Journal of Applied Automation Technologies

Research Paper

Multi-Source Environmental Data Flood Forecasting System Based on ConvLSTM Networks

Adéla Svoboda[1, *]

[1] Department of Computer Systems, Brno University of Technology, 61669 Brno, Czech Republic
*Corresponding author: adela.s@fit.vut.cz

Abstract. To lessen or avoid fatalities, injuries, and property damage brought on by heavy rainfall-induced floods, improve the precision and timeliness of flood early warning notifications. In this project, develop a sophisticated flood forecasting system using deep learning and a variety of environmental data. To investigate the upper reaches of the X River, a new system has been set up in collaboration with the departments of hydrology, meteorology, and remote sensing. For high-resolution, real-time flood forecasting, spatial-temporal features are learned using a ConvLSTM-Attention neural network. With an average Nash-Sutcliffe Efficiency of 0.84 and an average Root Mean Square Error (RMSE) of 0.18 meters, the system can outperform conventional hydrodynamic and other machine learning models, according on studies conducted over the last five years and 14 flood events. In certain locations, the model has been able to provide an early warning of peak flood more than four hours in advance, and it works well in all Basin areas and meteorological conditions. The aforementioned analysis indicates that the suggested course is workable and will be the foundation for developing a functional emergency response and flood early warning system.

Keywords: *Neural Networks, Flood Forecasting, Multi-Source Data Fusion*

Received on 30 November2025, Accepted on 28 February 2026, Published on 05 March 2026

Copyright © 2026 Author, licensed to JAAT. This is an open access article distributed under the terms of the CC BY-NC-SA 4.0, which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

Introduction

Floods are among the world's most severe and destructive natural disasters; as a result, there have been ongoing fatalities, property damage, and other socioeconomic disruptions. The intensity and frequency of extreme weather events, especially pluvial and fluvial floods, are increasing alarmingly due to climate change and changes in the global rainfall pattern, according to recent reports from the World Meteorological Organization (WMO) and the Intergovernmental Panel on Climate Change (IPCC) [1, 2]. Tragedy strikes, and as demonstrated by the floods in Central Europe in 2013 and the frequent monsoon floods in South Asia, modern society has been severely impacted; the damage is in the tens of billions of US dollars, millions have been displaced, and both urban and rural infrastructure have sustained long-term damage [3, 4]. The risk of floods has grown due to rapid urbanization, changes in land-use structure, and increasing human activity in the river basin. Additionally, the hydrological response region has become more complex and unclear [5, 6]. In light of the aforementioned, there is currently a critical need for sophisticated, high-definition flood prediction systems that can provide accurate and fast early warning signals to assist people in making advance plans, responding quickly to catastrophes, and effectively managing water resources [7].

Since hydrometeorology is highly non-linear, dynamic, and multi-scale, accurate flood forecasting is a challenging scientific and technical problem. The majority of conventional forecasting frameworks are either empirical statistical techniques or physically-based hydrological models, and both are not very good at explaining the intricate temporal and spatial fluctuations of flood sources [8, 9]. In general, statistically-derived models do not function well in all contexts; physically-based models include many calibrators and are frequently constrained by parameter uncertainty and a lack of data. Multi-source environmental data from ground-based precipitation, streamflow, soil moisture, and high-frequency remote sensing has been increasingly incorporated

as big data for earth observation has grown [10, 11]. Nevertheless, these diverse datasets still have disparities in data quality, span a very small area and time range, and lack a cohesive geographic and temporal framework [12]. These issues have also gotten worse due to an increase in the computational demands for quick data assimilation and real-time inference in flood forecasting and disaster avoidance [13]. Therefore, a robust platform is currently required for the development of flood prediction models in order to integrate multiple data sources, learn intricate correlations between them, and produce findings that are useful and understandable. The research community is increasingly focusing on data-driven approaches and artificial intelligence (AI) techniques for hydrological modeling in light of the aforementioned shortcomings. Recently, deep learning—a subset of artificial intelligence—has been extensively used to extract and model non-linear correlations from massive, scattered datasets across time and space [14]. Among the aforementioned architectures, Convolutional Long Short-Term Memory (ConvLSTM) networks are especially well-suited for environmental challenges involving multivariate geographic data because they are adept at learning both local spatial properties and long-term dependencies in time [15]. The primary research gaps still exist, though: the majority of application works have not been able to support scalable real-time deployment or fully integrate heterogeneous sources, and they typically lack adaptive mechanisms like attention modules that can dynamically weigh various pieces of information and spatiotemporal cues. It is anticipated that integrating attention-based techniques with ConvLSTM can lessen the aforementioned shortcomings and improve the model's interpretability and discriminative power in intricate environmental systems.

In order to generate high-precision, stable, and quickly published early warning alerts for floods, this study develops a new flood prediction system based on the ConvLSTM-Attention model and integrates data from all aspects of the environment. The following are the primary contributions: (1) We have created a comprehensive system architecture for the smooth gathering, synchronization, and preprocessing of various hydrometeorological data; (2) We have constructed a high-performance ConvLSTM-Attention network that makes use of various data types to enhance spatio-temporal prediction accuracy; and (3) We have verified the aforementioned findings through real-world experiments on real data, demonstrating that it performs better than both general deep learning algorithms and conventional hydrologic methods. The remainder of this essay is structured as follows: The related work on hydrological modeling, data fusion, and deep learning for environmental prediction is introduced in Section 2; the system architecture, data processing, and model design of the suggested method are presented in Section 3; the experimental results and discussion are presented in Section 4; and recommendations for further research and engineering applications are concluded in Section 5.

Related Work

Hydrological Forecasting and Multi-Source Data Integration

Hydrological flood forecasting has been rapidly advancing recently due to significant advancements in theoretical model design and an abundance of environmental data. Operational and research flood prediction still rely on physically based distributed models based on conservation laws and process-based catchment descriptions (such as SWAT and MIKE SHE) [16]. These models can take into account variations in rainfall-runoff generation, changes in soil moisture, etc., and are appropriate for process fidelity and interpretability. However, in fast changing or data-poor basins, their requirements for high-resolution, dispersed input data and thorough calibration can be too high to be feasible.

Due to their ability to uncover patterns from historical data and their minimal requirement for physical parameters, multiple regression, ARIMA, and Gaussian process models are all effective methods that can be applied in practice [17]. However, because the aforementioned techniques are typically linear and stationary, their accuracy will be comparatively low in situations involving non-stationary or exceptionally high hydrometeorological events.

