attention mechanisms with STA-GRU and STA-LSTM models. In their study, the STA-GRU model outperformed other traditional deep learning models in terms of overall generalizability and reliability, demonstrating superior performance in metrics such as MAPE, R^2, RMSE, and MAE over a 7-day forecasting horizon. They also highlighted that these models improved performance by effectively learning critical data features, particularly in the case of small datasets.

Also, this novel hybrid model was optimized using PSO. While the research results coincide with the following studies, which show that optimization-based deep learning models are effective in hydrological time series forecasts. In a study, Li et al. (2019) introduced a Bayesian network approach to address the challenge of node selection. Their experimental findings demonstrated that the proposed model, which utilizes directed correlation-based Bayesian networks, produces a more efficient, sparse model with improved prediction performance. Wu et al. (2023) used the BO model to optimize the BiLSTM model. It has also been shown that the VMD-CNN-AM-BOA-BiLSTM model, which was established by combining the hybrid structure with various signal processing techniques, produces promising outputs in flow prediction. The study supports the current study in that BO-based depth models show satisfactory results in hydrological predictions. Chong et al. (2022) compared BO, random search and hyperband algorithms in order to optimize various parameters of ANN, CNN and LSTM models in short lead time flow prediction. Analyses indicate that BO is also effective in hyperparameter optimization. The study outputs of Chong et al. (2022) and Ghimire et al. (2021) coincide with the study conducted to produce effective results in streamflow estimation of deep learning and BO-based algorithms. In their study, Lian et al. (2022) applied PCA and BO for the selection of input combinations and parameters of PCA-LSTM-BO, SVR and GBRT. The results showed that PCA-LSTM-BO is superior to other models in flow prediction. Lian et al.'s (2022) study overlaps with the current study in terms of improving the estimation results by ensuring the correct selection of hyperparameters of the BO algorithm in flow estimation. Jahandideh-Tehrani et al.'s (2021) daily flow forecast found that the PSO-ANN model produced superior outputs than GA-ANN. PSO metaheuristic algorithm exhibits superiority in parameter optimization in artificial intelligence models, as it shows significant advantages thanks to its ability to quickly capture optimal solutions, flexible retrieval and noise-sensitive modeling ability (de Moura Oliveira et al. 2018; Prakash et al. 2023). Thanks to these advantages, it also contributes to the effective solution of hydrological problems. In particular, integrating a hybrid model with convolutional layers makes possible successful feature extraction from spatial data, which is of great importance in the field of hydrology, where spatial patterns play a decisive role in water flow and distribution. By using convolutional layers, it can be obtained local dependencies, in hydrological data such as precipitation patterns, which are necessary for accurate streamflow predictions (Li et al. 2024). Moreever, the studies have been revealed that multi-head attention mechanisms enhance the efficiency of LSTM networks in capturing long-term dependencies in time series data, which is critical for tasks such as flood and flow prediction (Mangukiya et al.

2023). Additionally, optimizing this hybrid model using PSO is particularly important in hydrology, where hyperparameters often need to be fine-tuned to accurately reflect the complexities of hydrological processes. Therefore, in future studies, it is necessary to prove the global optimization ability by trying it in regions with different climate structures and in predicting different hydrological variables. Using machine learning techniques for parameter optimization in flow estimation is also recommended.

While the proposed Bayesian-ConvLSTM-PSO model demonstrates robust performance and significant improvements over conventional and state-of-the-art methods, it is not without limitations. One notable challenge is the computational intensity of the PSO algorithm. Due to its iterative nature and the evaluation of multiple particles in each iteration, PSO can be computationally demanding, particularly for larger datasets or models with extensive hyperparameter spaces. In this study, the computational burden was mitigated by limiting the swarm size and number of iterations. However, this may have restricted the exploration of broader parameter spaces. Future research could address this by exploring distributed computing techniques or hybrid optimization algorithms to enhance scalability.

