HYDRA-LSTM-GRU: Self-Attention-Enhanced Rainfall-Runoff Modelling for Indian River Basins

Sagar S Lachure[1,2*], and Ashish Tiwari[1]

[1]CSE Department, Visvesvaraya National Institute of Technology (VNIT), Nagpur, India
[2]CE & IT Department, Veermata Jijabai Technological Institute (VJTI), Mumbai, India
Email: sagarlachure@gmail.com (S.S.L.); at@cse.vnit.ac.in (A.T.)
*Corresponding author

Manuscript received November 29, 2024; revised February 24, 2025; accepted March 11, 2025; published December 17, 2025

***Abstract*—Floods are unpredictable natural disasters, often triggered by extreme weather conditions, particularly in countries like India. This study utilises machine learning to predict runoff, with a focus on attention-based models, including Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) architectures. We compare Attentional LSTM, PCA-based LSTM, Self-Attention LSTM, and Self-Attentional GRU models across various window lengths (1, 7, 14, and 30 days). The Attentional LSTM model consistently outperforms others, achieving the lowest Mean Squared Error (MSE=0.000476) and Mean Absolute Error (MAE=0.015475) while attaining the highest Nash-Sutcliffe Efficiency (NSE=0.978). A 14-day window size provides the best balance between predictive accuracy and computational efficiency. While other attention-based models, such as the Self-Attentional GRU (MSE=0.000654, MAE=0.018878, NSE=0.970), perform well, simpler models without attention mechanisms show significantly lower accuracy. Additionally, incorporating Principal Component Analysis (PCA) does not enhance performance, as the PCA-Based Attention LSTM (MSE= 0.000855, MAE= 0.022152, NSE=0.961). The model underperforms compared to the Attentional LSTM. Explainable AI (XAI) techniques such as LIME are used to analyse model behaviour, highlighting the importance of attention mechanisms in improving time-series predictions for hydrological applications.**

***Keywords*—PCA, Attention LSTM, LSTM-GRU-ATT, NSE, forecasting**

I. INTRODUCTION

Accurate forecasting of rainfall-runoff has become increasingly crucial in recent years due to the growing impact of climate change on river basins worldwide [1]. While traditional hydrological models are valuable, they often fail to capture the intricate nonlinear interactions inherent in hydrological systems [2]. The effectiveness of deep learning models, particularly Recurrent Neural Networks (RNNs) such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) [3], in addressing these challenges stems from their ability to model time-series data efficiently.

This study focuses on Indian river basins, where the variability in climate and hydrological patterns poses significant challenges for accurate rainfall-runoff predictions [4–8]. The LSTM-GRU model proposed in this study integrates LSTM and GRU architectures to enhance prediction accuracy. Additionally, attention mechanisms are incorporated to improve model performance further. The Attention framework, initially developed for natural language processing [9], has demonstrated significant improvements in approximating complex relationships in sequential data.

A comprehensive review of existing literature indicates that integrating the LSTM-GRU with an attention mechanism presents a novel and promising approach for application in Indian river basins. This study also reexamines Principal Component Analysis (PCA) to assess its impact on model performance. Given the diverse climatic conditions and socio-economic significance of Indian river basins, accurate runoff predictions are essential for effective water resource management, flood prevention, and mitigation strategies.

This work evaluates the effectiveness of the HYDRA-LSTM-GRU model in specific Indian river basins, ensuring that the study remains focused and practical. The study compares existing deep learning models and highlights their capability to enhance hydrological predictions. The primary objectives of this work are outlined below:

- **To design and develop a novel PCA-based self-attentional LSTM with a Softmax layer.** This involves creating a hybrid model combining PCA and LSTM with self-attention mechanisms to enhance the model's ability to capture complex temporal dependencies and nonlinear relationships in rainfall-runoff processes specific to Indian river basins.
- **To design and implement a self-attention-enhanced LSTM-GRU model.** This involves integrating LSTM and GRU networks with self-attention mechanisms to enhance the model's ability to capture intricate temporal dependencies and nonlinear relationships in rainfall-runoff processes within Indian river basins.
- **To customise the model for the Indian River basins.** The model is adapted to account for the unique hydrological characteristics and seasonal variations of Indian river basins, including the dynamics of the monsoon. It will be validated using historical rainfall and runoff data from specific Indian river basins to ensure accuracy & reliability.
- **The goal is to optimise the model for predictive performance and scalability.** The model parameters are fine-tuned to maximise predictive accuracy, with a focus on forecasting runoff during extreme weather events. The model will also be scaled for application across multiple Indian river basins, taking into account regional hydrological variability and data availability.
- **To apply LIME XAI for model interpretability.** Explainable AI (XAI) techniques, such as LIME, will be used to enhance the transparency of model predictions and identify the key parameters that contribute most to the prediction process.

A. *Related Work*

Introduction to the Modelling of Rainfall-Runoff. Iterative rainfall-runoff modelling is essential in water resource management, especially in areas susceptible to severe weather phenomena and climate change [10, 11]. Many conventional hydrological models heavily rely on physical processes and require large datasets to achieve precision. Nonetheless, these models face constraints due to the intricate nature of natural systems and the accessibility of granular data [12]. Data-driven paradigms, including deep learning models, have gained significant importance in recent years due to their ability to capture intricate, nonlinear correlations in time series data without requiring a comprehensive set of input variables [13]. Recurrent Neural Networks (RNNs) specifically develop Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRU) to process sequential data in hydrological modelling domains [14]. Hydrology has widely employed Extended Support Vector Machine (LSTM) networks, showcasing significant improvements in runoff prediction across various basins [15]. According to Kratzert *et al.* [16], LSTM showed superior performance compared to conventional models in forecasting streamflow in several basins in the United States. They attributed this to its ability to manage long-term dependencies in hydrological data. Furthermore, Ma *et al.* [17] and Chen *et al.* [18] demonstrated that GRU, a streamlined version of LSTM, has achieved similar results, especially in situations with insufficient data. In the Shaxi River Basin, they used GRU to forecast river runoff. Modelling Self-Attention in Time Series Forecasting, Vaswani *et al.* [19] first used the Self-Attention (SA) mechanism for tasks in natural language processing. It has shown much promise in making deep learning models more useful by capturing how things depend on each other over time. This approach enables the model to focus solely on relevant segments of the input sequence, resulting in improved prediction accuracy. Adding SA to LSTM and GRU models for hydrological purposes has led to the creation of hybrid architectures, such as LSTM-SA and GRU-SA [20].

Yen *et al.* [21, 22], Chen *et al.* [23], Mohammed *et al.* [24] and Garg *et al.* [25] conducted research demonstrating that incorporating SA into these models can significantly improve the accuracy of runoff forecasts by deepening the model's understanding of the time-dependent interactions between rainfall and runoff.

Implementation in India's river basins poses distinct challenges for rainfall-runoff modelling due to their varied meteorological conditions, diverse topographies, and socioeconomic importance. Prediction models with high accuracy are essential for efficient water management, flood prevention, and mitigation measures [26, 27]. Despite significant advancements in modelling methodologies, Indian settings have limited the use of hybrid deep- learning models. The HYDRA-LSTM-GRU model is introduced to bridge this gap, integrating LSTM, GRU, and SA processes to enhance prediction skills in complex settings. Incorporating LSTM, GRU, and self-attention mechanisms into a hybrid model marks a significant breakthrough in hydrological modelling. Specially designed for the Indian River Basins, the HYDRA-LSTM-GRU model can provide more precise and dependable forecasts, thereby assisting in the efficient management of water resources amid growing climatic fluctuations. We also proposed our novel PCA-based LSTM-Attention structure, backed by a softmax, to compare the results with those of our proposed model.

