A hybrid rainfall-runoff model: integrating initial loss and LSTM for improved forecasting

Wei Wang[1], Jie Gao[1], Zheng Liu[2] and Chuanqi Li[1]*

[1]School of Civil Engineering, Shandong University, Jinan, China, [2]Jinan Water Resources Engineering Service Center, Jinan, China

Accurate rainfall-runoff modeling is crucial for disaster prevention, mitigation, and water resource management. This study aims to enhance precision and reliability in predicting runoff patterns by integrating physical-based models like HEC-HMS with data-driven models, such as LSTM. We present a novel hybrid model, Ia-LSTM, which combines the strengths of HEC-HMS and LSTM to improve hydrological modeling. By optimizing the "initial loss" (Ia) with HEC-HMS and utilizing LSTM to capture the effective rainfall-runoff relationship, the model achieves a substantial improvement in precision. Tested in the Yufuhe basin in Jinan City, Shandong province, the Ia-LSTM consistently outperforms individual HEC-HMS and LSTM models, achieving notable average Nash-Sutcliffe Efficiency (*NSE*) values of 0.873 and 0.829, and average R^2 values of 0.916 and 0.870 for calibration and validation, respectively. The study shows the potential of integrating physical mechanisms to enhance the efficiency of data-driven rainfall-runoff modeling. The Ia-LSTM model holds promise for more accurate runoff estimation, with wide applications in flood forecasting, water resource management, and infrastructure planning.

KEYWORDS

rainfall-runoff modeling, hybrid model, initial loss (Ia), HEC-HMS, LSTM

1 Introduction

Rainfall-runoff modeling is essential in hydrology, especially for tasks like reservoir management, flood forecasting, and water resource planning (Chen and Adams, 2006; Young and Liu, 2015). Despite significant progress, accurately predicting runoff remains a big challenging due to the complex, nonlinear, and dynamic nature of the rainfall-runoff process (Wang et al., 2006; Xie et al., 2019). This complexity is further compounded by various influencing factors, including rainfall patterns, initial soil moisture, terrain, land cover, and infiltration (Wang and Ding, 2003; Perera et al., 2019). Sudden rainstorms further emphasize the need for a comprehensive understanding of primary rainfall patterns (Xie et al., 2023a; Xie et al., 2023b). The impact of urban imperviousness on runoff and flooding dynamics has also emerged as a crucial factor in recent studies (Shukla et al., 2020; Mehr and Akdegirmen, 2021).

Vegetation and soil properties play a significant role in regulating the hydrological cycle, impacting various processes such as interception, infiltration, evaporation, and surface depression storage (Shukla et al., 2018). Notably, initial loss or initial abstraction (Ia) represents the rainfall occurring before the initiation of surface runoff. Ia is influenced by factors like vegetation cover, soil infiltration capacity, and antecedent moisture condition in the soil. Its magnitude is closely tied to both climatic conditions and moisture level in the

watershed, making accurate estimation Ia for runoff determination and flood management (Zheng et al., 2020).

Rainfall-runoff models are broadly classified into physically-based models and data-driven models (Devia et al., 2015; Bartoletti et al., 2018; Mohammadi et al., 2022). Physically-based models, such as the Hydrologic Engineering Center-Hydrologic Modeling System (HEC-HMS) (Feldman, 2000), Xinanjiang (XAJ) model (Zhao, 1992), soil and water assessment tool (SWAT) (Arnold et al., 1998), MIKE-SHE (Jaber and Shukla, 2012), and HSPF (Bicknell et al., 1997), employ mathematical equations to represent hydrological processes. While these models provide valuable insights, their development demands a deep understanding of hydrological processes and extensive basin parameters, leading to a complex and time-consuming development process (Fenicia et al., 2008; Chen et al., 2022). The Hydrologic Modeling System (HMS), designed by the Hydrologic Engineering Center (HEC) of the United States Army Corps of Engineers, is a widely adopted rainfall-runoff analysis tool worldwide. The physical processes are so complex in hydrological models that it is difficult to discover the information from the available inputs.

Data-driven models offer a compelling alternative, establishing relationships between input and output data without the need for detailed understanding of underlying physical processes (Noori and Kalin, 2016; Yaseen et al., 2016; Lees et al., 2021). These models rely on historical rainfall and runoff data, making them suitable for handling non-linear and stochastic systems (Hu et al., 2018; Kratzert et al., 2018; Gao et al., 2020). Prominent data-driven methods for rainfall-runoff modeling include artificial neural networks (ANN) (Haykin and Network, 2004), support vector machines (SVM) (Cortes and Vapnik, 1995), genetic programming (Savic et al., 1999; Danandeh and Nourani, 2018), random forests (Breiman, 2001), fuzzy logic (Hundecha et., 2001) and regression in the reproducing kernel hilbert space (RRKHS) (Safari et al., 2020). These models use historical data to identify patterns and associations, enabling them to make precise predictions or estimates based on observed data patterns.

In recent years, deep learning, as a type of data-driven modeling, has gained substantial attention in hydrology due to its adaptability and minimal data requirements (Beven, 2020; Gu et al., 2020; Zhou et al., 2023). Among various deep learning approaches, Long Short-Term Memory (Hochreiter and Schmidhuber, 1997) networks have proven their effectiveness in various hydrological applications, including rainfall prediction (Barrera-Animas et al., 2022), flood forecasting (Hu et al., 2018; Rahimzad et al., 2021), and river water table prediction (Kim et al., 2022). As emphasized by Kratzert et al. (2018), the strength of the LSTM models lies in their capacity to capture long-term dependencies between the input and output.

The integration of physically-based and data-driven models in rainfall-runoff modeling has received considerable interest, driven by their complementary strengths (Tian et al., 2018; Sun et al., 2019; Zhou et al., 2022). Several hybrid models have exhibited promise in this domain. For instance, the XAJ-LSTM model, proposed by Cui et al. (2021), combines the Xinanjiang (XAJ) conceptual model with LSTM neural networks for multistep-ahead flood forecasting. This hybrid model utilizes the model forecast results of XAJ as input variables for LSTM, thus enhancing the physical mechanisms of hydrological simulation. By incorporating discharge forecasts from the XAJ model, the XAJ-LSTM hybrid model overcomes the

limitations of LSTM's input variables, resulting in notably improved performance. Similarly, Gholami & Khaleghi (2021) conducted a comparative analysis of ANN and HEC-HMS models in rainfall-runoff simulation. Narayana Reddy and Pramada, (2022) integrated HEC-HMS with ANN to enhance daily discharge simulation and yearly peak discharge prediction. Farfan et al. (2020) used streamflow series forecasts from a conceptual model as input for back-propagation neural networks, leading to markedly improved streamflow predictions. Hitokoto and Sakuraba (2020) successfully integrated a rainfall-runoff model with a feed-forward artificial neural network to predict real-time water level processes. These instances highlight the effectiveness of hybrid models in enhancing predictive accuracy.

While previous research has made significant progress in rainfall-runoff modeling, there remains a critical need for innovative approaches to address the limitations of current models. Notably, the absence of physical mechanism poses a substantial obstacle in applying machine learning methods, which typically rely on labeled observations (Xie et al., 2021). The consideration of initial loss (Ia) within a deep learning network for rainfall-runoff simulation has received limited attention. Ia represents a crucial stage in the rainfall-runoff process. To address these challenges, this study proposes the hybrid rainfall-runoff model, integrating initial loss and LSTM. This integration harnesses the strengths of both physically-based and data-driven approaches, offering the potential for substantial advancements in accurately predicting and managing rainfall-induced runoff events.

The main objectives of this study are: 1) to develop the Ia-LSTM hybrid model, combining the advantages of the widely used hydrologic model, HEC-HMS, with the predictive capabilities of LSTM; 2) to conduct a comprehensive evaluation of the performance of the proposed hybrid model against the individual HEC-HMS and LSTM models. To assess the model's effectiveness, a case study is undertaken in the Yufuhe Basin, located in Jinan City, Shandong Province. The integration of the HEC-HMS model with LSTM enables a more comprehensive representation of the rainfall-runoff process, considering both the physical processes and historical data patterns. The incorporation of initial loss estimation and LSTM aims to improve the accuracy and reliability of runoff forecasting.

The contributions of this paper can be summarized as follows. First, it introduces the Ia-LSTM model, a novel rainfall-runoff model based on the integration of initial loss and LSTM. Second, the model is applied to the tasks of individual rainfall-runoff modeling in the Yufuhe basin, demonstrating its effectiveness.

The paper is organized as follows: Section 2 provides an overview of the study area and the data utilized. It also briefly describes the HEC-HMS model, LSTM network, and Ia-LSTM hybrid model. Section 3 presents the research results and discussions. Finally, Section 4 concludes the paper by summarizing the key findings.

