SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Hybrid Lstm-Xgboost Framework for Multi-Day Water Level Forecasting: A Residual Error Correction Approach

Premlata Vidyarthy[1], Shashi Ranjan[2], Amit Kumar[3], Ranjan kumar[4], Alka Kumari[5], Mariya Azmi[6], Kowkuntla Rama Krishna Reddy[7*]

[1]Assistant Professor, Civil Engineering Department, Government Engineering College Sheikhpura, Bihar, India pincode- 811105. E-mail: rajkiprem@gmail.com; Orcid: 0009-0004-8813-6689
[2]Assistant Professor, Civil Engineering Department, Nalanda College of Engineering, Chandi (Nalanda), Bihar, India, Pincode-803108. E-mail: shashi.ce18@nitp.ac.in; Orcid: 0000-0001-5912-7995
[3] Assistant Professor, Civil Engineering Department, Government Engineering College Nawada, Bihar, India, Pincode-805111. E-mail: amitkumar.pg18.ce@nitp.ac.in; Orcid: 0000-0001-7704-4771
[4]Assistant Professor, Civil Engineering Department, Government Engineering College Nawada, Bihar, India, Pincode-805111. E-mail: ranjankrdns@gmail.com; Orcid: 0009-0004-1824-7535
[5]Assistant Professor, Civil Engineering Department, Government Engineering College Nawada, Bihar, India, Pincode-805111. E-mail: alka23jan@gmail.com; Orcid: 0009-0009-7405-1912
[6]Assistant Professor, Civil Engineering Department, Government Engineering College Nawada, Bihar, India, Pincode-805111. E-mail: mariyaazmi11@gmail.com, Orcid: 0009-0009-1783-3758
[7] Assistant Professor, Civil Engineering Department, TKR College of Engineering and Technology Hyderabad, India, Pincode-500097. E-mail: ramakrishna1491@gmail.com; Orcid: 0000-0002-7970-403X

ABSTRACT
Accurate river water level forecasting is crucial for flood risk mitigation, reservoir operation, and sustainable water resource management under increasing hydrological variability. This study proposes a hybrid Long Short-Term Memory–XGBoost (LSTM–XGBoost) framework for multi-day water level forecasting in the Punpun River Basin, India. The framework combines LSTM's temporal sequence learning with XGBoost's nonlinear regression and residual correction capabilities. Forecasting performance was evaluated for 1, 3, 5, 7, and 10-day lead times using KGE, R^2, NSE, RMSE and MAE. The hybrid model consistently outperformed standalone LSTM and XGBoost models. At the 1-day horizon, it achieved KGE, R^2, and NSE values of 0.976, 0.971, and 0.971, respectively, with RMSE and MAE of 0.315 m and 0.172 m. Although accuracy declined with increasing lead time, the hybrid model exhibited slower performance degradation. Results confirm improved reliability and robustness in capturing nonlinear hydrological dynamics and peak water levels.

Keywords: Water level forecasting; LSTM; XGBoost; Hybrid model; Residual correction; Multi-lead time prediction

1. INTRODUCTION

Accurate prediction of river water levels plays a critical role in flood early warning systems, reservoir management and the sustainable planning of water resources (Hipni et al., 2013; Sapitang et al., 2020). The growing occurrence and severity of extreme hydrological events, largely influenced by climate variability, have highlighted the necessity for dependable forecasting systems that can generate precise predictions over varying lead times (Chaubey et al., 2021). Hydrological forecasting methods are generally classified into two main groups: physically-based models and data-driven models. Although physically-based models provide a process-oriented understanding of hydrological dynamics, their application often demands extensive parameter calibration, high computational effort and comprehensive information on catchment properties, which limits their practicality in regions with limited data availability (Ranjan & Roshni, 2023; Tripura et al., 2021).

In comparison, data-driven machine learning approaches have become increasingly popular as effective alternatives because of their ability to model complex nonlinear relationships between predictor variables and target outputs without requiring explicit representation of physical hydrological processes (Ahmed et al., 2022; Zakaria et al., 2021). Among the different machine learning techniques, Long Short-Term Memory (LSTM) networks have received considerable attention in hydrological time series forecasting owing to their strong capability to learn long-term temporal dependencies (Hochreiter & Schmidhuber, 1997; Kratzert et al., 2018). Unlike traditional recurrent neural networks (RNNs), LSTM models effectively address the vanishing gradient problem, allowing them to preserve information over longer sequences. This feature is particularly important for hydrological modeling, where previous catchment conditions strongly affect present hydrological states (Zhang et al., 2018; Hu et al., 2018).

Several studies have highlighted the strong performance of LSTM networks in a wide range of hydrological applications. For example, Kratzert et al. (2018) demonstrated that LSTM models outperformed the Sacramento Soil Moisture Accounting (SAC-SMA) model in rainfall–runoff simulations across 241 catchments in North America. Similarly, Hunt et al. (2022) found that LSTM models produced skillful streamflow forecasts that exceeded the performance of both raw and bias-corrected GloFAS forecasts for prediction lead times of up to five days. In another study, Sabzipour et al. (2023) compared LSTM with the physically-based CEQUEAU model in a snow-dominated Canadian catchment and reported that LSTM generated more reliable forecasts for lead times extending to nine days. Despite these promising outcomes, standalone LSTM models still face certain challenges, especially in accurately predicting extreme hydrological events and reducing systematic biases that become more pronounced over longer forecasting horizons (Cho & Kim, 2022; Lees et al., 2021).

Ensemble learning techniques, particularly Random Forest (RF) and extreme Gradient Boosting (XGBoost), have shown strong predictive capabilities by combining multiple decision trees to model complex nonlinear relationships while reducing the risk of overfitting (Breiman, 2001; Chen & Guestrin, 2016). These approaches have been widely and successfully employed in applications such as flood susceptibility mapping (Zhao et al., 2018; Chen et al., 2020), groundwater level forecasting (Mishra & Maurya, 2025), and streamflow prediction (Shortridge et al., 2016). Nevertheless, unlike LSTM networks, tree-based ensemble methods lack an inherent mechanism for capturing temporal dependencies, which restricts their suitability for sequential hydrological time series data. The performance metrics are widely recommended for hydrological model evaluation (Moriasi et al., 2007).

Recognizing the complementary advantages of these techniques, recent studies have increasingly investigated hybrid frameworks that integrate deep learning models with ensemble learning approaches. The core idea behind these hybrid systems is that LSTM networks are highly effective in capturing temporal dynamics and long-term dependencies, whereas tree-based models are better suited for learning residual errors and nonlinear corrections that may not be fully represented by LSTM models (Konapala et al., 2020; Yang et al., 2020). This residual correction strategy has been applied successfully in several hydrological studies. For instance, Cho and Kim (2022) developed a WRF-Hydro-LSTM hybrid model in which the LSTM component was used to predict the residual errors of the physically-based WRF-Hydro model, resulting in significant improvements in percent bias. Likewise, Prakash and Tripura (2024) reported that an LSTM-RF hybrid framework enhanced the Kling–Gupta Efficiency (KGE) from 0.948 to 0.985 for 1-day lead time water level forecasting in the Punpun River Basin.