Another potential limitation lies in the risk of overfitting associated with the Bayesian component of the model. By treating model weights as probability distributions, Bayesian inference inherently adds complexity to the model, which can lead to overfitting if priors are overly confident or the dataset is small. Nevertheless, additional strategies, such as Bayesian model averaging, hierarchical Bayesian models, or simpler architectures, could further reduce overfitting risks and improve generalization in future implementations.

Despite these challenges, the proposed model's strong performance across diverse datasets underscores its robustness and practical applicability, even under conditions of data scarcity and variability. Addressing these limitations in future work will help to refine the methodology and expand its utility in hydrological forecasting and other domains.

Conclusion

In this study, it was integrated the 1D convolutional layers, a convolutional Bayesian network, multi-head attention, and LSTM networks to create a novel hybrid model, which was then optimized using PSO. This model has been successfully used to enhance the accuracy and to reduce uncertainty of hydrological time series forecasting. Our proposed model has demonstrated competitive performance compared to state-of-the-art models across several metrics. Specifically, in the Ahullu station, our model achieved the lowest RMSE and a high R^2 score, outperforming other models in terms of accuracy and reliability. Similarly, at the Kizilli and Erenkaya stations, the proposed model showed strong performance, with notably lower RMSE and MAE values compared to the GRU, A-LSTM, and LSTM-RNN models. These results highlight the effectiveness of our hybrid model in capturing complex temporal dependencies and uncertainties in hydrological data. A thorough discussion of the study's findings is provided below:

- Findings reveal that the proposed model surpasses all other models by reflecting superior prediction accuracy. The novel hybrid Bayesian-ConvLSTM-PSO method can be used to obtain high accuracy hydrological forecasting.
- In the ablation study, it was obtained very high accuracy levels in all datasets, and R^2 values were obtained as 0.9750, 0.9950 and 0.9755 for Ahullu, Kızıllı and Erenkaya datasets, respectively.
- In the comparison of the proposed model with the state-of-the-art models, it has been observed that the hybrid model proposed in this study yields higher prediction accuracy compared to the GRU, A-LSTM, and LSTM-RNN models.

The developed hybrid model has been assessed exclusively using daily-scale flow data. Future research can be applied with shorter time intervals, such as hourly or even 15- and 30-min periods. Additionally, incorporating other hydrological variables such as temperature, evaporation, humidity, and snowmelt into the analysis could provide further insights into the proposed model's applicability in hydrological studies. Furthermore, if the hyperparameters of PSO and the established hybrid model fail to determine the optimal result, it may lead to overfitting problems and challenges in modeling large datasets. Also, the main limitation is its high computational cost. Consequently, various parameters were tested within the established hybrid model to achieve the optimal solution. This approach allows for maximizing the accuracy of flow estimation. Looking ahead, it is intended to strengthen the current approach by integrating global feature extraction methods in possible future studies. Prediction accuracy can be further increased by combining these methods with additional optimization of the PSO algorithm's parameters, such as particle size, spacing, learning factors and halting circumstances. Furthermore, hybridizing the PSO algorithm with new algorithms or integrating it with other optimization techniques can help create more reliable models. The advantages of PSO, such as its ability to search for the global search and superior convergence performance, make this algorithm superior to many others. Therefore, it is believed that hybridizing innovative metaheuristic algorithms with PSO in future studies will contribute to the literature. On the other hand, the latest

metaheuristic optimization methods like Walrus optimization algorithm (WaOA), Jellyfish search algorithm (JSO), and Pufferfish optimization algorithm (POA) can also be used to enhance the hybrid model. These developments will open the door to more comprehensive applications in hydrological research as well as improved daily streamflow predictions. These enhancements aim to strengthen the model's ability to recognize wider patterns and connections in the data, potentially resulting in improved forecasting accuracy and reliability. Continuous efforts will concentrate on refining these methods and assessing their effects on model performance, with the goal of further advancing predictive capabilities in hydrological forecasting.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11600-025-01570-0.

Acknowledgements The author is immensely grateful to the General Directorate of State Hydraulic Works for their support and cooperation in sharing the required data. We would like to extend our sincere thanks to both the Editor and the anonymous reviewers for their invaluable contributions to the content and development of this paper.