2 Materials and methods

This section provides an overview of the study area and data (Section 2.1), introduces the HEC-HMS model (Section 2.2), explains the LSTM model structure (Section 2.3), presents the

FIGURE 1
Elevation and distribution of rainfall stations in the study area.
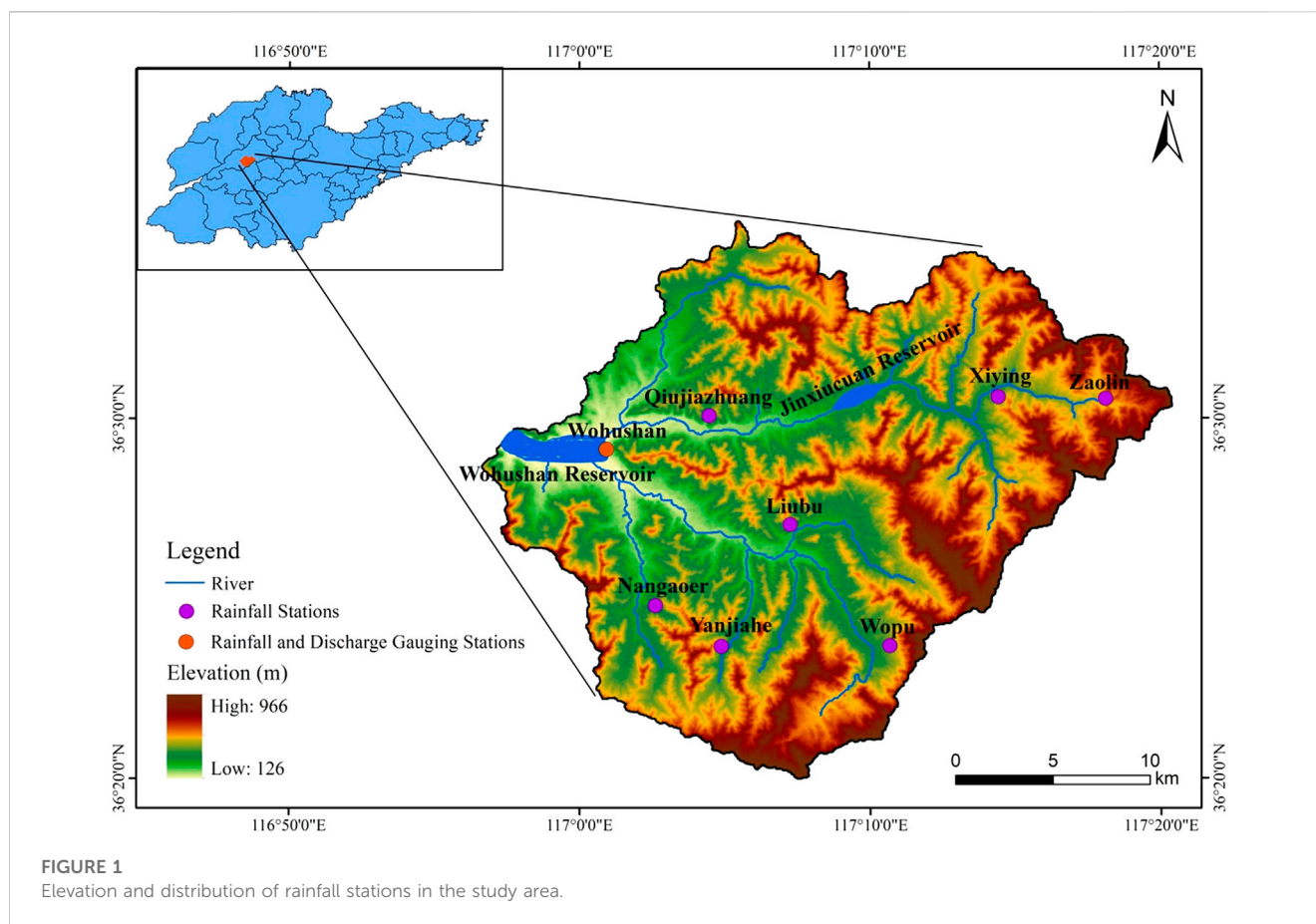
proposed framework based on the LSTM (Section 2.4), and outlines the evaluation metrics of model performance (Section 2.5).

2.1 Study area and data

This study focuses on the Yufuhe basin, located upstream of the Wohushan Reservoir in Jinan city, Shandong Province, China. Encompassing an area of 557km², the basin exhibits vulnerability to floods and droughts due to its unique natural and geographical conditions. Notably, both 2007 and 2013 witnessed large-scale floods resulting in significant economic losses in Jinan (Zhang et al., 2016). The basin plays a critical role in flood control and water management, featuring diverse topography including mountains, hills, and a complex river network.

The study area is characterized by a sub-humid continental monsoon climate, with an annual average temperature of 14.3°C and an average annual precipitation of 670.0 mm. Rainfall is concentrated within the flood season from June to September, marked by intense, short-duration rainfall events. The flood season accounts for approximately 70% of the annual precipitation, posing flood risks in the basin.

Within the Yufuhe basin, there are seven rain-gauge stations and one Wohushan stream flow gauge station located at the basin outlet. Figure 1 illustrates the location of the watershed, elevation, distribution of rainfall and flow gauging stations, as well as the streams. The land use and land cover (LULC) map for the Yufuhe basin in 2020 was sourced from the Institute of Geographic Sciences and Resources of the Chinese Academy of Sciences (http://www.resdc.cn/), offering a detailed representation at a 30-m resolution. The basin is characterized by abundant vegetation, with agricultural land accounting for approximately 38% and forests covering 35% of the total area (Figure 2).

Hourly flow runoff data from the Wohushan hydrological station and hourly precipitation data from seven gauges were collected from 1973 to 2020. After data preprocessing, 30 rainfall and runoff events, including 6136 one-hourly rainfall and runoff records, were selected for this study. Among these flood events, 20 were used for model calibration, and the remaining 10 were used for model validation.

2.2 HEC-HMS model

The HEC-HMS model, developed by the U.S. Army Corps of Engineers (USACE), can accurately predict streamflow, runoff volume, and other hydrologic parameters. It incorporates inputs such as land use, soil types, channel networks, and rainfall data. HEC-HMS offers variood, unit hydrograph method, Snyder unit hydrograph method, and others [(USACE 2000us hydrologic modeling methods, including the Soil Conservation Service (SCS) curve number meth]. These methods are selected based on the specific characteristics of the modeled watershed.
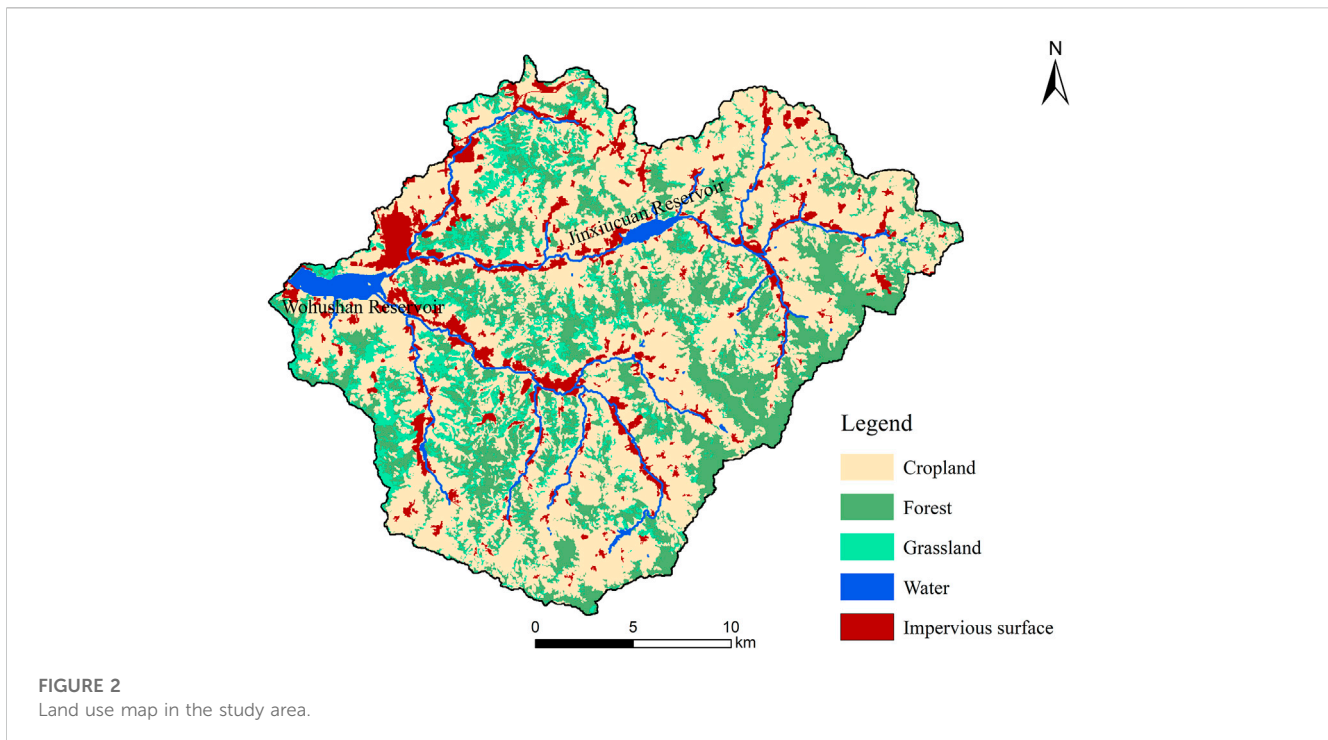
FIGURE 2
Land use map in the study area.