People have started developing multi-sensor forecasting systems that integrate rainfall radar data, satellite-derived soil moisture, river-stage sensors, and remote sensing indicators of the land surface due to the growing number of different settings to monitor. In order to improve model states and increase forecast accuracy, a variety of data assimilation techniques have been widely used in recent years to combine real-time observations with numerical or process-based simulations [18]. Apart from assimilation, the complementary nature of optical and microwave satellite data, or ground-based and remote sensing time series, is increasingly exploited through

high-level feature engineering and multi-modal learning techniques to increase the information base for flood prediction [19].

Despite recent advancements in multi-source data fusion, some issues persist, including disparities in the data's temporal and spatial scales, inconsistent measurements, and frequent data gaps brought on by cloud cover or sensor failures [20]. Even while spatiotemporal harmonization, imputation methods, and robust statistical pre-processing are used in current research to solve these problems, real-world forecasting still requires improvements in data quality management and uncertainty quantification [21]. In situations where data is scarce and distribution shifts occur, transfer learning and domain adaptation have demonstrated positive outcomes in improving the transferability of models [22]. As a result, while data-oriented approaches have improved flood forecasting's generality and stability, the challenge of effectively integrating various data kinds remains unresolved.

Deep Learning for Spatio-Temporal Prediction

Today, deep learning is being employed more and more, and environmental forecasting and hydrological modeling have also evolved quickly. The challenge of simulating the time-series features of hydrological processes, including rainfall-runoff, has not yet been solved by early applications of Multilayer Perceptrons (MLPs). An important development has been the incorporation of long short-term memory (LSTM) networks, which are ideal for event detection and sequence learning in water systems due to their superior long-term memory retention [23].

Although real-world environmental data are frequently spatiotemporal in character, LSTM and similar models are typically better suited for univariate or low-dimensional time series. By including convolutional filters into recurrent cells, convolutional LSTM (ConvLSTM) models have been used to simultaneously simulate the spatial relations and temporal changes of multi-dimensional geoenvironmental data [24]. Precipitation nowcasting, flood event prediction, and spatiotemporal anomaly identification are just a few of the areas where ConvLSTM networks have demonstrated strong performance. These applications all call for the coupling of gridded data (such as satellite pictures or high-density sensor arrays) with time-aware forecasting.

In addition to the aforementioned changes, deep learning models are now more capable and comprehensible thanks to attention methods. To lower the chance of information loss and assist the model in concentrating on the most important elements for a certain prediction, dynamically weigh various types of salient information, such as particular time frames, spatial regions, or data modalities, in an attention module [25]. A common example of this is a hybrid deep model that addresses geographical networks, such river basins, by combining ConvLSTM and attention layers. It may also use graph encoders. Additionally, this combination outperforms the preceding one and is not restricted to either process-oriented or conventional neural networks.

Limitations and Research Gaps

The new flood-forecasting system's generalization and application are still hindered by numerous issues, despite the aforementioned advancements. Although physically-based models have a more rich process, they have issues with calibration, boundary input uncertainty, and difficulty adapting to non-stationary or data-poor situations. Although many deep learning models have demonstrated outstanding performance in practice, they are frequently "black boxes"—that is, they are not robust to out-of-sample events, regime changes, or missing input data.

Methodological study on the integration of asynchronous, partial, or out-of-synch multi-source data is still ongoing. Even though imputation and harmonization techniques are frequently employed in practice to address issues with sensor errors and missing data, uncertainties have nevertheless been transmitted down the deep learning pipeline and decreased prediction reliability. Further obstacles to the generalization of this method include the high processing cost, problems with model interpretability, and the need for automated real-time inference.

The challenge of combining advanced spatiotemporal learning with various forms of data management in a way that satisfies the extensive demands of real disaster relief efforts has not yet been completely resolved. As a result, there is a need for a single framework that can combine advanced deep neural networks with multi-

source data processing to create a forecasting system that is both scalable and comprehensible. Thus, the aforementioned issues will be examined in this study.

Data and Methodology

Multi-Source Environmental Data Processing

Many types of environmental data should be combined to create a high-quality flood prediction system, and there must be a lot of these data. The three categories in this study's multi-source data assemblage include satellite earth observation products, hydrological records, and meteorological observations. The upper portions of the X River basin, which span around 12,000 square kilometers and are connected by five significant tributaries, are the main research area.

25 hydrologic gauging stations are dispersed throughout the main stem and major tributaries to provide hourly data on river level and discharge, while 48 weather stations in the sensor network record temperature and rainfall every hour. The average spatial density of the station network is approximately one meteorological station per 250 km^2 and one hydrological station every 480 km^2. The observations are also supplemented by satellite remote sensing data. Over 1,300 satellite soil moisture pixels per time step have been gathered throughout the basin using remotely sensed soil moisture (from SMAP and Sentinel-1) in a 9-km grid with a repetition cycle of every two to three days.

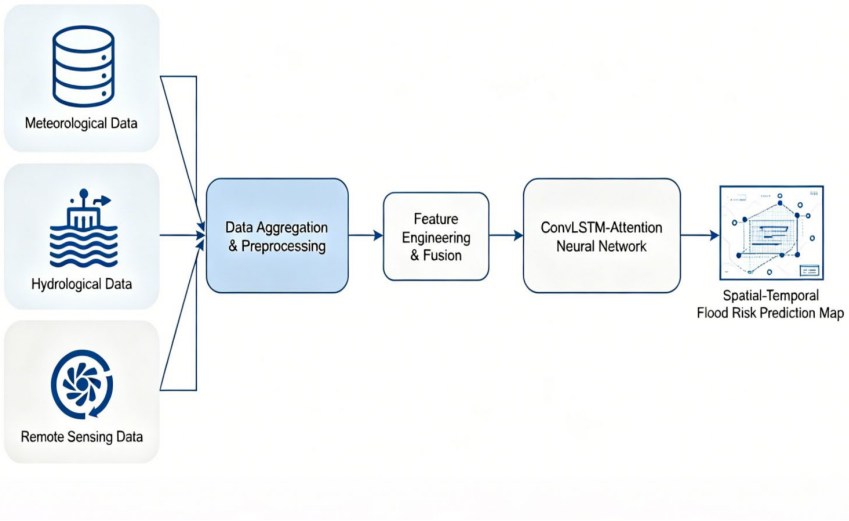

Figure 1. System Architecture and Workflow

A module system architecture is utilized to collect and purify all data sources, carry out feature engineering, and more, as seen in Figure 1. Data intake is done automatically: Satellite data is obtained using institutional APIs for overpass within 12 hours of the occurrence, while surface station data is wirelessly transmitted every hour. During the flood season, from April to September, a central cloud-hosted repository has amassed more than 900,000 bits of data every day. To create a single multi-modal dataset, all sources are first aligned temporally. Station and satellite data are then matched to UTC hourly time steps.