Despite these notable developments, important research gaps still exist. First, many hybrid modeling studies have concentrated primarily on short forecasting horizons of 1–3 days, while only a limited number have systematically examined how model performance deteriorates with increasing lead times. Second, the individual contributions of base models, such as LSTM and tree-based approaches, within hybrid architectures remain insufficiently explored. Third, although XGBoost offers advantages over Random Forest in terms of computational efficiency and regularization capability, its effectiveness as a residual correction tool for water level forecasting has not yet been comprehensively investigated. This study addresses these gaps by proposing and evaluating a hybrid LSTM-XGBoost framework for multi-day water level forecasting with lead times of 1, 3, 5, 7, and 10 days. The specific objectives are: (1) to develop a two-stage hybrid model where LSTM captures temporal patterns and XGBoost corrects residual errors; (2) to comprehensively evaluate model performance using multiple metrics (R^2, NSE, RMSE, MAE, KGE) across training and testing phases; (3) to compare hybrid performance against standalone LSTM and XGBoost baselines; and (4) to analyze performance degradation patterns with increasing lead time.

2. MATERIALS AND METHODS

2.1 Study Area and Dataset

Figure 1 illustrates the map of the Punpun River Basin, which is the focus of the present study. The river originates in the highlands of the Palamu district in the Indian state of Jharkhand at an elevation of approximately 300 meters, located at 24°11′ north latitude and 84°9′ east longitude. It eventually joins the Ganga River at Fatuha, about 25 km downstream of Patna. The Punpun River remains dry during the non-monsoon months; however, during the monsoon season, it experiences substantial discharge that often leads to flooding and inundation of adjacent areas. Consequently, water management in the basin is challenged by both extreme seasonal flow conditions high discharge during monsoon and negligible flow during dry periods.

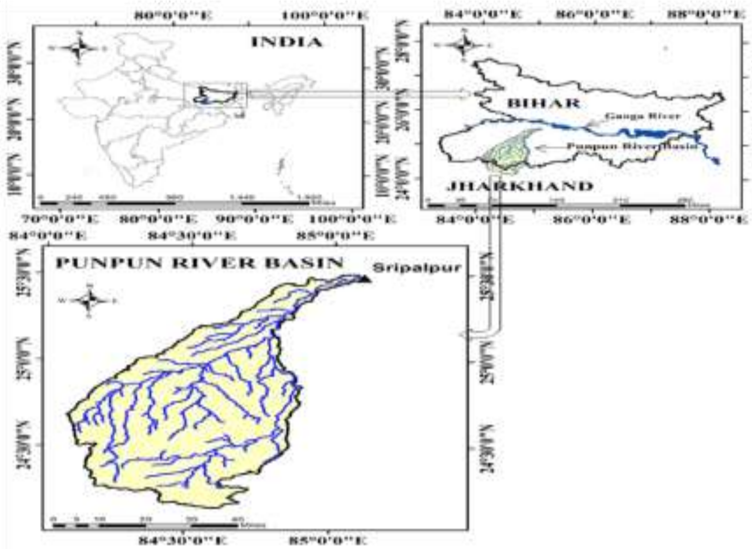


Figure 1 Study area of Punpun River basin

The basin receives significant rainfall between August and October, while the regional humidity remains moderate throughout the year. The average annual rainfall varies between 99 cm and 134 cm, with the upper catchment areas, particularly the Palamu district, receiving comparatively higher precipitation. Approximately 80–87% of the annual rainfall occurs during the monsoon season. In contrast, the lower part of the catchment experiences relatively uniform rainfall with little spatial variation. The Punpun River Basin study area covers about 7055 km^2. Geologically, the basin exhibits considerable diversity, consisting of recent alluvial deposits in the plains and formations of granite, gneiss, and charnokites in the upland regions.

2.2 Methodology

2.2.1 Long Short-Term Memory (LSTM) Network

The LSTM network, introduced by Hochreiter and Schmidhuber (1997), is a specialized recurrent neural network architecture designed to overcome the limitations of traditional RNNs in learning long-term dependencies. The core innovation of LSTM lies in its memory cell structure, which includes three gating mechanisms: the forget gate, input gate, and output gate. These gates regulate the flow of information into, through, and out of the memory cell, allowing the network to selectively retain or discard information over extended time sequences as depicted in Fig.2.

The forget gate determines which information from the previous cell state should be discarded:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(i)}$$

The input gate decides which new information should be stored in the cell state:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(ii)}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(ii)}$$

The cell state is updated by combining the retained previous state and the new candidate values:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(iv)}$$

Finally, the output gate determines the hidden state based on the updated cell state:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(v)}$$

$$h_t = o_t * \tanh(C_t) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(vi)}$$

where xt represents the input vector at time step t, ht denotes the hidden state at time step t, and Ct represents the cell state. The function σ is the sigmoid activation function, while tanh denotes the hyperbolic tangent activation function. W and b correspond to the weight matrices and bias vectors, respectively. The symbols ft, it and ot denote the forget gate, input gate and output gate activations, respectively, and Ct represents the candidate cell state

2.2.2 Extreme Gradient Boosting (XGBoost)

XGBoost, introduced by Chen and Guestrin (2016), is an advanced implementation of gradient boosted decision trees that has gained widespread adoption due to its exceptional predictive performance, computational efficiency, and built-in regularization capabilities. Unlike traditional gradient boosting that builds trees sequentially to correct errors of previous trees, XGBoost incorporates both L1 and L2 regularization to prevent overfitting and employs a sparsity-aware algorithm to handle missing data.

The objective function in XGBoost combines a loss function measuring prediction error with a regularization term penalizing model complexity:

$$\mathcal{L}(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(vii)}$$

Where $\mathcal{L}$(ϕ) denotes the overall objective function, yi,yi represents the loss function measuring the difference

between the predicted value yi and the actual value yi, and fk is the regularization term used to control model complexity.
In this study, XGBoost was employed in two capacities: (1) as a standalone regressor for direct water level prediction, and (2) as a residual correction model to predict errors from LSTM predictions.

2.2.3 Hybrid LSTM-XGBoost Framework

The proposed hybrid framework operates through a two-stage residual correction mechanism, as illustrated conceptually in Figure 2. The fundamental premise is that LSTM effectively captures the dominant temporal patterns and long-term dependencies in water level time series, but systematic residual errors persist. XGBoost, with its strength in modelling complex nonlinear relationships and interactions, is then employed to learn and predict these residual errors.
The LSTM model is trained on the historical dataset to produce base predictions yLSTM(t+n) for water level at lead time t+n:

$$\hat{y}_{LSTM}(t+n) = f_{LSTM}(X_t, X_{t-1}, \dots, X_{t-p}) \dots\dots\dots\dots\dots\dots\dots\dots\dots \text{(viii)}$$

whereXt−1 represents input features at time t (including historical water levels, discharge, and rainfall variables), and p is the lookback window length.
Residual errors are computed as the difference between observed water levels and LSTM predictions:

$$e(t) = y_{obs}(t) - \hat{y}_{LSTM}(t) \dots\dots\dots\dots\dots\dots\dots\dots\dots \text{(ix)}$$

These residuals represent systematic errors that the LSTM model fails to capture. An XGBoost model is then trained to predict these residuals using the same input features:

$$\hat{e}(t+n) = f_{XGBoost}(X_t, X_{t-1}, \dots, X_{t-p}) \dots\dots\dots\dots\dots\dots\dots\dots\dots \text{(x)}$$

The final hybrid prediction is obtained by combining the LSTM base prediction with the XGBoost residual correction:

$$\hat{y}_{hybrid}(t+n) = \hat{y}_{LSTM}(t+n) + \hat{e}(t+n) \dots\dots\dots\dots\dots\dots\dots\dots\dots \text{(xi)}$$

This two-stage approach leverages the complementary strengths of both architectures: LSTM captures sequential dependencies and long-term patterns, while XGBoost corrects remaining systematic biases and captures nonlinear patterns that LSTM may miss.