The HEC-HMS model comprises four main components: the basin model, meteorological model, control specifications, and time series model. The rainfall runoff process is delineated through four modules: loss, transformation, routing, and baseflow. Detailed information on the model's structure and processes can be found in the Technical Reference Manual (USACE-HEC, 2000) and the User's Manual of HEC-HMS.

2.2.1 Initial and constant loss method

The initial and constant loss method estimates surface losses in rainfall runoff modeling and is suitable for watersheds with limited soil data. This method requires two parameters: initial loss and constant rate. Initially, all rainfall is absorbed until the specified initial loss volume is attained, after which rainfall is lost at a constant rate. It considers antecedent moisture conditions and losses prior to reaching ultimate infiltration capacity. This method assumes a single soil layer for estimating moisture content changes, making it ideal for event simulation, particularly in data-scarce watersheds. The initial loss is influenced by antecedent moisture conditions and losses before reaching the ultimate infiltration capacity. It is worth noting that the initial loss parameter should be calibrated using observed data, although it is often estimated based on the soil moisture state at the beginning of the simulation and an assumed active layer depth. Throughout the simulation, a constant maximum potential rate of precipitation loss, f_c, is assumed.

The net rainfall, P_{et}, at time t, is calculated using the following equation (USACE, 2000b):

$$P_{et} = \begin{cases} 0 & if \; \sum P_i < I_a \\ P_i - f_c \Delta t & if \; \sum P_i > I_a \; and \; P_i > f_c \\ 0 & if \; \sum P_i > I_a \; and \; P_t < f_c \end{cases} \quad (1)$$

where P_{et} represents the net rainfall (mm), I_a denotes the initial loss (mm), P_i represents cumulative rainfall from time t to $t+\Delta t$ (mm), and f_c represents the average infiltration rate (mm/h).

Optimal values of the initial loss and the constant loss rate are determined during the calibration of HEC-HMS model, primarily to match the depths of effective precipitation and direct runoff.

2.2.2 Direct runoff calculation

The Snyder unit hydrograph method is used to estimate surface direct runoff resulting from excess precipitation. It utilizes a standardized unit hydrograph incorporating parameters like peak lag time, peak flow, and total duration. These parameters play a crucial role in understanding the hydrological response of a watershed to rainfall events.

The standard unit hydrograph relates rainfall duration (t_r) to basin lag time (t_p) as follows:

$$t_p = 5.5 t_r \quad (2)$$

The Snyder Unit hydrograph method requires specifying input parameters such as the basin lag time (t_p) and peak coefficient (C_p). Peak lag time is calculated using the following formula:

$$t_p = C \, C_t \, (L \, L_C)^{0.3} \quad (3)$$

in which L the length of the main stream from outlet to the divide (km); Lc is the length along the main stream to the nearest point of the watershed centroid; Ct is a coefficient (usually 1.8–2.2); C is a conversion constant (0.75 for SI units).

2.2.3 Baseflow calculation

Baseflow calculation involves accounting for the flow through a channel or the influence of groundwater in a hydrological system. HEC-HMS offers two methods for baseflow
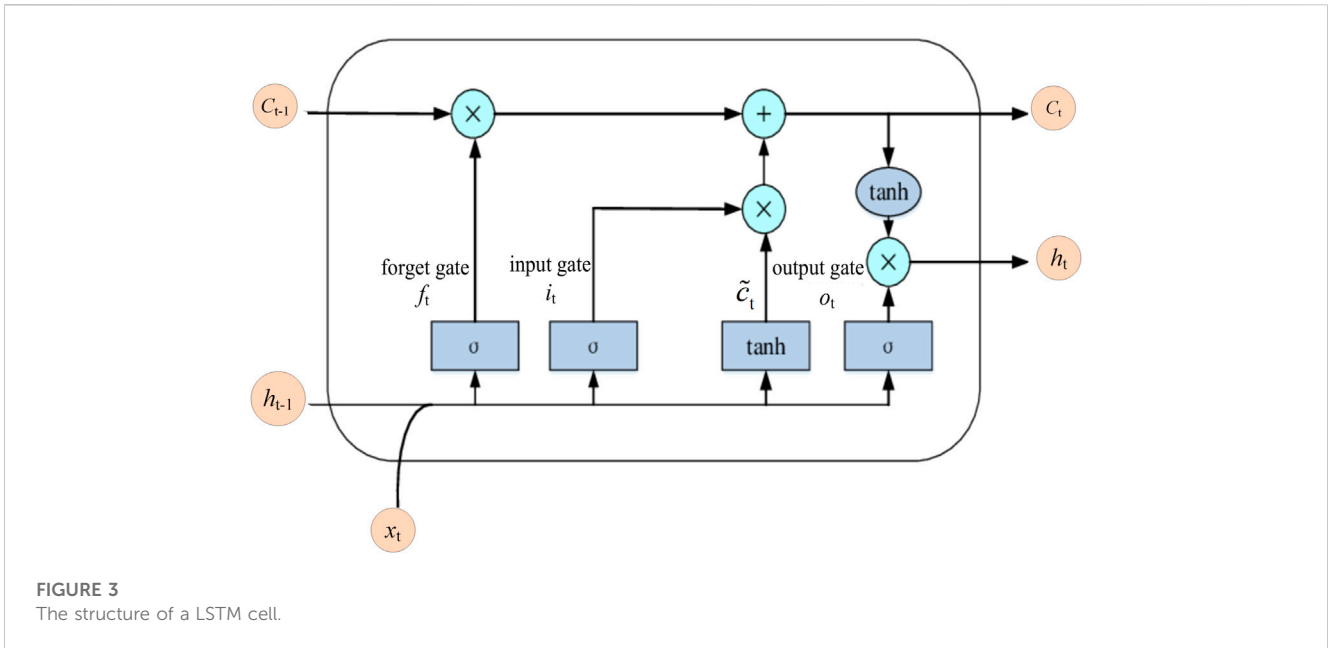
FIGURE 3
The structure of a LSTM cell.

calculation: recession and constant monthly. The recession method, utilized in this study, represents the drainage process from natural storage within a watershed. It employs an exponential decay function (Knebl et al., 2005) to relate the baseflow (Q_t) at a specific time (t) to an initial value (Q_0). The equation is defined as:

$$Q_t = Q_0 K^t \tag{4}$$

where K represents the exponential decay constant.

2.2.4 Flood routing

Flood routing in HEC-HMS provides various options for routing flood hydrographs through different reaches. The Muskingum method is commonly used for general flood routing.

In this study, the Muskingum method is adopted to compute the outflow from each reach during flood routing. This method is based on the following equation:

$$Q_{j+1} = C_1 I_{j+1} + C_2 I_j + C_3 Q_j \tag{5a}$$

where

$$C_1 = \frac{\Delta t - 2KX}{2K(1-X) + \Delta t} \tag{5b}$$

$$C_2 = \frac{\Delta t + 2KX}{2K(1-X) + \Delta t} \tag{5c}$$

$$C_3 = \frac{2K(1-X) - \Delta t}{2K(1-X) + \Delta t} \tag{5d}$$

where C_1, C_2 and C_3 are the routing coefficients for the concerned reach; I_j, I_{j+1} are the inflows to the reach at the beginning and end of the computation interval $\triangle t$, respectively, Q_j and Q_{j+1} correspond to the outflows from the reach at the beginning and end of computation interval, respectively. K denotes the travel time through the reach, and X is the Muskingum weighting factor ($0 \leq X \leq 0.5$). The coefficients C_1, C_2, and C_3 must satisfy the condition that their sum equals 1.0.

2.2.5 Parameter optimization methods

Calibrating the parameters of HEC-HMS model is a crucial step for improving the agreement between model results and observed data. The primary objective is to determine the most appropriate parameter values that yield the closest match between computed and observed hydrographs. This involves quantifying the match using an objective function, which compares the simulated and observed flow data. The objective function serves to assess the accuracy of the model's performance.

To execute parameter calibration, HEC-HMS provides two search methods: the Univariate Gradient algorithm (UG) and the Nelder-Mead algorithm (NM). These algorithms assist in minimizing the objective functions and determining the parameter values that provide the best fit.

In this study, the Peak-Weighted Root Mean Square Error (PWRMSE) function is chosen as the objective function for parameter calibration. The Nelder-Mead algorithm is employed to optimize the model parameters and obtain the most suitable values, ensuring accurate simulation results.

2.3 Long short-term memory (LSTM) network

The Long Short-Term Memory (LSTM) network was selected due to its exceptional ability to handle extended data sequences, a challenge commonly faced by conventional Recurrent Neural Networks (RNNs) (Hochreiter and Schmidhuber, 1997). In hydrological modeling, where processes like rainfall-runoff relationships exhibit complex temporal patterns, LSTM' capability to capture long-term dependencies is crucial.