Research Gap: Although prior research has demonstrated the effectiveness of deep learning models for hydrological forecasting, significant gaps remain in their application to Indian river basins. Most studies have focused on individual deep learning architectures without exploring hybrid models combining LSTM, GRU, and self-attention mechanisms. Moreover, the integration of Principal Component Analysis (PCA) for dimensionality reduction has not been extensively evaluated in conjunction with self-attentional architectures. Existing studies also lack model interpretability techniques such as Explainable AI (XAI) to analyse prediction behaviour. Our study aims to fill these gaps by developing a robust hybrid model tailored to the Indian hydrological landscape, optimising it for real-world applications, and utilising XAI techniques like LIME to enhance transparency.

The rest of the paper is organised as follows: Section II discusses the methodology in detail, including dataset preparation and proposed methodology. Section III presents the results and discussion, and Section IV concludes this article with a discussion of future work.

II. Methodology

This Section discusses the study area, dataset preparation, and proposed methodology in detail.

A. Study Area and Dataset Description

This study focuses on the Krishna River, one of the major rivers in the Indian river basin, which flows through the states of Maharashtra, Karnataka, Telangana, and Andhra Pradesh. The Krishna River is of immense hydrological and socio-economic significance, supporting agriculture, drinking water supply, and hydroelectric projects. The river is regulated by several reservoirs and dams, including the Almatti Dam, Tungabhadra Dam, and Nagarjuna Sagar Dam, which influence its runoff and discharge characteristics.

Fig. 1. Study area map showing the Krishna River basin and locations of meteorological and hydrological stations.

The dataset used in this study is sourced from the Dartmouth Flood Observatory (DFO) River and Reservoir Watch system (GFDS Area ID 2011), located at a latitude of 16.021° and longitude of 80.21°, in the Krishna Basin. The dataset includes essential hydrological variables such as river flow, 7-day runoff, flood intensity, and historical discharge patterns. To enhance predictive modelling, meteorological data from the Indian Meteorological Department (IMD), covering temperature and precipitation from January 1998 to January 2023, have been integrated. The combined dataset allows for comprehensive modelling of rainfall-runoff dynamics, taking into account meteorological influences.

To ensure the robustness of our analysis, a study map (Fig. 1) has been created, highlighting the locations of meteorological and hydrological stations used for data collection. This map provides insights into the spatial distribution of data sources and their potential influence on model predictions.

B. Dataset Preprocessing

Missing meteorological data can introduce biases and reduce the reliability of runoff predictions. To address this, we employed statistical imputation techniques to fill data gaps before training the model. Specifically, missing values were handled using linear interpolation for continuous gaps and mean imputation for sporadic missing entries:

- Linear Interpolation: Given a missing value at time step *t*, it was estimated using adjacent known values:

Advanced Methods for Large Gaps: In cases of extended missing periods, we experimented with Multiple Imputation using k-nearest neighbours (KNN), where missing values were estimated based on the weighted average of the k most similar observations. This method ensures smooth transitions in time-series data without introducing artificial variations.

$$x_t = x_{t-1} + \frac{(x_{t+1} - x_{t-1})}{(t+1-(t-1))} \times (t - (t-1)) \quad (1)$$

This method ensures smooth transitions in time-series data without introducing artificial variations.

- Mean Imputation: For minor, randomly missing data points, we used the mean of the respective feature over a defined period:

$$x_t = \frac{1}{N} \sum_{i=1}^{N} x_i \quad (2)$$

where N is the total number of available values in that feature and similar observations in the dataset.

C. Forecasting and Model Inputs

The hydrology stations used in this study are influenced by reservoirs and dams, which regulate water flow and impact natural runoff patterns. To account for these variations, we analyse both upstream and downstream stations, distinguishing between natural and regulated flow conditions.

The selection of input variables is based on their direct influence on runoff prediction. We use four key meteorological and hydrological variables:

- **Rainfall** (mm): Primary driver of surface runoff.
- **Minimum Temperature (Tmin)** (°C): Affects evapotranspiration rates.
- **Maximum Temperature (Tmax)** (°C): Influences hydrological processes such as snowmelt (if applicable).
- **Runoff/Discharge** (m^3/s): Target variable for prediction.

These variables were selected due to their strong correlation with runoff generation, as confirmed by preliminary statistical analysis and knowledge of the hydrological domain. While additional parameters such as soil moisture, wind speed, and relative humidity could provide further insights, their availability and reliability in historical datasets remain limited.

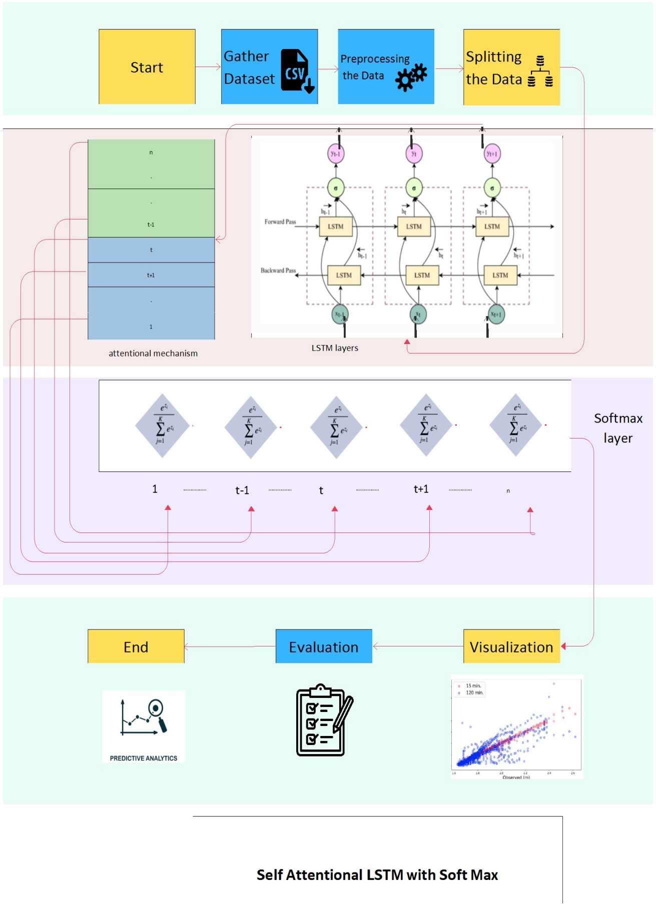

Fig. 2. PCA-Based LSTM-Attention model with a Softmax layer, as proposed.