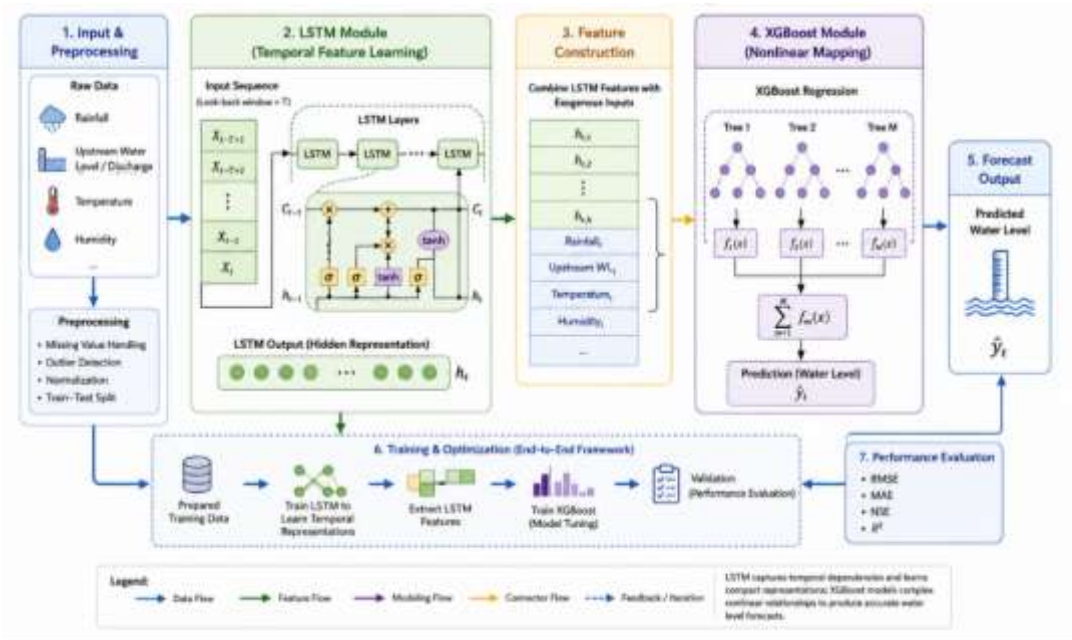


Figure 2. Schematic representation of the proposed LSTM-XGBoost hybrid framework for water level forecasting

2.2.4 Input Features

The prediction models were developed using hydrological, meteorological and temporal input variables. The primary input variables included daily rainfall and runoff, which directly influence river water level fluctuations. In addition, temporal variables such as day of the month and month of the year were extracted from the date column to represent seasonal variability and annual hydrological patterns.
To capture the delayed hydrological response of the river basin, lagged variables were also incorporated into the model. Specifically, the previous 10 days of rainfall, runoff and water level observations were used as additional predictors. These lag features enabled the models to learn the influence of antecedent hydrometeorological conditions on future water levels.

2.2.5 Hyperparameter Optimization

The LSTM model was developed using two hidden layers containing 64 and 32 neurons, respectively, with a dropout rate of 0.2 applied after each layer to reduce overfitting. The model was trained using the Adam optimizer and mean squared error (MSE) as the loss function with a batch size of 32 and a maximum of 100 epochs. Early stopping with a patience value of 10 epochs was employed to retain the best-performing model. For residual correction, the XGBoost model was configured with 300 trees, a maximum depth of 5, a learning rate of 0.05, a subsample ratio of 0.8 and a column sampling ratio of 0.8. The dataset was divided chronologically into 80% training and 20% testing data, and the models were developed for 1, 3, 5, 7 and 10 days lead time water level forecasting.

2.2.6 Performance Evaluation Metrics

Model performance was evaluated using five complementary metrics that assess different aspects of **predictive accuracy:**

Coefficient of Determination (R^2): Measures the proportion of variance in observed values explained by predictions:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i^{obs} - y_i^{sim})^2}{\sum_{i=1}^{n}(y_i^{obs} - \bar{y}^{obs})^2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(xi)}$$

Nash-Sutcliffe Efficiency (NSE): Widely used in hydrological modeling, NSE indicates how well predictions match observed values relative to the mean:

$$NSE = 1 - \frac{\sum_{i=1}^{n}(y_i^{obs} - y_i^{sim})^2}{\sum_{i=1}^{n}(y_i^{obs} - \bar{y}^{obs})^2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(xii)}$$

Root Mean Square Error (RMSE): Measures the square root of average squared differences, penalizing large errors:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i^{obs} - y_i^{sim})^2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(xiii)}$$

Mean Absolute Error (MAE): Measures average absolute differences, providing a robust error metric:

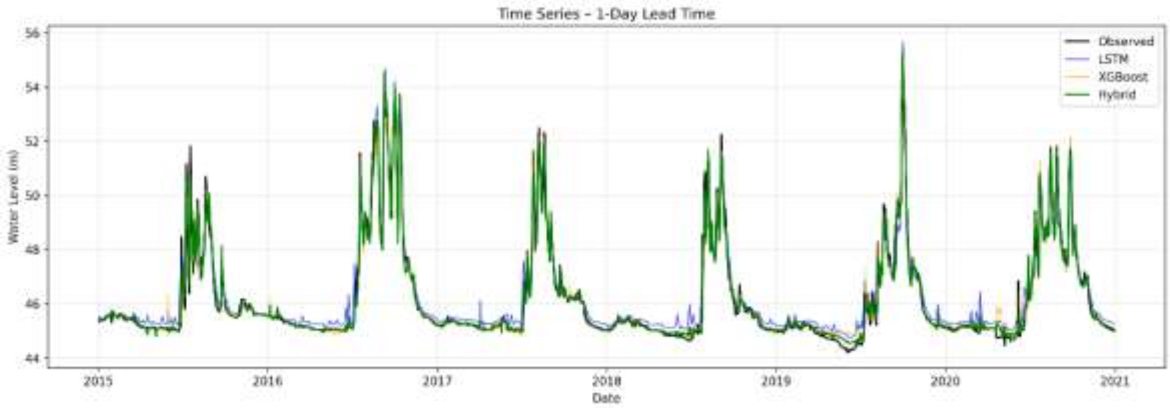
$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i^{obs} - y_i^{sim} | \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(xiv)}$$

Kling-Gupta Efficiency (KGE): Provides a diagnostically balanced assessment by decomposing error into correlation, bias, and variability components (Gupta et al., 2009):