Specifically, LSTM excels in preserving vital information over extended periods, allowing it to accurately model complex water-related processes. This type of deep learning model is designed to address challenges encountered by traditional RNNs, such as gradient exploding or vanishing problems. It achieves this

through specialized gate mechanisms that control information flow, proving highly effective in processing sequential data.

The basic unit of the LSTM network includes a memory and three types of gates: input gate, forget gate, and output gate. These gates play a crucial role in managing memory and capturing relevant features by controlling information flow within the LSTM unit. Figure 3 provides a visual representation of the structure of an LSTM cell.

The forget gate, represented by f_t, determines how much of the previous memory to discard, based on the current input x_t and the previous cell state c_{t-1}. The input gate, represented by i_t, controls the information to be stored in the cell state c_t. The output gate, represented by o_t, filters the output variable h_t. The equations for the gates are given as follows (Kratzert et al., 2018):

$$f_t = \sigma\left(W_{hf}h_{t-1} + W_{xf}x_t + b_f\right) \tag{6}$$

$$i_t = \sigma\left(W_{hi}h_{t-1} + W_{xi}x_t + b_i\right) \tag{7}$$

$$o_t = \sigma\left(W_{ho}h_{t-1} + W_{xo}x_t + b_o\right) \tag{8}$$

$$\tilde{c}_t = \tanh\left(W_{hc}h_{t-1} + W_{xc}x_t + b_c\right) \tag{9}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \tag{10}$$

$$h_t = o_t \odot \tanh\left(c_t\right) \tag{11}$$

where xt denotes the input, f_t is a forget gate, i_t is an input gate, o_t is an output gate, c_t is the cell state at time t; σ is Sigmoid function, ⊙denotes the element-wise multiplication of two vectors, bf, bi, bo, and bc are the corresponding bias; W_{hf}, W_{xf}, W_{hi}, W_{xi}, W_{ho}, W_{xo}, W_{hc} and W_{xc} are the network weights matrices; *tanh* is hyperbolic tangent function; h_{t-1} is the output of hidden state of previous step; and x_t is the input.

To train the LSTM model, it is crucial to configure the hyperparameters that govern the training process (Tian et al., 2018). Several hyperparameters, including learning rate, loss function, optimizer, dropout rate, batch size, and number of epochs, were tested and evaluated to determine the optimal values that give the best evaluation metrics. The final selected hyperparameters were as follows: a time step of 10, 256 neurons in the hidden layer, dropout rate of 0.20, and a batch size of 32. The Root Mean Square prop (RMSprop) optimizer with a decay coefficient of 0.8 and a learning rate of 0.0001 was utilized for model training. The training process involved 1000 iterations. The mean squared error (MSE) served as the loss function, measuring the average squared difference between the predicted values and the actual values.

To ensure accurate data analysis and enhance the efficiency and performance of the model, it is essential to preprocess the input data and map their attribute values to the range [0, 1]. Normalizing the input variables eliminates the influence of magnitude, thereby improving the accuracy and efficiency of network learning.

In this study, the rainfall and runoff data were preprocessed using min-max normalization method, which can be defined by Eq 1:

$$x_{norm} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{12}$$

where x_{norm}, x_i, $x_{\min}$, $x_{\max}$ represents the normalized, observed, minimum and maximum values of rainfall or runoff, respectively. This normalization process ensures that the input variables are scaled appropriately and enables effective analysis and learning by the network.

2.4 Ia-LSTM hybrid model

This study proposes an Ia–LSTM model to improve the accuracy of hourly runoff discharge predictions using LSTM. The model incorporates HEC-HMS model for dataset generation, using the effective rainfall data series obtained by subtracting the initial loss (Ia) from the total rainfall data. By considering the influence of Ia, the LSTM model is trained to predict flow discharge sequences, resulting in improved precision in rainfall-runoff predictions. Figure 4 illustrates the overall workflow of the Ia-LSTM the hybrid model. The Ia-LSTM hybrid model optimizes the determination of Ia using HEC-HMS and considers factors such as infiltration, vegetation interception, and evaporation that impact rainfall-runoff dynamics.

The development of the Ia-LSTM hybrid model involves the following steps:

(1) Data preparation: Historical rainfall-runoff data for the study area are collected and organized into rainfall-runoff data sequences.
(2) Dataset generation: The HEC-HMS model is used to optimize and accurately estimate the initial loss (Ia) by considering factors such as rainfall-runoff, land use, soil type, and DEM data. The effective rainfall data is derived by subtracting Ia from the total rainfall. This step involves generating a dataset comprising effective rainfall-runoff pairs.
(3) LSTM model construction and training: The LSTM model is constructed, with effective rainfall data serving as the input variable and the corresponding runoff data as the output. The model is trained to capture the hidden mapping relationship between the inputs and outputs. Throughout the training process, various parameter combinations are explored to identify the optimal settings that enhance performance and efficiency.
(4) LSTM model forecasting: The trained LSTM model is used to predict runoff by inputting the effective rainfall data sequence. As a reference, the LSTM model is also trained on the original rainfall-runoff sequence. The performance of the LSTM rainfall-runoff prediction model, accounting for Ia, is evaluated and compared.

2.5 Evaluation metrics of model performance

The performance of the developed models is assessed using four widely used metrics in other hydrological studies: Nash-Sutcliffe efficiency (*NSE*), root mean square error (*RMSE*), relative error of peak discharge (*REP*), and coefficient of determination (R^2).

NSE is extensively used for evaluating rainfall-runoff simulation (Kumar et al., 2016). It quantifies the agreement between simulated and observed data by comparing their variances. *NSE* is calculated using the following formula:
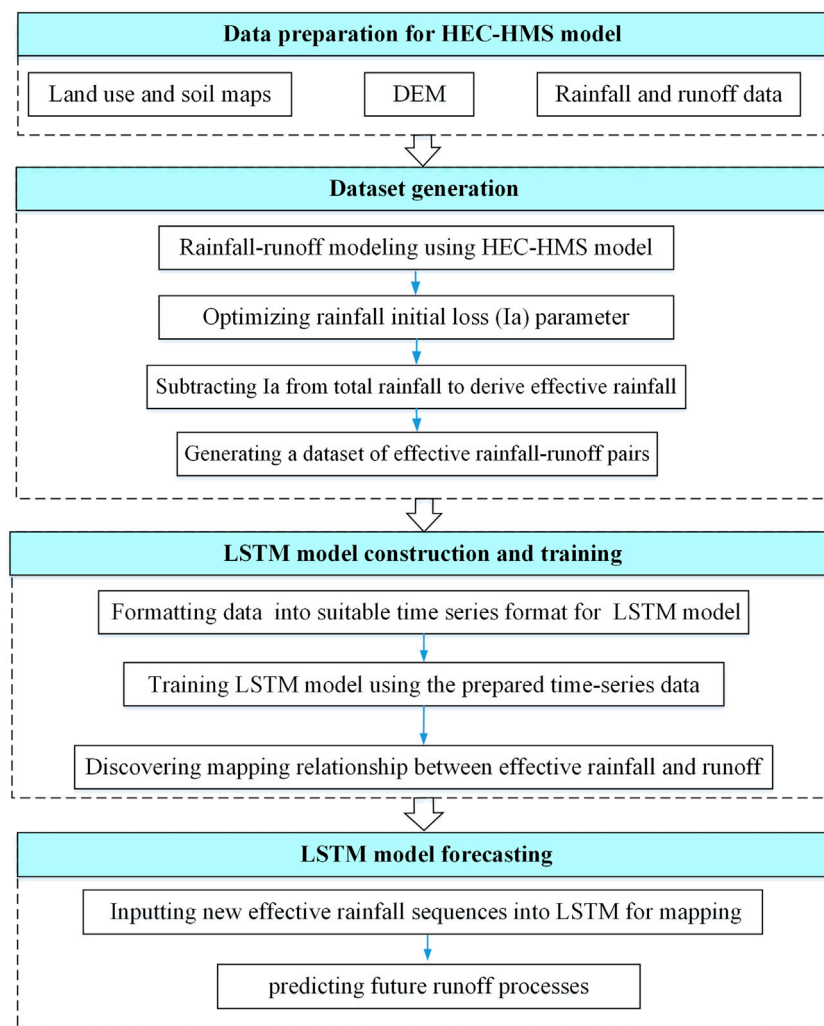
FIGURE 4
The flowchart of Ia-LSTM Hybrid Model.

$$NSE = 1 - \frac{\sum_{i=1}^{N} (O_i - P_i)^2}{\sum_{i=1}^{N} (O_i - \bar{O})^2} \qquad (13)$$

where O_i and P_i represent the observed and predicted runoff at the time step i, respectively; $\bar{O}$ is the average observed runoff, and N is the total number of observations. NSE ranges from $-\infty$ to 1, with 1 indicating a perfect match between the predictions and observations.

Root mean square error ($RMSE$) measures the effectiveness of the model and is the average of the squared difference between model simulated and observed values. $RMSE$ is used to represent the model's ability to predict flood events. $RMSE$ can be calculated by:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (O_i - P_i)^2} \qquad (14)$$

A lower $RMSE$ indicates better model simulation performance, with an $RMSE$ of 0 indicating an exact match simulated and observed values.