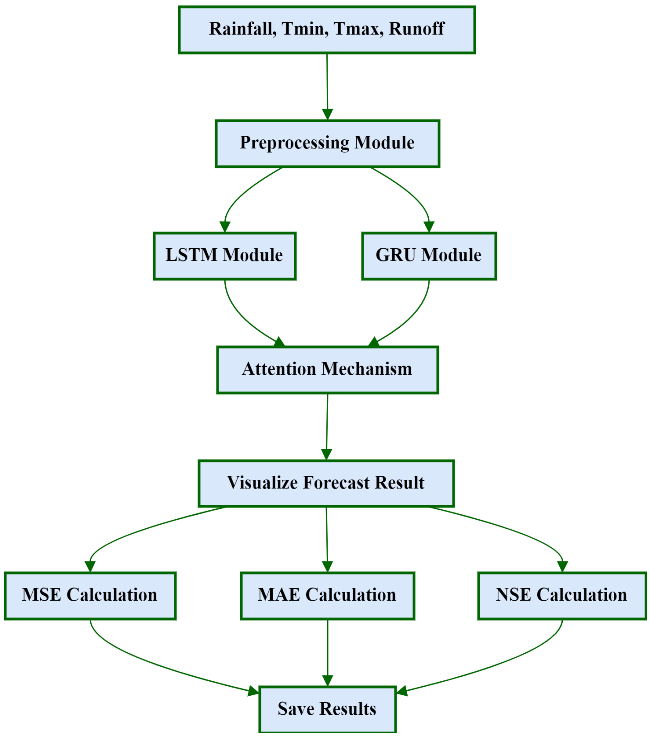

Fig. 3. LSTM-GRU-Attention model overall working flow.

In Fig. 2, we present the PCA-based LSTM-Attention model enhanced with a Softmax layer. This model is designed to accurately predict discharge by first processing the input data through Principal Component Analysis (PCA) to reduce its dimensionality. The resulting components are then passed through the Long Short-Term Memory (LSTM) layer, which captures the temporal dependencies of the sequence data. The final output is generated by the Softmax layer, as shown in the figure, which enables classification or regression based on the processed features. This model's architecture is beneficial for time-series prediction tasks, where capturing both long-term dependencies (via LSTM) and focusing on the necessary time steps (via the attention mechanism) are crucial for accurate forecasts. Following this, we introduce the hybrid LSTM-GRU-Attention model, as illustrated in Fig. 3. In this model, we preprocess the input data by passing it through both the LSTM and GRU modules in parallel, thereby capturing different aspects of the temporal relationships.

Hidden states from both models are combined and sent through an attention module, which helps focus on the most relevant parts of the sequence. The output is then processed using various evaluation metrics to evaluate the model's performance. The combination of LSTM, GRU, and attention mechanisms in these models enables a comprehensive approach to modelling complex, time-dependent data, thereby improving predictive accuracy by focusing on essential features and learning from both short-term and long-term dependencies.

D. Maximum Likelihood Estimation Using OLS

Table 1. Results of OLS regression. The dependent variable is daily runoff

	Coefficient	Std. Error	t-Statistic	P-value
const	0.0005	0.000	1.365	0.172
Avg-Dis	0.0003	1.15e-08	2.91e+04	0.000
rain	-6.74e-06	3.54e-06	-1.903	0.057
tmin	1.482e-06	1.45e-05	0.102	0.919
tmax	-1.514e-05	1.49e-05	-1.017	0.309
R-squared	1.000			
Adjusted R-squared	1.000			
F-statistic	2.786e+08			
Prob(F-statistic)	0.00			
Log-Likelihood	40818.00			
AIC	-8.163e+04			
BIC	-8.159e+04			
Durbin-Watson	1.908			
Omnibus	2869.982			
Prob(Omnibus)	0.000			
Jarque-Bera (JB)	425.633			
Prob(JB)	3.76e-93			
Skew	0.015			
Kurtosis	1.943			
Cond. No.	4.66e+04			

The Ordinary Least Squares (OLS) regression analysis results show a strong fit, with an R-squared value of 1.000, indicating that the model explains nearly all the variability in 'Daily Runoff'. However, this high value suggests potential overfitting. The F-statistic is large, with a p-value of 0.00, confirming the significance of the predictors shown in Table 1. The coefficient for 'Avg-Dis' (0.0003) is highly significant (p<0.001), while 'rain' has a slight negative impact on 'Daily Runoff' coefficient (i.e.,6.74e-06) and is marginally insignificant (p=0.057). The temperature Variables ('tmin' and 'tmax') have minimal impact and are statistically insignificant. The Durbin-Watson statistic (1.908) suggests no significant autocorrelation in the residuals, but the Omnibus and Jarque-Bera tests indicate non-normality. The model's high Condition Number (4.66e+04) also suggests possible multicollinearity, indicating that the model may require refinement to address issues related to overfitting, multicollinearity, and residuals.

E. Variance Inflation Factor

The Variance Inflation Factor (VIF) analysis shows no significant multicollinearity among the predictors in the regression model. 'Avg-Dis', 'rain', 'tmin', and 'tmax' all have low VIF values, indicating stable and reliable estimates. The high VIF for the constant term ('const') is not unusual but may suggest potential scaling issues. Data centring could help reduce the VIF for the constant without affecting the interpretation of other variables. Overall, the model is free from serious concerns regarding multicollinearity.

F. Mathematical Modelling

Given inputs rainfall_t, $t_{\min_t}$, and $t_{\max_t}$ at time step t, we proceed with the following steps:

- **Feature Preparation:** Combine these inputs into a feature vector x_t for each time step.

$$x_t = \text{rainfall}_t,\ t_{\min},\ t_{\max} \tag{3}$$

- **LSTM/GRU Update:** Use the LSTM or GRU model to update the hidden state h_t and output $Y_t = \text{avgdischarge}_t$.
- **Attention Mechanism (if applied):** Use attention scores to weigh the importance of past states or features.

Training these models on historical data will enable them to learn the relationships between rainfall, temperature, and discharge, allowing for predictions of average discharge based on future input conditions.

LSTM: The LSTM updates the hidden state h_t and cell state c_t based on input x_t as:

$$ht, ct = \text{LSTM}(xt, ht-1, ct-1) = ot \odot \tanh(ft \odot ct-1 + it \odot c\tilde{}t) \tag{4}$$

where f_t, i_t, $c\tilde{}_t$, and o_t are the forget, input, candidate, and output gates, respectively.

GRU: The GRU updates the hidden state h_t based on input x_t as:

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tanh\,(W_h \cdot [r_t \odot h_{t-1}, x_t] + b_h) \tag{5}$$

where z_t, r_t, and $\tilde{h}_t$ are the update gate, reset gate, and candidate hidden state, respectively.

Attention Mechanism: The attention mechanism calculates the context vector c based on query q, key K, and value V as:

$$c = \sum_{t=1}^{T} \frac{\exp\,(q.k_t)}{\sum_{t'} \exp\,(q.k_{t'})}\ .v_t \tag{6}$$

Interaction of Input Variables: Given input features $X_t = [\text{rainfall}_t,\ t_{\min t},\ t_{\max t}]$ and target variable $Y_t = \text{avgdischarge}_t$, the LSTM, GRU, and attention models use these inputs to predict the target as:

$$Y_t = Model(X_t) \tag{7}$$

where the model can be an LSTM, GRU, or an attention-based mechanism.

Algorithm 1: PCA-Based Attentional LSTM

Input: Sequence $X=[x_1,x_2,\ldots,x_T]$
Initialise: LSTM weights, PCA parameters
for each time step t **do**
Compute LSTM cell states c_t and hidden states h_t
Obtain attention vectors v_t from LSTM outputs
Apply PCA to reduce v_t to v_t'
Compute attention scores α_t and context vector c_t using v_t'
Update LSTM hidden state incorporating c_t
end for
Output: Final hidden state h_T

Algorithm 1 combines a PCA-based approach with LSTM and attention mechanisms. The PCA step reduces the dimensionality of the hidden states from the LSTM, and the attention mechanism helps focus on the most relevant time steps in the sequence. By incorporating both PCA and attention, the model is expected to improve its ability to predict discharge with reduced noise and greater interpretability.