Author contributions HCK contributed to Supervision, Methodology, Validation. SA contributed to Software, Methodology, Conceptualization. MEE contributed to Writing—reviewing & editing, Writing—original draft, Visualization. FO contributed to Writing—reviewing & editing, Data curation. OMK contributed to Formal analysis, Resources, Writing—original draft. AY contributed to Writing—reviewing & editing, Conceptualization, Formal analysis. All authors read and approved the final manuscript.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK).

Data availability The data will be available upon reasonable request.

Declarations

Conflict of interest The authors declare no conflicts of interest.

Consent for publication All authors have consent to publish this paper.

Ethical approval The manuscript complies with all the ethical requirements. The paper was not published in any journal.

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

References

Agrawal AK, Chakraborty G (2021) On the use of acquisition function-based Bayesian optimization method to efficiently tune SVM hyperparameters for structural damage detection. Struct Control Health Monit 28:e2693. https://doi.org/10.1002/stc.2693

Ahmad S, Jia H, Ashraf A, Yin D, Chen Z, Xu C, Chenyang W, Jia Q, Xiaoyue Z, Israr M, Ahmed R (2023) Water resources and their management in Pakistan: a critical analysis on challenges and implications. Water-Energy Nexus 6:137–150. https://doi.org/10.1016/j.wen.2023.10.001

Ahmed Y, Rahman MdM, Alam MS, Miah MI, Choudhury MdSH, Alharbi OA, Akhtar P, Rahman SM (2024) Harnessing neural network model with optimization for enhanced ciprofloxacin antibiotic adsorption from contaminated water: a transparent and objective framework. J Water Process Eng 65:105724. https://doi.org/10.1016/j.jwpe.2024.105724

Bates BC, Dowdy AJ (2024) Discerning the influence of climate variability modes, regional weather features and time series persistence on streamflow using Bayesian networks and multiple linear regression. Int J Climatol 44:997–1013. https://doi.org/10.1002/joc.8368

Brentan B, Mota F, Menapace A, Zanfei A, Meirelles G (2024) Optimizing pump operations in water distribution networks: balancing energy efficiency, water quality and operational constraints. J Water Process Eng 63:105374. https://doi.org/10.1016/j.jwpe.2024.105374

Chen Y-H, Krishna T, Emer JS, Sze V (2017) Eyeriss: an energy-efficient reconfigurable accelerator for deep convolutional neural networks. IEEE J Solid-State Circuits 52:127–138. https://doi.org/10.1109/JSSC.2016.2616357

Chen Y, Zhang H, Peng J, Ma S, Xu T, Tang L (2024) Forecasting the potential of reclaimed water using signal decomposition and deep learning. J Water Process Eng 65:105770. https://doi.org/10.1016/j.jwpe.2024.105770

Chong KL, Huang YF, Koo CH, Sherif M, Ahmed AN, El-Shafie A (2022) Investigation of cross-entropy-based streamflow forecasting through an efficient interpretable automated search process. Appl Water Sci 13:6. https://doi.org/10.1007/s13201-022-01790-5

de Moura Oliveira PB, Oliveira J, Cunha JB, 2018. Trends in Gravitational Search Algorithm BT - Distributed Computing and Artificial Intelligence, In: 14th International Conference, In: Omatu, S., Rodríguez, S., Villarrubia, G., Faria, P., Sitek, P., Prieto, J. (Eds.), Springer International Publishing, Cham, pp. 270–277.