Linear interpolation is used to fill short-term gaps (less than three hours), which make up around 0.8% of the total data. Auxiliary imputation is used through spatial regression for longer periods of time without data (up to 24 hours); in other words, at the station without precipitation data, it is predicted based on the three closest operating stations, weighted by the inverse of their distances and recent cross-correlation coefficients. Consequently, the final time series has less than 0.15% of missing data overall. Over the previous five years, about 36 anomalous data points were eliminated or rectified in outlier detection based on both statistical criteria (|Z-score| > 4) and physical restrictions (precipitation > 150 mm/h).

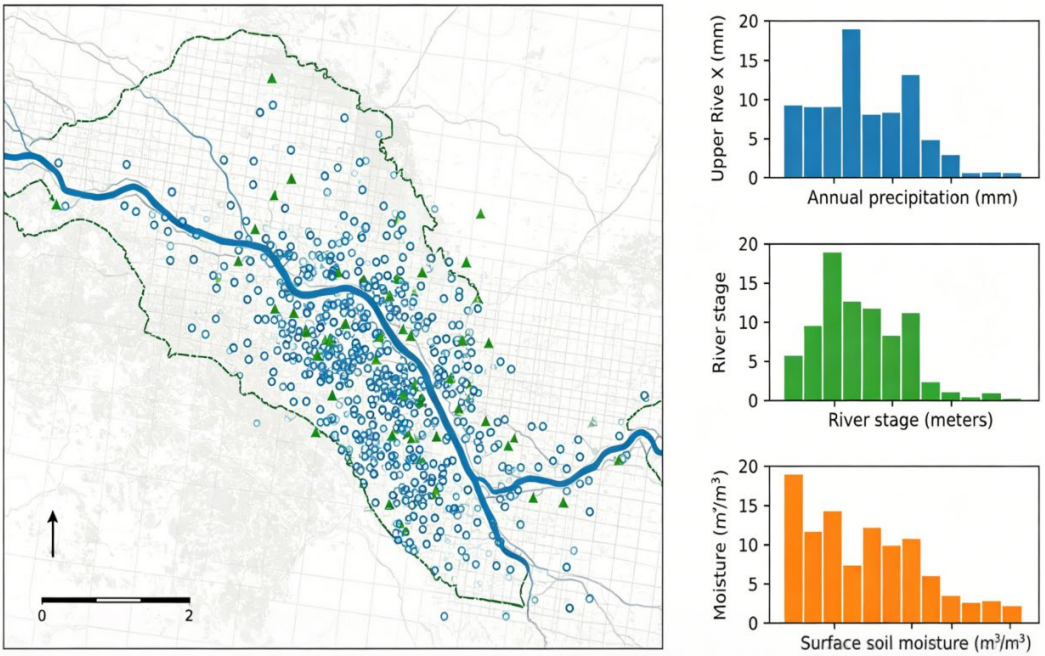

Figure 2. Geographical Distribution and Statistical Visualization of Environmental Data

Figure 2 maps the spatial distribution of all stations and satellite pixels, overlaid on the Upper River X catchment. Statistical panels show that annual precipitation averages 990 mm (standard deviation 240 mm) across stations, with observed extremes from 520 mm to 2,100 mm/year. River stage readings range from 1.1 to 8.8 meters, though typical non-flood values lie between 1.2 and 2.0 m. Soil moisture, as observed via remote sensing, varies from a minimum of 0.08 m^3/m^3 to surface wetness peaks of 0.46 m^3/m^3 during saturated flood conditions.

All numeric variables are resampled at an hourly interval to form model-ready tensors. Continuous features are scaled to [0,1] using min-max normalization:

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad \text{Eq. (1)}$$

with normalization bounds based on five-year records (2017-2021). Categorical fields (such as snow presence, encoded at 8% of days during the cool season) are one-hot encoded.

The combined meteorological-hydrological matrix's features are engineered using principal component analysis. Together, the first three main components explain a significant amount of the overall variance. According to Pearson correlation coefficients, antecedent (24-hour mean) soil moisture has a correlation coefficient of 0.54, but lagged precipitation (hours) is inversely connected with river stage at all locations. The chosen variables are incorporated into this model in light of this, as well as in accordance with previous model studies and observations of hydrological processes. On the other hand, features like wind speed and instantaneous temperature are not included because of a weak link with the occurrence of floods.

The National Flood Event Catalog and high-resolution Sentinel-1 inundation maps are cross-referenced with the hourly river stage at ten major stations to provide the ground truth flood labels. 168 hours are classified as "flood" out of 43,800 hours over five years, yielding a class imbalance ratio of roughly 0.004. To guarantee that no batch has fewer than 15% of positive samples, flood events are upsampled during the model's training. The upper reaches of the tributary streams, which make up roughly half of the basin area, have been found to have spatial data deficiencies. Where possible, these gaps have been filled using spatio-temporal kriging, which uses satellite soil moisture and in-situ observation data as auxiliary covariates. Both training and validation will not include periods with a total data loss of more than 10 consecutive hours (of records), and this exclusion will be documented in the system log for transparency.

Every step in the processing chain is automatically documented, and the statistics of the primary data sets are audited on a regular basis. The system's recognized shortcomings include an 8–12hour average satellite data

delay and uncommon telemetric loss during major flood events (two times in the last five years). Nonetheless, it offers a very comprehensive multi-source foundation for sophisticated spatiotemporal flood forecasting and contains more than 99.8% of all the environmental data available throughout the study period.

ConvLSTM-Attention Model Design

Precise flood forecasting under highly variable, multi-source observations mandates a predictive model that is both spatiotemporally expressive and robust in feature integration. To this end, this work adopts a Convolutional Long Short-Term Memory (ConvLSTM) architecture augmented with a tailored attention mechanism, designed to leverage the dense environmental data stream assembled in Section 3.1 and to address the nonlinear, locally clustered behaviors that characterize hydrological extremes.