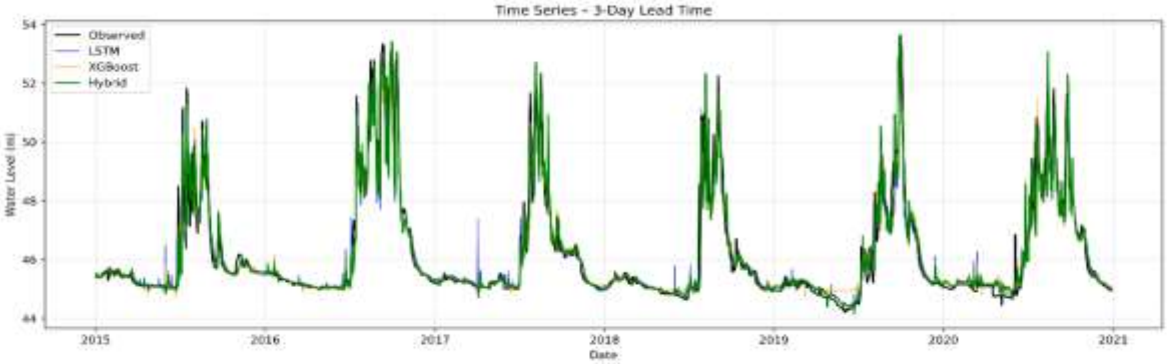
$$KGE = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(xv)}$$

3. RESULTS AND DISCUSSION

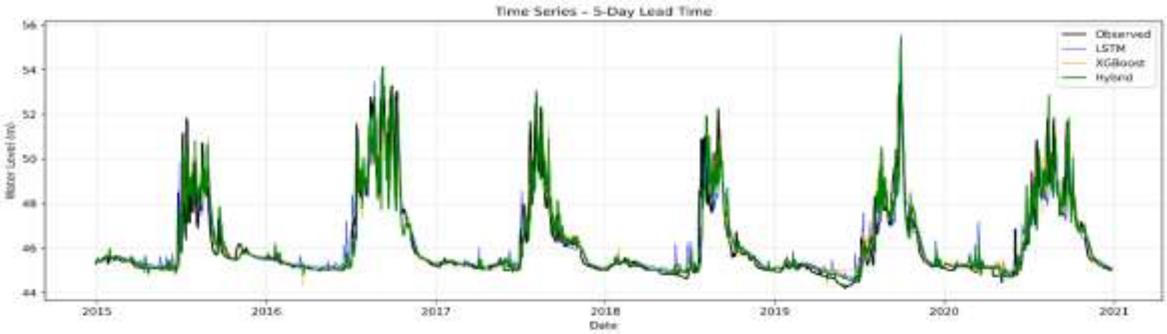
The study employed three forecasting approaches, namely the LSTM model, XGBoost model, and a hybrid LSTM–XGBoost model for predicting river water level at different forecasting lead times. The predictive capability of the developed models was evaluated using statistical indicators including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), coefficient of determination (R^2), and Nash–Sutcliffe Efficiency (NSE). The training and validation loss curves indicated that all models successfully learned the temporal characteristics of the dataset as shown in Figure 3 (a)-(e). However, significant differences were observed in convergence behavior and generalization performance among the models. The hybrid LSTM–XGBoost model exhibited faster convergence and lower validation loss compared to the standalone LSTM model, indicating improved learning stability and reduced overfitting. In contrast, the LSTM model showed relatively slower convergence with larger fluctuations in validation loss, particularly at higher forecasting lead times.



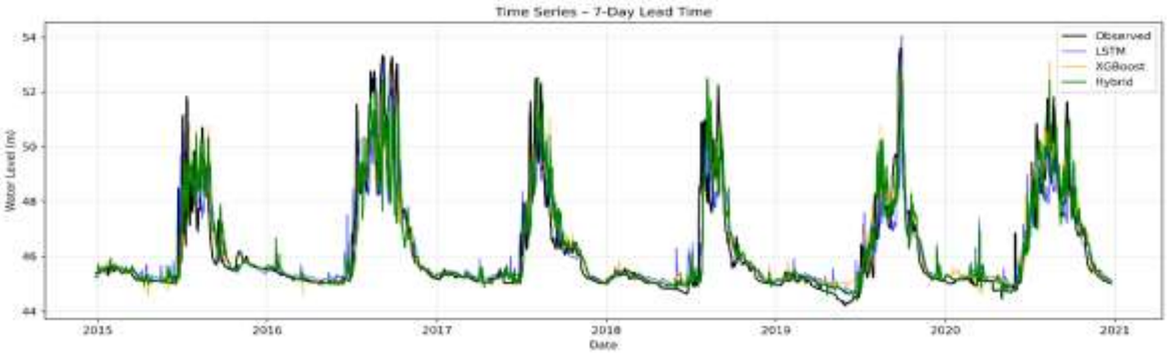
a)



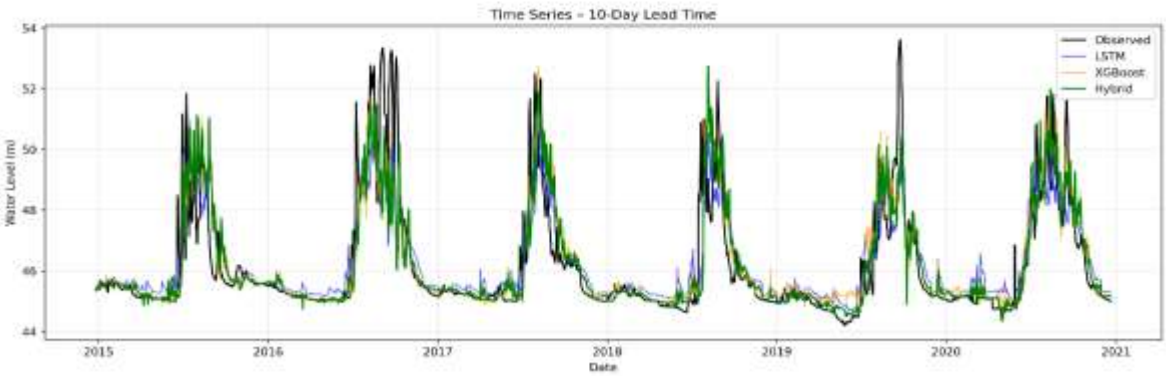
b)



c)



d)



e)

Figure 3 Plots showing result of predicted model with observed water level at different lead time a) 1-Day b) 3-Day c) 5-Day d) 7-Day e) 10-Day

The comparative performance analysis revealed that the hybrid Long Short-Term Memory–XGBoost model consistently outperformed the standalone LSTM model across all forecasting lead times and evaluation metrics Figure 4. The percentage improvement analysis demonstrated substantial enhancement in predictive capability, particularly for error-based metrics. The Kling–Gupta Efficiency (KGE) improved progressively from approximately 3.7% at 1-day lead time to about 17.4% at 10-day lead time, indicating improved agreement between observed and predicted water levels during longer forecasting horizons. Similarly, the coefficient of determination (R^2) showed improvements ranging from nearly 1.9% to 3.4%, confirming the superior variance explanation capability of the hybrid framework.