Algorithm 2: LSTM-GRU-Attentional Model

Input: Sequence $X=[x_1,x_2,\ldots,x_T]$
Initialise: LSTM, GRU, and attention weights
for each time step t **do**
Compute LSTM hidden state h_t^L
Compute GRU hidden state h_t^G
Combine LSTM and GRU hidden states
$h_t^{Combined}=W_1 \cdot h_t^L + W_2 \cdot h_t^G$
Compute attention scores α_t
Compute context vector c_t using h_t
end for
Output: Final attention-weighted context vector c_T

Algorithm 2 uses both LSTM and GRU models in parallel. The hidden states from both models are combined to form a" combined" hidden state, which is then passed through an attention mechanism to assign importance to different time steps. The result is an attention-weighted context vector, which is used for making predictions about average discharge. This model leverages the strengths of both LSTM and GRU, utilising attention to capture complex relationships in time-series data.

III. Result

In this section, we discuss the results and present the model based on the preceding literature. Furthermore, we conduct a range of experiments related to various models. We experimented with multiple parameters in both LSTM and GRU, adjusting specific hyperparameters to optimise the model's performance in terms of both speed and accuracy.

Additionally, we tested the model with all parameters, performed correlation analysis, and experimented with both all and selected parameters. As a result of these activities, we discovered that the results did not vary significantly, leading us to refine our entire methodology and the parameters we use for our research.

The Attentional LSTM typically exhibits the lowest Mean Squared Error (MSE), suggesting its superior predictive accuracy 5. While the PCA-based LSTM Self-Attention model delivers satisfactory results, the Self-Attentional GRU model encounters difficulties when dealing with bigger window widths. Once again, the Attentional LSTM demonstrates the lowest Mean Absolute Error (MAE), indicating that it has the lowest average prediction errors. Significant degradation in the performance of the Self-Attentional GRU is observed as the window size increases. The Attentional LSTM model achieves the highest NSE values, suggesting a strong ability to closely approximate observed data. Although the other models generally perform satisfactorily, the Attentional LSTM consistently surpasses them. Based on the above data, it can be concluded that the Attentional LSTM model with a 14-day window size is the most effective. Due to its ability to balance precision and resilience over various window lengths, it is a very suitable option for the specified application.

A. Model Comparison

The Attentional LSTM model significantly outperforms all other models in forecasting runoff within a 14-day timeframe, achieving the lowest Mean Squared Error (MSE), Mean Absolute Error (MAE), and the highest Normalised Squared Error (NSE). Machine learning models that include attention mechanisms, such as Self-Attentional GRU and LSTM-GRU-ATT, also show outstanding performance, underscoring the significance of these mechanisms in capturing temporal relationships. In contrast, models lacking attention mechanisms, such as LSTM and GRU without attention, demonstrate notably inferior performance, underscoring their limitations in effectively forecasting runoff (Figs. 4 & 5). To summarise, the Attentional LSTM and other attention-based models are optimal for predicting runoff in this application.

The LIME (Fig. 6) analysis provides a clear explanation of the model's forecast for a specific hazard. The anticipated value is 0.08, which is significantly influenced by two Principal Component Analysis (PCA) features: PCA Feature 1 and PCA Feature 2. Both characteristics contribute negatively to the forecast, indicating that the projected runoff declines as their values decrease. PCA Feature 1 has a more pronounced negative impact as its value of -0.35 is much lower than the threshold of -0.17. This considerable contribution makes it a crucial factor in reducing the estimated runoff. Analysing these PCA features is beneficial for comprehending their significance in the model's decision-making process.

B. Comparison with Existing Mode

The Tang. *et al.* [5] proposed BPX-TCN, with Kling-Gupta efficiency reaching its maximum value of 0.91. BPX-TCN-attention has resulted in improved flood simulations and multi-step predictions. The [6] obtained an LSTM model with an RMSE of 0.05 and an MAE of 0.007. The model forecasted flood runoff for periods of one day and one week. When Ojha *et al.* [7] utilise all available input data (12 years of daily precipitation, weather temperature, and wind speed), along with 64 hidden units, the LSTM model structure achieves maximum accuracy. Only 12 years of daily rainfall were sufficient to train a model with excellent performance (NSE=0.84). The LSTM-SS framework, as given by Khurana *et al.* [9], outperforms benchmark models for runoff forecasts. The Nash-Sutcliffe model's median efficiency for 1-day-ahead runoff is 0.85. In one of the papers [8], the Random Forest model surpasses other machine learning models in predicting rainfall-runoff. The RF model received the highest NSE average score of 0.795. We have compared these models to our suggested model and found them superior. That is why it is not possible to draw any firm conclusions about the superiority of the proposed model just by comparing it against a small subset of models.

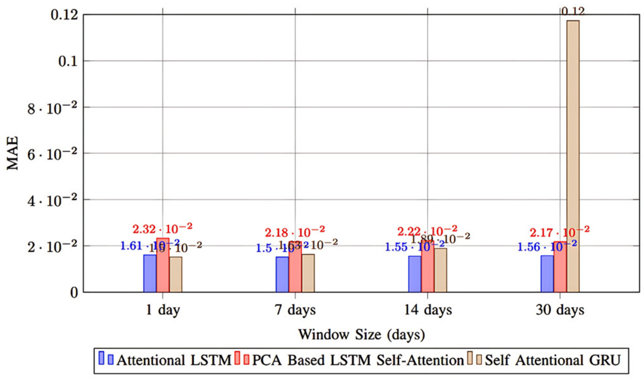

Fig. 4. Comparison of MAE for different models across window sizes.

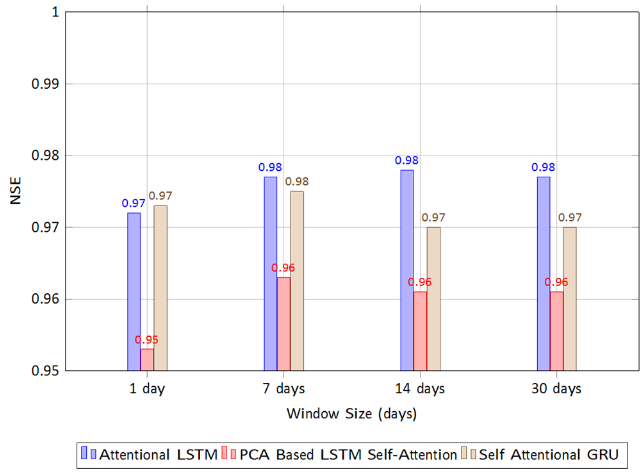

Fig. 5. Comparison of NSE for different models across window sizes.

C. Proposed Model Result

Above, you can see a graphic displaying the findings from the proposed model. Figs. 7, 8, 9, and 10 illustrate the various forms of the results discussed in this section. The bar charts in Figs. 7 and 8 compare the actual (blue) values with the expected (orange) values. The overfitting problem makes the projected data spikes visible in the fit, even though both models attempt to forecast the actual data. To provide a better understanding, we have included the suggested model output for both seven and fifteen days, along with the model's accuracy for each specific length. Figs. 7 and 8 present the rainfall-runoff forecast results, a critical aspect of hydrological modelling. Figs. 9 and 10 show the predicted runoff flow values. However, these values appear smaller than expected, potentially due to unit discrepancies, as runoff flow is measured in millimetres. Verification of conversion factors and scaling in the dataset is necessary to ensure an accurate representation of the data.