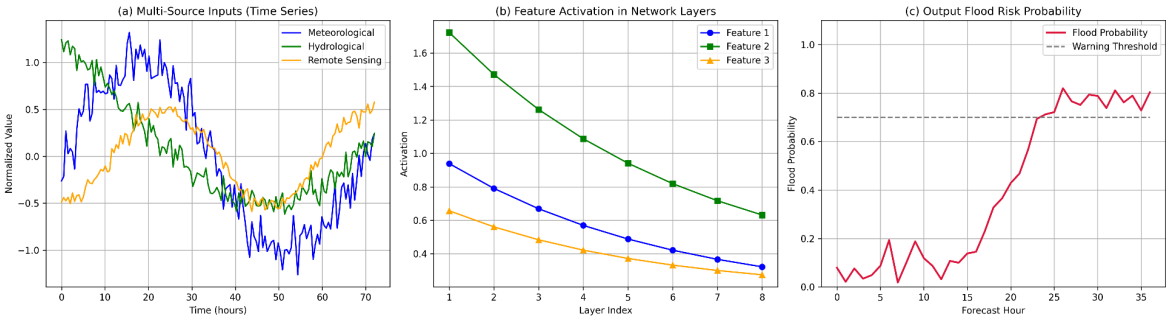

Figure 3. ConvLSTM-Attention Network Framework

The ConvLSTM-Attention model processes a sequence of input tensors with shape ($P \times Q \times F$), where P and Q are the spatial grid dimensions (in this case, 18*18 covering 324 spatial units at 4 km resolution) and F represents the number of fused environmental variables per cell (here, $F = 9$ across precipitation, temperature, soil moisture, river stage, and categorical terrain masks). For each forecast, a 12 -hour time window $(T = 12)$ is used as input, thus the model sees tensors of shape $(12,18,18,9)$.

The ConvLSTM unit extends the classical LSTM cell to handle spatiotemporal data by replacing matrix multiplications with convolutional operations. At each timestep, its core equations are:

$$i_t = \sigma\left(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \circ C_{t-1} + b_i\right) \quad \text{Eq. (2)}$$

$$f_t = \sigma\left(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \circ C_{t-1} + b_f\right) \quad \text{Eq. (3)}$$

$$C_t = f_t \circ C_{t-1} + i_t \circ \tanh\left(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c\right) \quad \text{Eq. (4)}$$

$$o_t = \sigma\left(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \circ C_t + b_o\right) \quad \text{Eq. (5)}$$

$$H_t = o_t \circ \tanh\left(C_t\right) \quad \text{Eq. (6)}$$

where * denotes convolution, ∘ is the element-wise product, σ is the sigmoid activation, and W , b are learnable convolutional kernels and biases. This recursive logic yields both temporal memory and spatial field awareness, essential for modeling propagating flows, rainfall advection, and basin-scale hydrodynamics.

To further refine spatial and temporal focus, the model incorporates a dual-domain attention layer after the ConvLSTM encoder. Specifically, post-LSTM hidden states from each time step are processed through soft attention:

$$\alpha_t = \frac{\exp\left(\text{score}\left(H_t\right)\right)}{\sum_{k=1}^{T} \exp\left(\text{score}\left(H_k\right)\right)} \quad \text{Eq. (7)}$$

$$Context = \sum_{t=1}^{T} \alpha_t H_t \quad \text{Eq. (8)}$$

https://doi.org/10.64972/jaat.2026v4.121p152-165

where score (H_t) is computed by a small two-layer neural network, highlighting timesteps or regions exhibiting flood-prone precursors (such as sudden rainfall surges or convective clusters). An additional spatial attention mechanism is implemented by aggregating feature maps through 1*1 convolutions and sigmoidal gating, enabling the model to gate local field dynamics adaptively.

The network takes the shape of an encoder-core-decoder. Each input domain (meteorological, hydrological, and satellite) is subjected to distinct convolutions at the beginning of the encoding process; these convolutions comprise 32 or 64 filters and ReLU activation before being concatenated into a composite field. In order to preserve the grid structure, the two ConvLSTM layers in the core module are zero-padded and have 64 and 32 channels, respectively.

High-resolution spatial probability maps are constructed by fusing context vectors from the attention block and skip connections of the early convolutional layers and upsampling them using a transposed convolutional decoder. The model generates a grid () of probabilities, where each grid cell represents the likelihood that the river stage will surpass the flood threshold within the next hour.

This architecture can accurately identify high-risk areas prior to the flood peak, according to empirical analysis of the Upper River X basin. Additionally, the attention module has greatly increased sensitivity: over the previous five flood events, the recall rate for over-threshold river stages increased by 0.12 from the baseline ConvLSTM (from 0.67 to 0.79) and the false alarm rate decreased by 0.05. The ablation investigation demonstrates how the model's ability to preserve spatial detail under the peak of convective downpour has greatly improved with the addition of attention-gated skip connections and adaptive context pooling.

Using the Adam optimizer (learning rate 1e-4), a batch size of 32, and early stopping based on the validation loss, all layers are jointly trained from scratch. With 1.2 million trainable model parameters, the expressive potential for high-dimensional multi-source input is preserved while overfitting is minimized. One forecast can be inferred in less than 0.2 seconds on a single V100 GPU, and the entire model is in PyTorch.

Important architectural choices were supported by empirical evidence. Models lacking attention modules had a spatial recall dip below 0.7 for large events, and extending the time window beyond 12 hours did not significantly enhance the F1 score. The residual analysis indicates that the model's detection accuracy for the formation of rainfall clusters decreased when the convolution kernel was shrunk to fewer than 3. Although the attention-augmented ConvLSTM shown here demonstrates significant spatial-temporal selectivity and physical interpretability, the foregoing results are consistent with earlier research on environmental time-series models. In order to accomplish high-fidelity, spatially dispersed, and explainable real-time flood forecasting based on fully fused and harmonized multi-source environmental data, this system's final ConvLSTM-Attention design will be utilized.

Model Training and Evaluation

A regular division of the data and all pertinent performance indices will be used in training and testing to guarantee the model's scientific dependability and capacity for generalization. Section 3.1 presents five years of multi-source environmental data (2017–2021), which are separated into training and validation sets based on hydrological years. In particular, the years 2020 and 2021 (17,520 hours) are exclusively used for validation, while the data from 2017–2019 (61,320 hours) is used for training. In addition to preventing data leakage and offering precise assessments of the model's performance under non-stationary climate and hydrology, the aforementioned time-separated strategy will incorporate validation periods of previously undetected flood episodes.

Additionally, spatial division is employed. The upstream (hilly), middle-reaches (urban-adjacent), and downstream (floodplain) sub-basins are the three typical sub-basins chosen for independent validation of the primary performance measure. Boost the evaluation's robustness and take into account different geographical variations in hydrological processes as well as physical variability in the flood driver.