Demiray BZ, Sit M, Mermer O, Demir I (2024) Enhancing hydrological modeling with transformers: a case study for 24-h streamflow prediction. Water Sci Technol 89:2326–2341. https://doi.org/10.2166/wst.2024.110

Eren Y, Küçükdemiral İ (2024) A comprehensive review on deep learning approaches for short-term load forecasting. Renew Sustain Energy Rev 189:114031. https://doi.org/10.1016/j.rser.2023.114031

Fang K, Kifer D, Lawson K, Feng D, Shen C (2022) The data synergy effects of time-series deep learning models in hydrology. Water Resour Res. https://doi.org/10.1029/2021WR029583

Fang J, Yang L, Wen X, Li W, Yu H, Zhou T (2024) A deep learning-based hybrid approach for multi-time-ahead streamflow prediction in an arid region of Northwest China. Hydrol Res 55:180–204. https://doi.org/10.2166/nh.2024.124

Gazioğlu C, Menteş EN, Kaya Ş, Tanık A (2019) Calculation of flood risk index for Yesilirmak Basin-Turkey. Int J Environ Geoinform 6:288–299. https://doi.org/10.30897/ijegeo.661533

Ghimire S, Yaseen ZM, Farooque AA, Deo RC, Zhang J, Tao X (2021) Streamflow prediction using an integrated methodology

based on convolutional neural network and long short-term memory networks. Sci Rep 11:17497. https://doi.org/10.1038/s41598-021-96751-4

Gogineni A, Chintalacheruvu MR (2024) Hydrological modeling and uncertainty analysis for a snow-covered mountainous river basin. Acta Geophys 72:3529–3545. https://doi.org/10.1007/s11600-023-01270-7

Granata F, Di-Nunno F (2024) Forecasting short- and medium-term streamflow using stacked ensemble models and different meta-learners. Stoch Environ Res Risk Assess 38:3481–3499. https://doi.org/10.1007/s00477-024-02760-w

Granata F, Di-Nunno F, Pham QB (2024a) A novel additive regression model for streamflow forecasting in German rivers. Results in Eng 22:102104. https://doi.org/10.1016/j.rineng.2024.102104

Granata F, Zhu S, Di-Nunno F (2024b) Advanced streamflow forecasting for Central European Rivers: the cutting-edge Kolmogorov-Arnold networks compared to transformers. J Hydrol 645:132175. https://doi.org/10.1016/j.jhydrol.2024.132175

Huang H, Wu M (2024) Short-term water demand prediction based on decomposition technique optimization and a multihead attention mechanism. J Hydroinf 26:1589–1605. https://doi.org/10.2166/hydro.2024.101

Hussain D, Hussain T, Khan AA, Naqvi SAA, Jamil A (2020) A deep learning approach for hydrological time-series prediction: a case study of Gilgit river basin. Earth Sci Inf 13:915–927. https://doi.org/10.1007/s12145-020-00477-2

Hutter F, Kotthoff L, Vanschoren J (2019) Automated machine learning. Springer Cham. https://doi.org/10.1007/978-3-030-05318-5

Istiqomah NA, Jung D, Khim J (2024) Biochar-based persulfate activation: rate constant prediction, key variables identification, and system optimization. J Water Process Eng 65:105839. https://doi.org/10.1016/j.jwpe.2024.105839

Jahandideh-Tehrani M, Jenkins G, Helfer F (2021) A comparison of particle swarm optimization and genetic algorithm for daily rainfall-runoff modelling: a case study for Southeast Queensland, Australia. Optim Eng 22:29–50. https://doi.org/10.1007/s11081-020-09538-3

Jahangir MS, Quilty J (2024) Generative deep learning for probabilistic streamflow forecasting: conditional variational auto-encoder. J Hydrol 629:130498. https://doi.org/10.1016/j.jhydrol.2023.130498

Jeon Y, Chang W, Jeong S, Han S, Park J (2024) A Bayesian convolutional neural network-based generalized linear model. Biometrics 80:57. https://doi.org/10.1093/biomtc/ujae057

Jia F, Zhu Z, Dai W, Le VV (2024) Short-term forecasting of streamflow by integrating machine learning methods combined with metaheuristic algorithms. Expert Syst Appl 245:123076. https://doi.org/10.1016/j.eswa.2023.123076

Jiang A, McBean E, Wang Y (2024) A sustainable environment requires sustainable water—a review of some water issues to learn from. Environ Rev. https://doi.org/10.1139/er-2024-0020