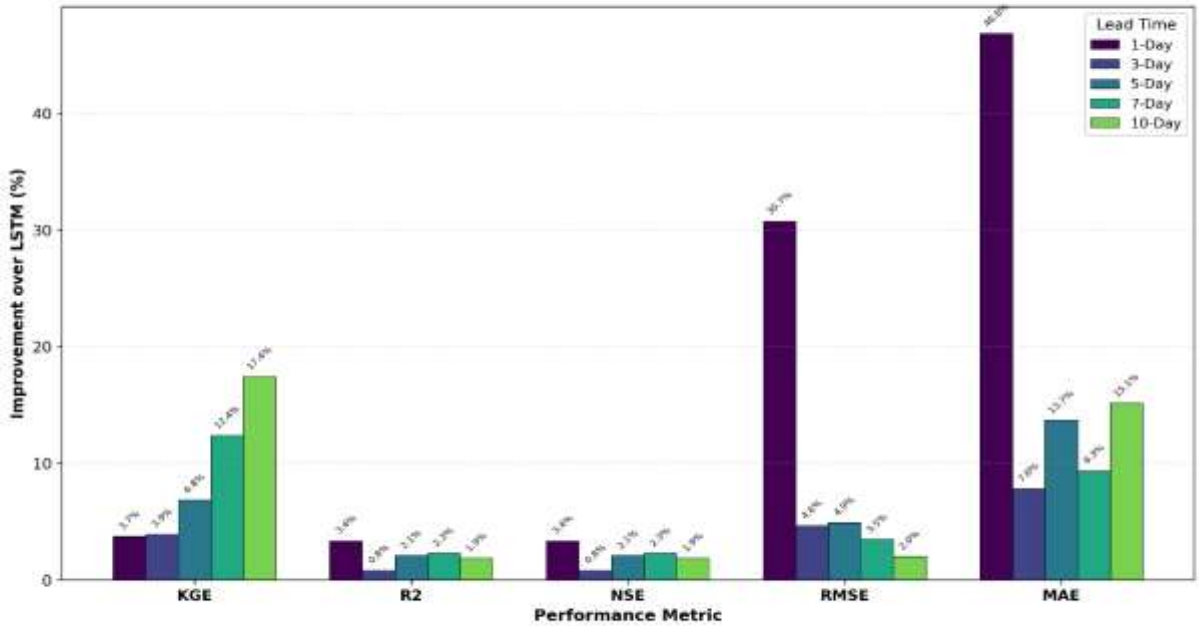


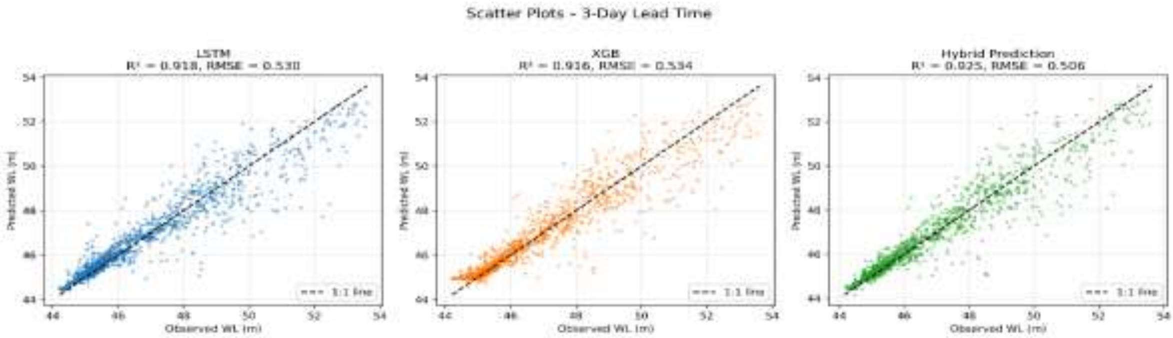
Figure 4 Representation of improvement of Hybrid model over standalone LSTM model.

The Nash–Sutcliffe Efficiency (NSE) also exhibited positive improvements across all lead times, with values ranging between approximately 1.9% and 3.4%, suggesting enhanced reproduction of temporal hydrological dynamics. The most significant improvements were observed in error reduction metrics. RMSE decreased by nearly 30.7% for the 1-day forecast and maintained positive reductions for extended lead times, indicating substantial minimization of forecasting uncertainty. Likewise, MAE showed the highest improvement, with a reduction of approximately 46.6% at 1-day lead time and improvements remaining above 7% even for longer forecasting intervals. These findings demonstrate that the integration of XGBoost with LSTM significantly enhanced forecasting accuracy, reduced prediction errors, and improved model robustness across short-, medium-, and long-term forecasting horizons.

Scatter plot analysis demonstrated clear differences in predictive behavior among the developed forecasting models across varying lead times (Fig.3). The hybrid LSTM-XGBoost model exhibited the strongest clustering of predicted values around the 1:1 reference line, indicating superior agreement between simulated and observed water levels and lower residual dispersion. The concentration of prediction points remained relatively stable even as forecasting horizon increased, highlighting the robustness of the hybrid framework in preserving temporal relationships under extended prediction conditions. The regression slopes of the hybrid model remained consistently close to unity with comparatively lower intercept deviation, suggesting minimal systematic bias and improved capability in reproducing both low-flow and high-flow hydrological conditions. The model particularly demonstrated enhanced representation of peak water level events, where nonlinear hydrological responses are generally difficult to capture accurately.

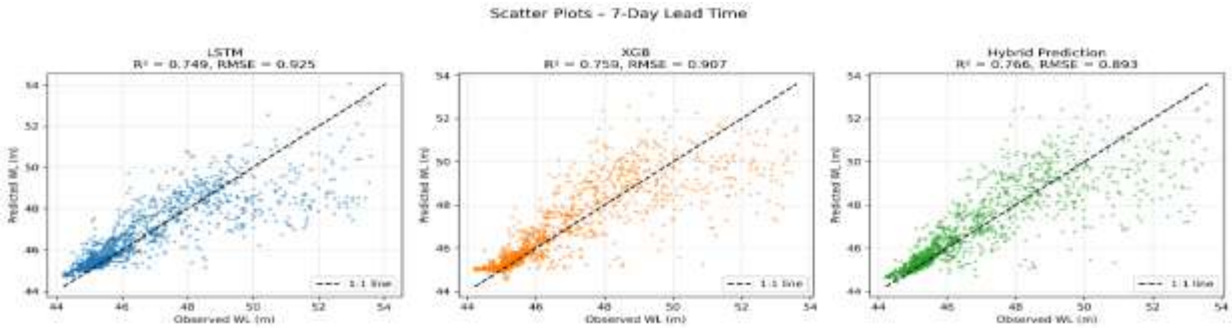
In contrast, the standalone LSTM model showed noticeably wider scatter around the reference line, especially during medium and high magnitude events. The dispersion became progressively larger with increasing lead time, indicating reduced capability in maintaining long-range temporal dependencies and increasing uncertainty accumulation during multi-step forecasting. Underestimation of peak events and occasional overprediction during recession phases suggested that the sequential memory structure of LSTM alone was insufficient to fully capture abrupt hydrological variability and nonlinear transitions in river water levels. Similarly, the XGBoost model exhibited moderate prediction spread under extreme hydrological conditions. Although XGBoost effectively captured nonlinear relationships in the dataset, its comparatively weaker temporal sequence learning capability limited its performance during rapidly varying hydrological events and longer forecasting horizons (Figure 5(a-e)).

a)