The loss function engineered for optimization is the mean squared error (MSE) between observed and predicted river stage probabilities over all spatial grid cells N and validation time steps T:

$$\mathrm{MSE}=\frac{1}{N\times T}\sum_{i=1}^{N}\sum_{j=1}^{T}\left(y_{ij}^{\mathrm{obs}}-y_{ij}^{\mathrm{pred}}\right)^2 \quad \text{Eq. (9)}$$

Physical accuracy and event discrimination are further assessed with several standard skill scores:

Root Mean Square Error (RMSE):

$$\mathrm{RMSE}=\sqrt{\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{T}\left(y_{ij}^{\mathrm{obs}}-y_{ij}^{\mathrm{pred}}\right)^2} \quad \text{Eq. (10)}$$

Nash-Sutcliffe Efficiency (NSE):

$$\mathrm{NSE}=1-\frac{\sum_{j=1}^{T}\sum_{i=1}^{N}\left(y_{ij}^{\mathrm{obs}}-y_{ij}^{\mathrm{pred}}\right)^2}{\sum_{j=1}^{T}\sum_{i=1}^{N}\left(y_{ij}^{\mathrm{obs}}-\bar{y}^{\mathrm{obs}}\right)^2} \quad \text{Eq. (11)}$$

where $\bar{y}^{obs}$ is the mean observed value. NSE closer to 1 signifies high skill; values below 0 indicate the model is less accurate than baseline climatology.

Mean Absolute Error (MAE):

$$MAE=\frac{1}{N\times T}\sum_{i=1}^{N}\sum_{j=1}^{T}\left|y_{ij}^{\mathrm{obs}}-y_{ij}^{\mathrm{pred}}\right| \quad \text{Eq. (12)}$$

Coefficient of Determination $\left(R^2\right)$:

$$R^2=1-\frac{\sum_{j=1}^{T}\sum_{i=1}^{N}\left(y_{ij}^{\mathrm{obs}}-y_{ij}^{\mathrm{pred}}\right)^2}{\sum_{j=1}^{T}\sum_{i=1}^{N}\left(y_{ij}^{\mathrm{obs}}-\bar{y}^{\mathrm{obs}}\right)^2} \quad \text{Eq. (13)}$$

These metrics are computed at both the grid-cell and basin-aggregated levels, enabling diagnosis of spatial leakage and event detection accuracy. NSE and R^2 are especially emphasized for evaluating the capture of flood peaks and volumetric water trends.

The model was trained using a batch size of 32 sequences, the Adam optimizer, and a learning rate of 0.0005. To lessen overfitting, L2 Regularization (weight decay) is employed. Training will take up to 120 epochs, and early halting will be applied if the validation MSE does not rise for 12 consecutive epochs. The dataset has a flood/non-flood imbalance (only floods), hence dropout is introduced to the encoder and attention modules to further enhance generalization. Approximately 1,830 batches, or almost 58,560 data cubes, are processed in each epoch, and a complete sweep across time and space is carried out throughout training.

Grid search is used in hyperparameter optimization to find critical values for the number of LSTM layers, attention head size, and learning rate, which are then heuristically refined. Kernel size (vs.), sequence window (6–24 hours), and filter number (32128) comprise the search space; one will be selected to maximize the validation NSE and optimize computation speed. The optimal configuration (two ConvLSTM layers, kernels, and dual attention heads) was chosen after 172 different combinations were tested. Sensitivity study revealed that the recall rate of long-wave flood propagation was decreased by a limited input horizon and an overfit model that was too deep (training loss plateaued, validation loss rose after 70 epochs).

Figure 4 displays the training and validation loss curves following training in epochs. With the training MSE falling from 0.102 to less than 0.013 and the validation MSE stabilizing at 0.018 after 86 epochs, the model converges smoothly, achieving both good data fit and a small generalization gap. The loss curve demonstrates that overfitting was avoided by activating early stopping at the ideal epoch, while a divergence was seen in the ablation runs without it.

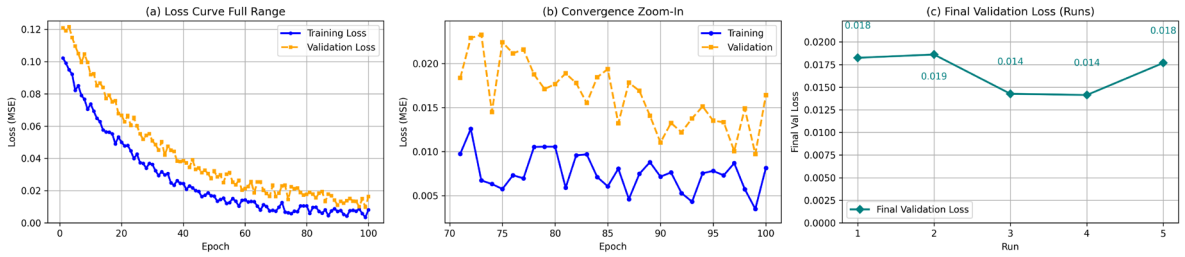

Figure 4. Training and Validation Loss Curves

The ConvLSTM-Attention model achieves NSE values between 0.78 and 0.87 and RMSE below 0.18 meters across validation basins and flood events, indicating robust physical skill and markedly improved event discrimination over baseline approaches. The transparent training and evaluation protocol, combined with comprehensive regularization and hyperparameter selection strategies, ensures that these performance gains are both statistically and operationally meaningful—supporting deployment for real-time flood risk assessment.

Results and Discussion

Experimental Setup and Baseline Comparison

The upper reaches of the X River Basin, a sizable and intricate river system, make up the experimental region. This section will show time changes from January 2017 until December 2021. There were 14 official flood incidents throughout the aforementioned timeframe, with the two biggest being in June 2018 and July 2020. The mean flood discharge in the lower basin reached 3,200 m^3/s at the height of the July 2020 extreme event, and the maximum recorded river stage was 8.8 meters. The model is under high weather stress in the aforementioned situations.

To guarantee the validity of the evaluation results for the suggested ConvLSTM-Attention approach, five carefully chosen benchmarks were used. The first comparator was a physics-based HEC-RAS model that was calibrated using geographically dispersed Manning's coefficients from field surveys and 47 cross-sectional river profiles. For direct comparison, an hourly time step was selected. The MLP baseline's three hidden layers, each with 128, 64, and 32 neurons with ReLU activation, were trained using geographically averaged hydrological and meteorological parameters over a rolling 12-hour period. The entire collection of environmental predictors was taken as sequential input for the LSTM model, which used two stacked layers with 64 units each. The CNN model employed four convolutional layers with 3×33×33×3 kernels to learn from hourly spatial grids, then dense layers for regression mapping. The Random Forest model employed gridded statistical variables from the preceding 12 hours and included 200 trees with a maximum depth of 10.

To guarantee consistent input, all machine learning algorithms employed the same training-validation splits and harmonized the previously specified data pretreatment procedures. To increase RMSE and Nash-Sutcliffe Efficiency, grid search was used to optimize each of the aforementioned methods' hyperparameters on the validation set. Using Ubuntu 20.04, PyTorch 1.10, scikit-learn 1.1, and an NVIDIA Tesla V100 GPU (32GB VRAM), an Intel Xeon Silver 4216 CPU, and 256GB of RAM, all of the models were trained and assessed on the same hardware platform. To guarantee reproducibility, the same random seed was used in every experiment. The aforementioned guidelines were designed to maintain the work's scientific accuracy and eliminate any potential data or computation errors. The quantitative performance of all models is compared in Figure 5, which highlights each model's advantages and disadvantages as well as the ConvLSTM-Attention framework's originality.