Katipoğlu OM (2023) Prediction of streamflow drought index for short-term hydrological drought in the semi-Arid Yesilirmak basin using wavelet transform and artificial intelligence techniques. Sustainability. https://doi.org/10.3390/su15021109

Kennedy J, Eberhart R, 1995. Particle swarm optimization, In: Proceedings of ICNN'95 - International Conference on Neural Networks. pp. 1942–1948 vol.4. https://doi.org/10.1109/ICNN.1995.488968

Kilichev D, Kim W (2023) Hyperparameter optimization for 1D-CNN-based network intrusion detection using GA and PSO. Mathematics 11:3724. https://doi.org/10.3390/math11173724

Kilinc HC, Ahmadianfar I, Demir V, Heddam S, Al-Areeq AM, Abba SI, Tan ML, Halder B, Marhoon HA, Yaseen ZM (2023) Daily scale river flow forecasting using hybrid gradient boosting model with genetic algorithm optimization. Water Resour Manage 37:3699–3714. https://doi.org/10.1007/s11269-023-03522-z

Kilinc HC, Haznedar B, Ozkan F, Katipoğlu OM (2024) An evolutionary hybrid method based on particle swarm optimization algorithm and extreme gradient boosting for short-term streamflow forecasting. Acta Geophys 72:3661–3681. https://doi.org/10.1007/s11600-024-01307-5

Lee D-G, Ahn K-H (2024) Improving medium-range streamflow forecasts over South Korea with a dual-encoder transformer model. J Environ Manage 368:122114. https://doi.org/10.1016/j.jenvman.2024.122114

Li Z, Jiang S, Li L, Li Y (2019) Building sparse models for traffic flow prediction: an empirical comparison between statistical heuristics and geometric heuristics for Bayesian network approaches. Transport b: Trans Dyn 7:107–123. https://doi.org/10.1080/21680566.2017.1354737

Li B, Sun T, Tian F et al (2024) Hybrid hydrological modeling for large alpine basins: a semi-distributed approach. Hydrol Earth Syst Sci 28:4521–4538. https://doi.org/10.5194/hess-28-4521-2024

Lian Y, Luo J, Wang J, Zuo G, Wei N (2022) Climate-driven model based on long short-term memory and Bayesian optimization for multi-day-ahead daily streamflow forecasting. Water Resour Manage 36:21–37. https://doi.org/10.1007/s11269-021-03002-2

Liu S, Lu D, Painter SL, Griffiths NA, Pierce EM (2023) Uncertainty quantification of machine learning models to improve streamflow prediction under changing climate and environmental conditions. Front Water. https://doi.org/10.3389/frwa.2023.1150126

Mangukiya NK, Sharma A, Shen C (2023) How to enhance hydrological predictions in hydrologically distinct watersheds of the Indian subcontinent? Hydrol Process 37:e14936. https://doi.org/10.1002/hyp.14936

Muhammad AU, Muazu T, Ying H, Ba AF, Tijjani S, Adam JM, Bello AU, Bala MM, Ali MH, Dabai US, Kumshe UMM, Yahaya MS (2024) Enhanced streamflow forecasting using attention-based neural network models: a comparative study in MOPEX basins. Model Earth Syst Environ 10:5717–5734. https://doi.org/10.1007/s40808-024-02088-y

Narayanan R, Ganesh N, (2024) A Comprehensive Review of Metaheuristics for Hyperparameter Optimization in Machine Learning, In: Metaheuristics for Machine Learning. pp. 37–72. https://doi.org/10.1002/9781394233953.ch2

Pinarlik M, Selek Z (2024) Evaluating operational performance and sustainability of water supply reservoirs in the Yesilirmak Basin, Turkey. Water Resour 51:449–461. https://doi.org/10.1134/S0097807823601632

Prakash R, Yadav S, Saha SK, (2023) A Metaheuristic Approach for the Design of Linear Phase FIR Differentiator, In: 2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES). pp. 744–748. https://doi.org/10.1109/CISES58720.2023.10183596