Suppose we experiment with the model and compare the results. In that case, we will be able to understand the impacts of using all the parameters together and those of using some of them separately. This work's fundamental discovery is that the principal component analysis (PCA) with all variables and the PCA with four variables produce the same findings. We discovered this by comparing the results of numerous experiments conducted in this particular instance. We can conclude that the suggested rainfall-runoff model outperforms the other models under test. Furthermore, we found that the proposed model significantly outperforms existing models in terms of generating MSE, MAE, and NSE values.

Table 2. Comparison of models for runoff prediction with a 14-day window size

Model	MSE	MAE	NSE
Attentional LSTM	0.000476	0.015475	0.978
PCA-Based Attention LSTM	0.000855	0.022152	0.961
Attentional GRU	0.000654	0.018878	0.970
LSTM (no attention)	0.014641	0.093	0.880
GRU (no attention)	0.015625	0.095	0.860
LSTM-GRU-ATT	0.000729	0.020	0.980

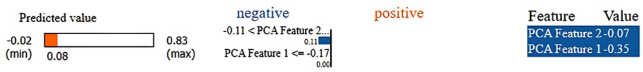

Fig. 6. Predictions of the LSTM-GRU-attention model for different forecast horizons.

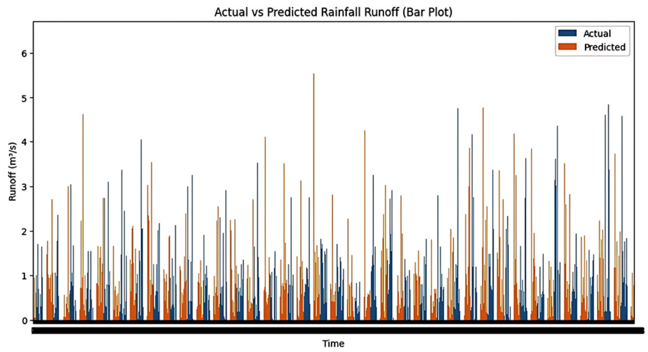

Fig. 7. LSTM-attention model output for window size=14

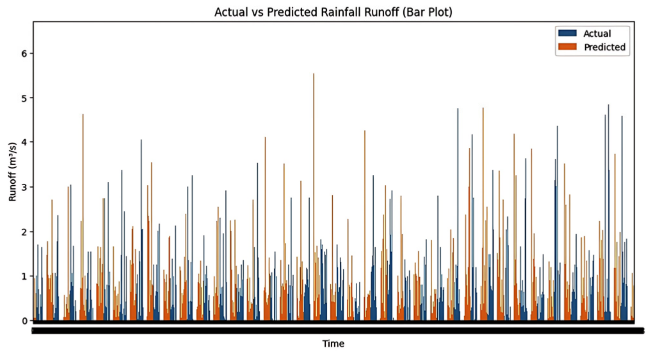

Fig. 8. PCA-based LSTM-attention model output for window size=14.

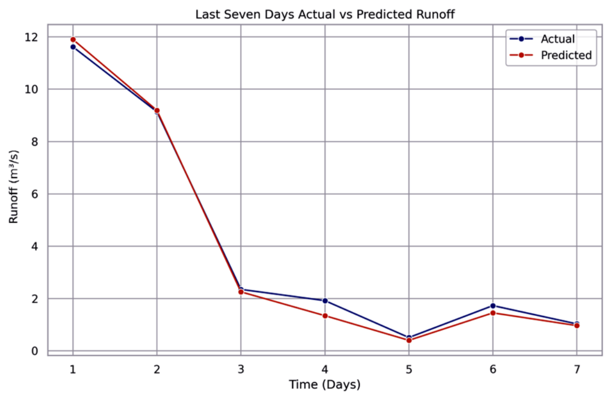

Fig. 9. LSTM-GRU-attention model seven-day prediction.

Our experiments demonstrate that the hyperparameters and other settings of our model outperform existing models in terms of processing time and resource usage. We also presented the windowing strategies to enable a deeper level of comprehension. During our experimentation with the window size, we discovered some intriguing data. We found that most models exhibit underfitting when the window size is one, but practically all models exhibit overfitting when the window size is thirty. To achieve the best possible results, window sizes of seven and fourteen are considered the optimum. Table 2 represents the output of all models with a window size of fourteen. Before employing this model, use the regression and IVF methods to analyse the data and identify potential risks, such as multicollinearity and other factors. Ultimately, we aim to gather additional information about the model by utilising the LIME model of XAI to enhance the level of trust in the provided model.

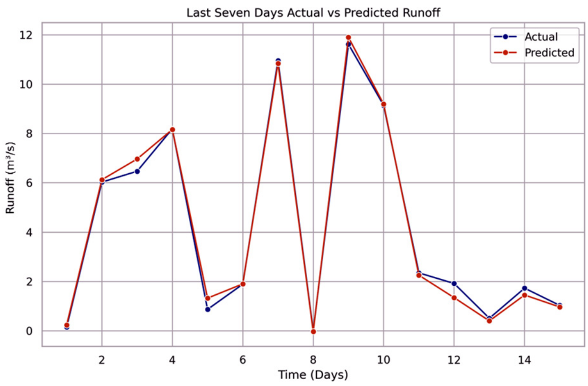

Fig. 10. LSTM-GRU-attention model fifteen-day prediction.

The Attentional LSTM achieves the lowest Mean Absolute Error (MAE), with values ranging from **0.0150 to 0.0161**, confirming its minimal average prediction errors. In contrast, the PCA-Based LSTM Self-Attention exhibits higher MAE values (**0.0217–0.0232**), while the Self-Attentional GRU struggles at larger window sizes, particularly at **30 days** (**MAE=0.1173**). The Attentional LSTM also achieves the highest Nash-Sutcliffe Efficiency (NSE), with a peak value of 0.978 at the 14-day window, demonstrating its superior ability to approximate observed data accurately. The results confirm that a **14-day window size** provides the best trade-off between precision and resilience.

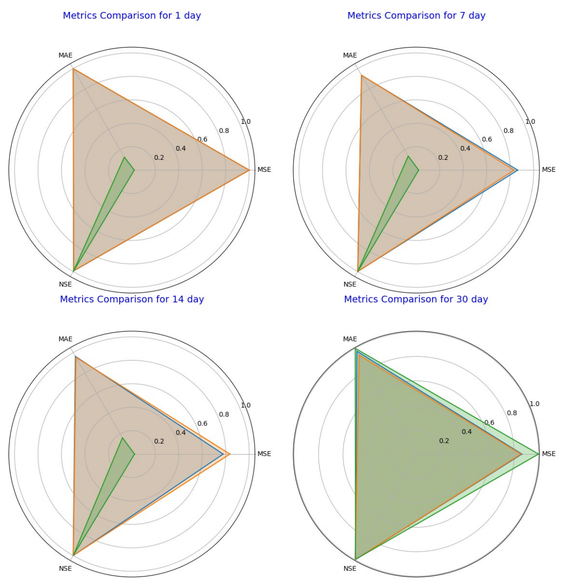

Fig. 11. Radar plot of all the metrics for all the window sizes.