REP assesses the accuracy and uncertainty associated with peak discharge estimation. It is calculated as:

$$REP = \frac{|O_p - P_p|}{O_p} \times 100\% \qquad (15)$$

where O_p and P_p represent the observed and predicted peak river flow discharge, respectively. A lower REP value indicates better performance, indicating that the model's predictions are closer to the actual observed results.

The coefficient of determination (R^2) quantifies the degree of correlation between the simulated and observed runoff (Kumarasamy and Belmont, 2018). It is calculated using the formula:

$$R^2 = \left(\frac{\sum_{i=1}^{N} (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^{N} (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^{n} (P_i - \bar{P})^2}} \right)^2 \qquad (16)$$

where $\bar{O}$ and $\bar{P}$ represent the average value of observed and predicted runoff, O_p and P_p are the observed and predicted runoff, respectively. R^2 values range from 0 to 1, with higher values indicating a better fit between the model outputs and the target outputs. Higher R^2 values suggest greater predictive power. An R^2 equal to 1 denotes an ideal
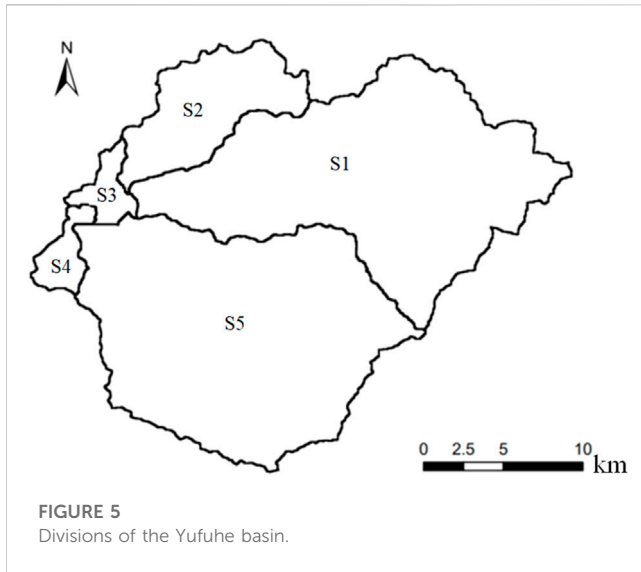
FIGURE 5
Divisions of the Yufuhe basin.

TABLE 1 Key characteristics of sub-basins in the Yufuhe basin.

Sub-basin	Area (km^2)	Average slope	Stream length (km)
S1	227.5	0.023	20.4
S2	58.6	0.020	7.9
S3	16.4	0.020	34.1
S4	11.9	0.095	5.6
S5	242.6	0.021	32.4

fit. Model performance is categorized as very good ($0.7 < R^2 < 1$), good ($0.6 < R^2 < 0.7$), satisfactory ($0.5 < R^2 < 0.6$), or unsatisfactory ($R^2 < 0.5$) (Ayele et al., 2017).

These metrics collectively provide a comprehensive assessment of the model's performance, encompassing simulation quality, accuracy of peak discharge predictions, and the correlation with observed data.

3 Results and discussion

3.1 Estimation of initial loss

To accurately estimate initial losses, the Yufuhe basin was divided into sub-basins (S1, S2, S3, S4, S5) as shown in Figure 5. Table 1 provides key characteristics of these sub-basins, including their areas, average slopes, and stream lengths. This subdivision allowed for a precise assessment of initial losses. The initial and constant loss method requires the specification of parameters including the percent impervious area, initial loss (Ia), and constant loss rate. The Thiessen polygon method was employed to estimate the average rainfall for the entire watershed, and specific runoff parameters for each sub-basin were determined.
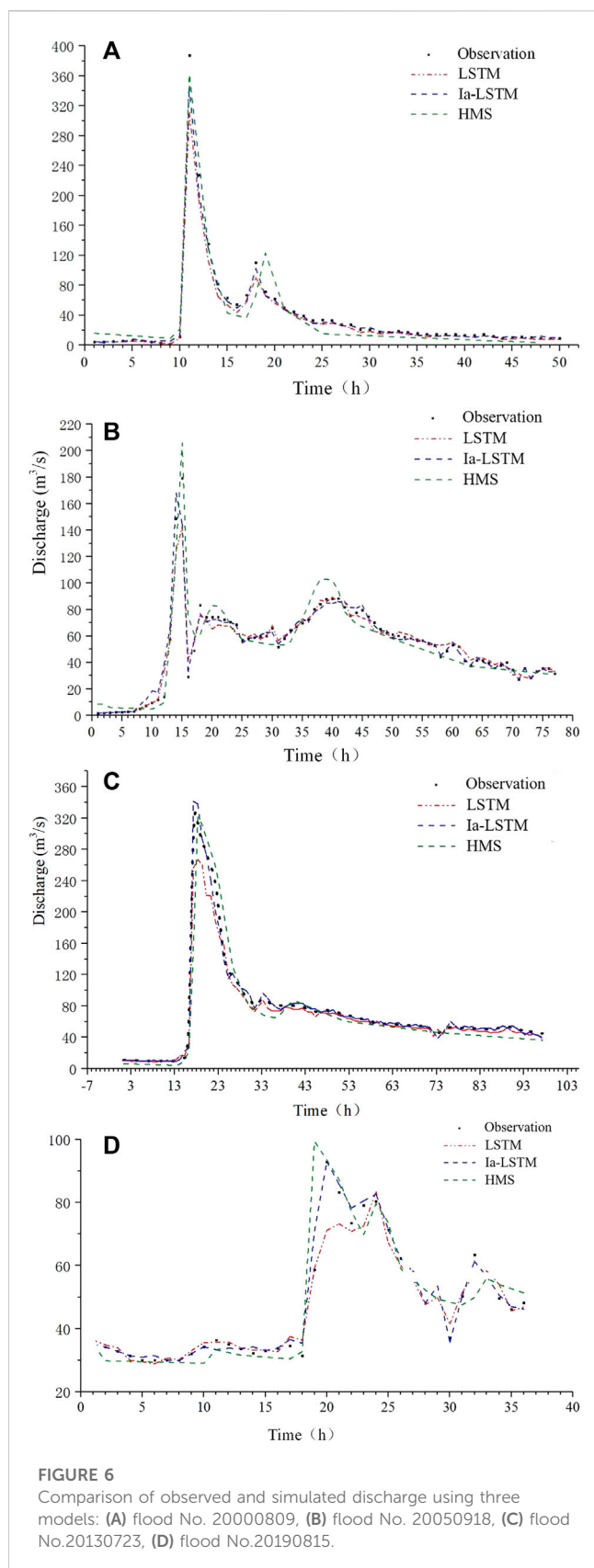
The value of initial loss (Ia) depends on the topography and land use conditions within the watershed. Typically, it is set at 10%–20%

TABLE 2 Optimization of initial loss (Ia) values for different flood events.

Flood No.	Precipitation (mm)	S1	S2	S3	S4	S5
19730715	101.0	19.1	21.5	18.5	21.3	18.3
19740801	105.1	18.8	19.4	19.8	17.2	19.4
19740812	47.3	9.8	12.6	11.3	15.0	13.5
19770716	38.5	15.8	13.5	16.1	12.9	16.1
19780701	97.4	14.4	17.3	14.7	16.6	14.0
19780713	51.5	8.8	9.7	9.4	10.6	8.0
19800629	169.2	24.1	22.9	22.7	19.3	19.3
19850729	240.8	12.3	12.7	11.9	11.8	12.7
19900709	31.8	14.3	17.0	16.3	15.1	14.3
19900720	58.2	7.0	6.5	10.8	8.1	96
19900814	73.4	12.5	14.0	14.7	11.5	10.2
19910727	65.2	17.1	19.4	16.5	17.5	18.6
19940702	90.9	18.1	17.5	20.6	19.0	17.3
19940712	53.8	8.3	9.8	9.6	8.7	7.8
19950815	18.2	12.4	13.3	11.4	12.2	12.6
19950902	65.5	10.2	8.9	9.5	10.4	10.8
19960803	94.2	12.4	13.3	11.4	12.2	12.6
19980702	37.9	10.9	11.2	9.3	9.9	9.9
19980802	113.1	21.3	17.9	20.7	20.7	21.1
20000809	162.8	13.0	16.3	14.2	13.7	14.7
20030828	58.0	8.3	9.8	9.6	8.7	7.8
20040729	65.1	12.1	14.9	15.6	17.2	18.0
20050918	156.5	18.1	15.6	16.3	16.7	17.9
20110811	50.4	19.2	15.4	17.3	19.9	14.6
20110831	42.3	10.9	11.2	9.3	9.9	11.9
20120714	82.7	17.0	16.2	16.6	20.7	17.1
20130723	124.5	10.8	13.5	11.6	13.5	12.2
20160715	24.0	12.0	16.5	14.2	13.3	11.4
20190815	123.0	9.5	11.4	10.2	9.3	12.0
20200807	56.0	13.9	15.6	13.3	12.5	12.2

of the total rainfall for forested areas. In this study, based on the soil and land use characteristics, the Ia value was determined as 30 mm, while the constant loss rate ranged from 0.30 mm/h to 1.16 mm/h.