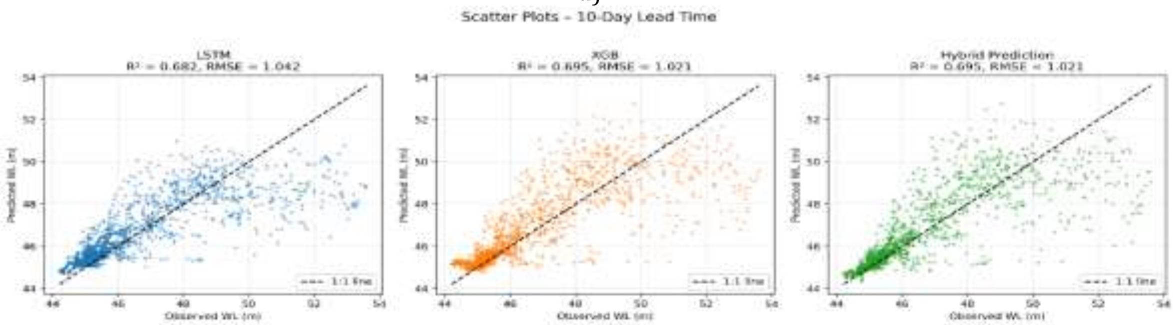


b)

c)



d)



e)

Figure 5 Scatter plot of observed vs predicted water level for different models at a) 1-Day b) 2-Day c) 3-Day d) 4-Day e) 5-Day lead time.

The statistical indicators derived from scatter plot analysis further validated the superiority of the hybrid framework. The hybrid model consistently achieved the highest coefficient of determination (R^2) values across all lead times, indicating stronger explanatory power and improved representation of observed variance in water level dynamics. Higher R^2 values reflected the model's enhanced ability to preserve temporal continuity and nonlinear fluctuations simultaneously. Similarly, the Nash–Sutcliffe Efficiency (NSE) values for the hybrid model remained closest to unity, confirming excellent predictive efficiency and reliable reproduction of hydrological behavior. The relatively lower RMSE and MAE values associated with the hybrid framework further indicated reduced prediction uncertainty and smaller deviation between observed and simulated values. Collectively, these results demonstrate that integrating LSTM with XGBoost substantially improves predictive stability, minimizes systematic forecasting error, and enhances the capability to model complex hydrological dynamics across short-, medium-, and long-term forecasting horizons.

The heatmap Figure 6 clearly demonstrates that the Hybrid Prediction model achieved the best overall forecasting performance across most lead times. Higher KGE, R^2, and NSE values together with lower RMSE and MAE values indicate superior prediction accuracy and stronger agreement with observed water levels. For the 1-day lead time, the hybrid model produced the highest KGE (0.976), R^2 (0.971), and NSE (0.971), while simultaneously achieving the lowest RMSE (0.315) and MAE (0.172), indicating excellent short-term forecasting capability.

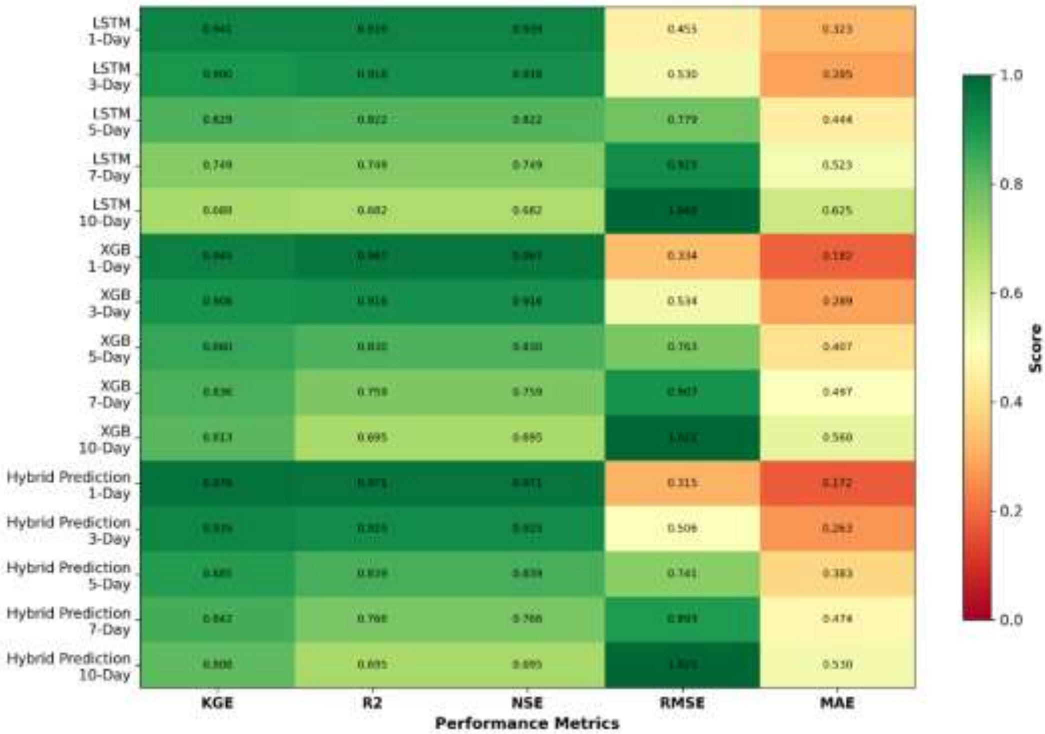


Figure 6 Performance heat map at different lead times

The XGBoost model also demonstrated strong performance, particularly during short- and medium-range forecasting horizons, whereas the standalone Long Short-Term Memory model showed comparatively lower efficiency and higher prediction errors. As lead time increased from 1-day to 10-day forecasts, the performance of all models gradually declined, reflected by decreasing KGE, R^2, and NSE values and increasing RMSE and MAE values. However, the hybrid framework maintained comparatively stable performance even at extended lead times, demonstrating stronger robustness and improved capability in preserving long-range temporal information. The results confirm that combining LSTM with XGBoost significantly enhances forecasting reliability and reduces prediction uncertainty compared to standalone models.

Lead time analysis revealed a progressive reduction in forecasting accuracy with increasing prediction horizons as shown in Figure 7. This trend was observed across all models and evaluation metrics. Short lead times generally produced lower prediction errors and stronger correlations with observed values, while extended lead times introduced higher uncertainty and accumulated forecasting errors. For the 1-day lead time, all models achieved satisfactory prediction performance, indicating strong short-term forecasting capability. The differences among models became more apparent at 5-day and 10-day lead times, where the hybrid model clearly outperformed the LSTM and XGBoost architectures.