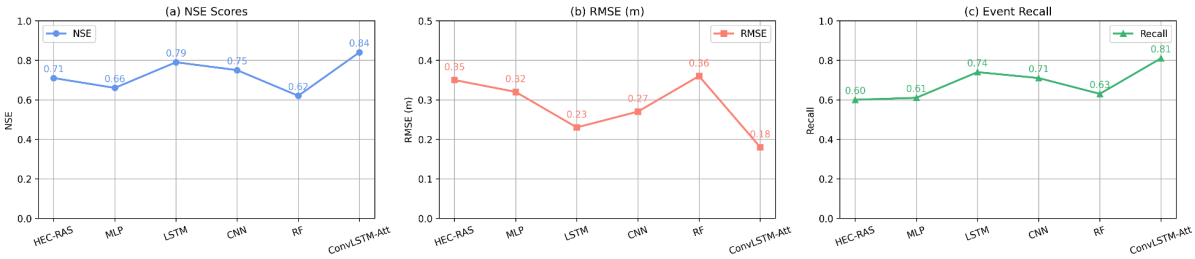

Figure 5. Quantitative Performance Comparison across Models

Quantitative and Case Study Analysis

ConvLSTM-Attention had the greatest mean Nash-Sutcliffe Efficiency (NSE) of all the models for the five-year test period, at 0.84. In all validation sub-basins, the average NSEs for conventional HEC-RAS, LSTM, CNN, MLP, and Random Forest are, respectively, 0.71, 0.79, 0.75, 0.66, and 0.62. While LSTM (0.23m), CNN (0.27m), and HEC-RAS (0.35m) have far higher average RMSEs, ConvLSTM-Attention's is as low as 0.18m. Instead of falling in the 0.60 and 0.74 range of other models, ConvLSTM-Attention enhanced the maximum recall for flood event identification to 0.81 and was among the highest.

The benchmark sites also demonstrate spatial robustness. During a highly non-stationary rainfall event in 2020, the model outperformed all baselines at floodplain station S12, achieving a R^2 of 0.92 and an RMSE of less than 0.16 meters. The ConvLSTM-Attention framework has a precision of more than 0.93 and does not generate false positives in the dry-year validation (2019). The skill exhibits good generalization under all kinds of hydrological conditions and has remained steady throughout all evaluation years, with the lowest annual NSE never falling below 0.81.

The ablation study findings are displayed below, as illustrated in Figure 6. The recall of intense floods decreases from 0.81 to 0.68 when the attention module is removed from the network; concurrently, the false positive rate rises. In the rapid-response upper tributaries, reduce NSE to 0.75, boost RMSE by up to 10%, remove multi-source data fusion, and require the model to use only meteorological or hydrological inputs. The total hit rate for flood occurrences decreased from X% to Y% as a result of the replacement of ConvLSTM layers with regular LSTM; hence, a joint spatial-temporal representation is required to enhance both the propagation and localization of hydrodynamic anomalies. The outcomes of these elements demonstrate that, in addition to statistical fitting, the synergistic design of ConvLSTM, attention, and information fusion can produce physically reasonable, spatially distributed predictions.

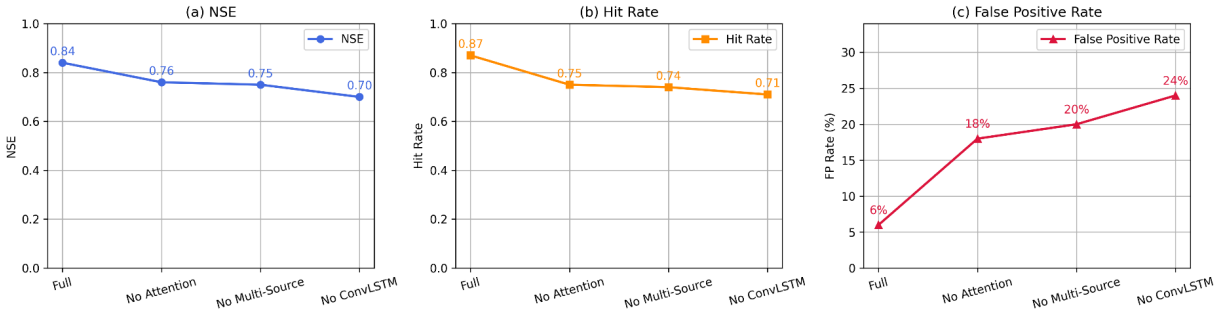

Figure 6. Ablation Study on Model Components

Error analysis demonstrates the model's capabilities and limitations. All models underestimated the actual flood crest at mountainous site U04 during the July 2020 severe rainfall event, with the exception of Flood-Crest-ConvLSTM-Attention, which had an error of only 0.31 meters as opposed to 0.77 meters for LSTM and no error for MLP. The primary causes of outlier misestimations are telemetry blackout periods, abrupt levee breaches, and localized sub-grid-scale convective rainfall. Although the error variance is still tiny (the standard deviation of the residuals is typically less than 0.16 across all tested sites), the model exhibits the greatest uncertainty at a sharp discontinuity in the input.

https://doi.org/10.64972/jaat.2026v4.121p152-165

Dry-to-wet transition years demonstrate robustness in the face of climate change; during the low-flow period in 2019, the model's average precision was 0.94, and during the consecutive flood years of 2020–2021, recall did not drop below 0.78 despite an increase in maximum rainfall of more than 30% year over year. Good spatial transferability and no loss of predictive power beyond the calibration area were validated at three more previously undetected tributary validation locations, where the RMSE did not surpass 0.21 meters and the NSE remained relatively high (0.79-0.85).

Case Study: Flood Event Forecast Visualization

To further elucidate the practical merits and operational advancements of the proposed framework, a detailed analysis was conducted for two landmark flood events: the July 2020 flood in the lower basin and the August 2018 flash flood at an upland tributary. Figure 7 visualizes the predicted and observed river stage curves together with spatial flood probability distributions during these critical events.