D. Comparison of all Model Results

Fig. 11 presents radar charts that compare model performance measures across several window sizes: 1 day, 7 days, 14 days, and 30 days. Each radar map features three axes that denote the metrics: NSE, MSE, and MAE. Comparison of Metrics by Window Size-Each subplot represents a distinct window size (e.g., 1 day, 7 days). A larger filled area in the radar plot indicates superior overall model performance. Representation of Axes: NSE: Assesses the model's predictive accuracy relative to the mean of the observed data. A value approaching 1 signifies superior performance. MSE: Represents the mean squared error between anticipated and actual values. Reduced values signify superior performance. MAE: Denotes the average magnitude of mistakes. Reduced values are preferable. Analysing the Form and Surface Area: The orange and green polygons juxtapose two models or configurations. For extensive window sizes (e.g., 30d), the green region appears to prevail, indicating consistent or enhanced performance across all criteria. The 1-day plot exhibits a comparatively minor region, potentially signifying diminished predictive efficacy for short-term forecasting. As the window size expands, the occupied area in the radar map increases, indicating enhanced performance for extended forecasts.

E. Discussion

The results demonstrate that the Attentional LSTM model surpasses alternative architectures in runoff prediction. It consistently maintains the highest NSE (**0.978**) while minimising MAE (**0.0150–0.0161**) across different window sizes. These findings highlight the importance of attention mechanisms in capturing intricate temporal dependencies, thereby improving hydrological forecasting accuracy.

A comparison with prior studies further validates the superiority of our approach. Tang *et al.* [5] introduced the BPX-TCN model, achieving a maximum Kling-Gupta Efficiency (KGE) of **0.91**. However, our Attentional LSTM outperforms this, with an NSE consistently exceeding **0.97**. Similarly, Garg *et al.* [25] developed an LSTM model with an RMSE of **0.05** and an MAE of **0.007**; however, their study was limited to short-term forecasts (1-day and 7-day). In contrast, our model maintains strong predictive performance over extended periods. Kim *et al.* [6] applied LSTM models to daily precipitation data, achieving an NSE of **0.84**, which is significantly lower than the performance of our model. Likewise, Yin *et al.* [8] proposed LSTM-SS, reporting a median NSE of **0.85**, which reinforces the notion that attention mechanisms and proper hyperparameter tuning further improve results. Ojha *et al.* [7] found that Random Forest achieved an NSE of only **0.795**, highlighting the advantages of deep learning-based methods in handling nonlinear hydrological patterns.

Despite these promising results, some limitations exist. The effectiveness of our model depends on the quality of the data, as missing values and noise can introduce uncertainty. Preprocessing methods, such as regression-based imputation and inverse variance weighting (IVW), are recommended to mitigate these risks. Additionally, attention mechanisms increase computational complexity, requiring higher processing power. Future work should explore lightweight architectures or hybrid approaches that integrate deep learning with physics-based hydrological models for improved efficiency. Another challenge is explainability. While we employed the LIME model from Explainable AI (XAI) to interpret predictions, further efforts are needed to align deep learning outputs with domain knowledge in the hydrological context.

Future studies should incorporate climate change variables, land-use data, and soil moisture levels to enhance model generalisation. Additionally, testing the model across different regions will validate its adaptability. Exploring hybrid models that combine deep learning with traditional hydrological modelling techniques could further improve prediction reliability. To strengthen future work, systematic hyperparameter tuning can be performed using Bayesian Optimisation or Genetic Algorithms to explore optimal parameter combinations efficiently. Additionally, sensitivity analysis can be conducted to assess the impact of different hyperparameter settings on model performance metrics. This discussion can be placed in the Methodology section under a new subsection, Hyperparameter Selection and Sensitivity Analysis, or in the Discussion section to highlight potential improvements.

F. Model Interpretability and Practical Usability

While LIME provides local interpretability by explaining individual predictions, deep learning models remain inherently complex, making it challenging for hydrologists to understand their decision-making process fully. To enhance transparency, integrating SHAP can provide a more comprehensive and global understanding of feature contributions. Additionally, developing rule-based surrogate models or simplified regression-based approximations alongside the deep learning framework can help bridge the gap between interpretability & accuracy. Visualising feature importance trends over time and incorporating expert-driven constraints into the model can further improve trust and usability in practical hydrological applications.

G. Model Optimisation for Real-Time Deployment

The integration of LSTM, GRU, and attention mechanisms enhances the model's ability to capture long-term dependencies in runoff prediction. However, this also increases computational complexity, making real-time deployment challenging, especially in resource-constrained environments. To optimise efficiency while maintaining predictive accuracy, several strategies can be applied, including model pruning, quantisation, and knowledge distillation. Pruning removes redundant neurons, while quantisation reduces precision, lowering memory usage and improving inference speed. Knowledge distillation can also help by training a smaller model (the student) using the original, complex model (the teacher) to reduce computational overhead.

Another practical approach is the use of hybrid architectures with efficient attention mechanisms, such as Linformer or Performer, which reduce memory complexity from $O(n^2)$ to $O(n)$, thereby improving inference feasibility. Additionally, feature selection techniques such as Autoencoders can replace PCA for dimensionality reduction, simplifying the model without sacrificing accuracy. Finally, deploying the model through edge computing or cloud-based frameworks (e.g., TensorFlow Lite, ONNX Runtime) can enhance adaptability, ensuring real-time performance even in dynamic environments.

H. Model Performance in Extreme Flood Events

Accurately predicting extreme flood events is crucial for disaster management. Yet, this study primarily focuses on overall runoff prediction accuracy without an explicit evaluation of the model's performance during extreme events. Flood events often involve sudden and highly nonlinear hydrological responses, which may not be well captured by deep learning models trained on historical average conditions.

One key challenge in predicting extreme floods is data imbalance—flood events are rare compared to normal runoff conditions, leading the model to be biased toward typical flow levels. This can result in an underestimation of peak flows, reducing the model's reliability in forecasting extreme events. Techniques such as oversampling flood event data, synthetic data generation, or specialised loss functions (e.g., weighted loss for high-runoff cases) could help improve flood prediction accuracy.

Additionally, feature selection and input data resolution play a crucial role. Variables such as antecedent soil moisture, extreme precipitation indices, and upstream reservoir discharges can significantly influence flood magnitudes. Future model refinements should incorporate these additional hydrological indicators to enhance predictive ability.

Another aspect to consider is temporal generalisation. The model may perform well on historical data but struggle with extreme weather patterns that are unprecedented due to climate change. Continuous retraining with updated data and integrating climate model projections could help improve adaptability.

1) Applicability to different river basins

The model's performance in different river basins with varying topographical and climatic conditions would depend on several factors, including data availability, hydrological variability, and the extent to which the model has been trained on diverse datasets. The Attentional LSTM, which demonstrated strong predictive accuracy within the tested dataset, may require retraining or fine-tuning when applied to new regions with different hydrological characteristics.