The optimization procedure involved using a search method to minimize an objective function and find optimal parameters. To determine the optimal values of Ia for different flood events, the parameters were optimized using the Nelder-Mead optimization algorithm, with the peak-weighted root mean square as the objective function. The resulting optimized values are presented in Table 2. The analysis of Table 2 reveals the following conclusions regarding the relationship between rainfall and initial loss values:

FIGURE 6
Comparison of observed and simulated discharge using three models: **(A)** flood No. 20000809, **(B)** flood No. 20050918, **(C)** flood No.20130723, **(D)** flood No.20190815.

Different sub-basins exhibit varying initial loss values for the same flood event. For example, in the flood event on 19730715 with a rainfall of 101 mm, the corresponding initial loss values for the

sub-basins are as follows: S1-19.1 mm, S2-21.5 mm, S3-18.5 mm, S4-21.3 mm, and S5-18.3 mm.

The magnitude of initial loss values is not solely determined by the amount of rainfall. Other factors, such as the surface condition, rainfall characteristics, topography and slope, soil type and moisture content, and antecedent rainfall, play a significant role in determining initial loss values. These factors interact and collectively influence the extent of initial rainfall loss.

There is no clear linear relationship between rainfall and initial loss values in Table 2. This suggests that the estimation of initial loss values cannot solely rely on the amount of rainfall. Instead, a comprehensive understanding and consideration of the various factors influencing initial loss is necessary for accurate estimation.

3.2 Performance comparison

Flood events No. 20000809 and 20050918 were selected for model calibration, while flood events 20130918 and 20190815 were used for model validation. Figure 6 presents a comparison of simulated and observed discharges for four flood events in the Yufuhe basin using three models: HEC-HMS, LSTM, and Ia-LSTM. The figure shows that these models can generally capture the overall runoff process during the rainfall-runoff forecasting. However, some discrepancies exist in accurately simulating localized peak values. Despite this, the predicted values exhibit consistent trends with the observed values.

Regarding the comparison of relative error of peak discharge (*REP*) for the three models, Table 3 shows that different models exhibit varying performance in simulating peak discharge for each flood event. The LSTM model exhibits relatively large relative errors, particularly exceeding 20% for the flood event on 20130918. In contrast, both the Ia-LSTM and HEC-HMS models demonstrate significantly smaller relative errors, with all four flood events falling within the acceptable range. Notably, the Ia-LSTM model outperforms the other models, with a mere 1.3% error for the peak discharge during the flood event on 20190815. On average, the HEC-HMS model has a relative error of 9.8%, while the Ia-LSTM model has 8.1% for the peak discharge across all four flood events. These findings highlight the superior performance of the Ia-LSTM model in simulating peak discharge.

Table 4 presents the errors in peak time for different flood events predicted by the HEC-HMS, LSTM, and Ia-LSTM models. For the flood event on 20190815, the LSTM model exhibited a peak time error of 2 h. Conversely, for the flood event on 20130918, both HEC-HMS model and LSTM had a peak time error of 1 h. In contrast, the Ia-LSTM model achieved accurate peak time predictions for three out of the four flood events, with a maximum peak time error of 1 h. Notably, the Ia-LSTM model outperformed the other models by accurately simulating the temporal pattern of peak discharge propagation.

Table 5 presents a comprehensive comparison of three models (HEC-HMS, LSTM, and Ia-LSTM) based on key performance metrics: Nash-Sutcliffe Efficiency (*NSE*), Root Mean Square Error (*RMSE*), and Coefficient of Determination (R^2). Notably, during the flood event on 20000809, all models demonstrated exceptional performance with *NSE* coefficients above 0.86, *RMSE* values ranging from 7.234 to 14.503, and R^2 coefficients

TABLE 3 Comparison of relative error of peak discharge (*REP*) for three models.

Flood No.	Observed flow (m³/s)	HEC-HMS		LSTM		Ia-LSTM	
		Predicted flow (m³/s)	Relative error %	Predicted flow (m³/s)	Relative error %	Predicted flow (m³/s)	Relative error %
20000809	387.22	362.62	−6.4	310.94	−19.7	345.86	−10.7
20050918	179.37	205.92	14.8	142.23	−20.7	164.65	−8.2
20130918	360.00	321.77	11.1	274.18	−23.8	314.46	−12.1
20190815	92.92	99.39	7.0	78.42	−15.6	94.13	1.3

exceeding 0.90. The HMS model showed heightened accuracy in predicting the flood event on 20190815, potentially due to its detailed consideration of the rainfall process. Additionally, the Ia-LSTM model consistently displayed commendable performance across various flood events, with *NSE* coefficients ranging from 0.755 to 0.923, *RMSE* values between 2.314 and 7.234, and R^2 coefficients from 0.798 to 0.941. Importantly, the Ia-LSTM model consistently outperformed the LSTM model, highlighting its effectiveness in flood prediction and modeling.

The Ia-LSTM model consistently outperforms in various flood events, showing lower *RMSE*, and higher *NSE* and R^2 values. This highlights its effectiveness in flood prediction, especially compared to the LSTM model, emphasizing the importance of initial loss incorporation for accurate simulations.

3.3 Impact of initial loss

The analysis of initial loss in the proposed hybrid model provides valuable insights for improving rainfall-runoff predictions. By integrating initial loss estimation with LSTM neural networks, the Ia-LSTM model captures the complex interactions among various hydrological components, including rainfall, vegetation, soil, and runoff. This integration allows for a more comprehensive representation of the rainfall-runoff process, leveraging the strengths of physically-based and data-driven modeling approaches.

Consistent results demonstrate the superiority of the Ia-LSTM hybrid model over the individual HEC-HMS and LSTM models in estimating peak discharge, predicting peak time, and achieving higher *NSE*, lower *RMSE*, and greater R^2 values. The incorporation of initial loss estimation enhances the model's ability to simulate runoff dynamics. This leads to improved accuracy and reliability. In the Yufuhe basin case study, the Ia-LSTM model demonstrates an average improvement of 6.05% and 13.7% in peak discharge estimation compared to the HEC-HMS model and LSTM, respectively.

These findings emphasize the importance of accurate initial loss estimation in rainfall-runoff modeling, particularly for flood management and forecasting. Accurate initial loss estimation provides a clearer understanding of the initial loss processes and their impact on runoff generation. Through the optimization of initial loss values obtained from the HEC-HMS model, the Ia-LSTM model achieves heightened accuracy and reliability in simulating rainfall-runoff dynamics.

3.4 Comparison with previous studies

The Ia-LSTM hybrid model represents a significant advancement in rainfall-runoff modeling. Previous studies in the Yufuhe basin have employed various methodologies and models. For instance, Zhang et al. (2016) developed a distributed flood forecasting model based on sub-basins, river reaches, and reservoirs, achieving high performance with a Nash-Sutcliffe Efficiency (*NSE*) exceeding 0.70 and a Relative Error of Peak Discharge (*REP*) below 10%. Similarly, Yang et al. (2013) focused on the application of the SWAT distributed hydrological model, yielding satisfactory results with *NSE* and R^2 exceeding 0.70, and a relative error in peak flow below 15%. Their work highlights the effectiveness of their model in capturing key influencing factors of floods within the Yufuhe basin.

In recent years, machine learning models, particularly those based on LSTM, have exhibited promise in runoff forecasting. For instance, Xiang and YanDemir, (2020) proposed an LSTM-sequence-to-sequence rainfall-runoff model, demonstrating notable predictive power for short-term flood predictions. The LSTM model produced *NSE* values of 0.72, 0.80, and 0.93 for the Tripoli, Independence, and Anamosa stations, respectively. Additionally, an LSTM network was applied to build a data-driven model for streamflow prediction in an urban watershed.

While deep learning algorithms may not fully capture the rainfall-runoff process, they can be used to discern streamflow patterns and to identify effective variables, making them the preferred choice for modeling in data-poor catchments.

In our study, the Ia-LSTM model outperforms previous models, exhibiting *NSE* coefficients ranging from 0.755 to 0.923, *RMSE* values between 2.314 and 7.234 m³/s, and R^2 coefficients from 0.798 to 0.941. These results signify substantial advancements in rainfall-runoff modeling.

This research builds upon earlier works by incorporating initial loss estimation and utilizing the powerful Ia-LSTM hybrid model. This approach significantly enhances accuracy and reliability in simulating rainfall-runoff dynamics, particularly in terms of estimating peak discharge and predicting peak time.

The findings of this study have important implications for flood forecasting and water resource management. The Ia-LSTM hybrid model demonstrates superior performance in simulating peak discharge and predicting peak time compared to individual HEC-HMS and LSTM models. This suggests its potential for accurate and reliable rainfall-runoff modeling, which is crucial for disaster prevention, mitigation, and water resource management.

Calibration and Validation	Flood No.	Observed peak time	HEC-HMS		LSTM		Ia-LSTM	
			Predicted peak time	Period time error	Predicted peak time	Period time error	Predicted peak time	Period time error
Calibration	20000809	24	23	−1	24	0	24	0
	20050918	15	15	0	15	0	14	−1
Validation	20130918	17	18	1	18	1	17	0
	20190815	20	19	−1	22	2	20	0

Additionally, the integration of initial loss estimation with LSTM neural networks represents a significant advancement in rainfall-runoff modeling. This approach captures complex interactions among various hydrological components, providing a more comprehensive representation of the rainfall-runoff process.