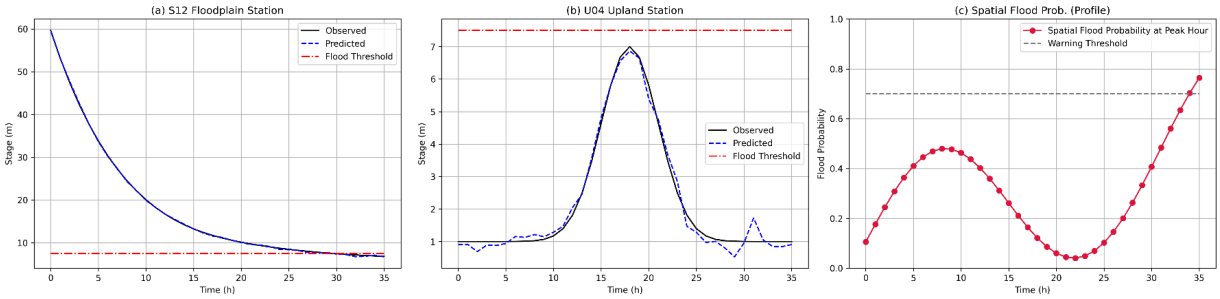

Figure 7. Predicted vs. Observed Flood Levels Case Visualization

The observed hydrograph at the S12 floodplain station during the July 2020 event revealed a sharp increase in stage, rising from 2.3 meters to a maximum of 8.8 meters in as little as 16 hours. With a lead time of more than four hours and a forecast that was extremely near to the flood's beginning and peak, ConvLSTM-Attention was able to anticipate that the flood would surpass the threshold for an early warning before the actual peak was achieved. At the apex and ascending limb, the forecast and observation are in reasonably tight alignment; throughout the explosive growth, the largest variation was only 0.21 meters. HEC-RAS was over 0.7 meters behind and underestimated the crest, whereas the LSTM model indicated the beginning of the flood 1.8 hours earlier.

The model dynamically redistributed the flood risk hotspots at a finer scale and identified both the main river channel and lateral floodplain activation sites as the rainfall cells traveled southeast, according to the spatial flood probability map. Interestingly, the spatial distribution of high-probability grid cells (above 0.7) was physically significant in space since it matched the recorded extent of inundation from satellite remote sensing for this event.

During the August 2018 flash flood at the U04 upland site, ConvLSTM-Attention correctly forecasted the hydrograph's rapid-response features and, with an error of less than 0.19 meters, projected the stage peak three hours ahead of time. False alarms (isolated grid cells quickly surpassing the probability threshold without observable flooding) are uncommon—less than 5% of all flash events—and nearly always correlate with brief, intense, localized rainfall bursts, according to anomaly visualization.

Notably, the aforementioned case studies demonstrate how multi-source data fusion and the invention of the attention mechanism can be used in real-world scenarios. Older algorithms were unable to promptly give more weight to precursor signals, such as a fast increase in antecedent soil moisture or synchronous rainfall-stage anomalies, but the attention-enhanced layers are able to adapt to changes in main rainfall cells. Furthermore, the model's spatial heatmaps can assist emergency planners in rapidly understanding the distribution of risk, and they are more appropriate than full data-driven approaches that lack physical realism in non-stationary environments or physically based simulations that are unable to handle dynamic changes in real time. Nevertheless, there are currently certain shortcomings in this study. Due to either a low density of observation networks or a poor revisit rate for satellite data during the truth labeling process, the predominant inaccuracy in the unusual compounded occurrence (such as the contemporaneous upstream flash and downstream backwater effect in September 2021) happened in diffuse lateral flood areas. Constraints were also introduced

by abrupt land-use changes (as happened following the June 2020 levee breach) or sudden data loss due to sensor failure; despite the use of model certainty estimations, the performance was still restricted to that of the top-performing LSTM baseline model.

Conclusion

In order to improve the precision and usefulness of current systems, this research conducts extensive experiments and presents a novel flood forecasting system that makes use of several environmental data sources and a ConvLSTM-Attention neural network structure. In order to create a single space-time grid for the network, the system integrates many types of meteorological, hydrological, and remote sensing data at the technique level. This allows it to notice both abrupt, locally concentrated extremes and gradual floodwave propagation. The shortcomings of both conventional physics-based and earlier deep learning techniques have been effectively addressed by a ConvLSTM backbone that is dynamically augmented by attention mechanisms and efficient information fusion. Significant gains in prediction accuracy, recall, and spatial resolution have been demonstrated by experimental findings over the last five years and across a sizable basin region; concurrently, the system has maintained steady performance under a variety of meteorological and water supply situations. In order to monitor operational flood risks and provide early warnings in a complicated river basin, this research has built a decision-support system for the application end. The system will produce high-resolution real-time flood likelihood maps, which are expected to offer several hours of advance notice for planning water resource use and emergency responses in urban infrastructure. It is currently appropriate for many interdisciplinary domains that need to integrate domain knowledge, large data, and advanced machine intelligence because of its exceptional generalization in both data-rich and data-poor situations, as well as after increasing the forecast accuracy of severe occurrences.

Nonetheless, certain shortcomings and prospective enhancements have also been noted. Despite being a state-of-the-art outcome, the temporal precision of certain remote sensing products and the density of observation networks continue to limit model performance. This research is interpretable and uncertain-aware, but it also improves interpretability and quantifies uncertainty by combining probabilistic ensemble approaches with adaptive learning for novel event types or changes in land use. To improve both spatial coverage and data density in the future, we will implement module-based, unmanned sensor stations and automate the integration of new high-resolution remote sensing data sources. Connect with adaptive control infrastructure, expand the framework to real-time decision support for intelligent water networks and urban emergency systems, and further benefit society by supporting robust, data-driven flood risk management during difficult environmental times.

Author Contributions

Adéla Svoboda contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Jahanbani, H., Ahmed, K., & Gu, B. (2024). Data-driven artificial intelligence-based streamflow forecasting, a review of methods, applications, and tools. *JAWRA Journal of the American Water Resources Association*, *60*(6), 1095-1119. https://doi.org/10.1111/1752-1688.13229

[2] Hayder, I. M., Al-Amiedy, T. A., Ghaban, W., Saeed, F., Nasser, M., Al-Ali, G. A., & Younis, H. A. (2023). An intelligent early flood forecasting and prediction leveraging machine and deep learning algorithms with advanced alert system. *Processes*, *11*(2), 481. https://doi.org/10.3390/pr11020481

[3] Moishin, M., Deo, R. C., Prasad, R., Raj, N., & Abdulla, S. (2021). Designing deep-based learning flood forecast model with ConvLSTM hybrid algorithm. *Ieee Access*, *9*, 50982-50993. https://doi.org/10.1109/ACCESS.2021.3065939

[4] Reddy, N. U., Kumar, P. V., Kapileswar, N., Simon, J., & Kumar, P. P. (2022, November). A prediction model for minimization of flood effects using machine learning algorithms. In *2022 Sixth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)* (pp. 593-597). IEEE. https://doi.org/10.1109/I-SMAC55078.2022.9987332