Differences in rainfall patterns, soil types, vegetation, and elevation can significantly influence runoff dynamics. A model trained on one region may not generalise well to another without necessary adjustments. The effectiveness of the model would also depend on the availability and quality of input data, such as historical precipitation, temperature, and river discharge records. If similar data sources are available in the new river basin, adaptation would be more feasible. To enhance adaptability, Principal Component Analysis (PCA)-based approaches can help extract relevant features for different climatic conditions, and transfer learning techniques could be explored to leverage knowledge from previously trained models. Before deployment, it is crucial to validate the model using local data and perform hyperparameter tuning to ensure optimal performance in the new environment. Ultimately, while the proposed model is robust, its direct application to different river basins may require modifications, retraining, and careful validation to maintain accuracy and reliability.

2) Impact of PCA on model performance

The study observed that incorporating Principal Component Analysis (PCA) did not significantly improve model performance, which raises the question of whether alternative dimensionality reduction techniques might yield better results. PCA operates under the assumption that the most informative features are those with the highest variance. However, in hydrological modelling, low-variance features may still hold critical information relevant to runoff prediction. This could explain why PCA-based feature selection did not enhance prediction accuracy in our experiments.

One possible explanation is that PCA may have discarded features crucial for temporal dependencies in the dataset, thereby reducing the model's ability to capture long-term correlations in hydrological processes. Alternative techniques, such as Autoencoders, which use non-linear transformations to learn feature representations, might be more effective in retaining important information while reducing dimensionality.

3) Limitations across different climatic zones

While the proposed Attentional LSTM model has shown promising results in the Krishna River Basin, its applicability to other climatic zones presents particular challenges. One key limitation is the model's reliance on region-specific hydrological patterns, which may not generalise well to river basins with different precipitation regimes, soil compositions, and seasonal variations. For instance, regions with extreme aridity or heavy monsoonal influences may exhibit runoff behaviours that differ significantly from those observed in the Krishna River Basin.

Another limitation is the availability and quality of training data. The model is trained using historical hydrometeorological data from the Krishna River Basin, which may not be representative of other climatic conditions. When applied to basins with varying rainfall patterns, snowmelt contributions, or evapotranspiration rates, the model may require extensive retraining with localised data to maintain accuracy.

Additionally, land use and topographical variations pose challenges to the direct transferability of models. The Krishna River Basin primarily consists of agricultural and semi-urban landscapes. However, basins in mountainous, coastal, or heavily forested regions may exhibit hydrological responses that the model has not been trained to capture effectively. Incorporating additional features, such as snow cover data in colder climates or tidal influences in coastal regions, may be necessary for better adaptability.

4) Potential for early warning flood prediction systems

The proposed Attentional LSTM model demonstrates strong predictive capabilities, making it a viable candidate for integration into an early warning flood prediction system. By leveraging real-time hydrological and meteorological data, the model can provide timely flood forecasts, enabling authorities and communities to take proactive measures to mitigate the impact of flooding.

However, several challenges must be addressed for real-world deployment. First, data availability and quality are crucial, as real-time flood forecasting requires continuous and high-resolution input data. Any gaps or inaccuracies in sensor data could compromise the model's performance. Second, model interpretability and trust remain key concerns, as decision-makers prefer models that provide clear explanations for their predictions. Implementing explainable AI techniques, such as SHAP or LIME, could enhance user confidence.

Another major challenge is computational efficiency. While deep learning models deliver high accuracy, their real-time application requires optimised algorithms and hardware infrastructure to ensure rapid processing and low latency. Furthermore, generalisations that impact across different river basins with varying climatic and topographical conditions must be carefully evaluated. Retraining the model on diverse datasets or employing transfer learning techniques could improve adaptability.

5) Real-time application and scalability of the model

A critical aspect of deploying machine learning models for hydrological forecasting is their ability to handle real-time data streams and adapt to changing environmental conditions. While this study demonstrates the effectiveness of deep learning models in predicting runoff, real-time implementation poses additional challenges, including data latency, computational efficiency, and adaptability to dynamic conditions.

For real-time applications, the model must efficiently process incoming sensor data, such as rainfall, temperature, and streamflow, with minimal lag. This requires integration with streaming data platforms (e.g., Apache Kafka, Google Dataflow) and the ability to handle high-frequency data updates. One approach to enhance real-time adaptability is incremental learning, where the model continuously updates its weights as new data arrives, rather than retraining from scratch.

Another vital factor is scalability; the model must be capable of handling large-scale hydrological networks across different river basins. Cloud-based deployment using distributed computing frameworks (e.g., TensorFlow Serving, AWS SageMaker) can facilitate real-time predictions across multiple locations simultaneously.

Furthermore, real-world hydrological systems are subject to non-stationarity, meaning that relationships between inputs and outputs may shift over time due to climate change or land-use modifications. To address this, adaptive models incorporating concept drift detection techniques (e.g., Kalman filtering, online learning algorithms) can help maintain accuracy over time.

6) Addressing overfitting and the need for regularisation

The study reports high NSE values, which suggest strong predictive performance. However, potential overfitting cannot be ruled out since no explicit regularisation techniques, such as dropout, early stopping, or L1/L2 regularisation, were applied during model training. Overfitting can lead to reduced generalizability, especially when dealing with complex hydrological patterns. Future work should explore these techniques to ensure the model maintains robust performance across diverse datasets and avoids learning noise or redundant patterns.

To further validate the model's reliability, cross-validation and testing across different hydrological conditions should be considered. A comparative analysis between training and validation loss trends can help identify overfitting, allowing for necessary adjustments. Additionally, incorporating Bayesian optimisation or grid search for hyperparameter tuning may aid in selecting an optimal regularisation strategy. These enhancements will ensure that the model remains both accurate and generalizable.

I. Impact of Additional Hydrological Features on Model Performance

Incorporating additional hydrological features, such as soil moisture, land-use data, and groundwater levels, could enhance the model's predictive capabilities by providing a more comprehensive representation of the watershed's hydrological processes. Soil moisture directly influences runoff generation and evapotranspiration, while land-use patterns affect infiltration rates and surface water flow. Integrating these features could help the model better capture the complex interactions between precipitation, soil properties, and river discharge.

However, adding new features also increases data requirements and computational complexity. The effectiveness of additional features should be assessed through feature importance analysis, such as SHAP values or permutation importance, to determine their contribution to model performance. If found beneficial, dimensionality reduction techniques (e.g., Autoencoders or Feature Selection methods) can be employed to retain only the most informative variables while minimising overfitting risks.

IV. Conclusion

This study demonstrates that factorial analysis is a practical pre-processing step for selecting appropriate models, as it provides critical insights into dataset collinearity and other factors that may contribute to underfitting and overfitting. Our findings confirm that deep learning models, particularly Long Short-Term Memory (LSTM) and its hybrid variants, significantly outperform traditional approaches in rainfall-runoff modelling. The results indicate that incorporating attention mechanisms improves predictive accuracy, with the LSTM-GRU-Attentional model achieving the best performance, yielding a Mean Squared Error (MSE) of 0.000729, a Mean Absolute Error (MAE) of 0.020, and a Nash-Sutcliffe Efficiency (NSE) of 0.980. Comparatively, the standard LSTM model without attention performed the worst, with an MSE of 0.014641, an MAE of 0.093, and an NSE of 0.880.

Furthermore, our proposed PCA-Based Attentional LSTM model, designed to enhance feature selection and reduce dimensionality, demonstrated competitive performance with an MSE of 0.000855, an MAE of 0.022152, and an NSE of 0.961. Although PCA increases dataset complexity, our analysis shows that its application yields results highly similar to those obtained using all variables, reinforcing its effectiveness in optimising data representation. The results further establish that attention-based mechanisms prove model accuracy, as evidenced by the Attentional.