The Ia-LSTM hybrid model shows promise for a wide range of applications, including flood forecasting, water resource management, and infrastructure planning. Its effectiveness in data-driven rainfall-runoff modeling with integrated physical mechanisms can significantly enhance the efficiency of flood prediction and management.

4 Conclusion

This study presents a hybrid rainfall-runoff model combining initial loss estimation with LSTM networks, significantly enhancing runoff forecasting accuracy. Effective runoff, obtained by subtracting initial loss from total rainfall through HEC-HMS simulations, was used as the input for the LSTM network. The Ia-LSTM hybrid model, integrating physically-based and data-driven modeling approaches, outperforms both individual HEC-HMS and LSTM models, as evidenced by case studies in the Yufuhe basin.

The integration of physically-based and data-driven modeling techniques in the Ia-LSTM hybrid model offers a comprehensive representation of the rainfall-runoff process. This integration significantly improves the model's ability to capture the complex dynamics of rainfall-runoff, resulting in enhanced peak discharge estimation. The optimized initial loss values derived from the HEC-HMS model contribute to the increased accuracy of the Ia-LSTM model.

The case studies conducted in the Yufuhe basin demonstrate the effectiveness of the Ia-LSTM model in simulating peak discharge and accurately predicting peak time for the flood events. The performance of Ia-LSTM model was evaluated with Nash-Sutcliffe Efficiency (NSE), root mean square error (RMSE), relative error of peak discharge (REP) and coefficient of determination (R^2). The Ia-LSTM model, in particular, shows an average improvement of 6.05% and 13.7% in peak discharge estimation compared to the HEC-HMS model and LSTM, respectively. The model achieves NSE values ranging from 0.755 to 0.923, RMSE values between 2.314 and 7.234 m^3/s, and R^2 coefficients from 0.798 to 0.941. This demonstrates the consistent outperformance of the Ia-LSTM model across various flood events, as indicated by lower RMSE, and higher NSE and R^2 values.

These findings highlight the importance of accurate initial loss estimation and the potential of hybrid modeling approaches in improving rainfall-runoff predictions. Accurate estimation of initial loss enables a better understanding of the runoff generation process and its influence on peak discharge. The integration of initial loss estimation with LSTM in the hybrid model contributes to its superior performance in simulating peak discharge and capturing the temporal pattern of peak flow propagation. These findings offer promise for enhancing the accuracy and reliability of hydrological forecasting models.

While LSTM has been effective in rainfall-runoff forecasting, there's room for improvement. Extending the output sequence length using historical rainfall-runoff data will significantly enhance long-term predictions.

Simplifying the complex process of initial loss estimation, which currently relies on HEC-HMS, is crucial. Future research can explore efficient techniques like the SCS curve method, considering factors

TABLE 5 Comparison of *NSE*, *RMSE* and *R^2*.

Flood No.	NSE			RMSE (m³/s)			R^2		
	HMS	LSTM	Ia-LSTM	HMS	LSTM	Ia-LSTM	HMS	LSTM	Ia-LSTM
20000809	0.890	0.861	0.923	14.503	13.088	7.234	0.923	0.909	0.956
20050918	0.784	0.786	0.823	7.635	6.614	6.050	0.822	0.725	0.875
20130918	0.862	0.820	0.902	10.202	15.657	6.783	0.916	0.875	0.941
20190815	0.741	0.552	0.755	4.120	8.800	2.314	0.805	0.626	0.798

such as soil type, pre-rainfall soil moisture, and the CN parameter. This streamlined approach makes initial loss estimation practical and applicable in real-world scenarios.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

WW: Writing–original draft. JG: Data curation, Formal Analysis, Writing–review and editing. ZL: Investigation, Writing–review and editing. CL: Funding acquisition, Methodology, Writing–review and editing.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Arnold, J. G., Srinivasan, R., Muttiah, R. S., and Williams, J. R. (1998). Large area hydrologic modeling and assessment part I: model development 1. *JAWRA J. Am. Water Resour. Assoc.* 34 (1), 73–89. doi:10.1111/j.1752-1688.1998.tb05961.x

Barrera-Animas, A. Y., Oyedele, L. O., Bilal, M., Akinosho, T. D., Delgado, J. M. D., and Akanbi, L. A. (2022). Rainfall prediction: a comparative analysis of modern machine learning algorithms for time-series forecasting. *Mach. Learn. Appl.* 7, 100204. doi:10.1016/j.mlwa.2021.100204

Bartoletti, N., Casagli, F., Marsili-Libelli, S., Nardi, A., and Palandri, L. (2018). Data-driven rainfall/runoff modelling based on a neuro-fuzzy inference system. *Environ. Model. Softw.* 106, 35–47. doi:10.1016/j.envsoft.2017.11.026

Beven, K. (2020). Deep learning, hydrological processes and the uniqueness of place. *Hydrol. Process.* 34 (16), 3608–3613. doi:10.1002/hyp.13805

Bicknell, B. R., Imhoff, J. C., Kittle, J. L., Jr, Donigian, A. S., Jr, and Johanson, R. C. (1997). *Hydrological simulation program—FORTRAN user's manual for version 11*. Report No. EPA/600/R-97/080. Athens, GA, USA: US Environmental Protection Agency.

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi:10.1023/a: 1010933404324

Chen, C., Jiang, J., Liao, Z., Zhou, Y., Wang, H., and Pei, Q. (2022). A short-term flood prediction based on spatial deep learning network: a case study for Xi County, China. *J. Hydrology* 607, 127535. doi:10.1016/j.jhydrol.2022.127535

Chen, J., and Adams, B. J. (2006). Integration of artificial neural networks with conceptual models in rainfall-runoff modeling. *J. Hydrology* 318 (1-4), 232–249. doi:10.1016/j.jhydrol.2005.06.017

Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297. doi:10.1007/bf00994018

Cui, Z., Zhou, Y., Guo, S., Wang, J., Ba, H., and He, S. (2021). A novel hybrid XAJ-LSTM model for multi-step-ahead flood forecasting. *Hydrology Res.* 52 (6), 1436–1454. doi:10.2166/nh.2021.016

Danandeh Mehr, A., and Nourani, V. (2018). Season algorithm-multigene genetic programming: a new approach for rainfall-runoff modelling. *Water Resour. Manag.* 32, 2665–2679. doi:10.1007/s11269-018-1951-3

Devia, G. K., Ganasri, B. P., and Dwarakish, G. S. (2015). A review on hydrological models. *Aquat. procedia* 4, 1001–1007. doi:10.1016/j.aqpro.2015. 02.126

Farfán, J. F., Palacios, K., Ulloa, J., and Avilés, A. (2020). A hybrid neural network-based technique to improve the flow forecasting of physical and data-driven models: methodology and case studies in Andean watersheds. *J. Hydrology Regional Stud.* 27, 100652. doi:10.1016/j.ejrh.2019.100652

Feldman, A. D. (2000). *Hydrologic modeling system HEC-HMS: technical reference manual*. Washington, D.C, USA: US Army Corps of Engineers, Hydrologic Engineering Center.

Fenicia, F., Savenije, H. H., Matgen, P., and Pfister, L. (2008). Understanding catchment behavior through stepwise model concept improvement. *Water Resour. Res.* 44 (1), 1–13. doi:10.1029/2006wr005563

Gholami, V., and Khaleghi, M. R. (2021). A simulation of the rainfall-runoff process using artificial neural network and HEC-HMS model in forest lands. *J. For. Sci.* 67 (4), 165–174. doi:10.17221/90/2020-jfs

Gu, H., Xu, Y. P., Ma, D., Xie, J., Liu, L., and Bai, Z. (2020). A surrogate model for the Variable Infiltration Capacity model using deep learning artificial neural network. *J. Hydrology* 588, 125019. doi:10.1016/j.jhydrol.2020.125019

Haykin, S., and Network, N. (2004). A comprehensive foundation. *Neural Netw.* 2, 41.