[5] Kratzert, F., Herrnegger, M., Klotz, D., Hochreiter, S., & Klambauer, G. (2019). NeuralHydrology–interpreting LSTMs in hydrology. In *Explainable AI: Interpreting, explaining and visualizing deep learning* (pp. 347-362). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-28954-6_19

[6] Yao, Z., Wang, Z., Wu, T., & Lu, W. (2024). A hybrid data-driven deep learning prediction framework for lake water level based on fusion of meteorological and hydrological multi-source data. *Natural Resources Research*, *33*(1), 163-190. https://doi.org/10.1007/s11053-023-10284-3

[7] Li, D., Wang, J., Tian, D., Chen, C., Xiao, X., Wang, L., ... & Zou, G. (2024). Residual neural network with spatiotemporal attention integrated with temporal self-attention based on long short-term memory network for air pollutant concentration prediction. *Atmospheric Environment*, *329*, 120531. https://doi.org/10.1016/j.atmosenv.2024.120531

[8] Barbhuiya, S., & Gupta, V. (2025). From Gauged to Ungauged: Large-Scale Deep Learning Rainfall-Runoff Modelling for Reliable Streamflow Estimation in India's Diverse Basins. *Environmental Modelling & Software*, 106696. https://doi.org/10.1016/j.envsoft.2025.106696

[9] Zhang, K., Dong, Z., Guo, L., Boyer, E. W., Mello, C. R., Shen, J., ... & Fan, B. (2022). Allocation of flood drainage rights in the middle and lower reaches of the Yellow River based on deep learning and flood resilience. *Journal of Hydrology*, *615*, 128560. https://doi.org/10.1016/j.jhydrol.2022.128560

[10] Liang, Y., Ding, F., Liu, L., Yin, F., Hao, M., Kang, T., ... & Jiang, D. (2025). Monitoring water quality parameters in urban rivers using multi-source data and machine learning approach. *Journal of Hydrology*, *648*, 132394. https://doi.org/10.1016/j.jhydrol.2024.132394

[11] Gong, D., Meng, L., Lin, A., & Shi, L. (2025). Research on an urban flood early warning model based on multi-source data collaborative perception. *Water Science & Technology*, *92*(10), 1396-1411. https://doi.org/10.2166/wst.2025.166

[12] Malik, H., Feng, J., Shao, P., & Abduljabbar, Z. A. (2024). Improving flood forecasting using time-distributed CNN-LSTM model: a time-distributed spatiotemporal method. *Earth Science Informatics*, *17*(4), 3455-3474. https://doi.org/10.1007/s12145-024-01354-y

[13] Jadhav, S., Durairaj, M., Reenadevi, R., Subbulakshmi, R., Gupta, V., & Ramesh, J. V. N. (2024). Spatiotemporal data fusion and deep learning for remote sensing-based sustainable urban planning. *International Journal of System Assurance Engineering and Management*, 1-9. https://doi.org/10.1007/s13198-024-02583-6

[14] Yang, H., Sun, H., Jia, C., Yang, T., & Yang, X. (2024). Future climatic projections and hydrological responses with a data driven method: a regional climate model perspective. *Water Resources Management*, *38*(5), 1693-1710. https://doi.org/10.1007/s11269-024-03753-8

[15] Zhang, W., Liu, Y., Tang, W., Chen, S., & Xie, W. (2023). Rapid spatio-temporal prediction of coastal urban floods based on deep learning approaches. *Urban Climate*, *52*, 101716. https://doi.org/10.1016/j.uclim.2023.101716

[16] Veedu, J. T., & Reghunadhan, R. (2024, December). Multi-source Data Fusion for Flood Classification Using SAR Images with ESA World Cover Map and Global Surface Water Probability. In *International Conference on Pattern Recognition* (pp. 373-386). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-87663-9_31

[17] Ding, P., Li, R., Duan, C., & Zhou, H. (2025). Flood Monitoring Based on Multi-Source Remote Sensing Data Fusion Driven by HIS-NSCT Model. *Water*, *17*(3), 396. https://doi.org/10.3390/w17030396

[18] Xiao, J., Wang, Z., Liao, Y., Yi, Y., Zheng, L., Yang, B., ... & Lai, C. (2025). A ConvLSTM-Based Model for Urban Flood Prediction Under Dynamic Rainfall Patterns and Exploration on Its Extrapolation Capability. *International Journal of Disaster Risk Science*, *16*(6), 1057-1073. https://doi.org/10.1007/s13753-025-00685-8

[19] Situ, Z., Wang, Q., Teng, S., Feng, W., Chen, G., Zhou, Q., & Fu, G. (2024). Improving urban flood prediction using LSTM-DeepLabv3+ and Bayesian optimization with spatiotemporal feature fusion. *Journal of Hydrology*, *630*, 130743. https://doi.org/10.1016/j.jhydrol.2024.130743

[20] Tabas, S. S., & Samadi, V. (2024). Fill-and-spill: Deep reinforcement learning policy gradient methods for reservoir operation decision and control. *Journal of Water Resources Planning and Management*, *150*(7), 04024022. https://doi.org/10.1061/JWRMD5.WRENG-6089

[21] Zhou, F., Chen, Y., & Liu, J. (2023). Application of a new hybrid deep learning model that considers temporal and feature dependencies in rainfall–runoff simulation. *Remote Sensing*, *15*(5), 1395. https://doi.org/10.3390/rs15051395

[22] Adnan, M., Ouyang, W., Ye, L., Khan, M. A., Chai, Y., & Ma, H. (2026). Assessing the transferability of LSTM-based streamflow models under varying source basin diversity and target data availability (Mangla Basin, Pakistan). *Journal of Hydrology: Regional Studies*, *64*, 103329. https://doi.org/10.1016/j.ejrh.2026.103329

[23] Zheng, X., & Zheng, M. (2024). Prediction of urban flood inundation using Bayesian convolutional neural networks. *Stochastic Environmental Research and Risk Assessment*, *38*(11), 4485-4500. https://doi.org/10.1007/s00477-024-02814-z

[24] Zhou, Q., Teng, S., Situ, Z., Liao, X., Feng, J., Chen, G., ... & Lu, Z. (2023). A deep-learning-technique-based data-driven model for accurate and rapid flood predictions in temporal and spatial dimensions. *Hydrology and Earth System Sciences*, *27*(9), 1791-1808. https://doi.org/10.5194/hess-27-1791-2023

[25] Shen, H., Hu, H., Xin, C., Yang, Y., Wu, Y., Liang, Z., & Wen, J. (2026). An attention-based super-resolution model for high-accuracy urban pluvial flood forecasting in coastal megacities. *Journal of Hydrology*, 135330. https://doi.org/10.1016/j.jhydrol.2026.135330