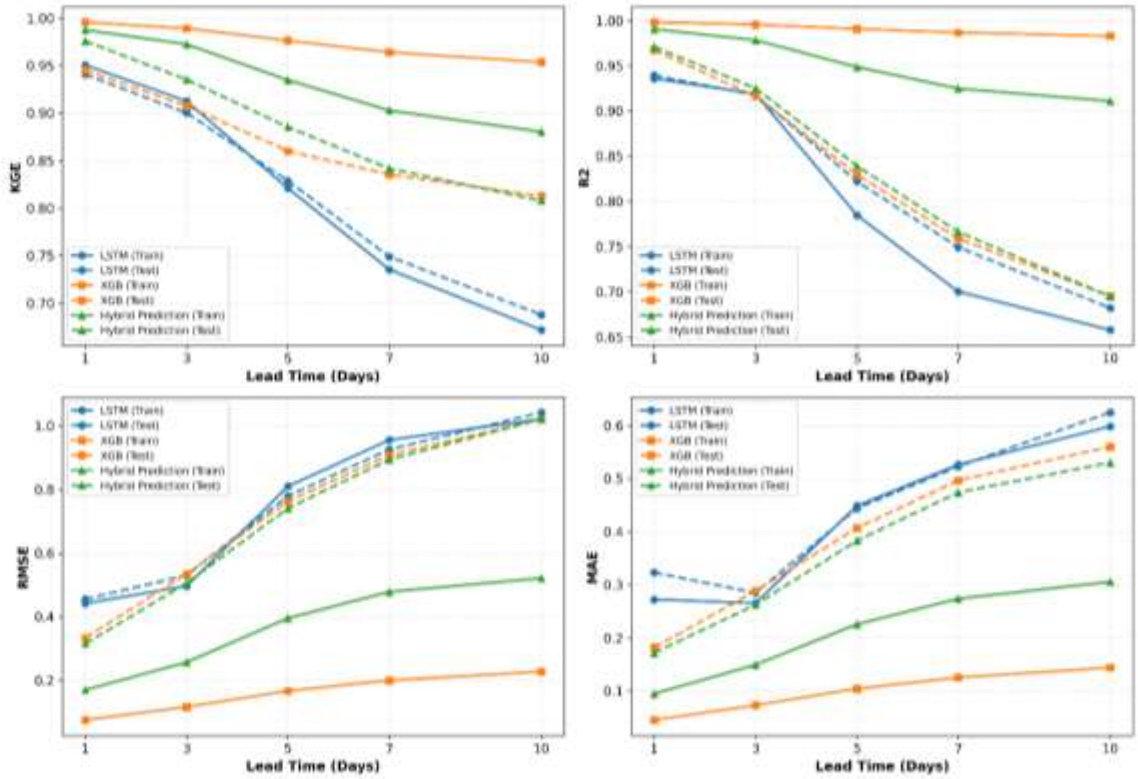


Figure 7 Line plot showing training vs testing comparison of all models

The degradation in performance with increasing lead time is expected in temporal forecasting systems due to the accumulation of uncertainty and reduced influence of recent observations. However, the hybrid model exhibited comparatively slower performance degradation, demonstrating its ability to preserve long-range temporal information more effectively. The line plots comparing RMSE, MAE, and R^2 across different lead times further confirmed this observation. Error metrics increased gradually with lead time, whereas correlation-based metrics declined. The LSTM model experienced the steepest deterioration, suggesting that recurrent memory structures are less efficient for long-range prediction tasks. In contrast, the hybrid model retained stable predictive capability over extended forecasting horizons. These findings indicate that attention-based architectures are more suitable for medium- and long-term forecasting applications where capturing distant temporal relationships is essential.

The comparative evaluation of error metrics across different forecasting lead times further confirmed the superiority of the hybrid Long Short-Term Memory–XGBoost model over the standalone LSTM and XGBoost models. The hybrid framework consistently produced the lowest RMSE and MAE values across nearly all lead times, indicating reduced forecasting uncertainty and improved prediction reliability. For the 1-day forecast horizon, the hybrid model achieved the minimum RMSE (0.32 m) and MAE (0.17 m), compared to RMSE values of 0.45 m and 0.33 m and MAE values of 0.32 m and 0.18 m for the LSTM and XGBoost models, respectively. Similar trends were observed for medium forecasting horizons, where the hybrid model maintained lower error magnitudes, achieving RMSE values of 0.51 m, 0.74 m, and 0.89 m for the 3-day, 5-day, and 7-day forecasts, respectively. At the extended 10-day lead time, all models exhibited increased prediction errors due to uncertainty accumulation; however, the hybrid framework still maintained comparatively lower MAE values (0.53 m) and comparable RMSE performance (1.02 m), demonstrating improved robustness under long-range forecasting conditions (Figure 8).

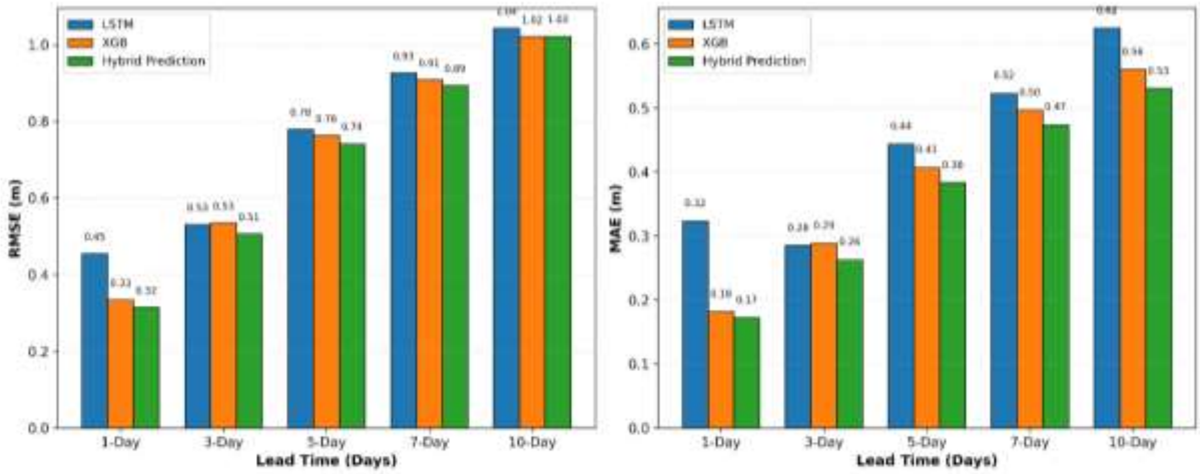


Figure 8 Error metrics comparison across lead times

The gradual increase in RMSE and MAE values with increasing lead time reflects the growing complexity and uncertainty associated with multi-step hydrological forecasting. Nevertheless, the hybrid model exhibited slower performance degradation compared to the standalone models, indicating stronger capability in preserving long-term temporal information and nonlinear hydrological relationships. The improved performance of the hybrid framework can be attributed to the complementary strengths of the constituent models, where LSTM effectively captures sequential temporal dependencies while XGBoost enhances nonlinear feature learning and residual error correction. This integrated learning strategy substantially reduced systematic prediction errors and improved forecasting stability across all lead times.

Overall, the results clearly establish the hybrid LSTM–XGBoost framework as the best-performing model among the evaluated approaches. Its superior statistical accuracy, lower prediction errors, stronger robustness under extended forecasting horizons, and enhanced capability in reproducing hydrological variability demonstrate its suitability for reliable river water level forecasting and operational hydrological applications.