Hitokoto, M., and Sakuraba, M. (2020). Hybrid deep neural network and distributed rainfall-runoff model for real-time river-stage prediction. *J. JSCE* 8 (1), 46–58. doi:10.2208/journalofjsce.8.1_46

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 9 (8), 1735–1780. doi:10.1162/neco.1997.9.8.1735

Hu, C., Wu, Q., Li, H., Jian, S., Li, N., and Lou, Z. (2018). Deep learning with a long short-term memory networks approach for rainfall-runoff simulation. *Water* 10 (11), 1543. doi:10.3390/w10111543

Hundecha, Y., Bardossy, A., and Werner, H. W. (2001). Development of a fuzzy logic-based rainfall-runoff model. *Hydrological Sci. J.* 46 (3), 363–376. doi:10.1080/02626660109492832

Jaber, F. H., and Shukla, S. (2012). MIKE SHE: model use, calibration, and validation. *Trans. ASABE* 55 (4), 1479–1489. doi:10.13031/2013.42255

Kim, D., Lee, J., Kim, J., Lee, M., Wang, W., and Kim, H. S. (2022). Comparative analysis of long short-term memory and storage function model for flood water level forecasting of Bokha stream in NamHan River, Korea. *J. Hydrology* 606, 127415. doi:10.1016/j.jhydrol.2021.127415

Knebl, M. R., Yang, Z. L., Hutchison, K., and Maidment, D. R. (2005). Regional scale flood modeling using NEXRAD rainfall, GIS, and HEC-HMS/RAS: a case study for the San Antonio River Basin Summer 2002 storm event. *J. Environ. Manag.* 75 (4), 325–336. doi:10.1016/j.jenvman.2004.11.024

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M. (2018). Rainfall–runoff modelling using long short-term memory (LSTM) networks. *Hydrology Earth Syst. Sci.* 22 (11), 6005–6022. doi:10.5194/hess-22-6005-2018

Kumar, P. S., Praveen, T. V., and Prasad, M. A. (2016). Artificial neural network model for rainfall-runoff-A case study. *Int. J. Hybrid Inf. Technol.* 9 (3), 263–272. doi:10.14257/ijhit.2016.9.3.24

Kumarasamy, K., and Belmont, P. (2018). Calibration parameter selection and watershed hydrology model evaluation in time and frequency domains. *Water* 10 (6), 710. doi:10.3390/w10060710

Lees, T., Buechel, M., Anderson, B., Slater, L., Reece, S., Coxon, G., et al. (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 Earth Syst. Sci.* 25 (10), 5517–5534. doi:10.5194/hess-25-5517-2021

Mehr, A. D., and Akdegirmen, O. (2021). Estimation of urban imperviousness and its impacts on flashfloods in Gazipaşa, Turkey. *Knowledge-Based Eng. Sci.* 2 (1), 9–17. doi:10.51526/kbes.2021.2.1.9-17

Mohammadi, B., Safari, M. J. S., and Vazifehkhah, S. (2022). IHACRES, GR4J and MISD-based multi conceptual-machine learning approach for rainfall-runoff modeling. *Sci. Rep.* 12 (1), 12096. doi:10.1038/s41598-022-16215-1

Narayana Reddy, B. S., and Pramada, S. K. (2022). A hybrid artificial intelligence and semi-distributed model for runoff prediction. *Water Supply* 22 (7), 6181–6194. doi:10.2166/ws.2022.239

Noori, N., and Kalin, L. (2016). Coupling SWAT and ANN models for enhanced daily streamflow prediction. *J. Hydrology* 533, 141–151. doi:10.1016/j.jhydrol.2015.11.050

Perera, T., McGree, J., Egodawatta, P., Jinadasa, K. B. S. N., and Goonetilleke, A. (2019). Taxonomy of influential factors for predicting pollutant first flush in urban stormwater runoff. *Water Res.* 166, 115075. doi:10.1016/j.watres.2019.115075

Rahimzad, M., Moghaddam Nia, A., Zolfonoon, H., Soltani, J., Danandeh Mehr, A., and Kwon, H. H. (2021). Performance comparison of an LSTM-based deep learning model versus conventional machine learning algorithms for streamflow forecasting. *Water Resour. Manag.* 35 (12), 4167–4187. doi:10.1007/s11269-021-02937-w

Safari, M. J. S., Arashloo, S. R., and Mehr, A. D. (2020). Rainfall-runoff modeling through regression in the reproducing kernel Hilbert space algorithm. *J. Hydrology* 587, 125014. doi:10.1016/j.jhydrol.2020.125014

Savic, D. A., Walters, G. A., and Davidson, J. W. (1999). A genetic programming approach to rainfall-runoff modelling. *Water Resour. Manag.* 13, 219–231. doi:10.1023/a:1008132509589

Shukla, A. K., Ojha, C. S. P., Garg, R. D., Shukla, S., and Pal, L. (2020). Influence of spatial urbanization on hydrological components of the upper ganga river basin, India. *J. Hazard. Toxic, Radioact. Waste* 24 (4), 04020028. doi:10.1061/(asce)hz.2153-5515.0000508

Shukla, A. K., Pathak, S., Pal, L., Ojha, C. S. P., Mijic, A., and Garg, R. D. (2018). Spatio-temporal assessment of annual water balance models for upper Ganga Basin. *Hydrology Earth Syst. Sci.* 22 (10), 5357–5371. doi:10.5194/hess-22-5357-2018

Sun, A. Y., Scanlon, B. R., Zhang, Z., Walling, D., Bhanja, S. N., Mukherjee, A., et al. (2019). Combining physically based modeling and deep learning for fusing GRACE satellite data: can we learn from mismatch? *Water Resour. Res.* 55 (2), 1179–1195. doi:10.1029/2018wr023333

Tian, Y., Xu, Y. P., Yang, Z., Wang, G., and Zhu, Q. (2018). Integration of a parsimonious hydrological model with recurrent neural networks for improved streamflow forecasting. *Water* 10 (11), 1655. doi:10.3390/w10111655

Wang, W., and Ding, J. (2003). Purification of boiling-soluble antifreeze protein from the legume Ammopiptanthus mongolicus. *Nat. Sci.* 1 (1), 67–80. doi:10.1081/PB-120018370

Wang, W., Vrijling, J. K., Van Gelder, P. H., and Ma, J. (2006). Testing for nonlinearity of streamflow processes at different timescales. *J. Hydrology* 322 (1-4), 247–268. doi:10.1016/j.jhydrol.2005.02.045

XiangYan, Z. J., and Demir, I. (2020). A rainfall-runoff model with LSTM-based sequence-to-sequence learning. *Water Resour. Res.* 56 (1). doi:10.1029/2019wr025326

Xie, K., Liu, P., Zhang, J., Han, D., Wang, G., and Shen, C. (2021). Physics-guided deep learning for rainfall-runoff modeling by considering extreme events and monotonic relationships. *J. Hydrology* 603, 127043. doi:10.1016/j.jhydrol.2021.127043

Xie, T., Zhang, G., Hou, J., Xie, J., Lv, M., and Liu, F. (2019). Hybrid forecasting model for non-stationary daily runoff series: a case study in the Han River Basin, China. *J. Hydrology* 577, 123915. doi:10.1016/j.jhydrol.2019.123915

Xie, X., Huang, L., Marson, S. M., and Wei, G. (2023b). Emergency response process for sudden rainstorm and flooding: scenario deduction and Bayesian network analysis using evidence theory and knowledge meta-theory. *Nat. Hazards* 117 (3), 3307–3329. doi:10.1007/s11069-023-05988-x

Xie, X., Tian, Y., and Wei, G. (2023a). Deduction of sudden rainstorm scenarios: integrating decision makers' emotions, dynamic Bayesian network and DS evidence theory. *Nat. Hazards* 116 (3), 2935–2955. doi:10.1007/s11069-022-05792-z

Yang, S., Xu, Z., Kong, Ke., Miao, S., and Zhang, S. (2013). A flow simulation based on SWAT model in Wohushan reservoir basin. *China Rural Water and Hydropower* (5), 11–18. doi:10.3969/j.issn.1007-2284.2013.05.003

Yaseen, Z. M., Jaafar, O., Deo, R. C., Kisi, O., Adamowski, J., Quilty, J., et al. (2016). Stream-flow forecasting using extreme learning machines: a case study in a semi-arid region in Iraq. *J. Hydrology* 542, 603–614. doi:10.1016/j.jhydrol.2016.09.035

Young, C. C., and Liu, W. C. (2015). Prediction and modelling of rainfall–runoff during typhoon events using a physically-based and artificial neural network hybrid model. *Hydrological Sci. J.* 60 (12), 2102–2116. doi:10.1080/02626667.2014.959446

Zheng, Y., Li, J., Dong, L., Rong, Y., Kang, A., and Feng, P. (2020). Estimation of initial abstraction for hydrological modeling based on global land data assimilation system–simulated datasets. *J. Hydrometeorol.* 21 (5), 1051–1072. doi:10.1175/jhm-d-19-0202.1

Zhang, L., Yang, Z., and Liu, G. (2016). A forecast model of distributed flood in Yufuhe basin and its application. *J. Water Resour. Water Eng.* 27 (3), 66–72. doi:10.11705/j.issn.1672-643X.2016.03.13

Zhao, R. J. (1992). The Xinanjiang model applied in China. *J. Hydrology* 135, 371–381. doi:10.1016/0022-1694(92)90096-e

Zhou, Q., Teng, S., Situ, Z., Liao, X., Feng, J., Chen, G., et al. (2023). A deep-learning-technique-based data-driven model for accurate and rapid flood predictions in temporal and spatial dimensions. *Hydrology Earth Syst. Sci.* 27 (9), 1791–1808. doi:10.5194/hess-27-1791-2023

Zhou, Y., Cui, Z., Lin, K., Sheng, S., Chen, H., Guo, S., et al. (2022). Short-term flood probability density forecasting using a conceptual hydrological model with machine learning techniques. *J. Hydrology* 604, 127255. doi:10.1016/j.jhydrol.2021.127255