LSTM achieves an NSE of 0.978, significantly outperforming non-attention-based models. Similarly, the Attentional GRU model exhibited strong performance with an MSE of 0.000654, an MAE of 0.018878, and an NSE of 0.970, confirming its effectiveness in time-series forecasting.

Despite these advancements, specific research gaps remain. While our hybrid LSTM-GRU-Attention model improves prediction accuracy, its computational complexity requires further optimisation for large-scale applications. Additionally, the study highlights the need for enhanced explainability in deep learning models for hydrological forecasting. Future work will focus on refining model architectures by incorporating Explainable AI (XAI) techniques, such as Local Interpretable Model-Agnostic Explanations (LIME), to understand the contributions of individual features better. Moreover, expanding the study to include real-time applications in Indian river basins will allow for broader validation and scalability across diverse hydrological conditions.

Conflict of Interests

The authors declare no conflict of interest.

Author Contribution

Ashish Sir proposes the initial idea and conceptualises it. I, Sagar Lachure, carry out the experimental work. The results are reviewed and approved by Ashish Sir. Based on the findings, a research paper is drafted. Revisions are made by the feedback received, and all authors approve the final version.

Acknowledgment

The authors wish to thank Jaykumar Sir for their help.

References

[1] S. Gosling, J. Zaherpour, N. Mount *et al.*, "A comparison of changes in river runoff from multiple global and catchment-scale hydrological models un- der global warming scenarios of 1 C, 2 C and 3 C," *Climatic Change*, vol. 141, pp. 577–595, 2017.

[2] M. Fabbri and G. Moro, "Dow jones trading with deep learning: The unreasonable effectiveness of recurrent neural networks," *Data.*, pp. 142–153, 2018.

[3] J. Torres, D. Hadjout, A. Sebaa, F. Mart´ınez-A´lvarez, and A. Troncoso, "Deep learning for time series forecasting: A survey," *Big Data*, vol. 9, pp. 3−21, 2021.

[4] T. Lees, M. Buechel, B. Anderson, L. Slater, S. Reece, G. Coxon, and S. Dadson, "Benchmarking Data-Driven Rainfall-Runoff Models in Great Britain: A comparison of LSTM-based models with four lumped conceptual models," *Hydrology and Earth System Sciences*, vol. 25, 2021.

[5] Z. Tang, J. Zhang, M. Hu, Z. Ning, J. Shi, R. Zhai, C. Liu, J. Zhang, and G. Q. Wang, "Improving rainfall-runoff simulation in Semi-Arid Basins by combining data segmentation and attention-based deep learning," ResearchGate, 2024.

[6] D. Kim, "Data collection strategy for building rainfall-runoff LSTM model predicting daily runoff," *Journal of Korea Water Resources Association*, 2021.

[7] S. Ojha, V. Singh, and T. Roshni, "Comparison of machine learning techniques for rainfall-runoff modeling in Punpun river basin, India," *International Journal of Advanced and Applied Sciences*, 2023.

[8] H. Yin, F. Wang, X. Zhang, Y. Zhang, J. Chen, R. Xia, and J. Jin, "Rainfall-runoff modeling using long short-term memory based step-sequence framework," *Journal of Hydrology*, 2022.

[9] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: state of the art, current trends and challenges," *Multimedia Tools and Applications*, vol. 82, pp. 3713–3744, 2023.

[10] K. Beven, *Rainfall-runoff Modelling: the Primer*, John Wiley & Sons, 2012.

[11] P. Sharma, P. Patel, and V. Jothiprakash, "Data-driven modelling framework for streamflow prediction in a physio-climatically heterogeneous river basin," *Soft Computing*, vol. 25, pp. 5951–5978, 2021.

[12] M. Clark, M. Bierkens, L. Samaniego, R. Woods, R. Uijlenhoet, K. Bennett, V. Pauwels, X. Cai, A. Wood, and C. Peters-Lidard, "The

evolution of process-based hydrologic models: historical challenges and the collective quest for physical realism," *Hydrology And Earth System Sciences*, vol. 21, pp. 3427–3440, 2017.
[13] M. Reichstein, G. Camps-Valls, B. Stevens *et al.*, "Deep learning and process understanding for data-driven Earth system science," *Nature*, vol. 566, pp. 195-204, 2019.
[14] M. Waqas and U. Humphries, "A critical review of RNN and LSTM variants in hydrological time series predictions," *MethodsX*, 102946, 2024.
[15] F. Kratzert, D. Klotz, C. Brenner, K. Schulz, and M. Herrnegger, "Rainfall–runoff modelling using long short-term memory (LSTM) networks," *Hydrology and Earth System Sciences*, vol. 22, pp. 6005–6022, 2018.
[16] F. Kratzert, D. Klotz, G. Shalev, G. Klambauer, S. Hochreiter, and G. Nearing, "Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets," *Hydrology and Earth System Sciences*, vol. 23, pp. 5089–5110, 2019.
[17] J. Ma, W. Gao, S. Joty, and K. Wong, "An attention-based rumor detection model with tree-structured recursive neural networks," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 11, pp. 1–28, 2020.
[18] Z. Chen, H. Lin, and G. Shen, "TreeLSTM: A spatiotemporal machine learning model for rainfall-runoff estimation," *Journal of Hydrology: Regional Studies*, vol. 48, 101474, 2023.
[19] A. Vaswani, "Attention is all you need," in *Proc—Advances in Neural Information Processing Systems*, 2017.
[20] Y. Yan, W. Zhang, Y. Liu, and Z. Li, "Simulated annealing algorithm optimized GRU neural network for urban rainfall-inundation prediction," *Journal of Hydroinformatics*, vol. 25, pp. 1358–1379, 2023.
[21] H. Yin, Z. Guo, X. Zhang, J. Chen, and Y. Zhang, "Runoff predictions in ungauged basins using sequence-to-sequence models," *Journal of Hydrology*, vol. 603, 126975, 2021.
[22] H. Yin, X. Zhang, F. Wang, Y. Zhang, R. Xia, and J. Jin, "Rainfall-runoff modeling using LSTM-based multi-state-vector sequence-to-sequence model," *Journal of Hydrology*, vol. 598, 126378, 2021.
[23] S. Chen, M. Ren, and W. Sun, "Combining two-stage decom- position based machine learning methods for annual runoff forecasting," *Journal of Hydrology*, vol. 603, 126945, 2021.
[24] A. Mohammed and G. Corzo, "Spatiotemporal convolutional long short-term memory for regional streamflow predictions," *Journal of Environmental Management*, vol. 350, 119585, 2024.
[25] N. Garg, S. Negi, R. Nagar, S. Rao, and K. R. Seeja, "Multi-variate multi-step LSTM model for flood runoff prediction: a case study on the Godavari River Basin in India," *Journal of Water and Climate Change*, 2023.
[26] H. Yin, F. Wang, X. Zhang, Y. Zhang, J. Chen, R. Xia, and J. Jin, "Rainfall-runoff modeling using long short-term memory-based step-sequence framework," *Journal of Hydrology*, vol. 610, 127901, 2022.
[27] S. Shrestha and S. Pradhanang, "Performance of LSTM over SWAT in Rainfall-Runoff Modeling in a Small, Forested Watershed: A Case Study of Cork Brook, RI," *Water*, vol. 15, p. 4194, 2023.