4. CONCLUSIONS, LIMITATIONS AND FUTURE SCOPE

The present study evaluated the performance of standalone Long Short-Term Memory (LSTM), XGBoost, and hybrid LSTM–XGBoost models for river water level forecasting across multiple lead times. The comparative analysis based on KGE, R^2, NSE, RMSE, and MAE demonstrated that the hybrid framework consistently outperformed the individual models under both short- and long-term forecasting conditions. The hybrid model achieved higher predictive efficiency with lower forecasting uncertainty, particularly during 1-day and medium-range forecasts, where it produced the lowest RMSE and MAE values and the highest correlation-based metrics. The results further revealed that integrating sequential temporal learning capability of LSTM with the nonlinear regression strength of XGBoost substantially improved the representation of hydrological variability and reduced systematic prediction errors. Although all models experienced gradual degradation in forecasting performance with increasing lead time, the hybrid framework maintained comparatively stable prediction capability, indicating stronger robustness in preserving long-range temporal dependencies. The study therefore confirms the effectiveness of hybrid deep learning–machine learning frameworks for reliable river water level forecasting and operational hydrological applications.

Despite the promising results, certain limitations remain in the present study. The developed models relied primarily on historical hydrological observations, while several important influencing factors such as rainfall intensity, upstream discharge variability, soil moisture conditions, land use change, and climatic indices were not explicitly incorporated into the modeling framework. This may have contributed to increased forecasting uncertainty during extreme hydrological events and peak flow conditions. In addition, the performance of all models declined at longer forecasting horizons due to uncertainty accumulation and reduced influence of recent observations. Future research may improve forecasting reliability through the integration of additional hydro-meteorological variables, spatial information, and real-time monitoring datasets. Advanced approaches such as ensemble learning, attention-based architectures, physics-informed neural networks, and uncertainty quantification methods may further enhance predictive accuracy and model interpretability. Furthermore, integration of remote sensing datasets and climate projections could support large-scale operational forecasting and climate-resilient water resource management systems.

Acknowledgements

The authors wish to thanks to anonymous reviewers and to all the members of Flood Management Improvement Support Centre, Water Resource Department Bihar, Govt. of Bihar for providing data and other relevant information that adds in facilitating this study.

Conflict of Interest
The authors have no conflicts of interest to declare that are relevant to the content of this article.

Funding
No external funding has been received for this research work.

References

1. Ahmed, A. N., Yafouz, A., Birima, A. H., Kisi, O., Huang, Y. F., Sherif, M., Sefelnasr, A., & El-Shafie, A. (2022). Water level prediction using various machine learning algorithms: a case study of Durian Tunggal river, Malaysia. Engineering Applications of Computational Fluid Mechanics, 16(1), 422-440.
2. Breiman, L. (2001). Random forests. Machine Learning, 45, 5-32.
3. Chaubey, P. K., Srivastava, P. K., Gupta, A., & Mall, R. K. (2021). Integrated assessment of extreme events and hydrological responses of Indo-Nepal Gandak River Basin. Environment, Development and Sustainability, 23(6), 8643-8668.
4. Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).
5. Chen, W., Li, Y., Xue, W., Shahabi, H., Li, S., Hong, H., ... & Ahmad, B. B. (2020). Modeling flood susceptibility using data-driven approaches of naïve bayes tree, alternating decision tree, and random forest methods. Science of The Total Environment, 701, 134979.
6. Cho, K., & Kim, Y. (2022). Improving streamflow prediction in the WRF-Hydro model with LSTM networks. Journal of Hydrology, 605, 127297.
7. Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of hydrology, 377(1-2), 80-91.
8. Hipni, A., El-shafie, A., Najah, A., Karim, O. A., Hussain, A., & Mukhlisin, M. (2013). Daily forecasting of dam water levels: comparing a support vector machine (SVM) model with adaptive neuro fuzzy inference system (ANFIS). Water resources management, 27(10), 3803-3823.
9. Hochreiter, S. (1997). Long short-term memory. Neural Computation MIT-Press. Hu, C., Wu, Q., Li, H., Jian, S., Li, N., & Lou, Z. (2018). Deep learning with a long short-term memory networks approach for rainfall-runoff simulation. Water, 10(11), 1543.
10. Hunt, K. M., Matthews, G. R., Pappenberger, F., & Prudhomme, C. (2022). Using a long short-term memory (LSTM) neural network to boost river streamflow forecasts over the western United States. Hydrology and Earth System Sciences, 26(21), 5449-5472.
11. Konapala, G., Kao, S. C., Painter, S. L., & Lu, D. (2020). Machine learning assisted hybrid models can improve streamflow simulation in diverse catchments across the conterminous US. Environmental Research Letters, 15(10), 104022.
12. Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using long short-term memory (LSTM) networks. Hydrology and Earth System Sciences, 22(11), 6005-6022.
13. Lees, T., Buechel, M., Anderson, B., Slater, L., Reece, S., Coxon, G., & Dadson, S. J. (2021). Benchmarking data-driven rainfall–runoff models in Great Britain: a comparison of long short-term memory (LSTM)-based models with four lumped conceptual models. Hydrology and Earth System Sciences, 25(10), 5517-5534.
14. Mishra, A. K., & Maurya, N. S. (2025). Applicability of random forest regressor for prediction of fluoride in groundwater. National Academy Science Letters, 1-5.
15. Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE, 50(3), 885-900.
16. Prakash, R., & Tripura, J. (2026). Comparative Analysis of Machine Learning Models for Water Level Prediction in the Urban Water Basin. National Academy Science Letters, 1-5.
17. Ranjan, A., & Roshni, T. (2023). Analysis of hydrological alteration and environmental flow in Sone river basin. Acta Geophysica, 71(2), 949-960.
18. Sabzipour, B., Arsenault, R., Troin, M., Martel, J. L., Bissette, F., Brunet, F., & Mai, J. (2023). Comparing a long short-term memory (LSTM) neural network with a physically-based hydrological model for streamflow forecasting over a Canadian catchment. Journal of Hydrology, 627, 130380.
19. Sapitang, M., Wanie, M. R., Kushiar, K. F., Ahmed, A. N., & El-Shafie, A. (2020). Machine learning application in reservoir water level forecasting for sustainable hydropower generation strategy. Sustainability, 12(15), 6121.
20. Shortridge, J. E., Guikema, S. D., & Zaitchik, B. F. (2016). Machine learning methods for empirical streamflow simulation: a comparison of model accuracy, interpretability, and uncertainty in seasonal

watersheds. Hydrology and Earth System Sciences, 20(7), 2611-2628.
21. Tripura, J., Roy, P., & Barbhuiya, A. K. (2021). Simultaneous streamflow forecasting based on hybridized neuro-fuzzy method for a river system. Neural Computing and Applications, 33(8), 3221-3233.
22. Yang, S., Yang, D., Chen, J., Santisirisomboon, J., Lu, W., & Zhao, B. (2020). A physical process and machine learning combined hydrological model for daily streamflow simulations of large watersheds with limited observation data. Journal of Hydrology, 590, 125206.
23. Zakaria, M. N. A., Malek, M. A., Zolkepli, M., & Ahmed, A. N. (2021). Application of artificial intelligence algorithms for hourly river level forecast: A case study of Muda River, Malaysia. Alexandria Engineering Journal, 60(4), 4015-4028.
24. Zhang, J., Zhu, Y., Zhang, X., Ye, M., & Yang, J. (2018). Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas. Journal of hydrology, 561, 918-929.
25. Zhao, G., Pang, B., Xu, Z., Yue, J., & Tu, T. (2018). Mapping flood susceptibility in mountainous areas on a national scale in China. Science of the Total Environment, 615, 1133-1142.