Ramkumar D, Jothiprakash V (2024) Forecasting influent wastewater quality by chaos coupled machine learning optimized with Bayesian algorithm. J Water Process Eng 61:105306. https://doi.org/10.1016/j.jwpe.2024.105306

Sahoo BB, Jha R, Singh A, Kumar D (2019) Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting. Acta Geophys 67:1471–1481. https://doi.org/10.1007/s11600-019-00330-1

Senel E, (2023) Semantic expression extraction from images with depthwise separable convolution and LSTM networks. Konya Technical University.

Seunghwan S, Chung M (2022) Evaluation of applicability of 1D-CNN and LSTM to predict horizontal displacement of retaining wall according to excavation work. Int J Adv Comput Sci Appl

(IJACSA) 13:86–91. https://doi.org/10.14569/IJACSA.2022.0130210
Shridhar K, Laumann F, Liwicki M, (2019) A Comprehensive guide to Bayesian Convolutional Neural Network with Variational Inference.
Wan W, Zhou Y, Chen Y (2024) Interpretable and explainable hybrid model for daily streamflow prediction based on multi-factor drivers. Environ Sci Pollut Res 31:34588–34606. https://doi.org/10.1007/s11356-024-33594-2
Wang Y, Huang Y, Xiao M et al (2023a) Medium-long-term prediction of water level based on an improved spatio-temporal attention mechanism for long short-term memory networks. J Hydrol 618:129163. https://doi.org/10.1016/j.jhydrol.2023.129163
Wang Y, Wang W, Chau K, Xu D, Zang H, Liu C, Ma Q (2023b) A new stable and interpretable flood forecasting model combining multi-head attention mechanism and multiple linear regression. J Hydroinf 25:2561–2588. https://doi.org/10.2166/hydro.2023.160
Wang W, Du Y, Chau K, Cheng C-T, Xu D, Zhuang W-T (2024) Evaluating the performance of several data preprocessing methods based on GRU in forecasting monthly runoff time series. Water Resour Manage 38:3135–3152. https://doi.org/10.1007/s11269-024-03806-y
Wei Y, Hashim HB, Lai SH, Chong KL, Huang YF, Ahmed AN, Sherif M, El-Shafie A (2024) Comparative analysis of artificial intelligence methods for streamflow forecasting. IEEE Access 12:10865–10885. https://doi.org/10.1109/ACCESS.2024.3351754
Wu J, Wang Z, Hu Y, Tao S, Dong J (2023) Runoff forecasting using convolutional neural networks and optimized bi-directional long short-term memory. Water Resour Manage 37:937–953. https://doi.org/10.1007/s11269-022-03414-8
Yilmaz V, Koycegiz C, Buyukyildiz M (2024) Performance of data-driven models based on seasonal-trend decomposition for streamflow forecasting in different climate regions of Türkiye. Phys Chem Earth, Parts a/b/c 136:103696. https://doi.org/10.1016/j.pce.2024.103696
Yin L, Wang L, Li T, Lu S, Tian J, Yin Z, Li X, Zheng W (2023) U-Net-LSTM: Time series-enhanced lake boundary prediction model. Land. https://doi.org/10.3390/land12101859
Yu Q, Tolson BA, Shen H, Han M, Mai J, Lin J (2024) Enhancing long short-term memory (LSTM)-based streamflow prediction with a spatially distributed approach. Hydrol Earth Syst Sci 28:2107–2122. https://doi.org/10.5194/hess-28-2107-2024
Zhang Y, Zhou Z, Deng Y et al (2024) Daily streamflow forecasting using networks of real-time monitoring stations and hybrid machine learning methods. Water 16:1284. https://doi.org/10.3390/w16091284
Zhou S, Guo S, Du B et al (2022) A hybrid framework for multivariate time series forecasting of daily urban water demand using attention-based convolutional neural network and long short-term memory network. Sustainability 14:11086. https://doi.org/10.3390/su141711086

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